



السجل العلمي



المؤتمر الدولي لعلوم وهندسة الحاسوب باللغة العربية

الدورة السابعة

الجزء الثاني

السجل العلمي

المؤتمر الدولي لعلوم وهندسة الحاسوب باللغة العربية

الدورة السابعة

١٤٢٣ هـ - ٢٠١١ م

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



الجلسة التاسعة

تصنيف الحروف عن طريق خوارزم التحسين بسرب الجسيمات

نواورية نبيلة¹،

¹ جامعة الكيبك بمنتريال، كلية العلوم/ قسم الحاسوب

CP. 8888, Succ. Centre-ville, Montreal, QC H3P3P8, Canada

البريد الإلكتروني: nouaouria.nabila@gmail.com

الخلاصة. هذه المقالة عبارة عن دراسة استخدام التحسين بسرب الجسيمات (ت. س. ج.) لتصنيف الحروف اللاتينية. الفكرة الرئيسية قائمة على البحث عن المواقع الأمثل للجسيمات المناظرة للنقاط الوسطى من الفئات. لأجل الحصول على تحكم أكثر للجسيمات استعملنا آلية حبس الجسيمات، كما استعملنا آلية تشتيت الرياح للحصول على أكثر تنوع. فضلا عن ذلك قمنا بمجموعة اختبارات على قاعدة بيانات الحروف اللاتينية، وقد تحصلنا على دقة اعتراف أفضل من تلك التي تحصلت عليها ادوات التصنيف المعروفة في هذا المجال.

الكلمات الجوهرية: الذكاء الاصطناعي، التحسين بسرب الجسيمات، التصنيف، حبس الجسيمات، تشتيت الرياح، الاعتراف بالحروف.

1 مقدمة

قد تنوعت البحوث في حقل التحسين بسرب الجسيمات (ت. س. ج.) ، ولا سيما في مجال التحسين المستمر للدوال. ومع ذلك إن عددا متزايدا من الأعمال تبحث في استخدامات جديدة لخوارزميات التحسين بسرب الجسيمات. بولي في [1] قام بإحصاء مجالات التطبيق الرئيسية حيث تم إستعمال ت. س. ج. بنجاح. تم تقسيم التطبيقات إلى 26 فئة مختلفة، وإن كان كثير منها ينتمي إلى أكثر من فئة واحدة. واستند التصنيف على تحليل أكثر من 1100 بحث منشور عن التحسين بسرب الجسيمات في قاعدة البيانات IEEE Xplore وكانت النتائج أن التجميع والتصنيف واستخراج البيانات يمثل 4.3٪ من إجمالي الإنتاج.

التصنيف بالإشراف على التعلم يعالج مشكلة عامة لإيجاد C-تقسيم معقول إلى أصناف من مجال الإدخال، بإعطاء C-قسم من مجموعة أمثلة تدريبية. في هذه الورقة، نقدم حل لهذه الإشكالية قائم على منهجية التحسين بسرب الجسيمات ، وتطبيقه على تصنيف الحروف اللاتينية. بما أن ت. س. ج. هو أساسا أداة تحسين، مشكلة استخراج C-قسم يمكن التعبير عنها باستخراج C مواقع الجسيمات الأمثل المقابلة لـ C نقطة وسطى من طبقات التصنيف. الخوارزم يقوم باجتياز مجال البحث عن طريق اختبار دالة لياقة الجسيمات وفقا لمعنى التشابه مع مجموعة الأمثلة التدريبية ، وسيتم الاحتفاظ بالجسيمات التي تحصل على أكبر قيمة لدالة اللياقة كنقاط وسطى. وقد قمنا بإختبار الخوارزم مع آليات مختلفة مثل الحبس والتشتت، والنتائج التي تحققت تتحدى تلك التي وجدت في الأدب العلمي.

في ما يلي، نقدم أولاً ملخص عن خلفية تحسين سرب الجسيمات. القسم 3 يفسر منهجية التصنيف التي تم استخدامها. القسم 4 يعرض مجموعة البيانات المستخدمة في الاختبارات وإستغلالاتها السابقة في المنشورات العلمية. ثم نقوم بعرض النتائج التجريبية ومناقشتها في القسم 5. وأخيراً، القسم 6 يلخص الورق.

2 الخلفية الطبيعية و الخوارزمية للتحسين بسرب الجسيمات

في نظام التحسين بسرب الجسيمات، كل جسيم مرشح ليمثل حلاً لمشكلة التحسين المطروحة [2]. عادة ما يؤثر على موقع الجسيم هو أفضل موقع زاره بذاته (أي تجربته الخاصة) وموقع أفضل الجسيمات في جواره (أي تجربة الجسيمات المجاورة). عندما يكون الجوار يتكون من السرب بأكمله، أفضل موضع في الجوار يشار إليه بـ "الأفضل الشامل" gbest [3]. عندما يستخدم أصغر جوار، يشار عادة إلى الخوارزم الناتج بـ "الأفضل المحلي" lbest. نوعية أداء كل جسيم أي مدى قرب الجسيم إلى موقع "الأمثل العام" هو قياس دالة اللياقة التي تختلف تبعاً لمشكلة التحسين المطروحة.

يقوم ت.س.ج على سرب يتكون من m فرد أوجسيم¹، كل منها يمثل حلاً ممكناً لمشكلة ذات N بُعد. النمط الوراثي له $2N$ من المعاملات، N الأول يمثل الإحداثيات المكانية لـ الجسيمات، والـ N الأخير مكونات سرعته في فضاء المشكلة ذات N بُعد. من وجهة النظر التطورية، يتحرك الجسيم بسرعة تتكيف داخل فضاء البحث ويحتفظ في ذاكرته بالموقع الأفضل الذي توصل إليه. يتم تغيير المعاملات عند الانتقال من كرة إلى أخرى على النحو التالي: السرعة $v_i(t+1)$ للجسيم i في اللحظة $t+1$ هي مزيج خطي لسرعته $v_i(t)$ في اللحظة t ، مع الفرق $b_i(t) - p_i(t)$ بين موقع أفضل حل تحصل عليه الجسيم في اللحظة t ، وموقعه الحالي، وكذلك مع الفرق $b_g(t) - p_i(t)$ بين أفضل موقع وُجد من طرف مجموع جسيمات السرب في أي وقت مضى والموقع الحالي للجسيم. لدينا:

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot U(0, 1) \otimes (b_i(t) - p_i(t)) + c_2 \cdot U(0, 1) \otimes (b_g(t) - p_i(t)) \quad (1)$$

حيث ترمز الأحرف الواضحة إلى متجهات، كما يدل الرمز $\otimes$ على عملية ضرب الاتجاهات نقطة بنقطة، $U(0, 1)$ هي دالة تقوم بإرجاع اتجاه عشوائي تأخذ مواقعه من توزيع موحد في $[0, 1]$ ، C_1 هو المعامل المعرفي، و C_2 المعامل الاجتماعي، w هو عامل القصور الذاتي الذي يأخذ قيمه في $[0, 1]$. ويجب أن تكون قيم السرعة ضمن نطاق v_{min} و v_{max} . يمكن القيام بتغيير للاحسن في الخوارزم الأصلي بتغيير قيمة w أثناء التنفيذ؛ بدءاً من w_{max} كقيمة قصوى، و إنقاصها خطياً عندما يزيد عدد التكرارات، حتى تصل إلى قيمة الحد الأدنى w_{min} :

$$w(t) = w_{max} - (w_{max} - w_{min}) \cdot t / T_{max} \quad (2)$$

¹ اخترنا المصطلح "جسيم" عوضاً عن "جزيئية" لقربه أكثر لمعنى "فرد" ينتمي إلى السرب.

في المعادلة السابقة، t و T_{max} هما عدد التكرارات الحالية، والحد الأقصى المسموح به، على التوالي. ثم يتم حساب موقع كل الجسيمات في الخطوة التالية عن طريق جمع الموقع الحالي وسرعته :

$$p_i(t+1) = p_i(t) + v_i(t+1) \quad (3)$$

وتتكرر هذه العمليات T_{max} مرة، أو حتى يتم التحقق من بعض معايير التوقف الأخرى. كمعيار نمذجي نجد تحقيق الخطأ الأدنى المرجو بالنسبة للحل الأمثل.
الخوارزم العام ملخص كالتالي :

البداية

لكل الجسيمات

إختيار عشوائي لموقع وسرعة الجسيمات

مادام (لم يتم التوصل إلى العدد الأقصى من التكرار) القيام بـ

لكل الجسيمات

حساب قيمة دالة اللياقة $\psi_i(t)$

إذا كان ($\psi_i(t)$) هي أفضل من أفضل قيمة لياقة $\psi(b_i(t))$ في التاريخ الجسيمات

إذا $\psi_i(t) = \psi(b_i(t))$ ، ونتخذ الجسيم الحالي كالأفضل الجديد $b_i(t)$

إذا لزم الأمر ، تحديث الجسيم الأفضل الشامل $b_g(t)$

حساب سرعة الجسيمات على أساس المعادلة (1) .

تحديث موقع الجسيمات على أساس المعادلة (3) .

تحديث عامل القصور الذاتي على أساس المعادلة (2) .

النهاية

في تطبيقات أكثر تطوراً يكون هنالك عدة طرق لتعديل آلية تحديث موقع الجسيمات. من بينها اخترنا إثنان لهذا العمل. أولهما قائم على آلية الحبس عن طريق الحد من تغير موقع الجسيم [3]. أي تعديل موقع الجسيمات على طول البعد k في $[1, N]$ بـ :

$$p_{i,k}(t+1) = \min(\max(p_{i,k}(t) + v_{i,k}(t+1), P_{\min}), P_{\max}) \quad (3a)$$

حيث $P_{\min}=0$ و $P_{\max}=1$. يمكن تعويض المعادلة (3) بالمعادلة (3a) في الخوارزم.

أما الآلية الثانية، وُصفت في [4] كطريقة فوضوية، و هي مستوحاة من ظاهرة تشتت الرياح الطبيعية. وتقوم على صياغة سرعة الرياح واتجاه الرياح كنموذج للتقلبات الجوية الممكن تأثيرها على تعديل موقع الجسيم كما المعادلة التالية :

$$vw(t+1) = vw(t) + v_{op} \cdot rand() + v_{su} \cdot rand() \quad (4)$$

حيث أن vw هي سرعة الرياح، v_{op} هو الاتجاه المعارض و يساوي -1 و v_{su} هو اتجاه الدعم و يساوي $+1$. آثار سرعة الرياح حالة من اثنين: يمكن تعارض حركة الجسيمات أو تدعيمها. تأثير المعارضة يبطل الجسيمات في الوصول إلى الحل "الأفضل الشامل"، في حين أن تأثير الدعم يزيد من سرعة الجسيمات في التوصل إلى الحل "الأفضل الشامل". يتم التحديث بمعادلة الرياح بشكل منفصل لكل الجسيمات. ومما يدعم ذلك حقيقة أن كون الجزينات منفصلة مكانيا من بعضها البعض، وبالتالي فهي تخضع لقوى ديناميكية مختلفة من الغلاف الجوي. عندما تكون قيم معارضة ودعم اتجاه الرياح، ولسرعتها على قدم المساواة فهي تُعتبر صياغة لجو ثابت. معادلة تحديث موقع على طول البعد k في $[1, N]$ ، هي كالتالي :

$$p_{i,k}(t+1) = p_{i,k}(t) + v_{i,k}(t+1) + vw_k(t+1) \quad (3b)$$

عند الجمع بين التشتيت و الحبس نتحصل على:

$$p_{i,k}(t+1) = \min(\max(p_{i,k}(t) + v_{i,k}(t+1) + vw_k(t+1), P_{\min}), P_{\max}) \quad (3c)$$

القيم الأولية من سرعة الرياح على طول اتجاه الرياح تلعب دورا هاما في تحديد التقارب الأخير من الجسيمات إلى الحل الأمثل. أيضا، هذا العامل يضمن البحث الأمثل في فضاء الحل.

3 التصنيف بخوارزم التحسين بسرب الجسيمات

منهاجية التصنيف المستخدمة في هذا العمل مستوحاة من [5] و [6]. والفكرة الرئيسية هي أن فئة التصنيف معرفة تماما بنقطتها الوسطى. عند إعطاء قاعدة البيانات مع C فئات و N معاملات الإدخال، يمكن أن يُنظر إلى مشكلة التصنيف على أنها العثور على أفضل قيمة إحداثيات النقاط الوسطى في الفضاء ذو N بعد أثناء مرحلة التدريب، ثم تصنيف متجه المُدخلات إلى أقرب نقطة وسطى.

نبدأ مع سرب من الجسيمات، كل واحد يمثل النقطة الوسطى المحتملة، واستخدامها في مرحلة التدريب للحصول على أفضل نقطة وسطى لكل فئة. خلال مرحلة القرار، يتم مقارنة النقطة الوسطى إلى نماذج الفئة عن طريق دالة اللياقة من أجل تأسيس دقة التعرف (أو بمعنى آخر نسبة أخطاء تصنيف)

لدينا x_j تمثل النموذج الأولي من الفئة j ، يتم تعريف دالة اللياقة ψ للجسيم i كمسافة أوكليد بين x_j والنقطة الوسطى الحالية المتمثلة في $p_i^{CL_{known}(x_j)}$. حيث:

$$\psi(i) = \frac{1}{D_{Train}} \sum_{j=1}^{D_{Train}} d(\bar{x}_j, \bar{p}_i^{CL_{known}(\bar{x}_j)}) \quad (5)$$

حيث D_{Train} ، هو عدد الحالات التي تتألف منها مجموعة التدريب. عند حساب المسافة، جميع المكونات في الفضاء ذي N بعد تُطبع حسب قيمها القصوى، ومجموع مكونات المسافة يُقسم على N ، بهذه الطريقة تضمن أن كل المسافات تنتمي إلى المجال $[0.0, 1.0]$ ، وكذلك ψ . ونظرا لدالة اللياقة المختارة، تصبح مشكلة التصنيف مشكلة إيجاد الحد الأدنى. أثناء مرحلة التدريب، كلما كانت

يتم حساب أداء التجربة كنسبة أخطاء التصنيف للحالات في مجموعة الاختبارات من قبل أفضل جسيم لكل فئة معينة.

4 مجموعة البيانات LET و الأعمال السابقة

LET عبارة عن قاعدة بيانات للحروف الكبيرة من الأبجدية اللاتينية، تحتوي على عدد من الخصائص الإحصائية لصورهم. وقد حصلنا على هذه البيانات من مستودع البيانات UCI Repository [7] مانحها الأصلي كان دي. جي. سلايتس. حيث استخدمها في مجال المصنفات الجينية من طراز هولند [8].

LET مجموعة من 20000 صور فريدة التي تم إنشاؤها بواسطة الصور المشوهة بشكل عشوائي من الأحرف الكبيرة الـ 26 المخطوطة بـ 20 من الخطوط التجارية المختلفة. وتمثل الخطوط الأصلية مجموعة كاملة من طبوع الحرف، بما في ذلك النصي، الشريف، المائل، والقوطي. قد تم تلخيص المعالم المتحصل عليها من الأحرف 20000 في 16 سمات عددية بدائية. قاعدة البيانات لديها الخصائص التالية :

- عدد من الحالات للتدريب = 16000
- عدد أبعاد الفضاء = 16.
- عدد الفئات = 26.
- عدد الحالات في مرحلة الفحص = 4000.

في [8]، قام الباحثون بدراسة نظام يستخرج قواعد التصنيف العام ضمن نموذج الإشراف على التعلم. يصفون عددا من نظم المصنفات التطويرية التي تخلق قوائم ثابتة الطول من القواعد شرط-عمل (أي قواعد تصنيف) التي يتم تطبيقها بالتوازي مع "رسائل" تمثل وجود أو عدم وجود ميزات معينة في البيئة الحالية. تمتلك هذه النظم ثلاثة عناصر رئيسية مميزة هي

- خوارزم يقارن أداء القواعد مع رسائل لتحديد القاعدة التي ينبغي تفعيلها ،
- خوارزم تعزيز قوة كل قاعدة، على أساس "اللياقة" في البيئة الحالية،
- و خوارزم إنشاء قواعد الذي يعمم النماذج أو يجمع بين القواعد الحالية لإنتاج أخرى جديدة.

من نتائج هذا البحث القدرة على تعلم الفئات الصحيح للحروف المرتبطة بالمتجهات المتكونة من 16 سمة مستخرجة من صور الحروف. أفضل دقة حُصل عليها ما يزيد قليلا عن 80٪. في عمل أكثر حداثة [9]، وُضع خوارزم جديد لحساب حدود خطية من نقطة مجموعات محدودة. هذا الخوارزم يتكون من مرحلتين أساسيتين: أولا ، يقوم بحساب هيبار مكعب تقريبي لكل فئة. كل المكعبات تعتبر مناطق غير محددة، حيث تختلط نقاط البيانات من مختلف الطبقات. تسمى نقاط البيانات التي تنتمي إلى مكعب واحد فقط بالمصنفة، ويتم إزالتها من باقي التحليل. في معظم الحالات، استخدام الهيبار مكعب يسمح بخفض كبير في عدد نقاط البيانات، وبالتالي، في الجهد

الحاسوبي. يدعى الخوارزم بـ Maxmin. كما قام الكُتاب بتجارب إضافية استخدموا فيها عددا من المصنفات المتاحة في WEKA suite (Waikato Environment for Knowledge Analysis) للمقارنة. و هي مجموعة من أدوات تعلم الآلة مكتوب بلغة جافا، وضعت في جامعة وايكاتو بنيوزيلندا.

الجدول 1 دراسة مقارنة عن مجموعة البيانات LET (مقتبس من [9])

الطريقة	المعنى أو المصدر	دقة التعرف
NB	Naive Bayes (with kernel) ساذج بايز (مع النواة)	% 74.12
IBk	Instance-based algorithm IBk (with k=5) الخوارزم المستند على الأمثلة	% 94.96
Logistic	Logistic Regression based classifier المصنف المستند على الانحدار	% 77.40
MLP	Multi-Layer Perceptron الشبكة العصبية الاصطناعية	% 83.20
LIBSVM	support vector machine classifiers Linear التصنيف الخطي بآلة المتجهات	% 82.40
J48	decision tree classifier J48 مصنف شجرة القرارات	% 87.70
PART	rule based classifier PART المصنف المستند على القواعد	% 87.32
Poly	a classifier based on polyhedral separability	% 88.68

	[10]. المصنف المستند على انفصال	
Maxmin	[9]	% 91.82

من أجل اختبار هذه الخوارزميات، اختاروا ممثل عن كل نوع من المصنفات : ساذج بايز (مع النواة)، الخوارزم المستند على الأمثلة، المصنف المستند على الانحدار اللوجستي ، الشبكة العصبية الاصطناعية Perceptron متعدد الطبقات، التصنيف الخطي بآلة المتجهات وJ48 مصنف شجرة القرارات (وهو بديل لـ C4.5) و المصنف المستند على القواعد بالإضافة إلى ذلك، وقاموا أيضا باختبار المصنف المستند على انفصال متعددة السطوح [10]. . أنتج مصنف الخوارزم المستند على الأمثلة أفضل دقة.

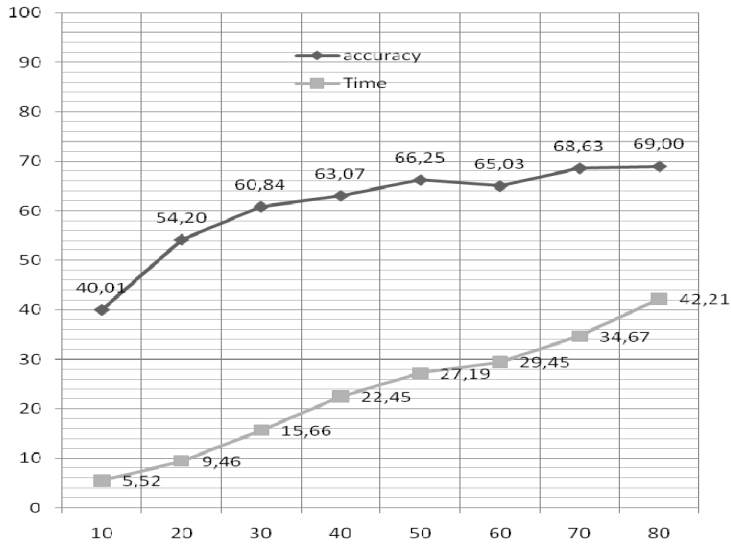
5 النتائج والمناقشة

تم الحصول على النتائج التي توصلنا إليها مع ثلاثة مناهج :

- تصنيف (ت.س. ج.) الخوارزم الكلاسيكي باستخدام المعادلة (3) لتحديث موقع الجسيم
يشار إليه بـ أ
- تصنيف (ت.س. ج.) على أساس آلية الحبس، وذلك باستخدام المعادلة (3A) للتحديث
المشار إليها باسم ب.باء.
- تصنيف (ت.س. ج.) باستخدام المعادلة (3C) لتحديث مواقع المشار إليها باسم جيم.

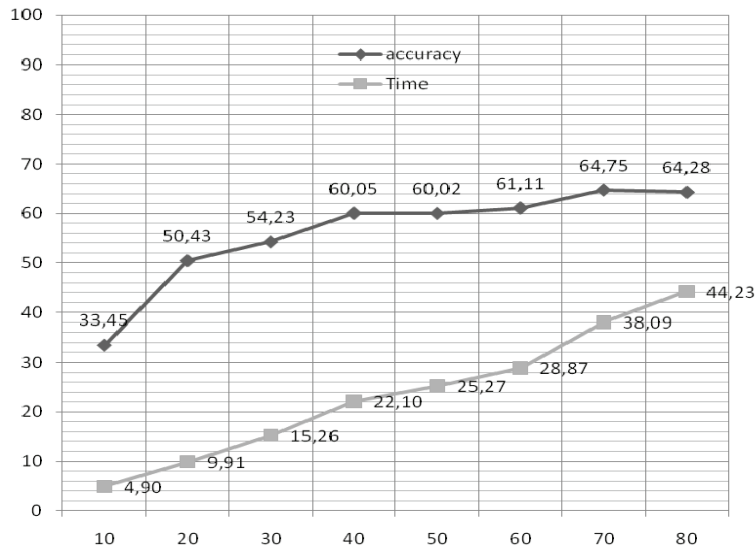
جميع النتائج المعروضة في هذا القسم تمثل متوسط 10 تجارب على معالج من نوع AMD AthlonTM 64 x 2 ثنائي النواة بسرعة 3800 + 2,00 غيغاهرتز ، مع 2,90 غيغابايت كذاكرة العشوائي .

نستخدم وقت التدريب لتقييم تعقيد الخوارزم .وينبغي أن نذكر أيضا أنه يستند هذا الإجراء على عتبة تسمح لإتخاذ القرار مرتبطة بالميدان .العتبة هي القيمة التي يتم بموجبها النظر في قيمة دالة اللياقة للاعتراف أم لا. لم يكن لدينا توجيهات للاختبار، حاولنا قيم مختلفة قبل ان نستقر على 0.005 ونتيجة لذلك، جميع النتائج التي تحصلنا عليها صحيحة بهامش خطأ قيمته 0.005 .
تم تعيين المعاملات بالقيم التالية $T_{max}=1000$ ، $v_{max}=0.5$ ، $v_{min}=-0.5$ ، $c_1=2.0$ ، $w_{min}=0.4$ ، $w_{max}=0.9$ ، $c_2=2.0$ و عدد متنوع من الجسيمات. تطبيق الطريقة إلى دقة التعرف من 69 ٪ لعدد الجسيمات 80 في فترة من الوقت التعلم 42.21 دقيقة (انظر الشكل 1.1أ).



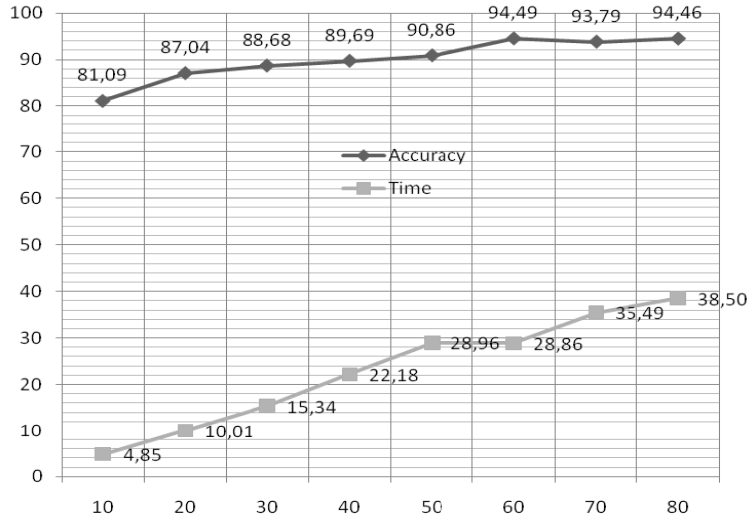
الشكل 1. أ. الدقة (%) ووقت التعلم (دقائق) للخوارزم الكلاسيكي دون الحبس

عند تغيير آلية تحديث الجسيمات مع الطريقة باء (انظر الشكل 1.ب)، تم الحصول على أسوأ أداء مع دقة 64.27% لنفس المعطيات، مع وجود تكلفة حساب 44.22 دقيقة. وهذا يعني أن الحبس لا يؤدي بشكل جيد دوماً.



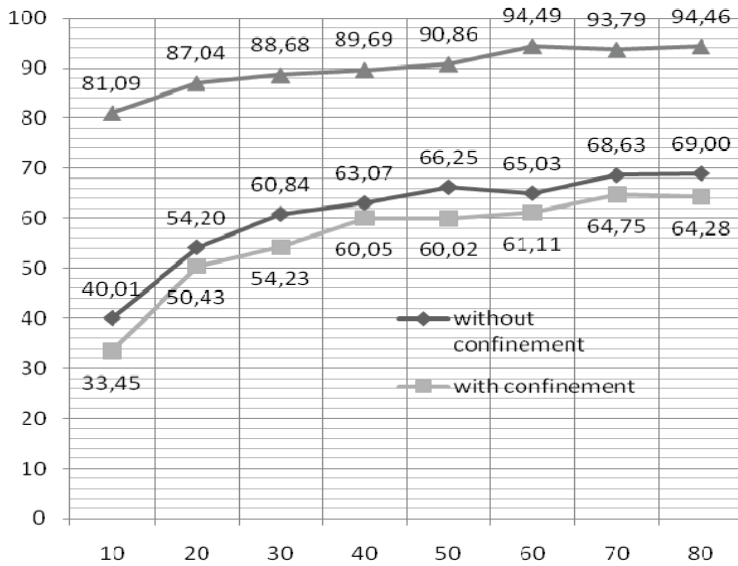
الشكل 1. ب. الدقة و الوقت للخوارزم بآلية الحبس

ثم طبقنا الطريقة ت (انظر الشكل 1.ت)، وتحصلنا على أفضل النتائج ، سواء من حيث الاعتراف أو الوقت 94.49 ٪ و 28.86 دقيقة فقط بـ 60 جسيم.



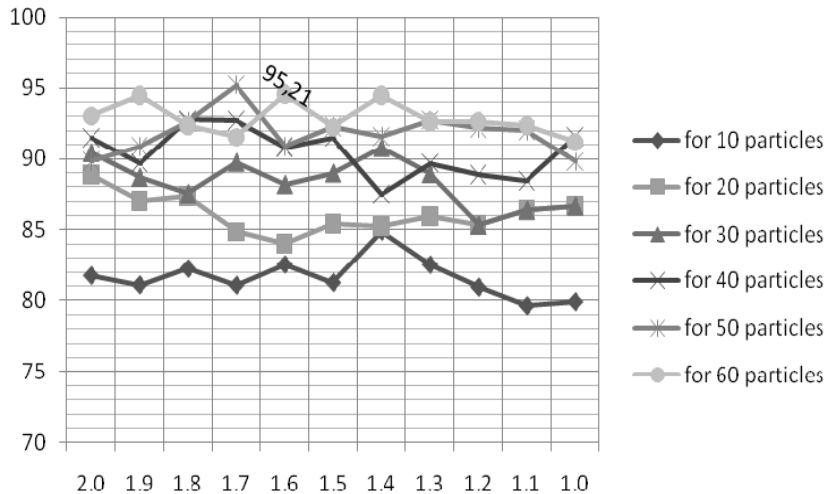
الشكل 1.ت. الدقة و الوقت للخوارزم بالحبس و التشتيت

مقارنة للطرق ألف وباء الطريقة جيم كان لها الأداء الأفضل و لو أنها تستغرق وقتاً أكثر قليلاً من غيرها عند وجود نفس حجم السرب (انظر الشكل 1.ج).



الشكل 1. ج. الدقة و الوقت للخوارزميات الثلاث

كما درسنا تأثير C_1 و C_2 ، العوامل المعرفية والاجتماعية، على معدل الدقة (انظر الشكل 2). وقد تباينت تلك القيم من 2.0 حتي 1.0. بقيمة 1.0 فإن أجزاء المعادلة (C_3) المعرفية والاجتماعية. والجمود وتشتت الرياح لها نفس الوزن.



الشكل 2. مجموعة النتائج بالتنشيت و الحبس لأحجام سرب مختلفة وعوامل اجتماعية ومعرفية مختلفة.

القيمة المثلى من حيث دقة التعرف كانت $C_1 = C_2 = 1.7$ تم الحصول على أفضل أداء لسرب مكون من 50 جسيم مع دقة = 95.20 % ووقت للتعلم 23.40 دقيقة.

6 الخاتمة

بالنظر إلى النتائج التي تحققت في قاعدة البيانات ، يمكن الاستنتاج بأن خوارزم التحسين بسرب الجسيمات هو أداة تصنيف مناسبة حتى لبيانات عالية الأبعاد متشكلة من عدد كبير من الحالات وفئات متعددة.

وعلاوة على ذلك، الخوارزم المقترح كان أفضل من البدائل التي ذكرت في المنشورات العلمية عند استخدام ألبتي الحبس و التشبث معاً، عن سرب يتكون من 50 جسيم بدقة من 95.20% ضد 94.96 % ، التي حصل عليها الخوارزم القائم على الأمثلة من أداة وىكا (انظر الجدول 1). وثمة ملاحظة أخرى مهمة هي أن آلية الحبس لا تضمن دقة جيدة دائماً، إذ يتم استخدامها لحصر الجسيمات في فضاء البحث بدلاً من تركها تنتشر مع خطر تجاوز الحدود، تشتت الرياح مع الحبس يؤدي إلى نتائج أفضل كما يتضح من النتائج إذ أنه يسمح باستكشاف الفضاء جيداً مع ضمان عدم تجاوز الحدود.

التحسين بسرب الجسيمات نهج برهن كفاءته عند التعامل مع مهمة التصنيف. النتائج المتحصل عليها تشير إلى أن التحسين بسرب الجسيمات هو مرشح مناسب وتنافسي للتصنيف ويمكن تطبيقه بنجاح في أكثر المجالات. ومع ذلك، مسألة مثيرة للاهتمام: "كيف سنتعامل مع البيانات الاسمية؟". وعلاوة على ذلك، عند التعامل مع مشكلة حقيقية يمكن لوصف البيانات أن يكون مزيجاً من سمات مستمر، منفصلة أو اسمية. إلى أي مدى التحسين بسرب الجسيمات يمكنه تصنيف هذه المجالات؟ الإجابة عن هذا السؤال سوف تكون التحدي القادم.

7 شكر وتقدير

الشكر يعود لمجلس البحوث في العلوم الطبيعية والهندسة بكندا (NSERC) الداعم المالي لهذا العمل.

المراجع

- [1]: Poli, R., Kennedy, J., Blackwell, T. Particle swarm optimization An overview. Journal of Swarm Intelligence, Ed. Springer New York, 2007.
- [2]: Kennedy, J., Eberhart, R. Swarm Intelligence. Ed. Morgan Kaufmann, 2001.
- [3]: Engelbrecht, A., P. Computational Intelligence An Introduction. John Wiley & Sons Editions, (2007).
- [4]: Chandramouli, K. and Izquierdo, E., "Image Classification using Chaotic Particle Swarm Optimization," in Proc. International Conference on Image Processing (ICIP '06), (2006).
- [5]: A., Abraham, Grosan, C. Swarm Intelligence in Data Mining. Ramos, V., (Eds.), Springer-Verlag Berlin Heidelberg 2006.

- [6] De Falco, I., Della Cioppa, A., Tarantino, E. Evaluation of Particle Swarm Optimization Effectiveness in Classification. LNAI3849, pp: 164-171, Springer-Verlag Berlin Heidelberg 2006.
- [7] Asuncion, A. & Newman, D.J., UCI Machine Learning Repository [http://www.ics.uci.edu/~mllearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, (2007).
- [8] P. W. Frey and D. J. Slate, "Letter Recognition Using Holland-style Adaptive Classifiers", in Machine Learning Vol 6 #2 March (1991).
- [9]: A. M. Bagirov, J. Ugon and D. Webb, "An Incremental Approach For The Construction Of A Piecewise Linear Classifier", In proceedings of The XIII International Conference "Applied Stochastic Models and Data Analysis" (ASMDA-2009) June 30-July 3, 2009, Vilnius, LITHUANIA, pp:507-511, (2009)
- [10]: Astorino, A. and Gaudioso, M. 2002. Polyhedral separability through successive LP, Journal of Optimization Theory and Applications, 112(2), 265–293.

المخلص باللغة الانجليزية

Title: A Particle Swarm Optimization Approach for Letter Recognition

Abstract. This work studies the use of Particle Swarm Optimization (PSO) for letter classification. The major idea is to retrieve optimal particle positions corresponding to the centroids of the classes. An experimental set up was realized and tested on the LET database, leading to recognition accuracies that were better than those obtained with well known classifiers.

Keywords: Particle Swarm Optimization, classification, letter recognition, confinement, wind dispersion.

المصطلحات

التحسين بمسرب الجسيمات	Particle Swarm Optimization
التصنيف	Classification
التعرف على الحروف	letter recognition
حبس	Confinement
تشتيت الرياح	wind dispersion

نمذجة مانت باستخدام الذكاء الصناعي

نغم حفطي سعيد، د. ميسم عبود، و أ. د. حامد الرويشدي
جامعه برونييل، مدرسه الهندسه والتصاميم،

Wireless Networks and Communications Centre
Brunel University, Kingston Lane, Uxbridge, Middlesex, UB8 3PH
Tel: +44 (0)1895 274000, Fax: +44 (0)1895 232806

البريد الالكتروني: (nagham.saeed, masysam.abbod, hamed.al-raweshidy)
@brunel.ac.uk.

الخلاصة. في مجال الاتصالات اللاسلكية ، شبكة الاتصالات العفوية (مانت) تتكون عند ربط مجموعة من العقد (الأجهزة في شبكه الاتصالات اللاسلكية يطلق عليها تسميه العقد) . هذه العقد لها القابليه على تشكيل شبكه اتصالات بدون مساعده الانترنت وبدون تخطيط مسبق. تتميز العقد بقابليتها على الحركه بشكل مستمر مع دوام التواصل مع الشبكه. الاتصال داخل الشبكه يتم باستخدام بروتوكول لاكتشاف المسار بين العقد .يعتبر تحدد المسار هو القضية الأساسية لشبكة الاتصالات العفويه الخاصه (مانت) لذلك العديد منه الباحثين ركزوا على ايجاد واقتراح المسار الامثل لبروتوكولات مانت ونتج هذا عن عده بروتوكولات. يستخدم هذا البحث الذكاء الصناعي (الشبكة العصبية الاصطناعية والشبكة العصبية الضبابية) لتكوين نماذج لاداء شبكه الاتصالات (مانت) باستخدام ثلاث بروتوكولات للمسار. البروتوكولات المستخدمه معروفه على نطاق واسع و هي Ad hoc on-demand Distance Vector و Dynamic Source Routing (DSR) و (AODV) Optimized Link State Routing (OLSR) هذا البحث يقارن نماذج الذكاء الصناعي مع نماذج المعادلات الرياضية التقليدية حيث تظهر النتائج النظرية ان نماذج الذكاء الاصطناعي أكثر دقة وتمثيل لشبكه مانت من المعادلات الرياضية.

الكلمات الجوهرية مانت(MANET) ، تكوين نماذج، الشبكة العصبية الاصطناعية، الشبكة العصبية الضبابية.

١ - المقدمة

الهدف الأساسي من أي بروتوكول لمانت (MANET) هو إنشاء الطريق الكفوء بين اي زوج من العقد القادر على تسليم الرسائل في الوقت المحدد. ينبغي أن يتم بناء الطريق مع الحد الأدنى من الحمل (overhead) وعرض النطاق الترددي (bandwidth consumption) [1]. أداء الشبكة يتأثر بشكل كبير بظروف الشبكة مثلاً: طبيعة البيانات المنقولة وعدد العقد وسرعة العقد، وما إلى ذلك [2]. بعبارة أخرى خلال فتره عمل شبكه مانت، قد تتأثر كفاءة الشبكة وقد يتعرض اداءها للانخفاض وفقاً لتغيرات في الظروف المحيطه بها. كذلك يتفاوت اداء الشبكة نسبته لبروتوكول المختار. لذلك بناء نموذج يمثل اداء الشبكة عند استخدام بروتوكول معين يسهل

عملية التخمين والتنبؤ بالبروتوكول الافضل لظروف معينه. خلال العقدين الماضيين دخلت تطبيقات الأنظمة ذكية في تخصصات عديدة وواسعه، ولكنها لم توظف بشكل واسع لحل مشاكل شبكات الاتصالات اللاسلكية مانت. يستخدم الذكاء الصناعي لتصميم برامج الحاسوب التي تتصرف بذكاء والباحثون بهذا المجال طورو النظريات الحسابية التي تبدأ مع تعريف واضحة المعالم والأوصاف للأهداف الحسابية والتي توفر المعلومات المعالجة للسلوكيات المستهدفة من حيث المدخلات والمخرجات، والخوارزميات [3].

يقدم هذا البحث نماذج لاداء شبكه اتصالات لاسلكية من خلال نمذجة أداء مانت لثلاثة بروتوكولات توجيه تبعاً للظروف المحيطة. تم تنظيم الباقي من هذا البحث على النحو التالي : الفصل ٢ يعرض ملامح رئيسية مختصره لبروتوكولات التوجيه المستخدمة في هذا البحث. الفصل ٣ يصف عملية تمثيل مانت ويحدد معايير أدائها. الفصل ٤ يلخص التكيف للشبكة العصبية الاصطناعية والشبكة العصبية الضبابية لتكوين نماذج مانت. يحتوي الفصل ٥ مقارنة بين نماذج مانت الناتجة عن المعادلات التجريبية والشبكة العصبية الاصطناعية والشبكة العصبية الضبابية. وأخيراً، الفصل ٦ يسلط الضوء على العمل الذي أنجز في هذا البحث مع افكار لتطوير البحث للمستقبل .

٢- بروتوكولات التوجيه

عدد كبير من البروتوكولات التوجيه الخاصة قد درست في السابق وأدائها تم مقارنته [4]. هذا البحث اختار ثلاثة من البروتوكولات المعروفة للاختبار. أجري التمثيل للشبكة وجمعت جميع البيانات الخاصة باداء الشبكة. البروتوكولات الثلاثة هي التالية:

DSR هو بروتوكول توجيه عند الطلب، ميزته هو ان المصدر (المرسل) هو الوجهه. أي المرسل يعرف المسار كاملاً قفزه بقفزه حتى المستلم النهائي. يتم تخزين هذه المسار في ذاكرة التخزين الخاصة بالمرسل. حزمه البيانات تحمل عنوان المسار في حقل خاص يسمى رأس الحزمة [5]

AODV هو بروتوكول توجيه عند الطلب، ميزته هو انه أساساً مكون من دمج DSR مع DSDV. حيث اقترح آلية اكتشاف المسار وصيانتها من DSR، واستخدام قفزه بقفزه في توجيه المسار والرقم التسلسلي من Destination Sequenced Distance Vector (DSDV) [6].

OLSR هو بروتوكول توجه المسار فيه على أساس المعرفة من اول نقطة إلى اخر نقطة، حيث يعتمد على الخوارزمية التقليدية link state. الحادثة في هذا بروتوكول هو أنه يقلل من عدد رسالات التحكم وعدد العقد المرسله باستخدام الاستراتيجي. [7] (MPR)

٣- تمثيل مانت

استندت النماذج على Opnet™ 11.5 modeler لتنفيذ بروتوكولات التوجيه الخاصة بمانت . مثلت الشبكة لمدة ساعة واحدة لكل بروتوكول ولكل حجم الشبكة، وقد كان معدل إرسال البيانات ١ ميغابت في الثانية. وكانت الحزم المرسله من مصدر واحد فقط إلى مستلم واحد فقط. كذلك ٤ مستويات سرعه قد طبقت: الثابتة (٠ م / ث) ، والمنخفضه (١ م / ث) والمتوسطة (١٠ م / ث)

والمرتفعه (٢٠ م / ث). سيناريوهات المحاكاة تأخذ بنظر الاعتبار ان المستخدمين يتحركون في اتجاهات مختلفة وبشكل عشوائي داخل المنطقة المحددة مسبقا. مناطق التغطية في السيناريو الأول ستكون منطقة مربعه (٥٠٠ م × ٥٠٠ م) ، ثم تزداد في كل مرحلة ٥٠٠ م حتى تصل مساحه منطقه التغطية في اخر سيناريو الى (٢ كم × ٢ كم) لكل بروتوكول توجيه مقيم، تم تنفيذ العديد من السيناريوات مع تغير حجم الشبكة، شبكات يتراوح حجمها من الصغيرة الى الكبيرة. خلال هذا العمل، وكانت أحجام شبكة ٥ ، ١٠ ، ١٦ ، و ٢٥ عقده على التوالي . ومقاييس الأداء المعتمدة للتحليل والمقارنه هي كالتالي :

- **التأخير (Delay):** متوسط الفاصل الزمني بالثانيه لتأخير وصول الحزمه لكامل الشبكة، أي الوقت الذي انقضى بين إرسال الحزمة من المصدر حتى استلامها .
- **معدل اقبال التوجيه (PDR):** النسبة بين متوسط كمية استلام الحزم بالثانيه في الشبكة بالكامل ومتوسط كمية ارسال الحزم بالثانيه في الشبكة بالكامل.
- **الحمل (Load):** كمية استلام الحزم (بت) في الشبكة بالكامل.

٤- النمذجة باستخدام الذكاء الاصطناعي

يهدف العمل في الحوسبة المرنة على الاستفادة من تقنيات أنظمة ذكية وتطبيقها، مثلا النمذجة بواسطه الشبكة العصبية الاصطناعية يمكن الاستفادة منها من خلال دمجها مع الخوارزميات التطورية لتكوين نظام واحد [8]. وهذا النظام ينبغي له أن يكون ذكي أي لديه القدرة على التعلم من التجربة حتى يتمكن من تحقيق اهدافه [9]. النمذجة تلعب دور مهم جدا في تصميم النظام الصناعي لأن النمذجة هي السعي للكشف عن ما يحدث فعلا في النظام الطبيعي. فالنموذج ليس فقط يمثل بعض ملامح النظام الطبيعي ولكن من المفترض ان صياغته يجب أن تكون متناسقة مع ما هو معروف عن النظام الطبيعي [10]. تم استخدام MATLABTM لنمذجة الشبكة العصبية الاصطناعية والشبكة العصبية الضبابية الاصطناعية. حيث يتم تغذية المعلومات من التمثيل السابق لمant المدخل هو ظروف الشبكة ممثلا بعدد العقد وسرعه التنقل والنتاج هو أداء الشبكة ممثلا بالتأخير ومعدل اقبال التوجيه والحمل.

١- النمذجة باستخدام الشبكة العصبية الضبابية

الشبكة المتعلمه المتكيفية (ANFIS) اسست على نظام الاستدلال الضبابي [11] . حيث يمكن ان تتعلم ANFIS مخطط المدخلات والمخرجات على أساس المعرفة البشرية في شكل القواعد الضبابية. (if-then) مستوحاة من نهج النمذجة لتاكاجي- سيكينو [12]. النماذج الثلاثية الابعاد المبينه بالرسوم الفرعيه (h, i and g) للاشكال 1, 2, and 3 Figure هو مبين في تقديم نماذج مانت باستخدام الشبكة العصبية الضبابية التي تم تدريبهم لتنتج نفس نتائج شبكة مانت. واستند النموذج على طبولوجيا نوع TSK ، إدخال ٢ متغير من نوع داله غاوسين، القواعد ٩ ، والتمرين تكرر ل ٣٠ مره. لكل بروتوكول توجيه تم تدريب ثلاث نماذج، كل نموذج يمثل واحدة من المخرجات الثلاث. وفي المجموع ، تم تطوير ٩ نماذج للبروتوكولات الثلاثة.

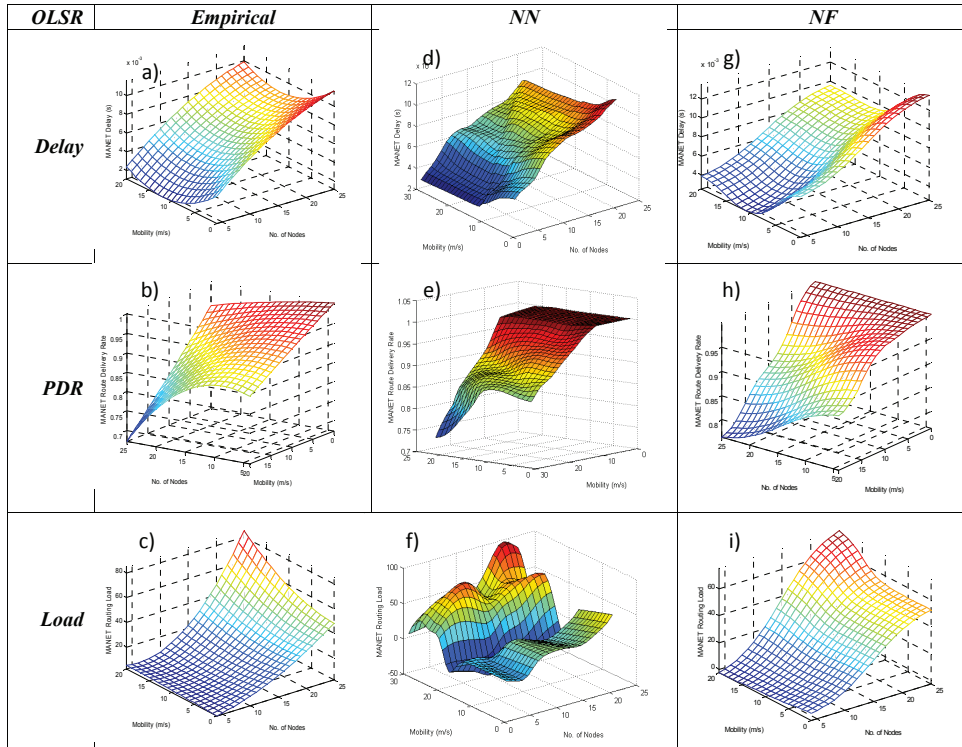


Figure 1. Modeling MANET performance for OLSR routing protocol

ب- النمذجة باستخدام الشبكة العصبية الاصطناعية

الشبكة العصبية الاصطناعية هي عبارة عن شبكة مكونة على غرار هيكل وسلوك الخلايا العصبية في الدماغ البشري، هي مكونة من خوارزميات حاسوبية قادرة على التكيف بالشكل المطلوب تبعاً للمعرفة المتكونة لها من خلال التجربة [13]. يمكن تدريب الشبكة للتعرف على الأنماط المعقدة وتصنيفها. تم تدريب الشبكة العصبية الاصطناعية والتي تمثل سلوك مانت وإدائها المدخلات كانت قاعده البيانات المستنتجة سابقاً من تمثيل مانت. لكل بروتوكول التوجيه، تم تدريب ثلاث نماذج للشبكة العصبية الاصطناعية حيث أن كل نموذج يمثل كل واحد من مخرجات الشبكة الثلاث. وفي المجموع، تم تطوير ٩ نماذج للبروتوكولات الثلاثة التي استخدمت يمكن الاستفادة منها في تحسين شبكة مانت. الشبكة العصبية الاصطناعية كانت شبكة الانتشار مكونة من طبقتين، الطبقة الأولى مكونة من ٤ خلية عصبية والثانية مكونة من خلية عصبية ٨ لمدة ٥٠ دوره.

من المشاكل المعروفة في الشبكة العصبية هي مشكله افراط في التدريب التي تتضمن 'ضوضاء' في عينة التولد. أحد الحلول لهذه المشكله هو تدريب مجموعه من الشبكات العصبية على نفس مجموعة البيانات ثم ايجاد المعدل لها. ١٠ نماذج تم إنشاؤها من ١٠ جلسات. حيث في كل دورة

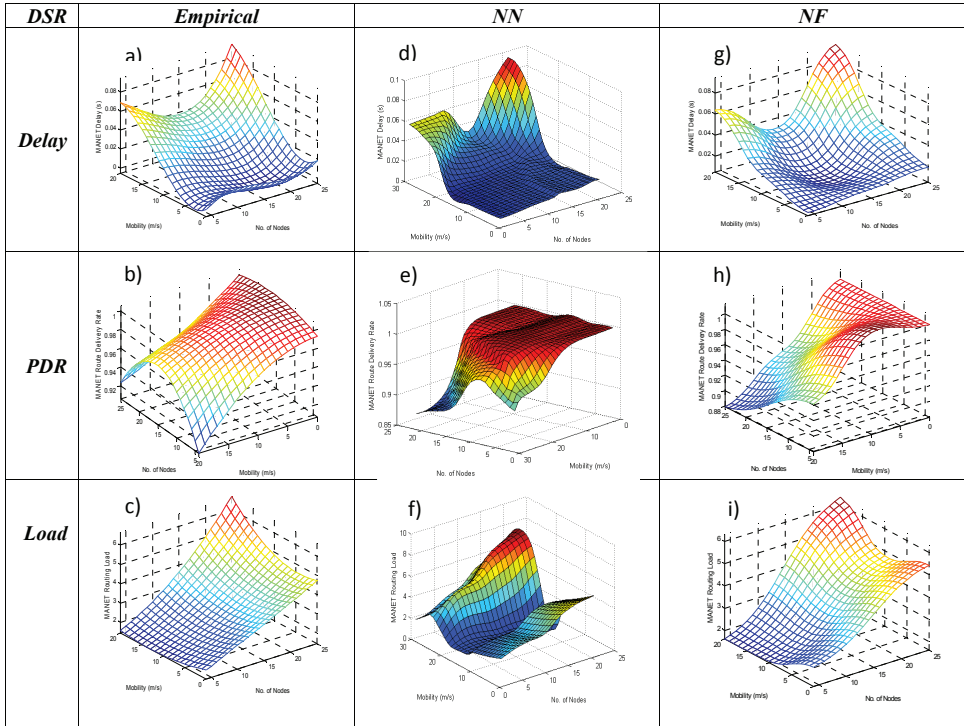


Figure 2. Modeling MANET performance for DSR routing protocol

تدريبية سوف ينتج نمودجا مختلف من هذه النماذج. النموذج النهائي يكون المعدل للنماذج العشرة [14] كما هو واضح في واعطت النماذج في الرسوم الفرعية (d, e and f) لاشكال 1, 2, and 3 والتي تمثل سلوك مانت وادائها.

٥- مقارنة بين النماذج الاصطناعية ونماذج المعادلات التجريبية

للحصول على نماذج للمعادلات تجريبية تم استخدام برامج Essential Regression software، حيث تم الحصول على Empirical equations. عند ادخل قاعده البيانات تم ترجمتها الى المعادلات ثم الى نماذج كما في الرسوم الفرعية (a, b, and c) لاشكال 1, 2, and 3، ويبين Table 1 ثلاث معادلات لكل بروتوكول توجيه التي تمثل أداء الإخراج حيث N هو عدد من العقد، و M هو سرعه التنقل، حيث ان كل معادله في هذا الجدول قد مثلت في الاشكال السابق ذكروها .

عند مقارنة النماذج مع بعضها، نجد ان نماذج الذكاء الصناعي اكثر تمثيلا لمانت واكثر دقة من نماذج المعادلات التجريبية، وخاصة نماذج الشبكة العصبية الاصطناعية الضبابية التي كانت أكثر سلاسة وأكثر واقعية من الشبكة العصبية الاصطناعية.

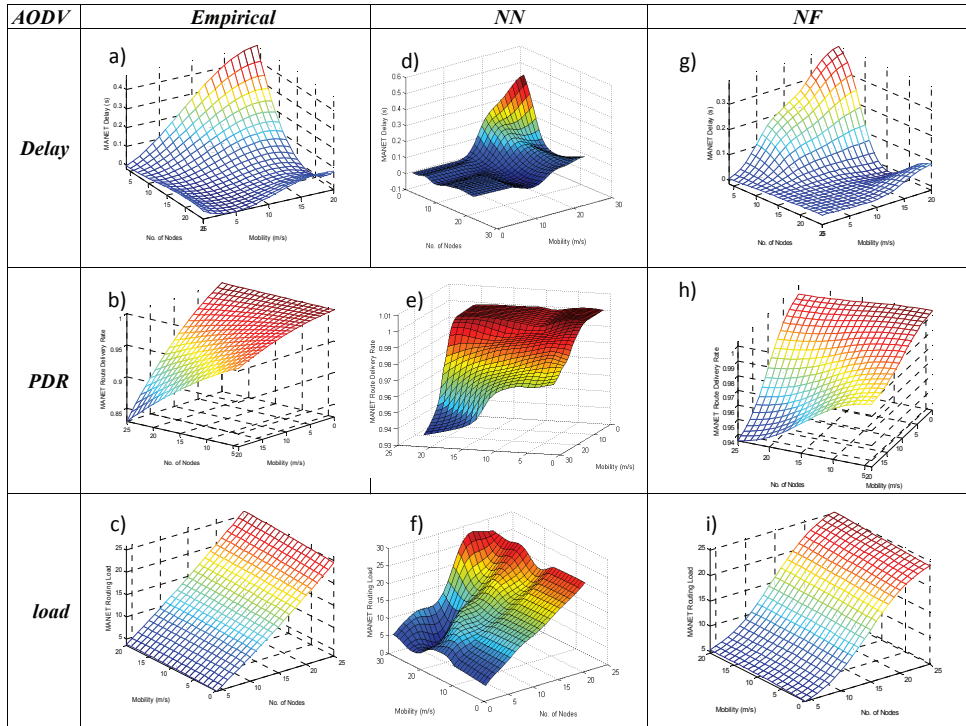


Figure 3. Modeling MANET performance for AODV routing protocol

في حالات معينة واحدة من بروتوكولات التوجيه تعطي أداء أفضل من مثيلاتها (البروتوكولات الأخرى) ولكن عندما تغير ظروف الشبكة أداء هذا البروتوكول سوف يتدهور في حين بروتوكول آخر ممكن ان يعطي أداء أفضل من سابقة. لذا لا يوجد بروتوكول واحد ممكن ان يقال عنه انه الأمثل لجميع التغيرات الحاصلة في شبكة مانت. لذلك النماذج الناتجة ممكن الاعتماد عليها لاختيار بروتوكول التوجيه الافضل المناسب لظروف معينة.

٦- الاستنتاجات

في هذا البحث تم نمذجة شبكة مانت باستخدام ثلاث بروتوكولات (OLSR, DSR, AODV) التمثيل تم لمدى ساعة واحدة باستخدام سيناريوهات متعددة لشبكات اتصال متعددة الأحجام (الصغيرة والمتوسطة والكبيرة) وذات سرعة متعددة (الثابتة، المنخفضة والعالية). من التمثيل

الشبكة جمعت ثلاث عوامل هامة تخص أداء الشبكة. ثم بدء توليد نماذج الذكاء الصناعي بالاعتماد على بيانات تمثيل الشبكة الاتصال. كذلك في هذا البحث تم مقارنه نماذج الذكاء الصناعي مع نماذج المعادلات التقليدية وتبين ان نماذج الشبكة العصبية الاصطناعية هي الأكثر دقة في تمثيل شبكه مانت لذلك ممكن مشاركتها في المستقبل في عملية صنع القرار واختيار بروتوكول التوجيه الأمثل حيث ممكن استخدامها مع الخوارزمية الوراثية (GA) أو سرب الجسيمات الأمثل (PSO) للمساعدة في تحديد البروتوكول الأمثل.

Table 1. Empirical Equation models for AODV, DSR and OLSR parameters

AODV	$\begin{aligned} Delay &= 0.00514 - 0.0037 N + 0.03298 M + 0.000556 N^2 - 0.00511 NM + \\ &0.00197 M^2 - 1.5 E - 05 N^3 + 0.00012 N^2 M + 3.07 E - 05 NM^2 - 8.1 E - 05 M^3. \\ PDR &= 0.999 + 8.72 E - 05 N - 0.000381 M + 1.32 E - 07 N^2 - 6.7 E - 05 NM \\ &- 5.5 E - 05 M^2 - 4.5 E - 06 N^2 M + 5.27 E - 06 NM^2. \\ load &= 0.188 + 0.825 N + 0.02728 M + 0.00599 N^2 + 0.00276 NM - 0.00122 M^2. \end{aligned}$
DSR	$\begin{aligned} Delay &= -0.02527 + 0.00942 N - 0.00199 M - 0.000704 N^2 - 0.00064 NM + \\ &0.000865 M^2 + 1.59 E - 05 N^3 + 2.06 E - 05 N^2 M + 3.24 E - 06 NM^2 - 2.5 E - 05 M^3. \\ PDR &= 0.983 + 0.00275 N - 0.00147 M - 8.3 E - 05 N^2 + 0.000369 NM - 0.000251 M^2 \\ &- 2.06 E - 05 N^2 M + 1.26 E - 05 NM^2. \\ load &= 1.418 + 0.124 N - 0.04696 M + 0.000308 N^2 - 0.00765 NM + 0.00172 M^2 \\ &+ 0.000266 N^2 M + 0.000272 NM^2. \end{aligned}$
OLSR	$\begin{aligned} Delay &= 0.00171 + 0.000631 N - 0.000494 M - 9.4 E - 06 N^2 + 2.05 E - 05 M^2. \\ PDR &= 1.01 - 0.00156 N - 0.00929 M - 5.13 E - 05 N^2 + 0.000593 NM \\ &- 4 E - 05 M^2 - 3.2 E - 05 N^2 M + 3.83 E - 06 NM^2. \\ load &= -0.463 + 0.433 N + 1.088 M + 0.0605 N^2 - 0.247 NM - 0.02128 M^2 \\ &+ 0.00835 N^2 M + 0.0043 NM^2. \end{aligned}$

٧- المصادر والمراجع

- [1] E. M. Royer and S. B. Chai-Keong Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks", *IEEE Personal Communications*, April 1999, pp 46-55.
- [2] T. H. Sulaiman and H. S. Al-Raweshidy, Patent "Context-Aware Routing Protocol", *Applicant: Brunel University*, Publication Number WO 2007/015089 A1.
- [3] T. Dean, J. Allen and Y. Aloimonos, "Artificial Intelligence Theory and Practice", *the Benjamin/ Cummings Publishing Company, Inc*, California, 1995.
- [4] M. Abolhasan, T. Wysocki and E. Dutkiewicz, "A review of routing protocols for mobile ad hoc networks", [http:// www.elsevier.com/locate/adhoc](http://www.elsevier.com/locate/adhoc), *Ad Hoc Networks* 2, 2004, pp 1-22.
- [5] J. Broch, D. Johnson and D. Maltz, "The dynamic source routing protocol for mobile ad hoc networks", *IETF Internet Draft*, <http://www.ietf.org/internetdrafts/draft-ietf-manet-dsr-01.txt>, December 1998.
- [6] G. Campos and G. Elias, "Performance Issues of Ad Hoc Routing Protocols in a Network Scenario used for Videophone Applications", *IEEE Proceedings of the 38th Hawaii International Conference on System Sciences*, 2005.
- [7] Laouti, P. M'uhlethaler, A. Najid and E. Plakoo, "Simulation Results of the OLSR Routing Protocol for Wireless Network", *Institut National de Recherche en Informatique et en Automatique*, INRIA, March 2002, Tech. Rep. 4414.
- [8] L. Zadeh, "Fuzzy Logic and Softcomputing", Plenary Speaker, *Proceedings of IEEE International Workshop on Neuro Fuzzy Control*, Muroran, Japan 1993.

- [9] W. Fritz, *E-book* <http://www.intelligentsystems.com.ar/intsynt>, 1997.
- [10] E. Bonabeau, M. Dorigo, G. Theraulaz, "Swarm Intelligence from Natural to Artificial Systems", *Oxford University Press*, United States of America, 1999.
- [11] J. R. Jang, "ANFIS: Adaptive-network, based fuzzy inference system", *IEEE transaction on Systems, Man and Cybernetics*, 1993, vol. 23, no. 3, pp 665-685.
- [12] G. Bersini, C. Bontempi and Decaestecker, "Comparing RBF and fuzzy inference systems on theoretical and practical basis", in F. Fogelman-Soulie', P. Gallinari (Eds.), *International Conference on Artificial Neural Networks*, Paris, 1995, vol. 1, pp169-174.
- [13] J. M. Zurada, "Introduction to Artificial Neural Systems", *West Publishing Company*, U.S.A, 1992.
- [14] C. M. Bishop, "Neural Networks for Pattern Recognition", *Clarendon Press*, Oxford, 1995.

الملخص باللغة الانجليزية

Title: Modeling MANET utilizing Artificial Intelligence

Abstract. In wireless communication, Ad hoc mobile network is a collection of mobile nodes that are dynamically and arbitrarily located in such a manner that the interconnections between nodes are capable of changing on a continual basis. In order to facilitate communication within the network, a routing protocol is used to discover routes between nodes. Routing is a fundamental issue for Mobile Ad Hoc Network (MANET). Many routing protocols have been proposed for MANET and some of them have been widely used. This paper utilizes Artificial Intelligent (Neural network and neuro-fuzzy) to model MANET routing for three different routing protocols: Dynamic Source Routing (DSR), Ad hoc on-demand Distance Vector (AODV) and Optimized Link State Routing (OLSR) and compare them with the traditional mathematical equation models. The results show that Artificial Intelligent (AI) models are more accurate and presenting to MANET than traditional mathematical equations.

Keywords: Mobile Ad Hoc Network (MANET), Artificial neural network, neuro-fuzzy network.

طريقة محسنة في إيجاد عتبة التجزئة للصور الرقمية اعتماداً على مبدأ حفظ العزم

رامي محمد جمعة، علي الزعرت، حسن اسماعيل مذكور
كلية علوم الحاسب والمعلومات، جامعة الملك سعود
ص.ب. 51178، الرمز البريدي 11543، الرياض، المملكة العربية السعودية

البريد الإلكتروني: rami@ksu.edu.sa • elzaart@ksu.edu.sa • mathkour@ksu.edu.sa

الخلاصة: تعتبر عملية إيجاد عتبة التجزئة من أهم تقنيات تقطيع الصور، تتميز بسهولة وفعاليتها في الحصول على أفضل جودة لتقطيع الصور. وتعد طريقة إيجاد عتبة التجزئة اعتماداً على مبدأ حفظ العزم من الطرق الشائعة في إيجاد عتبة التجزئة. نقدم في هذه الورقة العلمية تحسينات على طريقة تساي، حيث قمنا باستخدام التوزيع الطبيعي اللوغاريتمي الذي يعتبر أكثر ملائمة للبيانات المتناسقة والبيانات المزاحة لليمين، إضافة لاستخدام المشتق من الدرجة الأولى للحصول على قيمة عتبة التجزئة مما يتيح تحسين في سرعة الطريقة المطبقة.

الكلمات الجوهرية: تقطيع الصور، إيجاد عتبة التجزئة، حفظ العزم، التوزيع الطبيعي اللوغاريتمي.

1 مقدمة

يتم عادةً في أغلب تطبيقات معالجة الصورة، التمييز بين مجموعتين من عناصر الصورة تبعاً لتدرجها اللوني. المجموعة الأولى هي التي تتبع الكائنات وتختلف اختلافاً كبيراً عن المجموعة الثانية التي تتبع خلفية الصورة [11،1]. وتتمحور عملية إيجاد عتبة التجزئة في الصور حول إيجاد القيمة اللونية التي سيتم تبعاً لها تميز عناصر الصورة التابعة للكائنات عن نظيراتها التابعة لخلفية الصورة. حيث يعتبر كل عنصر يملك تدرج لوني أكبر من عتبة التجزئة يتبع للكائن وإلا فهو يتبع للخلفية. وهذا يجعل من الضروري أن تكون عملية إيجاد عتبة التجزئة دقيقة وفعالة. سيتمحور عملنا في هذه الورقة العلمية حول إضافة بعض التحسينات على طريقة حفظ العزم المقترحة من قبل تساي [2]. حيث قمنا بتطبيق التوزيع الطبيعي اللوغاريتمي لنمذجة الرسم البياني للصورة مما يعطي نتائج أكثر دقة، إضافة لتطبيق المشتق من الدرجة الأولى للحصول على القيمة المثالية للعتبة وهو ما يختصر كثيراً من الوقت في عملية إيجاد هذه القيمة.

سنناقش في هذه الورقة الأعمال السابقة المرتبطة بموضوع البحث، ثم سنقدم مبدأ حفظ العزم بالإضافة للتوزيع الطبيعي اللوغاريتمي. في القسم الثاني سوف نشرح الطريقة المقترحة، ثم في قسم النتائج نقدم بعض النتائج التي حصلنا عليها بعد تنفيذ الطريقة المقترحة إضافة لمقارنتها مع بعض النتائج السابقة. وفي النهاية الخاتمة والعمل المستقبلي.

1.1 الأعمال ذات الصلة

طُرح مبدأ حفظ العزم أول مرة من قبل ديلب وميتشل عام 1979 [2]. حيث تم تقديم المبدأ في تشفير اقتطاع البلوكات، والذي يستخدم مكتم حفظ العزم ثنائي المستوى بهدف ضغط الصور [12]. بعدها كان تساي [2] أول من قدم هذا المبدأ ليتم استخدامه في عملية إيجاد عتبة التجزئة. حيث اعتمد على الفرضية التي تقول بأن عزم التدرجات اللونية في الصور الأصلية تبقى ذاتها في الصور الناتجة عن التقطيع. وقد أعطت الطريقة المقترحة من قبل تساي أداء جيداً مقارنة مع التقنيات الأخرى [3]. و من ثم طرحت عدة تحسينات على طريقة تساي، فقد اقترح تشن وآخرون [4] تحسيناً على الطريقة في الحالة التي تكون فيها قمتي الرسم البياني للصورة متباعتين بشكل كبير.

يمكن تمثيل الرسم البياني للصورة كمزيج من عدة توزيعات احتمالية [5،6]، حيث أن صنف الرسم البياني قد يكون متناسقاً أو غير متناسق. معظم طرق إيجاد عتبة التجزئة تعتمد على توزيع غوشيان الذي يفترض أن البيانات تكون متناسقة [1]. وقد تم تحسين بعض طرق إيجاد عتبة التجزئة بحيث تم تطبيق توزيع احتمالي آخر يدعى توزيع غاما [6،7]، حيث يمكن باستخدام هذا التوزيع نمذجة البيانات المتناسقة وغير المتناسقة [6]. وقد تم أيضاً تطبيق طريقة تساي باستخدام توزيع غاما [7]. بعض الطرق قامت باستخدام التوزيع الطبيعي اللوغاريتمي الذي يستخدم لنمذجة البيانات المتناسقة والبيانات المزاحة لليمين [8]، وقد أعطت نتائج أفضل من توزيعي غوشيان وغاما [9].

1.2 إيجاد عتبة التجزئة باستخدام مبدأ حفظ العزم

يعرّف عزم الصورة بأنه متوسط وزني محدد الكثافات اللونية لعناصر الصورة. ويعرّف مبدأ حفظ العزم في معالجة الصور بأنه تحول الصورة إلى صورة أخرى مع حفظ عزوم الكتلة و/أو عزوم اللون للصورة الأصلية [10]. ويهدف التحول المبني على مبدأ حفظ العزم إلى تجميع القيم المختلفة للكثافات اللونية لعناصر الصورة في عدد معين من الأصناف. بحيث يمثل كل صنف من القيم في الصورة الناتجة باستخدام قيمة لونية واحدة.

لتكن لدينا الصورة الأصلية f حيث أن $f(x, y)$ قيمة الكثافة اللونية للعنصر الصورة (x, y) ، عندها يكون عزم الصورة من الرتبة i^{th} هو m_i المعطى بالعلاقة:

$$(1) \quad m_i = \frac{1}{n} \sum_x \sum_y f^i(x, y), \quad i = 1, 2, 3, \dots$$

حيث أن n هو عدد عناصر الصورة f .

ويحسب عزم الصورة من رسمها البياني من خلال العلاقة:

$$(2) \quad m_k = \frac{1}{n} \sum_{i=0}^{L-1} h(i) \cdot i^k, \quad k = 1, 2, 3, \dots$$

حيث أن $h(i)$ هو العدد الكلي من عناصر الصورة التي لها القيمة اللونية ذاتها i .

ويحسب العزم b_k من الرتبة k^{th} في الصورة الثنائية الناتجة عن التحويل بالعلاقة:

$$(3) \quad b_k(T) = P_1 \mu_1^k(T) + P_2 \mu_2^k(T), \quad k = 1, 2, 3, \dots$$

حيث أن P_1 و P_2 هي الاحتمالات المسبقة لصنفي الصورة (الكائن والخلفية)، T هي قيمة عتبة التجزئة المختارة، μ_1 و μ_2 هما المتوسطين لصنفي الرسم البياني للصورة.

في عملية إيجاد عتبة التجزئة باستخدام مبدأ حفظ العزم، اعتبر تساي [2] أن الصورة الأصلية هي نسخة مُموهة من الصورة الثنائية. واعتبر أن العزوم من الرتب الثلاث الأولى تبقى ذاتها بين الصورة الأصلية والصورة الثنائية، وبناءً عليه فإن القيمة المثالية للعتبة هي T التي تحقق العلاقة التالية:

$$(4) \quad T^* : m_i = b_i(T), \quad i = 1, 2, 3$$

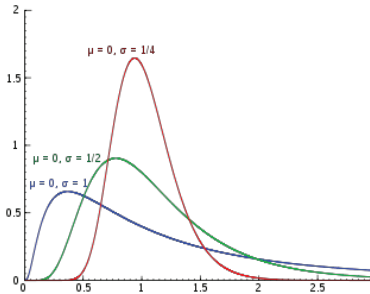
1.3 التوزيع الطبيعي اللوغاريتمي

التوزيع الطبيعي اللوغاريتمي هو توزيع احتمالي لمجموعة من المتغيرات العشوائية بحيث يكون اللوغاريتم الطبيعي لهذه المتغيرات موزع طبيعياً [8]. يستخدم التوزيع الطبيعي اللوغاريتمي لنمذجة البيانات المتناسقة والبيانات المزاحة إلى اليمين بشكل معتدل. يعطى تابع الكثافة الاحتمالي للتوزيع الطبيعي اللوغاريتمي بالعلاقة:

$$(5) \quad F(y, \mu, \sigma) = \frac{1}{x \cdot \sigma \sqrt{2\pi}} e^{\frac{1}{2} \left(\frac{\ln x - \mu}{\sigma} \right)^2}$$

حيث أن μ و σ هما على التوالي المتوسط والانحراف المعياري للتوزيع الطبيعي اللوغاريتمي، و x هي كثافة عنصر الصورة.

يوضح الشكل (1) ثلاث أشكال مختلفة لتابع كثافة التوزيع الطبيعي اللوغاريتمي مع قيم مختلفة لمتغيراته (لبارامترات).



الشكل 1: عدة نوابع كثافة للتوزيع الطبيعي اللوغاريتمي

يمكن تمثيل الرسم البياني لأصناف الصورة $f(x, y)$ باستخدام التوزيع الطبيعي اللوغاريتمي بالشكل التالي:

$$h(x) = p_1 f(x, \mu_1, \sigma_1) + p_2 f(x, \mu_2, \sigma_2)$$

حيث $f(x, \mu_1, \sigma_1)$ و $f(x, \mu_2, \sigma_2)$ يمثلان الصنف الأول والثاني على التوالي من الرسم البياني $h(x)$ للصورة. بالإضافة إلى أنه تم تقدير متغيرات (بارمترات) التوزيع الطبيعي اللوغاريتمي باستخدام مقدرات الاحتمالية القصوى على الشكل التالي:

$$(6), (7) \quad \mu_2 = \frac{\sum_{i=t}^{L-1} h(i) \log(i)}{\sum_{i=t}^{L-1} h(i)} \quad \mu_1 = \frac{\sum_{i=1}^t h(i) \log(i)}{\sum_{i=0}^t h(i)}$$

$$(8), (9) \quad \sigma_2^2 = \frac{\sum_{i=t}^{L-1} (\log(i) - \mu_2)^2}{\sum_{i=t}^{L-1} h(i)} \quad \sigma_1^2 = \frac{\sum_{i=1}^t (\log(i) - \mu_1)^2}{\sum_{i=0}^t h(i)}$$

$$(10), (11) \quad p_2 = \frac{\sum_{i=t}^{L-1} h(i)}{\sum_{i=0}^{L-1} h(i)} \quad p_1 = \frac{\sum_{i=1}^t h(i)}{\sum_{i=0}^{L-1} h(i)}$$

2 الطريقة المقترحة

قمنا في هذه الطريقة بتطبيق مبدأ حفظ العزم باستخدام التوزيع الطبيعي اللوغاريتمي. حيث يهدف استخدام التوزيع الطبيعي اللوغاريتمي إلى نمذجة الصورة بغية الحصول على قيمة عتبة أكثر دقة، على اعتبار أن التوزيع الطبيعي اللوغاريتمي يستخدم لنمذجة البيانات المتناسقة والبيانات المزاحة إلى اليمين بطريقة أكثر فعالية من توزيعي غاما وغوشيان. تتم عملية حساب عزم الصورة المدخلة باستخدام العلاقة (2). أما عزم الصورة الثنائية الناتجة فسيتم حسابها باستخدام العلاقة (3)، حيث أن μ_1 و μ_2 هما المتوسط لصنف الكائن من الرسم البياني للصورة والمتوسط لقسم الخلفية من الرسم البياني للصورة على التوالي، ويتم حسابهما عن طريق التوزيع الطبيعي اللوغاريتمي. افترض تساي [8] أن الصورة المدخلة ذات التدرجات اللونية هي عبارة عن نسخة مُموهة من الصورة المثالية الثنائية، وبأن العزوم الثلاثة الأولى للتدرجات اللونية تساوي نظيراتها في الصورة الثنائية وبالتالي:

$$|m_k - b_k(t)| < \varepsilon, \quad k = 1, 2, 3$$

ونقوم بتوسيعها لتشمل العزوم الثلاثة الأولى ومن ثم نقوم بتطبيق المشتق من الدرجة الأولى بالنسبة لـ T يكون:

$$(12) \quad \sum_{k=1}^3 \frac{\partial}{\partial t} b_k(t) = 0$$

وباستخدام العلاقات (6)، (7)، (10) و(11) وتعويضها في العلاقة (12) يمكننا الحصول على قيمة عتبة التجزئة المثالية وفق العلاقة التالية:

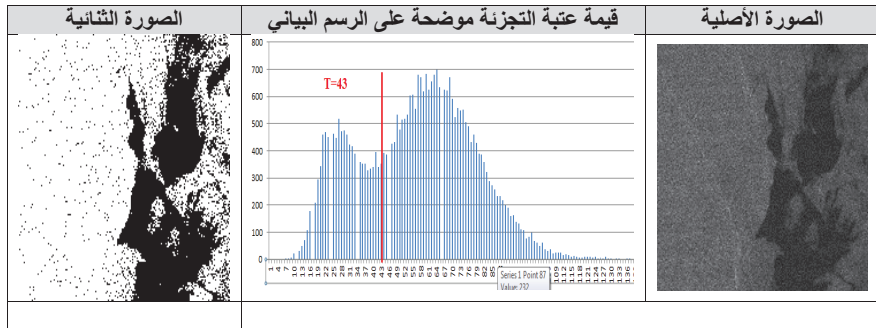
$$(13) \quad T^* = e^{\frac{\sum_2^3(k-1) \cdot (\mu_1^k - \mu_2^k)}{\sum_2^3 k \cdot (\mu_1^{k-1} - \mu_2^{k-1})}}$$

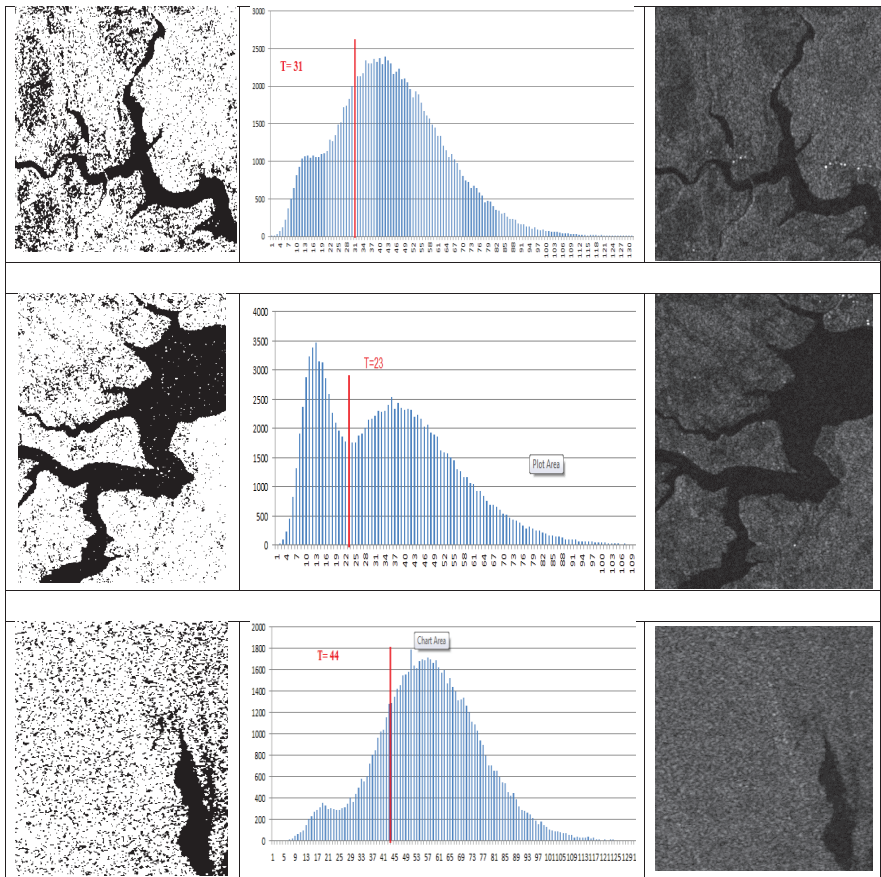
2.1 الخوارزمية المقترحة

1. قراءة الصورة المدخلة.
2. حساب الرسم البياني للصورة.
3. حساب القيمة الابتدائية T_0 ، باستخدام المتوسط للصورة كاملة (حسب التوزيع الطبيعي اللوغاريتمي)، من العلاقة (6).
4. حساب $\mu_2(T_0)$ ، $\mu_1(T_0)$ باستخدام العلاقات (6) ، (7).
5. حساب قيمة عتبة التجزئة T من العلاقة (13).
6. إذا كانت $|T - T_0| < \varepsilon$ عندها نكون قد وصلنا للقيمة المثالية للعتبة $T^* = T$ ، وإلا نضع $T_0 = T$ ، ونعيد عملية الحساب بدءاً من الخطوة 4.

3 النتائج

تم تنفيذ الطريقة المقترحة باستخدام اللغة البرمجية السي بلس بلس وباستخدام برنامج ماتلاب، وتم اختبارها على مجموعة من الصور الملتقطة باستخدام الرادار والتي تتكون من صنفين (كائن وخلفية). وقد أظهرت النتائج أن قيمة عتبة التجزئة المحسوبة وفقاً لطريقتنا تعتبر نتائج جيدة. في الجدول (1) نظهر القيم الناتجة لأربعة صور، بحيث نظهر الصورة الأصلية وقيمة عتبة التجزئة على الرسم البياني للصورة والصورة الثنائية الناتجة. وفي الجدول (2) تمت مقارنة قيم عتبة التجزئة للصوريتين الأوليتين من الجدول (1) مع قيم عتبة التجزئة المحسوبة وفق طريقة حفظ العزم وفق توزيعي غاما وغوشيان [7].





الجدول 1: نتائج الطريقة المقترحة على بعض صور الرادار

الصورة الأصلية	قيمة عتبة التجزئة		
	توزيع غوشيان	توزيع غاما	التوزيع الطبيعي اللوغاريتمي
الصورة الأولى	51	46	43
الصورة الثانية	50	46	31

الجدول 2: مقارنة قيم عتبة التجزئة الناتجة عن استخدام طريقة حفظ العزم مع التوزيعات الثلاثة الرئيسية غوشيان، غاما و التوزيع الطبيعي اللوغاريتمي

وبمعاينة النتائج في الجدولين نجد أن النتائج الحاصلة باستخدام طريقتنا مع التوزيع الطبيعي اللوغاريتمي تعتبر هي نتائج مباشرة وأفضل من سابقتها عند استخدام توزيعات أخرى.

4 الخاتمة والعمل المستقبلي

قمنا من خلال ورقة البحث باقتراح خطوتين بهدف تحسين عمل طريقة إيجاد عتبة التجزئة باستخدام مبدأ حفظ العزم. الخطوة الأولى هي استخدام التوزيع الطبيعي اللوغاريتمي لنمذجة الرسم البياني للصورة للحصول على نتائج أكثر دقة. أما الخطوة الثانية فهي توظيف المشنق من الدرجة الأولى للوصول إلى القيمة المثالية للعتبة. وقد أظهرت النتائج التجريبية التي قمنا بها على مجموعة من صور الرادار تحسناً في النتائج عن سابقتها. من المقترحات لتطوير هذه الطريقة وتوسيع نتائجها هي تطبيقها على صور متعددة الأصناف إضافة لتطبيقها على أنواع أخرى من الصور وبالأخص تطبيقها في مجال التمييز الضوئي للمحارف العربية.

المراجع

- [1] M. Sezgin, and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation". Journal of Electronic Imaging, vol. 13, no. 1, 2004, pp. 146-168.
- [2] W. Tsai, "Moment-preserving thresholding: A new approach," Computer Vision, Graphics, and Image Processing, vol. 29, Mar. 1985, pp. 377-393.
- [3] P. Sahoo, S. Soltani, and K. Wong, "A survey of thresholding techniques", Computer Vision, Graphics, and Image Processing. 1988, pp. 233-260.
- [4] I. Chen, M. Lai, and L. Wang, "An improvement in the moment-preserving thresholding method", International Journal of Imaging Systems and Technology, vol. 18, 2008, pp. 365-370.
- [5] R.J. Ferrari, R.M. Rangayyan, R.A. Borges, and A.F. Frère, "Segmentation of the fibro-glandular disc in mammograms using Gaussian mixture modelling," Medical & Biological Engineering & Computing, vol. 42, May. 2004, pp. 378-387.
- [6] A. El-Zaart, "Images thresholding using ISODATA technique with gamma distribution," Pattern Recognition and Image Analysis, vol. 20, 2010, pp. 29-41.
- [7] A. Al-Hussain and A. El-Zaart, "Moment-preserving thresholding using Gamma distribution," Computer Engineering and Technology (ICCET), 2010 2nd International Conference on, 2010, pp. V6-323-V6-325.
- [8] J. Aitchison, and J.A.C. Brown, "The lognormal distribution," (1957)
- [9] Yaser A. Reyad, Ali El-Zaart, Hassan Mathkour and Mansour Al-Zuair "Image Thresholding Using Split and Merge Techniques with Log-Normal Distribution" Canadian Journal on Image Processing and Computer Vision, Vol. 1 No. 3, July 2010, pp. 36-45.
- [10] S. Ferng, and W. Tsai, "Dynamic image compression using line edge features by moment-preserving technique," Proceedings of 1992 National Symposium on Telecommunications, Chungli, Taiwan, Republic of China, 1992, pp. 190-195.
- [11] R. Gonzalez, and R. Woods, "Digital Image Processing", 3rd ed. Prentice-Hall, 2008
- [12] E. Delp, and R. Mitchell, "Image compression using block truncation coding", Image compression using block truncation coding, IEEE Trans. Commu. COMM-27, 1979, pp. 1335-1341.

الملخص باللغة الانجليزية

Title:An Improved Moment-Preserving Method for Image Thresholding

Abstract. Thresholding is one of the most known image segmentation technique. Its simplicity and effeteness in finding an optimal threshold value give better segmentation quality. Moment-preserving thresholding proposed by Tsai is one of the most popular thresholding techniques. In this paper, we propose some improvements to Tsai's method by using the first derivative to get the threshold and by introducing Log-Normal distribution to moment-preserving thresholding, which enhances thresholding result in case of symmetric and positively skewed histogram modes.

Keywords: Segmentation, Thresholding, Moment-Preserving, Log-Normal distribution.

المصطلحات

Segmentation	تقطيع الصور
Thresholding	إيجاد عتبة التجزئة
Moment-Preserving	حفظ العزم
Gray level	تدرج اللوني
Object	كائن
Background	خلفية الصورة
Class	صنف
Histogram	الرسم البياني للصورة
Intensity	القيمة اللونية
Pixel	عنصر الصورة
Block TruncationCoding (BTC)	تشفير اقتطاع البلوكات
Quantizer	مكمم
Image compression	ضغط الصور
Symmetric	متناسق
Non-symmetric	غير متناسق
Log-Normal distribution	التوزيع الطبيعي اللوغاريتمي
Gaussian distribution	توزيع غوشياني
Gamma distribution	توزيع غاما
Binary Image	صورة ثنائية

Threshold	عتبة التجزئة
Prior Probability	احتمال مسبق
ProbabilityDensityFunction (PDF)	تابع كثافة احتمالي
Mean	متوسط
Standard Deviation	انحراف معياري
Maximum likelihoodestimators	مقدرات الاحتمالية القصوى
C++	سي بلس بلس
Matlab	ماتلاب
Blurred	مُموهة

تصميم نموذج للتعليم الإلكتروني لتقديم خدمات تربوية ملائمة

محمد غيلان¹، جابر البحديدي²، عبد الهادي فنان³
جامعة عبد المالك السعدي، كلية العلوم والتقنيات،
ص.ب. 416، الرمز البريدي 90000، طنجة، المملكة المغربية

البريد الإلكتروني: ghalamed@gmail.com¹، jaber.f15@gmail.com²، afennan@gmail.com³

الخلاصة. يأتي هذا البحث بعد دراسة تحليلية ونقدية لمجموعة من البحوث والمقالات التي تهتم بالتعليم الإلكتروني، فمعظمها يركز على حوسبة وإدارة المحتويات التعليمية على حساب العملية التربوية، وعليه فإن كل نظام قائم على هذا المبدأ يمنح للمتعلمين محتويات ثابتة و مسارات تعليمية محددة مسبقا لا توافق متطلبات أغلب المستفيدين.

لهذا السبب، اقترحنا توجيه بحثنا نحو تصميم نموذج للتعليم الإلكتروني يمكن المتعلمين من الحصول على خدمات تربوية تستجيب لحاجياتهم و تسائر فضولهم العلمي، كما أنها تواكب إيقاعاتهم في التعلم و تنزل إلى رغباتهم الخاصة.

من هذا المنطلق، تم تبني بيداغوجيا الأهداف كإطار مرجعي لمقاربة تكنولوجية اتخذت من الويب الدلالي و الهندسة الأنطولوجية و النظم المتعددة الوكلاء أساسا لتزويد نظامنا الإلكتروني بقدرات الاستدلال المنطقي تمكنه من تقديم خدمات تربوية ملائمة وإنشاء مسارات مشخصة تراهن على جودة الأداء و خلق إيقاعات جديدة للتعلم تأخذ بعين الاعتبار ملمح المتعلم (مستواه المعرفي، رغباته، كفاياته، مراكز اهتماماته، نمط تعلمه، الخ).

الكلمات الجوهرية: التعلم الإلكتروني، خدمات تربوية، الهندسة الأنطولوجية، الويب الدلالي، النظام المتعدد الوكلاء، ملمح.

1 مقدمة

إن الوعي بأهمية الويب الدلالي في بيئة التعلم والتكوين عن بعد أصبح يتزايد لدى المهتمين بإدماج تكنولوجيا المعلومات والاتصال في العمليات التعليمية/ التعليمية. فهذا المقال يروم تقديم نموذج للتعلم الإلكتروني يمكن من تحقيق الأهداف التالية:

- تجويد الخدمات التربوية وتكييفها حسب حاجيات المتعلمين
- توليد مسارات تعليمية مشخصة
- تشجيع التعلم الذاتي

من أجل هذا، قمنا باعتماد مقاربة تزاوج بين الأنطولوجيا والوكلاء البرمجيين في نسق تعاوني لتصميم نموذج متطور يمكن من تسهيل خدمات العاملين في مجال التعلم عن بعد ويمكن من تحقيق الغايات التربوية الرامية إلى الأهداف السالفة الذكر.

في هذا الإطار، اعتمدنا تقنيات الويب الدلالي من أجل تجويد وتسريع عملية البحث عن المعلومات وذلك باعتماد مبدأ الوصف الدلالي للموارد التربوية الذي تتيحه الهندسة الأنطولوجية، إذ أن هذه الأخيرة فرضت نفسها بالإحاح للارتقاء بالخدمات التربوية بشكل تصبح معه التعليمات قابلة للاستثمار من قبل وكيل برمجي.

إن النظام المقترح سيمكن من خلق مسارات تعليمية تنسجم مع إيقاعات التعلم وتستجيب للأهداف التي يسعى المتعلم لتحقيقها باستهلاكه للخدمة المقترحة ضمن سياق تعليمي مشروط بالتقويم القبلي والبُعدي لضمان نجاح العملية التربوية.

باختصار، هذا النظام سيمكن الفاعلين الأساسيين (المتعلمين والمدرسين و المصممين التربويين) من العمل في إطار تعاوني مع الوكلاء البرمجيين لإنشاء عدة للتكوين تمكن المستفيد من التعلم حسب ملمحه.

بعبارة أكثر دقة، إن الهدف الذي نسعى إليه من خلال هذه الورقة هو تجهيز كل نظام للتعلم الإلكتروني بمجموعة من القدرات والاستدلالات المنطقية التي تسمح بتكييف الموارد والخدمات المقدمة للتعلم حسب ملمحه (رغباته، معارفه، كفاياته، مراكز اهتماماته، مستواه المعرفي ونمط تعلمه...). ومن هذا الاعتبار كانت الحاجة ملحة إلى ضرورة اعتماد نظام يربط ما هو تربوي بما هو تكنولوجي في إبداع تكوين يتلاءم وخصوصيات كل متعلم.

2 المقاربة التربوية المعتمدة

لقد ارتأينا انطلاقاً من المقاربة البيداغوجية المرتكزة على الأهداف [1] إعداد تصور لمسارات تربوية للتعلم الإلكتروني و الذي ينسجم مع ملامح المتعلمين، لذلك توخينا الانطلاق من هذه التراتبية التي تستند على ثلاثة مستويات للأهداف البيداغوجية نصوغها كما يلي:

- الأهداف العامة
- الأهداف النوعية
- الأهداف الإجرائية

2.1 بنية مجزوءات التكوين

ولا بأس هنا، من التنكير بالآليات اشتغال البيداغوجيا القائمة على الأهداف، والتي تركز على تفكيك مجزوءات التعليم المعقدة لعناصر جزئية وبسيطة لتيسير العملية التعليمية/التعليمية بما فيه محطة التقويم.

إن تجزئ التعليمات يسمح لنا بمعاينة ثلاثة مستويات لتجريد النموذج التعليمي:

- الأجزاء : وهي تلك المرتبطة بالأهداف العامة
- الفصول : وهي المرتبطة بالأهداف النوعية
- وحدات التعليمات الجزئية القابلة للتقويم والتي أعطيناها اسم الخدمات التربوية، وهي تتعلق بالأهداف الإجرائية. وهذه الخدمات تنقسم إلى قسمين:

✓ الخدمات التربوية للتعلم

✓ الخدمات التربوية للتقويم

وتعمل هذه الخدمات التربوية على استقبال كل أشكال وسائل التواصل المسموح بها من قبل لغة HTML بما في ذلك مقطع من نص أو صورة أو صوت أو مقاطع من فيديو... ومن أجل توظيف خدمات التعلم التي يزود بها النظام المتعلمين على اختلاف مستوياتهم، فإن هؤلاء مطالبون بترجمة مقاصدهم وحاجاتهم في شكل أهداف تعليمية.

ولكي نضمن نجاح هذه العملية التي تخدم رغبات المتعلمين بشكل مباشر، فإننا نحتاج إلى تصميم لنظام منسجم، متعدد الأنطولوجيا ومتعدد الوكلاء نصوغه حسب القواعد التالية :

- التفاوض على أهداف التعلم
- تقدير الشروط القبلية بعد عملية التفاوض
- تخطيط تسلسل الخدمات التربوية للتعليم والتقويم
- بحث وتصفية، وكذا تقديم مسار وسائط فائقة للخدمات
- تقويم التعلم.

إن تسلسل الأهداف البيداغوجية من قبل النظام المعتمد، يتحقق انطلاقاً من شبكة للشروط المسبقة التي يقترحها واضع المجزوءة التعليمية، بمعنى أننا نتصور شبكة من العلاقات المنظمة لكل هدف تعليمي في علاقته بالآخر، فإذا افترضنا أننا بصدد تعلم الهدف الثاني فإن هذا الأخير لا يتحقق نجاحه إلا بنجاح الهدف التعليمي الأول. وهذه الميزة تعني أن النظام يستطيع اختيار هدف تعليمي جديد يخدم الهدف الذي لم يتحقق، ولإجراء هذه العملية، يتم اقتراح تزويد المتعلم بمجموعة من الخدمات المتصلة بالهدف المقترح.

2.2 الخدمات التربوية

في هذه المقاربة، نعتبر أن مفهوم الخدمة التربوية هو مكون يملك خصائص الاستقلالية الذاتية، وقابلية إعادة التوظيف، بالإضافة إلى إمكانية التقاسم. إنه نوع من الخدمة التي تملك توصيفا دلاليا يسمح بقابلية استعمال الخدمات ضمن عدة سياقات تعليمية.

وبعبارة أخرى، إن مفهوم الخدمة التربوية هو شق من العملية التي تسمح لنا بتحقيق الهدف البيداغوجي الإجرائي، ونفس الهدف التعليمي يمكن تحقيقه بطرق مختلفة حسب ملمح كل متعلم على حدة، ومن ثم، فإن مفهوم العملية هو مظهر آخر من مظاهر الخدمات التربوية، مادام يأخذ بعين الاعتبار مختلف الطرائق والاستراتيجيات التربوية داخل ما يصطلح عليه بالمسارات التربوية.

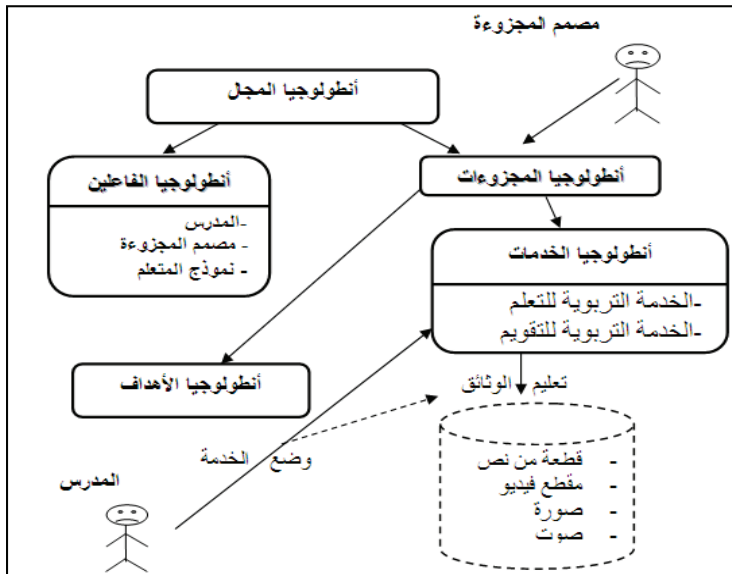
وكل خدمة تربوية تتحدد بملمح يسمح بتوصيف مظهرها العام أو ما نعتبر عنه ب(واجهة الخدمة) وتتضمن الهدف التعليمي وسياقه وطريقة الاشتغال عليه. وملمح الخدمة هذا سيمكن من معرفة مدى تطابق الخدمة التربوية مع ملمح المتعلم وهدفه التعليمي[2].

3 المقاربة متعددة الأنطولوجيا

لكي نختصر ونجمع مختلف التعاريف الواردة عن المفهوم، يمكننا القول بأن الأنطولوجيا في مجال الإعلاميات هي طريقة لتمثيل مجال معرفي معين على شكل مفاهيم منظمة بعلاقات هيكلية أهمها العلاقة (هو/ هي)، على سبيل المثال، الخدمة التربوية للتعليم هي خدمة تربوية [3].

إن الهدف من البنية المقترحة (الشكل 1) هو تجهيز نظامنا بقدرات تمكنه من الاستدلال المنطقي على الأوصاف الدلالية المحصل عليها من أجل التكيف التلقائي للموارد والخدمات المقترحة حسب خصوصية المتعلم، وذلك وفقا لملمحه. في هذا النموذج، سنركز الاهتمام على أنطولوجيا المجال المدرس والتي تنقسم إلى أنطولوجيتين فرعيتين: الأولى للمجزوءات، والثانية للفاعلين.

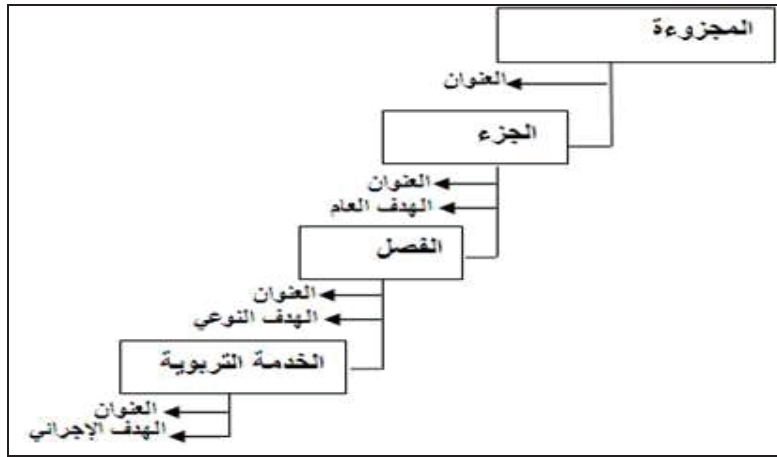
تتكون أنطولوجيا المجزوءات بدورها من أنطولوجيتين فرعيتين: أنطولوجيا الأهداف وأنطولوجيا الخدمات التربوية للتعليم والتقويم. وتسمح أنطولوجيا الخدمات بالتعليق على الوثائق وتعليمها وذلك بربطها بوصف دلالي دقيق. وتعتبر الدلالة عنصرا أساسيا، فهي من جهة، تؤمن مقاسمة وإعادة استعمال الخدمات، ومن جهة أخرى، تعمل على تسهيل عملية البحث والتركيب، وذلك اعتمادا على ملمح المتعلم الذي تم وصفه في نموذج.



الشكل 1. الهندسة المتعددة الأنطولوجيا

3.1 أنطولوجيا المجزوءات

تستعمل هذه الأنطولوجيا لبناء مجزوءة للتكوين حسب ثلاث مستويات لأهداف التربية (الشكل 2).

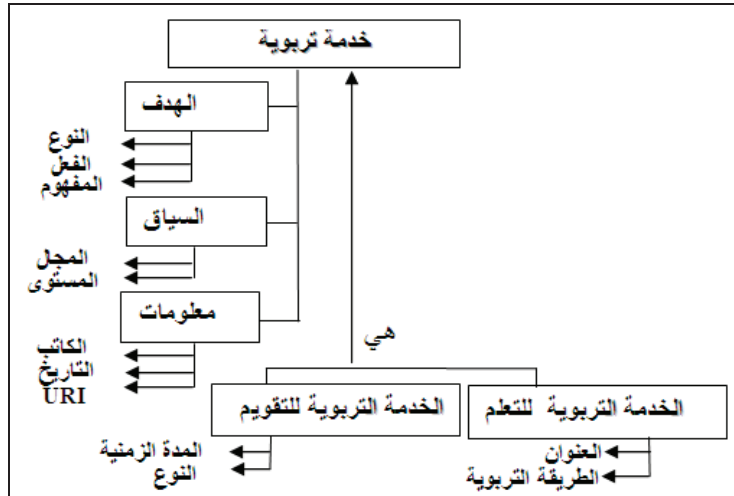


الشكل 2. أنطولوجيا مجزوءة التعليم

3.2 أنطولوجيا الخدمات التربوية

تمكن هذه الأنطولوجيا من الوصف الدلالي لخدمتين تربويتين (الشكل 3):

- الخدمة التربوية للتعلم والتي تسعى لتحقيق هدف إجرائي.
 - الخدمة التربوية للتقويم وتسمح باختبار مدى تحقق هدف مرجو من طرف الخدمة التربوية للتعلم.
- وفي سياق آخر، يمكن لمجموعة من الخدمات التربوية للتقويم أن تشكل الأساس لتوليد اختبار كامل لمدى التحكم في هدف عام أو نوعي.



الشكل 3: أنطولوجيا الخدمات التربوية

3.3 أنطولوجيا الأهداف البيداغوجية

تكمن هذه الأنطولوجيا الفرعية في ترتيب و تقييؤ الأهداف حسب صنافه بلوم [BLOOM 4] (الإلمام، الفهم، التطبيق، التحليل، التركيب و التقويم)، كما تساهم في الوصف الجزئي للخدمات التربوية (جزء الهدف)، ومن جهة أخرى، ستساعد المتعلمين على التعبير عن حاجياتهم في شكل أهداف تعليمية. [5]

3.4 أنطولوجيا الفاعلين

تتكون هذه الأنطولوجيا [5,6,7] من ثلاثة فاعلين رئيسيين : المتعلمين، المدرسين، ومصممي المجزوءات. وتشكل هذه الأخيرة أساسا لنمذجة المتعلم داخل نموذج يتضمن المعلومات الضرورية حول ملمحه، مسهله بذلك شخصنة عملية تعلمه (ملائمة وتكييف الخدمات التربوية حسب توقعاته). ويجب أن يكون هذا النموذج غنيا من الناحية الدلالية و قابلا لإعادة الاستعمال في نظام آخر. كما يجب أن يكون قابلا للتفسير من قبل الحاسوب ويتطور مع تطور التعلم، و ذلك بإغنائه تدريجيا مع عملية التعلم والتقويم.

4 مقارنة تعدد الوكلاء المتعاونين

يهتم هذا الجزء بتصميم وكلاء برمجيين قادرين على الانتظام بفعالية من أجل التعاون على إتمام المهام الموكلة إليهم، وذلك باستعمال كل من قدراتهم الفردية وكذا الإمكانيات المتاحة لهم من خلال النظام الذي يشكلونه (تواصل، تعاون، تعلم...) [2,9,10].

إن الهدف من هذا النظام المتعدد الوكلاء هو خلق مسارات الخدمات التربوية للتعلم وأخرى للتقويم تستجيب لمتطلبات وحاجيات المتعلم، وذلك باعتماد سياسة التواصل والتعاون بين مختلف الوكلاء البرمجيين (الشكل 4).

4.1 الوكيل المرافق

يتكون هذا الوكيل من عنصرين رئيسيين : الواجهة وقاعدة أفعال.

- تمكن الواجهة المتعلم من مباشرة العمليات - حسب حاجياته - ومن مشاهدة النتائج، كما تسمح له بالإطلاع على ملمحه.

- وتتكون قاعدة الأفعال من سلسلة من الوظائف البرمجية (LISP، PROLOG أو JAVA) التي يتم تشغيلها عن بعد بواسطة الواجهة. يسمح هذا الجزء الأخير بإنشاء ملمح المتعلم لحظة تسجيله و كذا تحديث ملمحه عند نهاية عملية التعلم و/أو التقويم أو بالاعتماد على الأفعال التي يقوم بها هذا الأخير.

يستعمل هذا الوكيل البروتوكول http للتواصل مع المتعلم (متصفح الويب)، ولغة FIPA-ACL للتواصل مع الوكيل المدير.

4.2 الوكيل المدير

يشكل هذا الأخير صلة وصل بين الوكيل المرافق وباقي وكلاء النظام، كما يعمل هذا الأخير، على تحديد الوكلاء الذين بإمكانهم المساهمة في إتمام المهمة المطلوبة (توليد مسار التعلم أو المعالجة، توليد مسار التقويم، الخ) وذلك بعد تحليل الطلب الوارد على شكل رسالة FIPA-ACL والذي يحتوي على كل من العملية المطلوب إنجازها وكذا المعلومات الضرورية على المتعلم [2].

4.3 الوكيل المولد لمسارات الخدمات التربوية للتعلم

يستقبل من الوكيل المدير طلبا FIPA-ACL يتضمن الهدف البيداغوجي للمتعلّم المعبر عنه من طرف الوكيل المرافق وذلك وفقا لحالتين :

- حالة مسار الخدمات التربوية للتعلم بهدف المعالجة : تأتي هذه العملية في نهاية اختبار غير ناجح بالاعتماد على ملمح المتعلم وتقرير الوكيل الكاشف للمباعدات الذي يعرض الأهداف التي لم يصل تحققها إلى الحد الأدنى المطلوب.
- حالة مسار الخدمات التربوية للتعلم: في هذه الحالة، يتم التعبير عن الهدف من طرف المتعلم مباشرة، وتعاد صياغته من طرف الوكيل المرافق من أجل إغناؤه بملمح المتعلم بغية التكيف والشخصنة.

في جميع الحالات، يقوم الوكيل المولد لمسارات التعلم بسؤال قاعدة المسارات للتحقق من الوجود المسبق للمسار المطلوب، في حالة وجوده، يرسل هذا الأخير جوابا إلى الوكيل المدير مباشرة. أما في حالة عدم وجوده، فإنه يسأل أنطولوجيا الأهداف لتحديد نوع الهدف: عام، نوعي أو إجرائي. عندما يكون الهدف مجردا، فإن هذا الأخير يوضحه بأهداف إجرائية لتحويلها بعد ذلك إلى وكيل النصفية على شكل طلب يتضمن ملمح المتعلم. وبعد استقبال الجواب يعمل الوكيل المولد لمسارات التعلم على تشكيل المسار المطلوب على شكل وثيقة وسائط فائقة يتم الحصول على خدماتها وكذا تنفيذها بفضل ل URIs الخاص بها.

4.4 وكيل التصفية

يقوم بسؤال أنطولوجيا الخدمات لاختيار خدمة التعلم أو التقويم الملائمة، وذلك اعتمادا على التوافق بين ملمح هذه الأخيرة وملمح المتعلم وكذا الهدف الإجرائي المرسل إليه. ويعتبر هذا الوكيل، هو المسؤول عن الشخصنة، كما أن دوره الرئيسي يتمثل في عملية البحث والترتيب وتقليص لائحة الخدمات المطلوب إرجاعها للوكيل المعني (الوكيل المولد لمسارات التعلم أو الوكيل المولد لمسارات التقويم) الذي يولد بدوره لائحة المسارات ويرتبها حسب ملمح المتعلم [2].

4.5 الوكيل الكاشف للمباعدات

يعمل هذا الوكيل على استعمال وظائف برمجية تمكن من تقييم درجة تطابق المنتج المنتظر والمنتج المحصل عليه من طرف المتعلم أثناء عملية التقويم. ويتم تقديم نتيجة هذا التقويم في تقرير يستخدمه الوكيل المولد لمسار المعالجة المركزة.

في هذا البحث، قدمنا نموذجاً للتعليم الإلكتروني والذي يبنى على المقاربة بالأهداف والهندسة الأنطولوجية، من أجل تلبية حاجيات المتعلم وذلك عن طريق إمداده بمسارات مشخصة ذات جودة. إن خلق هذا النوع من المسارات، يتطلب تعاوناً بين عدة وكلاء من نفس النظام أو بين وكلاء نظم متعددة للتعليم الإلكتروني والذين هم في تعاون مع بعضهم البعض. يمكننا القول، بأن تواجد هذه المقاربة التعاونية والتعاضدية سيكون ممكناً بفضل خلق نظام توجيهي مركزي يهدف إلى تحديد أنظمة التعلم الإلكتروني التي بإمكانها المساهمة في خلق مسارات أكثر توافقاً، وذلك بناء على مجموعة من المعايير المتفق عليها. ويمكن لهذه النقطة الأخيرة أن تكون موضوعاً لدراسة لاحقة.

مما سبق يتبين لنا أن النموذج المقترح يؤمن الوظائف التالية :

- مرافقة المتعلم طوال فترة تعلمه
- تحديد الخدمات التربوية التي تركز على المقاربة البيداغوجية بالأهداف
- تكوين مسارات مشخصة بطريقة ديناميكية، والتي تستجيب لتوقعات المتعلمين

المراجع

- [1] D.HAMELINE Objectifs pédagogiques, en formation initiale et en formation continue. parution 2005.
- [2] J. El Bouhdidi, M. Ghailani, O. Abdoun, A. Fennan, A New Approach based on a Multiontologies and Multi-agents System to Generate Customized Learning Paths in an E-Learning Platform, I.J.C.A, 2010
- [3] Gayo Diallo, Une Architecture à Base d'Ontologies pour la Gestion Unifiée des Données Structurées et non Structurées, Thèse : Université Joseph Fourier – Grenoble I École Doctorale MSTII. 2008.
- [4] Bloom, B. 1975. "Taxonomie des objectifs pédagogiques". Tome1. Presses de l'Université du Québec.
- [5] N.ZNIBAR .Construction de parcours pédagogiques individualisés: une approche orientée service. TICE Méditerranée 2007
- [6] Do Ngoc Kien, Moteur de composition pour le système d'information Sémantique et adaptatif. 2006.
- [7] Choquet, C.2007. Ingénierie et réingénierie des EIAH.
- [8] Broisin J., « Un Environnement Informatique pour l'Apprentissage Humain au Service de la Virtualisation et de la Gestion des Objets Pédagogiques », thèse, Université Paul Sabatier Toulouse 3, 2006.
- [9] D. Monticolo, V. Hilaire, S. Gomes and A. Koukam, "A Multi Agents Systems for Building Project Memories to Facilitate Design Process", International Journal in Integrated Computer Aided Engineering, accepted in july 2007.
- [10] C. CARABELEA, O. BOISSIER et A. F LOREA : Autonomie dans les systèmes multi-agents : tentative de classification. Actes des 11èmes Journées Francophones sur les Systèmes Multi-Agents (JFSMA), Novembre 2003.
- [11] D.Monticolo, Une approche organisationnelle pour la conception d'un système de gestion des connaissances fondé sur le paradigme agent.thèse, Université de Technologie de Belfort Montbéliard et Université de Franche Comté Ecole doctorale Sciences Pour L'ingénieur et Microtechniques (S.P.I.M.). 26 Février 2008.
- [12] Abel MH., Dieng-Kuntz R., Héryn D., Lenne D., Moulin C., Pompidor P. Langages pour le Web Sémantique et pour le E-Learning. Journée thématique AFIA : Web sémantique

- [13] pour le e-Learning. Plateforme AFIA. Nice, 30 mai. pp. 97-122. – 2005
 M. Cremene, M. Riveill, C. Martel, C. lophon, C. Miron, Adaptation dynamique de services, 1ère Conférence Francophone sur le Déploiement et la configuration de Logiciels, Grenoble, 2006.

المخلص باللغة الانجليزية

Title: Design of an e-learning model providing appropriate pedagogical services

Abstract. This search paper comes after a critical and analytical study of a group of research and articles that put an interest on E-Learning, which most of them are focused on computerization and management of educational content at the expense of the educational process. Thus, each system based on this principle gives learners fixed contents and predefined learning paths which do not agree with the requirements of most users.

For this reason, we have proposed directing our research towards the design model for E-Learning that learners can have access to, especially to educational services responsive to their needs and keep pace with their scientific curiosity. This model also keeps pace with their rhythms to learn and satisfy to their private desires.

From this standpoint, we adopted the pedagogy of objectives as a frame of reference for a technological approach. This last took the semantic web, the ontological engineering and the multi-agents systems as a basis to provide our e-learning system reasoning capabilities, which enable it to offer appropriate pedagogical services and create personalized paths betting on the quality of performance and the generation of new rhythms of learning by taking into account the profile of the learner (level of knowledge, preferences, competences, centers of interest, style of learning, etc.).

Keywords: e-learning, pedagogical services, ontological engineering, multi-agents systems, Semantic Web, profile

المصطلحات

التعلم الإلكتروني	E-Learning
الهندسة الأنطولوجية	Ontological Engenieering
الويب الدلالي	Semantic Web
الخدمات التربوية	Pedagogical Services
الخدمات التربوية للتعليم	Pedagogical Services of Learning
الخدمات التربوية للتقويم	Pedagogical Services of Evaluation

النظام المتعدد الوكلاء	Multi-Agents System
ملح	Profile
الأنطولوجيا	Ontology
بيداغوجيا الأهداف	pedagogy of objectives
مجزوءات التعليم	modules of teaching
وكيل برمجي	Software Agent
المقاربة المتعددة الأنطولوجيا	Multi-Ontologies Approach
مسارات تربوية للتعليم	Pedagogical Paths of Learning
الوكيل المرافق	Guide Agent
الوكيل المدير	Manager Agent
صناعة بلوم	BLOOM Taxonomy
متصفح الويب	Web Browser
الوكيل المولد لمسارات الخدمات التربوية للتعليم	Pedagogical Services Paths Generator Agent of Learning
وكيل التصفية	Filter Agent
الوكيل الكاشف للمباعدات	Gaps Detector Agent
الوكيل المولد لمسارات الخدمات التربوية للتقويم	Pedagogical Services Paths Generator Agent of Evaluation
وثيقة وسائط فائقة	Hypermedia Document
مسارات مشخصة	Personalized paths

تصميم حاكمات خمد متناسقة لأنظمة القوى الكهربائية باستخدام الخوارزميات الجينية المستمرة

محمد على عبيدو¹، السيد محمد الألفي²

¹ كلية الهندسة، جامعة الملك فهد للبترول والمعادن
الرمز البريدي 31261، الظهران، المملكة العربية السعودية
² كلية علوم وهندسة الحاسب، جامعة الملك فهد للبترول والمعادن
الرمز البريدي 31261، الظهران، المملكة العربية السعودية

البريد الإلكتروني: {mabido, alfy}@kfupm.edu.sa

الخلاصة: يقدم هذا البحث أسلوباً لتحسين اتزان أنظمة القوى الكهربائية عن طريق تصميم حاكمات خمد. وقد تم صياغة المشكلة تحت مجموعة واسعة من ظروف التحميل وتكوينات النظام كمشكلة أمثلة بدالة هدف تعتمد على القيم المميزة (eigen values). وتم تطبيق الخوارزميات الجينية المستمرة للبحث عن المعاملات المثلى للحاكمات. كما تم تقييم خصائص الخمد لأساليب التحكم المقترحة بدلالة معامل عزم الخمد عند ظروف تحميل مختلفة، وتم اختبارها في نظام للقوى الكهربائية ضعيف الاتصال. وقد بينت النتائج فعالية ومثانة الأساليب المقترحة عند ظروف التحميل المختلفة.

الكلمات الجوهرية: أنظمة القوى الكهربائية، الخوارزميات الجينية المستمرة، مثبتات، حاكمات الخمد، تحليل القيم المميزة.

1 مقدمة

من المعروف أنه عند توصيل نظم قوى كبيرة معاً بخطوط ضعيفة نسبياً يلاحظ وجودذبذبات بتردد منخفض. وقد تستمر تلك الذبذبات وتتمو لتسبب فصل لتلك النظم إذا لم تستخدم نظام للخمد بشكل مناسب [1, 2]. ويكثر في هذه الأيام استخدام معدات القوى الكهربائية لمثبتات تقليدية PSS. لكن نظراً لتغير معاملات الآلة مع التحميل فإن سلوكها يختلف عند ظروف التشغيل المختلفة. ولأن تلك المعاملات تتغير بطريقة معقدة نوعاً ما قد لا يكفي استخدام مثبتات بمعاملات ذات قيم ثابتة عند ظروف تشغيل معينة عندما تتغير ظروف التشغيل بشكل كبير. ولذا يجب أن يوفر PSS درجة معينة من المثانة (robustness) للتغيرات في معاملات وتجهيزات النظام وظروف التشغيل. طبقت طرق الأمثلة H_{∞} [3] للتصميم المثبتين لـ PSS؛ ومع ذلك يشكل اختيار دوال الوزن في تلك الطرق مشكلة. ومن ناحية أخرى تكون درجة المثبت المعتمد على H_{∞} بنفس ارتفاع درجة النظام قيد الدراسة. تؤدي تلك المشاكل إلى محدودية الاستفادة من تلك الطرق.

توجد دراسة تحليلية مفصلة لتأثير معاملات المثبت PSS التقليدي على الأداء الديناميكي لنظام القوى الكهربائية في [4]. وقد بينت الدراسة أن الاختيار المناسب لمعاملات المثبت يؤدي إلى أداء مُرضٍ للنظام. إضافة لذلك تم أيضاً توضيح أن PSS التقليدي يعطي أداء خمد مُرضياً على نطاق واسع لظروف التحميل [5]. وعلى الرغم من أن PSS يوفر إشارات مرتجعة تكملية لاتزان النظام، فإنه يعاني من مسؤوليته عن التسبب في إحداث تغيرات كبيرة في الجهد وقد يؤدي إلى التشغيل عند معامل قدرة متقدم عند وجود تشويشات حادة. ولتحسين أداء نظم PSS التقليدية تم اقتراح العديد من الطرق؛ منها على سبيل المثال استخدام حاكمات المنطق الضبابي المثالية [13]، [14]. كما توجد دراسة مقارنة بين نظامين من أنظمة الخمد (PSS، TCSC) في [15].

ولقد أدى التقدم المعاصر في صناعة الإلكترونيات القوى الكهربائية إلى تطوير نظم مرنة لنقل التيار المتردد (FACTS)، وتمتاز تلك النظم بأنه يمكن تنظيم تدفق القدرة الكهربائية لتقليل الذبذبات ذات التردد المنخفض وبالتالي تحسين اتزان النظام الكهربائي. وحديثاً تم استخدام أجهزة FACTS في أنظمة قوى عملية [6]. ويوجد تحليل مُفصل لخدم الذبذبات الكهروميكانيكية في أنظمة القوى الكهربائية باستخدام أجهزة FACTS في [7]. كما تم مناقشة تأثير أحمال خطوط النقل وخصائص الأحمال على الخمد في تلك الأجهزة [7]. ويناقش [8] عزم الخمد الناتج من استخدام أجهزة FACTS حيث تم تحليل والتحقق من عدة نقاط مهمة من خلال المحاكاة. ويتناول [9] عرض لمتطلبات مشكلة التناقص بين مثبتات PSS وأجهزة FACTS المختلفة.

في هذا البحث قمنا باقتراح أسلوب لتحسين اتزان أنظمة القوى الكهربائية عن طريق تصميم حاكمات خمد متناسقة. وقمنا بتطبيق الخوارزميات الجينية المستمرة للبحث عن المعاملات المثلى لتلك الحاكمات. كما قمنا بتقييم واختبار مدى فعاليتها عند مجموعة واسعة من ظروف التحميل المختلفة.

2 نمذجة أنظمة القوى الكهربائية

2.1 المُولد

ركزنا في هذه الدراسة على نظام ناقل لانهائي لآلة واحدة كما هو مبين بالشكل 1. وفي هذا النظام يكون المولد مزود بـ PSS وخط النقل مزود بـ TCSC ومعاوقة الخط $Z = R + jX$ وللمُولد حمل محلي بسماحية $Y_L = g + jb$. تم تمثيل المُولد بنموذج من الدرجة الثالثة يتضمن التأرجح الكهروميكانيكي والجهد الداخلي [1, 2]. يمثل التأرجح بالمعادلات التالية:

$$\dot{\delta} = \omega_b (\omega - 1) \quad (1)$$

$$\dot{\omega} = (P_m - P_e - D(\omega - 1)) / M \quad (2)$$

حيث P_m و P_e تمثلان قدرة الدخل والخرج للمُولد على التوالي، M تمثل ثابت القصور الذاتي، D تمثل معامل الخمد، δ تمثل زاوية العضو الدوّار، ω تمثل سرعة العضو الدوّار، ω_b تمثل سرعة

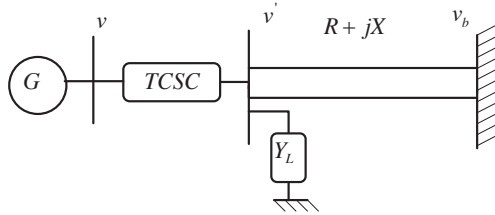
التزامن. يمكن التعبير عن قدرة الخرج للمولد بدلالة مركبات الإحداثيات d و q لتيار العضو المُنتِج (armature) i والجهد الطرفي v كما يلي:

$$P_e = v_d i_d + v_q i_q \quad (3)$$

ويحدد الجهد الداخلي E'_q من

$$\dot{E}'_q = (E_{fd} - (x_d - x'_d) i_d - E'_q) / T'_{do} \quad (4)$$

حيث E_{fd} تمثل جهد المجال، T'_{do} تمثل الثابت الزمني لمجال الدائرة المفتوحة، x'_d و x_d تمثلان الممانعة والممانعة العابرة للمولد باتجاه الأحداثي d على التوالي.



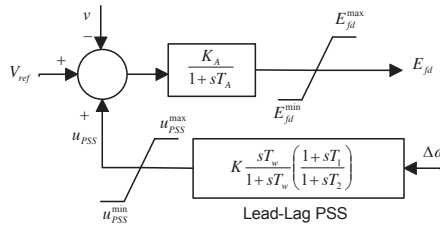
الشكل 1. نظام ناقل لانتهائي لآلة واحدة

2.2 نظام الإثارة و PSS

يمثل الشكل 2 نظام الإثارة المستخدم في هذه الدراسة، ويمكن وصفه بالمعادلة التالية:

$$\dot{E}_{fd} = (K_A (V_{ref} - v + u_{PSS}) - E_{fd}) / T_A \quad (5)$$

حيث K_A و T_A تمثلان الكسب والثابت الزمني لنظام الإثارة على التوالي، V_{ref} تمثل الجهد المرجعي. وكما هو مبين بالشكل 2 تم وضع حاكم متقدم - متأخر (Lead-Lag PSS) في مسار التغذية المرتجعة لتوليد إشارة الاتزان u_{PSS} .



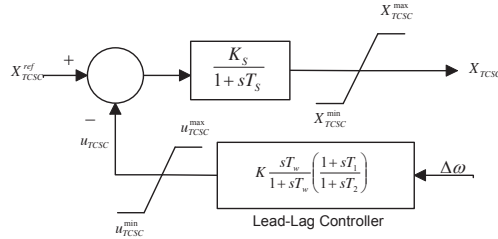
الشكل 2. نظام IEEE للإثارة بحاكم PSS

2.3 مثبت TCSC

يبين الشكل 3 مخططاً لمثبت TCSC بمُعوّض متقدم - متأخر، وتحدد قيمة الممانعة X_{TCSC} من

$$\dot{X}_{TCSC} = (K_s (X_{TCSC}^{ref} - u_{TCSC}) - X_{TCSC}) / T_s \quad (6)$$

حيث X_{TCSC}^{ref} تمثل الممانعة المرجعية لـ TCSC، K_s و T_s تمثلان الكسب والثابت الزمني لـ TCSC. وكما هو مبين في الشكل 3 تم وضع حاكم متقدم - متأخر تقليدي في مسار التغذية المرتجعة لتوليد إشارة الاتزان u_{TCSC} .



الشكل 3. مثبت TCSC بمُعوّض متقدم-متأخر

2.4 النموذج الخطي

عادةً ما يُستخدم النموذج الخطي التزايدي حول نقطة التشغيل الاسمية عند تصميم حاكمات الخمد للنمط الكهروميكانيكي [1، 2]. ينتج من عملية التحويل الخطي لنظام معادلات الحالة الآتية:

$$\begin{bmatrix} \dot{\Delta\delta} \\ \dot{\Delta\omega} \\ \dot{\Delta E'_q} \\ \dot{\Delta E'_{fd}} \end{bmatrix} = \begin{bmatrix} 0 & 377 & 0 & 0 \\ -\frac{K_1}{M} & -\frac{D}{M} & -\frac{K_2}{M} & 0 \\ -\frac{K_4}{T'_{do}} & 0 & -\frac{K_3}{T'_{do}} & \frac{1}{T'_{do}} \\ -\frac{K_A K_5}{T_A} & 0 & -\frac{K_A K_6}{T_A} & -\frac{1}{T_A} \end{bmatrix} \begin{bmatrix} \Delta\delta \\ \Delta\omega \\ \Delta E'_q \\ \Delta E'_{fd} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & -\frac{K_p}{M} \\ 0 & -\frac{K_q}{T'_{do}} \\ \frac{K_A}{T_A} & -\frac{K_A K_v}{T_A} \end{bmatrix} \begin{bmatrix} u_{PSS} \\ \Delta X_{TCSC} \end{bmatrix} \quad (7)$$

باختصار:

$$\dot{X} = AX + HU \quad (8)$$

وهنا يكون متجه الحالة $X = [\Delta\delta, \Delta\omega, \Delta E'_q, \Delta E'_{fd}]^T$ ، ومتجه التحكم $U = [u_{PSS}, \Delta X_{TCSC}]^T$ ، وتمثل K_1-K_6, K_p, K_q, K_v ثوابت لعملية التحويل الخطي.

3.1 تعريف النمط الكهروميكانيكي

يمكن استخدام معادلات الحالة للنموذج الخطي لتحديد القيم المميزة للمصفوفة A للنظام. من ضمن تلك القيم يوجد نمط للتذبذب مرتبط بالقصور الذاتي للآلة. ولكي تكون المثبتات فعالة فمن المهم جداً تعريف القيمة المميزة المرتبطة بالنمط الكهروميكانيكي. في هذه الدراسة استخدمنا طريقة عوامل المشاركة [10].

3.2 تصميم حاكمات الاتزان

يبين الشكلان 2 و 3 أحد النظم التي يكثر استخدامها على نطاق واسع لمثبتات الإثارة وتلك المعتمدة على TCSC. وفي كلا النظامين عادةً ما يتم تحديد قيم ثابتة مسبقاً للثوابت T_{11} و T_{22} . والمطلوب تحديد قيم للكسب K وللثابت الزمني T_1 لكل مثبت.

تم الأخذ بعين الاعتبار في هذه الدراسة عدة حالات عندما تكون معاملات النظام محددة بشكل دقيق وعندما لا تكون المعاملات دقيقة مع أحمال مختلفة تمثل التحميل الاعتيادي والخفيف والعالي والتحميل بمعامل قدرة متقدم. ويهدف التصميم إلى تحسين الخمد في النظام على مدى التشغيل الكامل بأحمال مختلفة. ولذلك تم تعريف دالة الهدف التالية التي تعتمد على القيم المميزة:

$$J = \max \{ \text{Real}(\lambda_i) \} \quad (9)$$

حيث λ_i تمثل القيمة المميزة للنمط الكهروميكانيكي عند ظروف التحمل i . وتهدف عملية الأمثلة إلى تحقيق القيمة الصغرى للدالة J مع تحقيق القيود على حدود قيم المعاملات. ولهذا يمكن صياغة التصميم كمسألة الأمثلة التالية:

$$\text{Minimize } J \quad (10)$$

Subject to

$$K^{\min} \leq K \leq K^{\max} \quad (11)$$

$$T_1^{\min} \leq T_1 \leq T_1^{\max} \quad (12)$$

$$T_3^{\min} \leq T_3 \leq T_3^{\max} \quad (13)$$

ولحل تلك المشكلة استخدمنا الخوارزميات الجينية المستمرة للبحث عن قيم المعاملات المثلى أو القريبة من المثلى. وقد قمنا بتصميم الحاكمات PSS و TCSC بشكل مستقل أولاً ثم بشكل متناسق بعد ذلك و قمنا بدراسة فعاليتها في جميع الحالات.

3.3 حساب معامل عزم الخمد

لتقييم مدى فعالية المثبتات المقترحة تم حساب معامل عزم الخمد. يمكن تحليل العزم إلى المركبات التزامن والخمد كما يلي:

$$\Delta T_e(t) = K_{syn} \Delta \delta(t) + K_d \Delta \omega(t) \quad (14)$$

حيث K_d و K_{syn} تمثلان معاملات عزم التزامن والخدم على التوالي. من الجدير بالذكر أن K_d هي مقياس للنمط الكهروميكانيكي للذبذبات [12]. ولكي نحسب K_{syn} و K_d نحسب أولاً الفرق بين انحراف العزم الحقيقي والعزم الناتج من جمع كلا المركبتين كالآتي:

$$E(t) = \Delta T_e(t) - (K_{syn} \Delta \delta(t) + K_d \Delta \omega(t)) \quad (15)$$

عندئذ يتم حساب K_{syn} و K_d بإيجاد القيم الصغرى لمجموع مربع الفروق على فترة المحاكاة t_{sim} كالآتي:

$$\sum_{i=1}^N [E_i]^2 = \sum_{i=1}^N [\Delta T_{ei} - (K_{syn} \Delta \delta_i + K_d \Delta \omega_i)]^2 \quad (16)$$

حيث $t_{sim} = N \times T_{samp}$ ، T_{samp} تمثل فترة أخذ العينات. ولذا يجب أن تحقق تلك المعاملات:

$$\frac{\partial}{\partial K_{syn}} \sum_{i=1}^N [E_i]^2 = 0 \text{ and } \frac{\partial}{\partial K_d} \sum_{i=1}^N [E_i]^2 = 0 \quad (17)$$

وبالتالي:

$$\sum_{i=1}^N \Delta T_{ei} \Delta \delta_i = K_{syn} \sum_{i=1}^N [\Delta \delta_i]^2 + K_d \sum_{i=1}^N [\Delta \omega_i \Delta \delta_i] \quad (18)$$

$$\sum_{i=1}^N \Delta T_{ei} \Delta \omega_i = K_d \sum_{i=1}^N [\Delta \omega_i]^2 + K_{syn} \sum_{i=1}^N [\Delta \omega_i \Delta \delta_i] \quad (19)$$

وبحل المعادلتين (18) و (19) نحصل على K_{syn} و K_d .

4 تحقيق الأسلوب المقترح

الخوارزميات الجينية هي خوارزميات بحث تعتمد على آليات الانتقاء الطبيعي ومبدأ البقاء للأصلح. ومن أهم سماتها كأسلوب لتصميم نظم التحكم هو تطلبها للحد الأدنى من المعرفة عن المحطة قيد الدراسة. فهي تقوم بأمثلة مؤشر الأداء باستخدام العلاقات بين المدخلات والمخرجات فقط، ولذا تحتاج لمعلومات أقل بكثير من طرق التصميم الأخرى. كما أنه بسبب توجيه عملية البحث باتجاه تحسين أداء معين، تكون المحصلة النهائية هي وحدة تحكم تحقق في نهاية المطاف معايير الأداء المطلوبة. كذلك نظراً لأن الخوارزميات الجينية لا تحتاج إلى وجود علاقة رياضية صريحة بين أداء النظام وعملية التحديث للبحث، فإنها توفر منهجية أمثلة أكثر عمومية من الطرق التحليلية التقليدية.

وبسبب مشاكل التمثيل الثنائي عند التعامل مع فضاء بحث متصل متعدد الأبعاد، قمنا بتطوير طريقة البحث عن المعاملات المثلى لنظم التحكم المقترحة معتمدة على استخدام الخوارزميات الجينية المستمرة [9]. وقد قمنا بتعريف دالة الهدف لتقييم الحلول المختلفة وذلك بتحويل نموذج النظام إلى نموذج خطي عند ظروف التحميل المختلفة على المدى المحدد وتعريف النمط الكهروميكانيكي وحساب معامل الخدم. في التجارب المختلفة التي أجريناها قمنا بتشغيل الخوارزمية الجينية عند قيم مختلفة للمعاملات وكانت القيم التالية هي الأكثر فعالية:

- معدل طفرة يساوي 0.01.
- معدل تبادل يساوي 0.9.
- حجم المجتمع يساوي 100.
- الحد الأقصى لعدد الأجيال يساوي 500.
- إنهاء البحث إذا لم يتغير الحل الأمثل لمدة 50 جيلاً.

5 النتائج والمناقشة

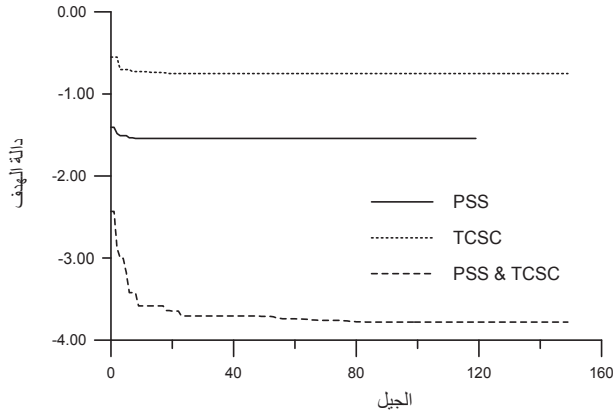
5.1 مثبتات الاتزان المقترحة وظروف التحميل المختلفة

في هذه الدراسة تم تحديد القيم المثلى لمعاملات نوعين من الحاكمات (PSS، TCSC) على نطاق واسع لظروف تشغيل وقيم غير دقيقة لمعاملات النظام. وقد أخذنا بعين الاعتبار أربع حالات مختلفة لظروف التحميل: اعتيادية (نمطي)، خفيفة، عالية، ذات معامل قدرة متقدم (leading power factor). وفي كل حالة قمنا بدراسة النماذج المختلفة في حالتين: ثبات معاملات النظام عند قيم محددة، وجود عدم دقة في قيم معاملات النظام. ويبين الجدول 1 الحالات المختلفة قيد الدراسة. وبناءً على ذلك يكون عدد النقاط تحت الدراسة لعملية التصميم هو 16.

الجدول 1. حالات الدراسة المختلفة

حالة التحميل	(P, Q) بوحدات pu	عدم دقة معاملات النظام
اعتيادي	(1.0, 0.015)	لا توجد
خفيف	(0.3, 0.1)	30% زيادة في ممانعة الخط X
عال	(1.1, 0.1)	30% نقصان في الثابت الزمني للمجال T'_{do}
معامل قدرة (pf) متقدم	(0.7, -1.3)	25% نقصان في القصور الذاتي للآلة M

تم تطبيق الأسلوب المقترح في نظام قوى كهربائية ذات اتصال ضعيف. وتوجد البيانات التفصيلية للنظام المستخدم في [1]. ويبين الشكل 4 تقارب دالة الهدف لكل من مثبتات الاتزان PSS و TCSC عند تصميمهما بشكل منفرد أو عند تصميمهما معاً بشكل متناسق (PSS&TCSC). وكما يظهر في هذا الشكل يكون التصميم المتناسق أفضل بكثير من استخدام كلٍ من المثبتات بشكل منفرد. ويبين الجدول 2 القيم النهائية للمعاملات المثلى لكل من مثبتات الاتزان المقترحة المختلفة.



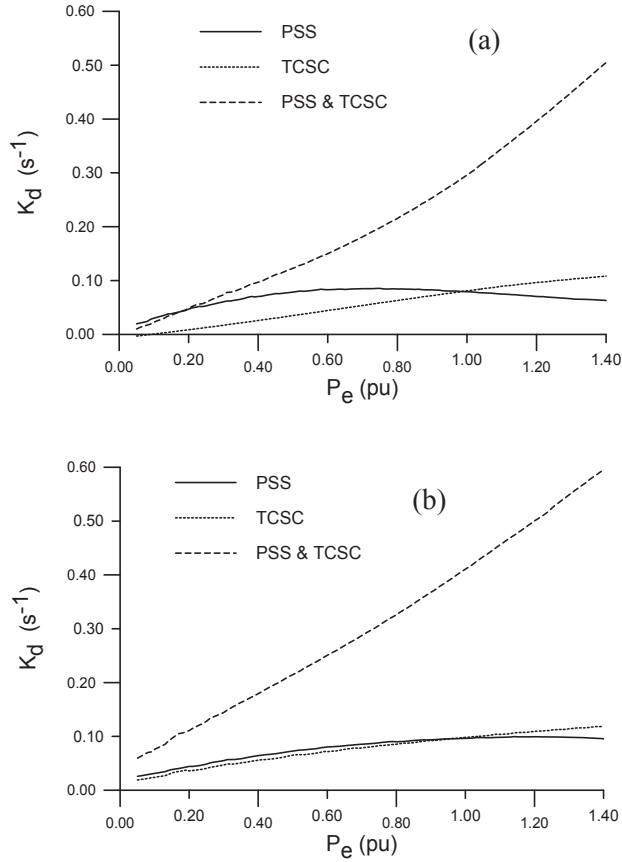
الشكل 4. تقارب دالة الهدف

الجدول 2. المعاملات المثلى للمثبتات المختلفة المقترحة

تصميم متناسق		تصميم مستقل		المعامل
TCSC	PSS	TCSC	PSS	
100	70.702	87.359	35.889	K
0.2026	0.1059	0.0402	0.2624	T_1
0.1	0.1	0.1	0.1	T_2

5.2 معامل عزم الخمد

تم تقدير معامل عزم الخمد لكي ندرس مدى فعالية مثبتات الاتزان المقترحة عند استخدامها بشكل منفرد أو بشكل متناسق. يبين الشكل 5 قيم K_d في مقابل التغيير في التحميل في الحالات: PSS فقط، TCSC فقط، PSS و TCSC معاً. ومن الملاحظ أن TCSC يعطي قيم سالبة لمعامل الخمد عند الأحمال المنخفضة عندما يكون معامل القدرة متقدماً. ويمكن ملاحظة أيضاً أن PSS أفضل من TCSC ولا يعاني من تلك المشكلة. وقد أدى استخدام التصميم المتناسق إلى تلافي حدة تلك المشكلة. كما أنه يوفر خمد أكبر بكثير من مثبتات الاتزان المنفردة مما يؤدي إلى تحسين ملحوظ في اتزان النظام.



الشكل 5.3. معامل الخمد عند تغير التحميل (a) $Q = -0.4$ (b) $Q = 0.4$

5.3 تحليل القيم المميزة والمحاكاة اللاحظية

ومن أجل شمولية الدراسة والتحقق من أداء التصميمات المختلفة تم اختبار جميع مثبتات الاتزان المقترحة عند التشويشات وظروف التحميل التالية:

- (a) أحمال اعتيادية $P = 1$ و $Q = 0.015$ وحدة قدرة (pu) وبخطاً سداسي الدورة.
- (b) أحمال خفيفة $P = 0.3$ و $Q = 0.015$ وحدة قدرة (pu) وبخطاً سداسي الدورة.
- (c) أحمال عالية $P = 1.1$ و $Q = 0.4$ وحدة قدرة (pu) وبخطاً ثلاثي الدورة.

تبين الجداول 3 و4 و5 القيم المميزة للنظام عند عدم استخدام المثبتات المُقترحة وعند استخدامها تحت تلك الحالات الثلاثة على التوالي. يمثل الصف الأول في كل من تلك الجداول الثلاث القيم المميزة للنمط الكهروميكانيكي. ومن الواضح أن هناك تحسن كبير في اتزان النظام مع استخدام المثبتات المُقترحة. ويظهر أيضاً تفوق التصميم المتناسق على التصميمات المنفردة عند جميع النقاط التي درست من خلال تحسن قيم معاملات الخمد للأنماط الكهروميكانيكية في جميع الحالات وذلك بإزاحتها إلى يسار مستوى التردد (s-plane). كذلك قمنا بعمل محاكاة للتصميمات المُقترحة المختلفة لتقييم مدى فعاليتها في النطاق الزمني (time domain). ويبين الشكل 6 استجابة النظام عند التسويشات والأحمال المختلفة قيد الدراسة. ومن خلال هذا الشكل يتبين أن التصميم المتناسق يُعطي أفضل خصائص خمد ويحسن بشكل كبير الاتزان بعد أول ذبذبة.

الجدول 3. القيم المميزة للمثبتات المُقترحة المختلفة عند التحميل القياسي

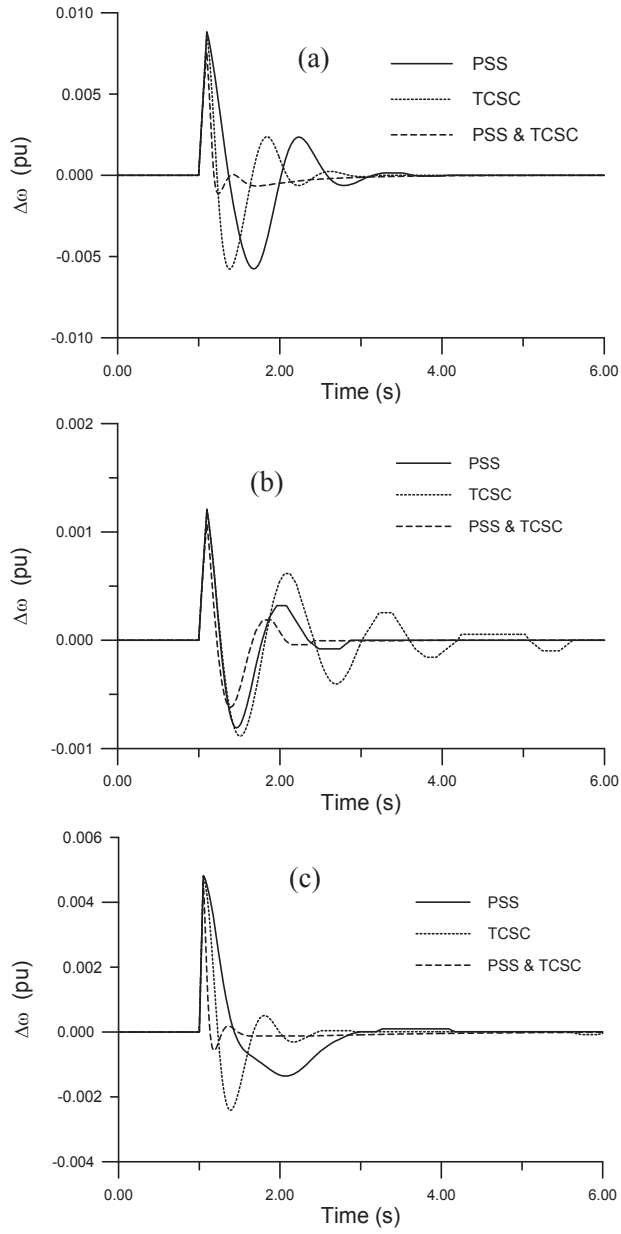
بدون تحكم	PSS فقط	TCSC فقط	تصميم متناسق
$+0.295 \pm j4.957$	$-2.233 \pm j9.616$	$-2.641 \pm j7.878$	$-7.846 \pm j7.025$
$-10.39 \pm j3.287$	$-2.595 \pm j2.699$	$-4.297 \pm j2.842$	$-10.47 \pm j12.6$
--	$-20.53, -0.209$	$-23.03, -13.28$	$-12.15, -10$
--	--	-0.209	$-1.387, -0.234$
--	--	--	-0.2

الجدول 4. القيم المميزة للمثبتات المُقترحة المختلفة عند التحميل الخفيف.

بدون تحكم	PSS فقط	TCSC فقط	تصميم متناسق
$-0.009 \pm j4.850$	$-2.043 \pm j5.060$	$-0.721 \pm j5.346$	$-2.836 \pm j7.009$
$-10.09 \pm j3.834$	$-4.455 \pm j5.719$	$-9.79 \pm j3.64$	$-16.21 \pm j4.732$
--	$-17.2, -0.204$	$-21.12, -8.05$	$-10, -8.221$
--	--	-0.203	$-3.869, -0.211$

الجدول 5. القيم المميزة للمثبتات المُقترحة المختلفة عند التحميل العالي

بدون تحكم	PSS فقط	TCSC فقط	تصميم متناسق
$+0.485 \pm j3.69$	$-2.938 \pm j9.038$	$-2.757 \pm j8.304$	$-8.528 \pm j6.028$
$-10.58 \pm j3.696$	$-1.92 \pm j2.251$	-23.93	$-11.51 \pm j15.76$
--	$-20.46, -0.215$	$-10.64, -8.159$	$-10, -9.434$
--	--	$-1.926, -0.227$	$-0.555, -0.32$
--	--	--	-0.2



الشكل 6. استجابة سرعة العضو الدوّار (a) حمل اعتيادي، (b) حمل خفيف، (c) حمل عال

6 الخاتمة

قدم هذا البحث دراسة لتصميم حاكمات لنظم القوى الكهربائية يعتمد على التطبيق المستقل والمنسق لمثبتات PSS و TCSC. وقد تم تطوير دالة هدف تعتمد على القيم المميزة (eigen values) لزيادة الخمد في النظام. وتم استخدام الخوارزميات الجينية المستمرة للبحث عن قيم المعاملات المثلى للمثبتات. بالإضافة لذلك تم تقييم مدى كفاءة المثبتات المقترحة وتم تقييم خصائص الخمد للأساليب المقترحة من خلال تقدير معامل عزم الخمد. ومن خلال اختبار المثبتات المقترحة في نظام قوى كهربائية ذات اتصال ضعيف وظروف تحميل مختلفة تبين مدى فعاليتها ومثابرتها لتحسين اتزان النظام. كما تبين تفوق المثبتات المتناسقة على الأساليب الأخرى التي تستخدم المثبتات بشكل مستقل.

7 شكر وتقدير

يتوجه الباحثان بالشكر لجامعة الملك فهد للبترول والمعادن بالمملكة العربية السعودية لما قدمته من دعم لهما أثناء القيام بهذا البحث.

المراجع

- [1] J. Machowski, J. Bialek, and J. Bumby, *Power System Dynamics: Stability and Control*, 2nd Edition, Wiley-Blackwell, 2008.
- [2] P. W. Sauer and M. A. Pai, *Power system Dynamics and Stability*, Prentice Hall, 1998.
- [3] T. C. Yang, "Applying H_{∞} Optimisation Method to Power System Stabilizer Design Parts 1&2," *International Journal of Electrical Power & Energy Systems*, Vol. 19, No. 1, 1997, pp. 29-43.
- [4] P. Kundur, M. Klein, G. J. Rogers, and M. S. Zywno, "Application of Power System Stabilizers for Enhancement of Overall System Stability," *IEEE Transactions on Power Systems*, Vol. 4, No. 2, 1989, pp. 614-626.
- [5] M. A. Abido and Y. L. Abdel-Magid, "Robust Design of Multimachine Power System Stabilizers Using Tabu Search Algorithm," *IEE Proceedings-Generation, Transmission, Distribution*, Vol. 147, No. 6, November 2000, pp. 387-394.
- [6] N. G. Hingorani, "Flexible AC Transmission," *IEEE Spectrum*, April 1993, pp. 40-45.
- [7] M. Noroozian and G. Anderson, "Damping of Power System Oscillations by Use of Controllable Components," *IEEE Transactions on Power Systems*, Vol. 9, No. 4, 1994, pp. 2046-2054.
- [8] H. F. Wang and F. J. Swift, "A Unified Model for the Analysis of FACTS Devices in Damping Power System Oscillations Part I: Single-machine Infinite-bus Power Systems," *IEEE Transactions on Power Systems*, Vol. 12, No. 2, 1997, pp. 941-946.
- [9] M. A. Abido and Y. L. Abdel-Magid, "Analysis and Design of Power System Stabilizers and FACTS Based Stabilizers Using Genetic Algorithms," in *Proceedings of Power System Computation Conference, PSCC-2002, Spain, June 24-28, 2002*.

- [10] Y. Y. Hsu and C. L. Chen, "Identification of Optimum Location for Stabilizer Applications Using Participation Factors," *IEE Proceedings, Part C*, Vol. 134, No. 3, May 1987, pp. 238-244.
- [11] A. M. A. Hamdan, "An Investigation of the Significance of Singular Value Decomposition in Power System Dynamics," *International Journal of Electrical Power and Energy Systems*, Vol. 21, 1999, pp. 417-424.
- [12] E. A. Failat. M. Bettayeb, H. Al-Duwaish. M. A. Abido, and A. Mantawy, "A neural Network Based Approach for On-line Dynamic Stability Assessment Using Synchronizing and Damping Torque Coefficients," *Electrical Power Systems Research*, Vol. 39, No. 2, 1996, pp. 103-110.
- [13] N. A. Singh, K. A. Muraleedharan, and K. Gomathy, "Particle Swarm Intelligence Tuned Fuzzy Controller for Damping Modal Oscillations of Power System," *IEEE Region 10 Conference*, (TENCON 2008), Nov. 2008.
- [14] M. Dubey and A. Dubey, "Genetic Algorithm Based Design of Fuzzy Logic Power System Stabilizers in Multimachine Power System," *World Academy of Science, Engineering and Technology*, vol. 66, pp. 975-982, 2010.
- [15] A. A. Ismail, H. K. M. Youssef, and H. M. Zeineldin "Comparative Study Between TCSC and PSS in Damping Electro-Mechanical Oscillations," *Proceedings of the 14th International Middle East Power Systems Conference*, (MEPCON'10), Cairo University, Egypt, December 2010.

الملخص باللغة الانجليزية

Title: Design of Damping Controllers in Electrical Power Systems Using Continuous Genetic Algorithms

Abstract. In this paper, we investigate the enhancement of the electrical power systems stability through the robust coordinated design of stabilizers. The problem is formulated as an optimization problem with an eigenvalue-based objective function under a wide range of loading conditions and system configurations. Then we developed and applied a continuous genetic algorithm to search for the optimal controller parameters. The damping characteristics of the proposed control schemes are evaluated in terms of the damping torque coefficient over a wide range of loading conditions and the proposed stabilizers were tested on a weakly connected power system. The simulation results showed the effectiveness and robustness of the proposed schemes over the considered loading conditions.

Keywords: Electrical power systems, Continuous genetic algorithms, Stabilizers, Damping controllers.

المصطلحات

Admittance	السماحية
Armature	العضو المنتج
Binary representation	تمثيل ثنائي
Controllers	حاكمات
Damping	خمد
Damping torque coefficient	معامل عزم الخمد
Disturbance	تشويش
Eigenvalues	قيم مميزة
Electrical power system	نظم القوى الكهربائية
Gain	الكسب
Generator	مولد
Genetic algorithms	الخوارزميات الجينية
Heavy load	حمل عالٍ
Impedance	المعاوقة
Inertia	القصور الذاتي
Lagging power factor	معامل قدرة متأخر
Leading power factor	معامل قدرة متقدم
Light load	حمل خفيف
Rotor speed	سرعة العضو الدوار
Normal load	حمل اعتيادي
Reactance	ممانعة
Response	استجابة
Stability	اتزان، استقرار
Stabilizer	مثبت اتزان
Stabilizing signal	إشارة الاتزان
State equations	معادلات الحالة
Time constant	الثابت الزمني
Transient reactance	الممانعة العابرة

سنبلة: منهج دلالي لتحسين استعمالية المعرفة المكتسبة خلال دورة حياة المنتج الصناعي

محمد فؤاد سريتي

جامعة الإمام محمد بن سعود الإسلامية، كلية علوم الحاسب والمعلومات
ص.ب. 93926، الرمز البريدي 11683، الرياض، المملكة العربية السعودية

البريد الإلكتروني: mfsriti.ccis@gmail.com

الخلاصة. إدارة دورة حياة المنتج هو مجال صناعي واسع جدا ويقدم تحديات جديدة لهندسة المعرفة. يواجه هذا المجال عدة مشكلات وخاصة في مشاركة، وتبادل، وإعادة استعمال المعرفة المتعلقة بالمنتج. فحل هذه المشكلات فقد أصبح اليوم غير كافي اللجوء إلى إدارة البيانات والمعلومات التقنية للحصول على مناهج متقدمة لتحديد المعلومة المفيدة وتقديمها للشخص المناسب وفي الوقت المناسب. ضمن هذه التحديات ولحل المشاكل المذكورة نشأ منهج سنبلة لتمثيل المعارف في شكل موحد وديناميكي (لدمج وتبادل البيانات) باستعمال تقنية العنكبوتية الدلالية والأنطولوجية ولضمان استعمالية المعارف بطريقة أفضل (تصفح المعارف بطريقة متعددة الأوجه) وهذا باستغلال النموذج هايبرتوبيك.

الكلمات الجوهرية: العنكبوتية الدلالية، أنطولوجية، هندسة المعرفة، إدارة دورة حياة المنتج.

1 مقدمة

منذ الثورة الصناعية والمحيط الإقتصادي والصناعي يخضعان لتحولات وتغيرات مهمة حيث أن تطور الإنتاج والتقنية لعبا فيها دوران مهمان. سعيا وراء الأرباح، ركزت الشركات في بادئ الأمر على إدارة المشروع للحصول على المنتج بكميات كبيرة ثم ركزت بعد ذلك على إدارة الجودة للحصول على المنتج الجيد والتميز بين المنافسين. سرعان ما عمّ تطبيق هذه المناهج وقل الربح المتعلق بها. ويرجع هذا أيضا إلى "العولمة، والطاقة الإنتاجية الفائضة في بعض الصناعات، وتطوير مهن جديدة وخاصة في مجال الخدمات" [1]. فكل هذه العوامل أدت بالشركات للتوجه إلى البحث العلمي والابتكار من أجل تقديم المنتج الأفضل ويتوقف هذا [2] على قرارات الشركة ومعارفها، فأصبح التميز يقتضي إذا [3] التحكم في رؤوس أموالها الفكرية (المعرفة) والعمل على إنمائها.

فلهذا الغرض تبحث الشركات المصنّعة اليوم عن إجابة للسؤال التالي: "بدأ مشروع جديد، فما هي المشاريع السابقة التي قدّمت حلول كلية أو جزئية للمشكلة الحالية؟".

فلإجابة عن هذا السؤال نقدم من خلال هذا البحث حلولاً تأخذ بعين الاعتبار إدماج مصادر المعرفة المختلفة والمستعملة داخل الشركة أو خارجها (من طرف شركائها)، توحيد بنية تمثيل المعارف، سهولة تحديث نموذج البيانات، وأخيراً تبويب المعارف وسلسلة البحث عنها وسرعة العثور عليها.

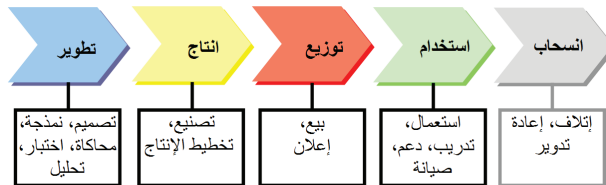
فيما يلي وفي الباب الثاني نقوم باستعراض المنتج خلال دورة حياته. وبعد ذلك نطرح في الباب الثالث المشاكل التي تعرقل استعمالية المعرفة المتعلقة بالمنتج. وفي البابين الرابع والخامس نتطرق إلى العنكبوتية الدلالية والنموذج هايبرتوبيك كتقنيتان لحل المشاكل المطروحة. نقترح بعد ذلك منهج ومنصة لتنفيذ الحلول المقترحة وذلك في البابين السادس والسابع. أما في الباب التاسع فتتطرق إلى دراسة تجريبية للمنهج المقترح. وأخيراً نختم البحث في الباب التاسع.

2 المنتج خلال دورة حياته

يمر المنتج الصناعي خلال دورة حياته بعدة مراحل (أنظر الشكل 1) من أبرزها: مرحلة التطوير، مرحلة الإنتاج، ومرحلة التوزيع كما يمكننا أيضاً إضافة مرحلتين الاستخدام (والتي تتضمن الصيانة) والانسحاب (وتمثل نهاية حياة المنتج). نستعرض في بداية هذا الباب مرحلة تطوير المنتج، وبعدها دورة حياته، وفي آخر الباب خصائصه.

2.1 تطوير المنتج

نركز في هذه الدراسة على مرحلة التطوير والتي يتدخل فيها كلٌّ من موظفي الشركة وشركائها. تعتبر مرحلة التطوير أهم مراحل المنتج الصناعي، لأنه يتم فيها تصميم المنتج وفقاً لكثير من القيود كمواصفات العملاء، وإجراءات الشركة، ومعايير قياسية للجودة، وغيرها من القيود. اعتماداً على هذه الموارد المعلوماتية يتم تعريف المنتج بواسطة أداة التصميم بمساعدة الحاسب بإنشاء نماذج ثلاثية الأبعاد. كما أن الملفات الرقمية التي تم إنشاؤها بواسطة هذه الأداة يجب أن تُدار وتشارك بطريقة فعالة لكي يُأخذ بعين الاعتبار مختلف نسخ هذه الملفات، ومختلف مركبات المنتج وتشكيلاته.



الشكل 1. دورة حياة المنتج

لضمان أحسن إدارة وأحسن مشاركة، تبنت الشركات المصنعة اليوم استراتيجية إدارة دورة حياة المنتج المذكورة فيما يلي.

2.2 إدارة دورة حياة المنتج

يُنظر لإدارة دورة حياة المنتج (يرمز لها فيما بعد بـ: إ.د.ح.م.) من قبل سيمداتا [3] كـ "منهج استراتيجي للشركة والذي يطبق مجموعة متسقة من الحلول التي تدعم الإنشاء التعاوني، التسيير، النشر واستخدام المعلومات المعرّفة للمنتج في جميع أنحاء الشركة الموسعة من مرحلة التصميم إلى نهاية حياة المنتج – شاملة للأشخاص، والعمليات، ونظم الحاسب".

من هذا التعريف نستخلص أن إ.د.ح.م. هي إستراتيجية تتضمن في نفس الوقت منهجية وتقنية. تسمح هذه الإستراتيجية بتحسين كفاءة الموظف وذلك بالتخفيض من التعقيد للدخول والحصول على معلومات الشركة. وقد تسمح أيضا بتحسين التعاون المتعدد التخصصات وذلك بتطوير وسائل المشاركة والتنسيق بين موظفي الشركة وبين الشركة وشركائها.

عموما، فإن إ.د.ح.م. تحاول تلبية جميع المتطلبات حول البيانات المتعلقة بالمنتج. وعلى ذلك فهي تشمل إدارة الوثائق، إدارة العمليات، وغيرها من التطبيقات لمساندة العمل التعاوني.

2.3 خصائص المنتج

غالبا ما يُمثل المنتج في منهج إ.د.ح.م. بمجموعة من العناصر. ونعرّف العنصر في هذا المجال كـ "وحدة أساسية مسيّرة في الشركة. هذه الوحدة تجمع وتُغلف المعلومات المتعلقة بمجموعة فرعية لمنتج" [4]. هذه المجموعة الفرعية من المعلومات تكمن في صنفين هما:

1. معلومات تصف المنتج، أو الوثائق، أو الوحدة السابق ذكرها، وتكون من أنواع شتى، منها: أشخاص (مالك، آخر تعديل، مجموعة المستخدمين، دور أو وظيفة، إلخ)، تاريخ (الإصدار، مؤشرات تصحيح، تاريخ الصنع، التعديل، الموافقة عليها، النشر، إلخ)، خصائص قابلة للقياس (الوزن والكتلة والكثافة، المادة، اللون، إلخ).
2. وثائق مرتبطة بالمنتج، ومنها: نماذج هندسية ثلاثية الأبعاد، رسومات (تخطيط، مقاطع، إلخ)، قواعد العمل، معايير قياسية، مواصفات تقنية ووظيفية، طلب تعديل وغيرها من الوثائق.

كما يمكن لكل وحدة أن تتكون من عدة وحدات أخرى. يتعلق تعريف الوحدة بحاجيات ومنتجات كل شركة صناعية. فنجدها مثلا في مجال تصنيع السيارات تمثل قطعة ميكانيكية وأما في مجال الصناعات الكيميائية تمثل جزيء كيميائي وهكذا.

3 استعمالية المعرفة

في بداية كل مشروع في شركة ما، يكون من الأهمية بمكان القدرة على عثور المعلومات ذات الصلة والتي تم إنتاجها وتخزينها في المشاريع السابقة لإعادة استعمالها في المشروع الحالي. لهذا، فمن الضروري توفير الوصول إلى مختلف نظم الشركة وشركائها. سنطرح في هذا الباب المشاكل التي تعرقل إعادة استعمال المعارف المكتسبة سابقا والمتلخصة في: تنوع النظم المستخدمة، تطور

تنظيمات وتركيبات البيانات من حين إلى آخر، وفي الأخير، وظائف البحث المحدود والتي تفرض على المستعمل أن يكون على علم مسبق بالمحتوى قبل البحث.

3.1 تنوع النظم

إدارة دورة حياة المنتج تدعم في الوقت الراهن العديد من الأنشطة الموزعة والأنظمة الغير متجانسة وذلك لإنشاء وتخزين ومشاركة المعلومات المتعلقة بالمنتج. و يتطلب هذا وفقاً [5] آليات قوية لدمج وتبادل المعلومات. فبالنظر، فمن الضروري مواجهة شبكة كبيرة من الشركات، وتنوع تمثيل البيانات، وتأمين الاتصالات.

الأسباب المؤدية إلى استعمال أنظمة مختلفة وغير متجانسة بين الشركة وشركائها، وحتى بين أقسام الشركة نفسها، هي عديدة ولا يتسع ذكرها في هذا المقال. وكما هو معلوم أن على هؤلاء الشركاء والأقسام التعاون والعمل على كائنات مشتركة. ولكن يتجلى لنا بكل وضوح أن لكل كائن عدة تمثيلات منفصلة عن بعضها البعض من قسم إلى آخر ومن شركة إلى أخرى.

يقدم هذا الوضع في سلامة المعلومات المتعلقة بهذه الكائنات وتناسق البيانات. فالحاجة تقتضي إلى إتاحة مختلف الشركاء الوصول إلى مختلف هذه المعلومات والبيانات المشتركة مع العلم أنها مخزنة في أنظمة مختلفة.

لحل هذه المشكلة، يمكننا جعل أنظمة إ.د.ح.م. قابلة للتشغيل المتبادل والعمل بانسجام باستعمال أساليب الدمج وتبادل البيانات.

3.2 تطور تنظيمات البيانات

المشكلة الثانية التي تواجه إعادة استعمال المعارف هي التحديث المستمر في تنظيمات وتركيبات البيانات في الشركة. بصفة عامة، تكنولوجيا إ.د.ح.م. تطور نموذج البيانات المرتبط بتنظيم المؤسسة وعملياتها. وبعد ذلك يمكن لهذا النموذج الخضوع لبعض التحديثات كإضافة فئة جديدة من المنتج أو التعامل مع شريك جديد وهذه الإضافة تقتضي بدورها تعديل بعض عمليات الشركة ودمج نظام الشريك مع نظام الشركة.

ولتطوير نموذج البيانات تسلك الشركة الخطوات التالية:

1. إعادة تصميم النموذج
2. دراسة أثر وتنفيذ التغيرات على أنظمة الشركة الموجودة
3. وبعد ذلك، ترحيل البيانات الحالية إلى النمذجة الجديدة

هذه التحديثات هي أمر مكلف للشركة وذلك من الناحية الفنية. فلهاذا الحاجة هنا هي جعل نموذج البيانات مرناً وديناميكياً للأخذ بعين الاعتبار وبسرعة التغيرات والتطورات الطارئة على هذا النموذج.

3.3 وظيفة البحث محدودة

منهجية وتقنية إدارة دورة حياة المنتج تركز بالأحرى على السير الحسن للمشروع الجاري التنفيذ. لإعادة استعمال البيانات من مشاريع ماضية فالسبيل الوحيد للوصول إليها يكون باستعمال وظائف البحث بما في ذلك البحث المتعدد المعايير والبحث في النص.

عموماً، وظائف البحث الحالية تفترض أن المستخدم يعرف مسبقاً ما يريد البحث عليه: يجب أن يعرف العنصر، نوع العنصر، أو قيم خصائص العنصر أو محتوى الوثيقة التي يبحث عليها. ولكن يكون صعباً جداً على المستعمل تحديد مجموعة من العناصر أو الوثائق التي قد تهمة أو تعينه على مهمة معينة بدون أن يكون مطلع مسبقاً على خصائصها أو محتواها. من الناحية المثالية، إذا كانت عناصر المنتج والوثائق الإلكترونية المتعلقة به مفهومة ومصنفة على مواضيع مفيدة، فيمكن للمستعمل تصفح قائمة (أو تسلسل هرمي) لهذه المواضيع ثم العثور على العناصر والوثائق التي تهمة. وعلى هذا فإنه من المهم جداً تزويد تقنية إ.د.ج.م. بوظيفة تصفح محتواه وفقاً لمصالح ووجهات نظر مستعمليه ومواضيع احترافية ذات صلة باختصاص الشركة.

4 العنكبوتية الدلالية كتقنية دمج وتوفيق البيانات

لحل مشكلة اختلاف النظم السابق ذكرها نلجأ إلى منهج مشاركة وتبادل البيانات والذي أصبح من ضروريات الشركة الموسعة من أجل تطوير منتجات جديدة وبسرعة [6]. في الواقع، تبادل البيانات هو من أساسيات إدارة دورة حياة المنتج [7] وهو حجر أساس لعدة تطبيقات كمشاركة المعلومات ودمج التطبيقات وإعادة استعمال المعارف. وفقاً لـ [8] فإن الدمج الذي يتم من خلاله تبادل البيانات يعرف بأنه العملية التي من خلالها دمج عدة مصادر مستقلة وموزعة وغير متجانسة في مصدر واحد للبيانات. ولهذا يجب استعمال تمثيل موحد لتسهيل عملية الدمج والتبادل. كما يجب أيضاً التوفيق والمطابقة بين التمثيلات المنفصلة لبيانات المنتج والموجودة في نظم متعددة.

4.1 أساليب دمج وتبادل البيانات

هناك العديد من أساليب دمج وتبادل البيانات ولكن نقتصر هنا بذكر التصنيف الذي نقترحه لهذه الأساليب¹ ونقسمها إلى ثلاثة أصناف:

1. استعمال قواعد البيانات
 2. استعمال معايير قياسية للتبادل
 3. استعمال الأنطولوجية والعنكبوتية الدلالية
- من ميزات الصنف الأول أنه يتحلى بكل محاسن قواعد البيانات كالاستمرارية والفعالية غير أنه يفقد هذا الصنف الكثير من المرونة في التعديلات.

¹ للتفصيل في أنواع أساليب الدمج ومقارنة محاسنها ومساوئها يرجى الرجوع إلى [4].

أما الصنف الثاني يكون باستعمال معايير التبادل. ولكن هذه المعايير غالبا ما تخص مجال صناعي معين وتستعمل ملفات ضخمة وتحتاج إلى زمن كبير للمعالجة.

أما الصنف الثالث يستعمل معايير من العنكبوتية الدلالية مفتوحة على كل المجالات وتقدم تمثيلات مرنة ومساعدة على تطبيق وتوافق البيانات.

4.2 المركبات

في طريقنا للبحث على أفضل تمثيل للبيانات المتعلقة بدورة حياة المنتج، سندرس هنا تقنية العنكبوتية الدلالية ومختلف جوانبها التي تدعمها لتمثيل، وتبادل ومشاركة المعلومات المتعلقة بالمنتج.

العنكبوتية الدلالية وفقا لمبتكرها [9] هو "امتداد للعنكبوتية الحالية والذي تملك من خلاله الموارد المعلوماتية على وصف غني لمحتواها سامحا للمستخدم البشري والآلة العمل معا بطريقة أفضل". وبلفظ آخر، تنص العنكبوتية الدلالية على جعل المحتوى الدلالي للموارد قابل للفهم ليس من طرف المستعملين فحسب بل أيضا من طرف البرامج. وللوصول إلى هذا الهدف، تقترح العنكبوتية الدلالية المركبات التالية:

- البيانات الوصفية: هي مجموعة من العلامات أو المعلومات تساعد على وصف واستغلال الموارد
 - الموارد: هي الوثائق الرقمية أو أي مورد يحتوي على المعلومات
 - اللغات: هم عبارة عن معايير قياسية لتمثيل الموارد وذلك استنادا على البيانات الوصفية
 - الأنطولوجية: تعتمد على لغات العنكبوتية الدلالية لنمذجة وتمثيل شكلي لمجموعة من المفاهيم المتعلقة بمجال معرفة معين (مثال مجال الهندسة أو الطب) وهذا التمثيل الشكلي يسهل استغلاله من طرف التطبيقات المتعلقة بهذا المجال.
- سنهتف فيما يلي بعرض مفصل للغات والأنطولوجية.

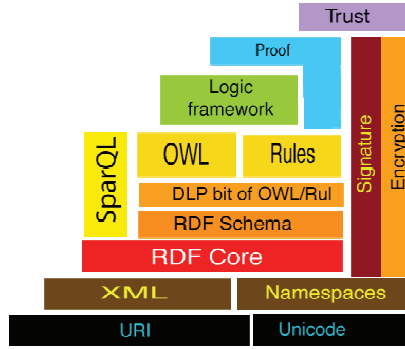
4.3 المعايير القياسية

تعرف العنكبوتية الدلالية مجموعة من اللغات وهي عبارة عن معايير قياسية على مستويات مختلفة انطلاقا من ترميز الحروف حتى يصل إلى توفير الأنطولوجيات لنظم الاستنتاج التلقائي (أنظر الشكل 2 الذي يعرض "كعكة" العنكبوتية الدلالية المقترحة من طرف تيم برنرز لي).

أبرز هذه المعايير وأهمها هما اثنان: لغة ترميز موسعة (يرمز لها فيما بعد بـ: ل.ت.م.) و / إطار توصيف الموارد (يرمز له فيما بعد بـ: إ.ت.م.) .

فبالنسبة للغة ل.ت.م. [10] تعتبر كلفة أساسية في النمذجة داخل العنكبوتية الدلالية وتتميز بمرونة كبيرة في التعبير عن محتوى الموارد. في سياق وصف موارد المتعلقة بدورة حياة المنتجات،

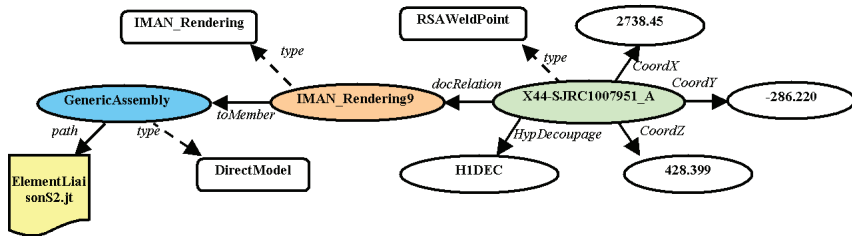
يستخدم ل.ت.م. على نطاق واسع [11]، [12]، [13]، [14] لدعم البيانات الغير متجانسة، وتطبيقات الدمج وتبادل المعلومات.



الشكل 2. مكدس العنكبوتية الدلالية²

ولكن في الواقع ل.ت.م. موجه للتمثيل التركيبي للمحتوى. وعلاوة على ذلك، فإنه لا يأخذ في الاعتبار المحتوى الدلالي لموارد إدارة دورة حياة المنتج [15].

وأما بالنسبة للغة إ.ت.م. [16] تستعمل غالبا في تمثيل مخططات لوصف بطريقة شكلية لموارد العنكبوتية وهذا التمثيل يمنح وصف دلالي لتلك الموارد. كما نُصِّ في [17] أن القيمة الحقيقية ل.ت.م. تكمن في سهولة النموذج الذي يقترحه والمتكون من ثلاثية: موضوع ومسند وهدف. وبعبارة أخرى، كل مورد (موضوع) يستطيع أن يكون له خصائص (مسندات) مع قيم (أهداف) لهذه الخصائص. واستنادا إلى هذا المنظور البسيط، فإن العناصر وخصائصها التي تتناولها نظم إ.د.ح.م. يمكن لها أن تصاغ بدقة.



الشكل 3. تمثيل بيانات المنتج بلغة إ.ت.م.

المثال في الشكل 3 يوضح ثلاثة موارد من إ.د.ح.م. الأول من اليمين تمثل عنصر، والثاني علاقة، وأما الثالث فهو عنصر مركب. نركز في هذا المثال على العنصر الأول فنلاحظ أنه يملك عدة

² [http://www.w3.org/2005/Talks/1110-iswc-tbl/#\(12\)](http://www.w3.org/2005/Talks/1110-iswc-tbl/#(12))

ميزات كالنوع، والإحداثيات س، ع، ص، وغيرها. فهذه طريقة تخطيطية لعرض الموارد وخصائصها على شكل ثلاثيات في لغة إ.ت.م.

كما أنه من المهم جدا [18] معرفة أن الأوصاف التي يمثلها إ.ت.م. يمكن لها أن تضبط بواسطة مخطط /إطار توصيف الموارد (يرمز له فيما بعد بـ: م.إ.ت.م.). يتركب م.إ.ت.م. من مجموعة مفردات، كالصنف والخاصية. وهذه المفردات تعمل مباشرة على المستوى الدلالي لنموذج البيانات بدلا من المستوى التركيبي، مما يجعل إ.ت.م. أبلغ من ل.ت.م. في وصف المعلومات ذات الصلة بدورة حياة المنتج.

وقبل أن نشرع في التكلم على الأنطولوجية، نريد التنويه على أن م.إ.ت.م. يعرف لغة بسيطة للأنطولوجية. ومع ذلك، إذا احتجنا إلى دلالة دقيقة وأكثر وضوحا فمن الأفضل أن تستعمل لغة وجودية العنكبوتية لقدرتها على تمثيل قيود أكثر تعقيدا [19].

4.4 الأنطولوجية

بعد دراستنا لعدة نماذج للبيانات المستعملة من طرف نظم إ.د.ج.م. وجدناها تشترك في ثلاثة مكونات: (1) أنواع العناصر و (2) خصائص كل نوع و (3) أنواع العلاقات التي يمكن أن توجد بين العناصر.

وتمثل الأنطولوجية بدورها شبكة دلالية، أو بعبارة أخرى، مجموعة من المفاهيم تخص مجال معين ومعرفة بالبيانات الوصفية ومتراطة بواسطة علاقات نوعية. ومن هنا يمكننا أن نقول أنه من السهل تمثيل نموذج البيانات المتعلق بنظام إ.د.ج.م. بالأنطولوجية بدون فقدان لأي معلومة.

والذي يسعنا أن نقول هو أن أهم خصائص الأنطولوجية المفيدة لـ إ.د.ج.م. هي: (1) أنها قابلة للمشاركة والاستخدام من طرف عدة تطبيقات، (2) أنها قابلة للتحديث بدون تكلفة في أي لحظة خلال دورة حياة المنتج، و أخيرا (3) أنها قابلة للتحويل من معيار قياسي إلى آخر.

4.5 الفوائد

استعمالنا للعنكبوتية الدلالية في بحثنا سيسهل علينا عمليتا الدمج والتبادل باستعمال تمثيل موحد، ومرن وديناميكي. كما أنه يجعل من الممكن قابلية تحديث تنظيم البيانات بدون خوف من التأثير على الموجود.

أما بالنسبة إلى التوفيق بين البيانات الموجودة في مصادر مختلفة فإن بعد الحصول على تمثيل النموذج وبياناته من كل مصدر على شكل أنطولوجية، فيمكننا تعيين تنسيقات بين مفاهيم كل أنطولوجية³.

³ نحيل القارئ مرة ثانية إلى [4] للتفصيل بخصوص التنسيق و التوفيق بين البيانات.

5 هايبرتوبيك كتقنية لتصفح المعارف

ذكرنا من قبل أنه من بين عوائق استعمالية المعرفة وظائف البحث المحدودة وقلنا أنه من المهم جدا تزويد تقنية إ.د.ح.م. بوظيفة تصفح محتواه وفقا لمصالح ووجهات نظر مستعمليه.

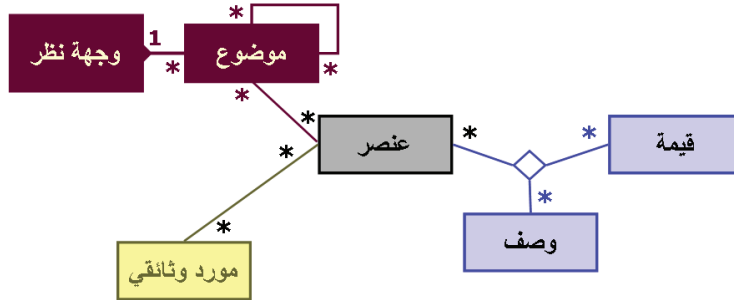
5.1 المبدأ

فقد جاءت فكرة إنشاء هايبرتوبيك في [20] و [21] لتزويد العنكبوتية الدلالية ببعد جديد ألا وهو البحث على موارد والعثور عليها بدون أن يعرف محتواها مسبقا. ويكون هذا بواسطة تصفح مخططات تصنيفية (شبكات دلالية أو أنطولوجيات) والتي قد أنشأت تدرجيا وبالتعاون بين مستخدميه.

فإن هايبرتوبيك (نرمز له فيما بعد ب: هـ.ت.) هو لغة تمثيل تستعمل لتعريف مخططات تصنيفية. ويعتبر أيضا كنموذج لتمثيل المعارف المتعلقة بمجال معين من المعارف لغرض إنشاء نظام تصفح متعدد الأوجه .

5.2 المركبات

مركبات هايبرتوبيك هي ست: وجهة نظر، موضوع، عنصر، وصف، قيمة، ومورد وثائقي.



الشكل 4. مكونات هايبرتوبيك

يمثل الشكل 4 مركبات هـ.ت. مع مختلف العلاقات الموجودة بينها.

إن الوحدة الأساسية والمدارة في هـ.ت. هي /العنصر وهو عبارة عن تعيين وضع أو إنتاج معالج بطريقة جماعية. ويرتبط معنى هذا العنصر بالمجتمع الذي يعالجه. مثال على ذلك، العنصر المعروف في إ.د.ح.م. قد يتطابق تماما مع العنصر هـ.ت.

قد يسند العنصر إلى وجهة نظر بمساعدة أوصاف إرشادية ويسمى كل منها موضوع. فإذا، يرتبط العنصر بعدة جهات نظر باستعمال المواضيع ويمكن للمواضيع أيضا أن تربط مع بعضها البعض.

يوصف العنصر بعدة/أوصاف عادية وكل وصف تسند له قيمة معينة. ويمكننا توصيل هذا العنصر بمورد وثائقي أو أكثر.

5.3 الفائدة

يعتبر هايبرتوبيك نظام ذو أهمية بالغة في تسهيل البحث وإعادة استعمال المعارف المكتسبة في الماضي. ويوفر وسيلة مساعدة للمستخدمين على تبويب المعارف وذلك من خلال مركباته. باستعمال هرت. لإعادة تنظيم محتوى نظام إ.د.ح.م. على شكل جهات نظر مختلفة يرجع بالفائدة الكبيرة على المستخدمين في التطلع واكتشاف المحتوى.

6 منهج سنبله

إن العمل البحثي المقترح في هذه الورقة يقدم مساهمات في حل المشاكل المعقدة لاستعمالية المعارف المكتسبة خلال دورة حياة المنتج. وتلخص المساهمات هذه في المنهج سنبله (سلسلة تبادل وإعادة استعمال ومشاركة المعلومات والمعارف المتعلقة بالمنتج) المستمد من مناهج هندسة المعارف والمعرف بما يلي:

1. منهجية من ستة خطوات (تعيين، التقاط، تشكيل، دمج، توزيع وتحديث المعرفة) لتحويل المعارف إلى رأسمال فكري.
 2. مستويين للنمذجة، الأول لتمثيل المعارف والثاني لعرض وتصفح المعارف.
 3. ثلاثة مركبات تقنية لتحقيق الاستخراج الدلالي للبيانات، والمطابقة والتوفيق بين البيانات، وتمثيل المعارف وحفظها في مستودع موحد.
- قبل أن نشرع في تفاصيل المنهج سنبله نفصل أولا الهدف من هذا المنهج.

6.1 الهدف

وفقا لـ [22] فإن "هندسة المعرفة تقترح مفاهيم، ومناهج وتقنيات تسمح بنمذجة، وتشكيل، واكتساب المعارف في التنظيمات لهدف تفعيل، هيكل وإدارة بالمعنى عام". وأما بالنسبة لـ [23] فإن هندسة المعرفة تتخصص في تصميم نظم توجيه ووصول والتي تسهم في إثراء دعامات المعرفة لكي تسمح توجيه مستعمل هذه الدعامات. من المؤكد أن دعامات المعرفة التي نتكلم عليها في هذا البحث هي دعامات رقمية، ولهذا تعتبر [24] هندسة المعرفة كهندسة الدعامات الرقمية للمعرفة.

المنهج سنبلة المقترح مستوحى من مناهج هندسة المعرفة لإدخال طريقة جديدة لـ:

- تسهيل تبادل البيانات أساسا بين نظم إ.د.ح.م.
- إثراء المحتوى الحالي لنظم إ.د.ح.م.
- توفيق نماذج البيانات لنظم إ.د.ح.م.

والذي يجب التنويه عليه هنا هو أن منهج سنبلة يهتم بالدلالة الموجودة بين البيانات المتعلقة بدورة حياة المنتج ولا تعالج المحتوى الدلالي للبيانات الهندسية.

6.2 المنهجية

تتكون منهجية هندسة المعرفة التي يقترحها سنبلة من ستة خطوات موضحة في الشكل 5 (اعتمادا على الخطوات المقترحة في [25] و [26]).



الشكل 5. منهجية سنبلة

خطوات المنهجية هي كالآتي:

1. **تعيين وتحديد مصادر المعرفة:** مصادر المعرفة يمكن أن تكون أداة، تقنيات، مناهج ونماذج للتمثيل. في الدراسة الحالية نظم إ.د.ح.م هي مصادر المعرفة.
2. **التقاط المعارف:** يكون باستخراج المعارف من مصدرها كلها أو بعضها. وهذا يُترجم في عملنا بعملية الاستخراج.
3. **تشكيل المعارف:** تحت شكل موحد ومرن وبعد ذلك وضعها في مستودع مشترك. وهذا يترجم بعملية التشكيل.
4. **دمج المصادر:** بإنشاء تنسيقات بين الكائنات من المستخرجة من مختلف المصادر. ونلجأ لهذا الدمج عندما نعين عدة مصادر تمثل كائنات مشتركة ولكن بطريقة منفصلة. وهذا يترجم بعملية الدمج.

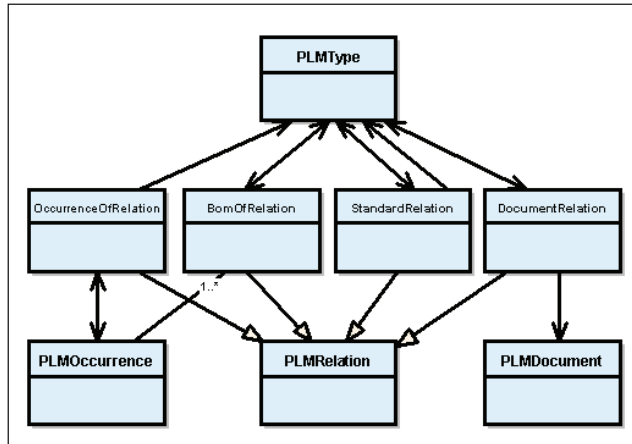
5. **توزيع المعرفة:** توفير الوسائل للمستخدمين للوصول وإثراء المعارف. وهذا سيكون وفقا لنموذج هابيرتوبيك.

6. **تحديث المعرفة:** من مصدرها الأصلي بانتظام وتلقائية. وهذا يترجم بعملية التحديث. كل العمليات المذكورة ستفصل فيما يلي.

6.3 نمذجة البيانات

مهما كان نوع نظام إ.د.ح.م. فإنه ينظم محتواه على مستويين:

1. مستوى النموذج: يعبر على الهيكلية بحيث أننا نجد فيه تعريف أنواع العناصر والخصائص التي تصفها وأيضا أنواع العلاقات والوثائق.
2. مستوى البيانات: يعبر على المحتوى الحقيقي بحيث أننا نجد فيه العناصر وخصائصها والعلاقات بين هذه العناصر والروابط للوثائق.

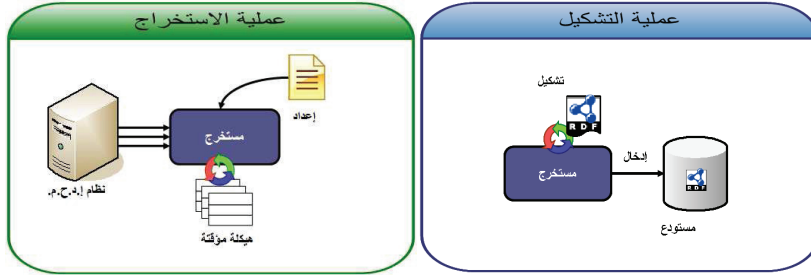


الشكل 6. نموذج البيانات

الشكل 6 يوضح لنا المستوى الأول وهو في الحقيقة عبارة عن نموذج ما وراء. يعني أن هذا النموذج هو نموذج عالي يستعمل في عملية التشكيل لإنشاء الأنطولوجية المناسبة لكل نموذج مستخرج من إ.د.ح.م.

6.4 مستخرج دلالي

عمليتا الاستخراج والتشكيل المذكورتين في تعريف المنهجية تنجزا من طرف المستخرج الدلالي. هذا المستخرج هو عبارة عن برنامج يسمح لنا باستخراج البيانات المرغوب فيها من مصدر معين وذلك بعد أن يتصل بهذا المصدر. كما أنه يجب لكل نوع من المصادر تطوير مستخرج خاص به وتعيين إعداده.



الشكل 7. عمليتا الاستخراج والتشكيل.

عند عملية الاستخراج، كما يوضحه الرسم الأيسر في الشكل 7، يمكن استعمال هيكل مؤقتة للبيانات وذلك من أجل إقامة صلات بين البيانات بطريقة أفضل، والتحقق من كمالها ومطابقتها للمصدر الأول. كما يمكن لعملية التشكيل أن تجرى في نفس الوقت مع عملية الاستخراج. بعد استخراج البيانات من المصدر، تقوم عملية التشكيل (الرسم الأيمن في الشكل 7) بتحويل هذه البيانات وتمثيلها في شكل من الأفضل أن يكون متين، شكلي ومحايد (مستقل عن أي قيد مهما كان تقنيا أو منهجيا). في حالة تطبيقنا فإننا اخترنا إطار توصيف الموارد (إ.ت.م). كلغة تمثيل وقد عللنا هذا الاختيار مسبقا.

تعريف نوع عنصر **RSAWeldPoint**

```
<rdf:Class rdf:about="http://www.cadesis.com/i-semantec/dp-
schema#RSAWeldPoint">
  <rdf:isDefinedBy rdf:resource="http://www.cadesis.com/i-semantec/dp-schema#" />
  <rdf:label>RSAWeldPoint</rdf:label>
  <rdf:subClassOf rdf:resource="http://www.cadesis.com/i-semantec/dp-
schema#PLMType" />
</rdf:Class>
```

تعريف الوصف **CoordX** من الفئة **RSAWeldPoint**

```
<rdf:Property rdf:about="http://www.cadesis.com/i-semantec/dp-schema#CoordX">
  <rdf:isDefinedBy rdf:resource="http://www.cadesis.com/i-semantec/dp-schema#" />
  <rdf:label>CoordX</rdf:label>
  <rdf:domain rdf:resource="http://www.cadesis.com/i-semantec/dp-
schema#RSAWeldPoint" />
  <rdf:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal" />
</rdf:Property>
```

الشكل 8. تعريف نوع عنصر بواسطة م.إ.ت.م.

كما نذكر أن نموذج البيانات يترجم إلى أنطولوجية و يعرف بمخطط إ.ت.م. (الشكل 8) وتعرف البيانات التي تمثل محتوى قاعدة المعارف بالشكل إ.ت.م. (الشكل 9).

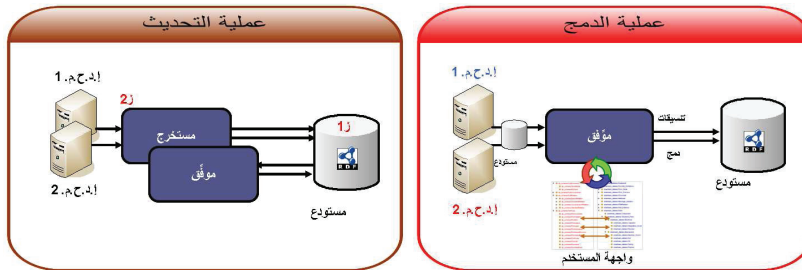
بعد الإنتهاء من عملية التشكيل، نكون قد حصلنا على مورد جديد للمعارف محفوظة في هذا المورد بطريقة نوعا ما معيارية. وهذا يساهم لا شك بشكل معتبر في قابلية التشغيل المتبادل بين التطبيقات التي ستستعمل هذا المورد.

```
<dps:RSAWeldPoint rdf:about="http://www.cadesis.com/i-semantic/docprocess#X44-SJ
RC1007951_A">
<dps:CoordX>2738.45</dps:CoordX>
<dps:CoordY>-286.220</dps:CoordY>
<dps:CoordZ>428.399</dps:CoordZ>
<dps:creation_date>05-Nov-2007 22:15</dps:creation_date>
<dps:last_mod_date>06-Nov-2007 13:43</dps:last_mod_date>
<dps:HypDecoupage>H1DEC</dps:HypDecoupage>
<dps:object_desc>SJRC1007951</dps:object_desc>
<dps:Diversite>C44.</dps:Diversite>
<dps:EtatImport>N</dps:EtatImport>
<dps:object_name>X44-SJRC1007951</dps:object_name>
<dps:uid>hLIIdynIPTmc$3A</dps:uid>
<dps:docRelation rdf:resource="http://www.cadesis.com/i-semantic/docprocess#IMAN_
Rendering9" />
</dps:RSAWeldPoint>
```

الشكل 9. تعريف عنصر بواسطة إ.ت.م.

6.5 موفقي النموذج

يستعرض نموذج البيانات توصيف مفصل لمجموعة من المفاهيم. النماذج المسخرجة من مختلف نظم إ.د.ج.م. على شكل أنطولوجيات نادرا ما تتشابه ولتجميع المحتويات المختلفة في مستودع مشترك يجب التوفيق بين هذه النماذج. ففي هذا السياق فإن مهمة عملية الدمج (الرسم الأيمن في الشكل 10) التي يقوم بها موفقي النماذج هي إنشاء تنسيقات بين مفاهيم النماذج المختلفة. وللعلم، فإن هذه العملية لا يمكن لها أن تكون تلقائية تماما أي تحتاج إلى تدخل المستخدم.

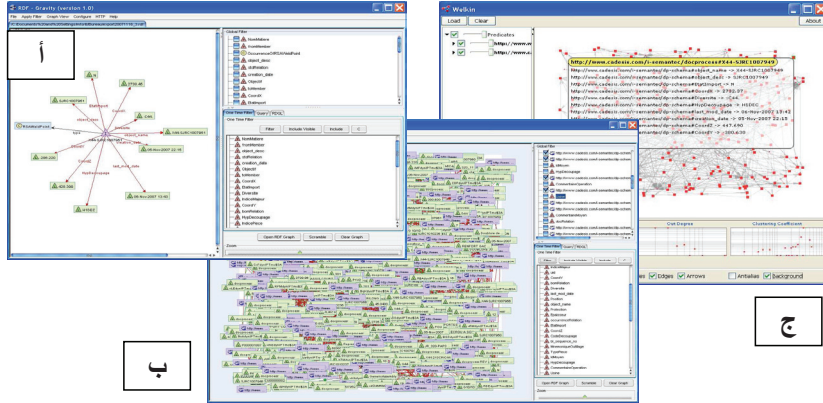


الشكل 10. عمليات الدمج والتحديث.

لإنشاء التنسيقات بين البيانات فإن الموقع يستعرض البيانات من المستودع وبعدها تعاد هذه البيانات مرة ثانية للمستودع وتحفظ مع التنسيقات.

6.6 تصفح المحتوى

لتصفح المعارف المحفوظة في المستودع لا يسعنا ذلك باستعمال متصفح إ.ت.م. أصلي لأن ذلك لا يعين المستعمل مطلقاً في البحث. لقطات الشاشة الثلاثة في الشكل 11 تبرز بوضوح قصور هذا النوع من الأدوات في استعراض المحتويات.



الشكل 11. التصفح بشكل الأصلي للمحتوى إ.ت.م.⁴

لمساندة المستعملين في البحث وتصفح المعارف واستغلال وإثراء المحتوى الدلالي في المستودع يستلزم إجراء نمذجة لربط محتوى المستودع مع مركبات هايبيرتوبيك. فيكون لدينا فيما يلي المفهوم من إ.د.ج.م. وما يقابله من مركب هـ.ت.:

- عنصر = عنصر
 - خاصية = وصف عام
 - رابط الوثائق = موارد وثائقية
 - الربط بين العناصر: فهذا الربط لم يكن موجوداً في النموذج هايبيرتوبيك قبل عملنا هذا. فبهذا نكون قد طورنا هـ.ت. لأخذ بعين الاعتبار هذا النوع من الروابط.
- وبعد هذه النمذجة يكون لدينا عرض المعارف على النحو الموضح في الشكل 12.

⁴ أ و ب تعرضا تصفح باستعمال RDF Gravity: <http://semweb.salzburgresearch.at/apps/rdf-gravity/index.html> و أما ج فياستعمال Welkin: <http://simile.mit.edu/welkin/guide.html>



6.7 المستودع المشترك

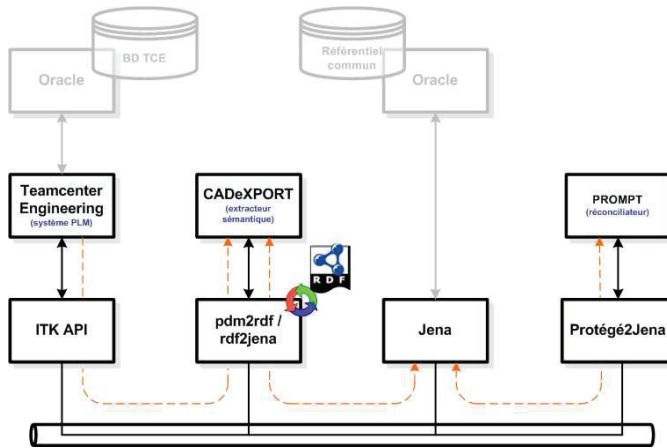
7 منصّة أي-سيمنتاك

لا يسعنا فيما تبقى أن نفصل في كل مكون للمنصة وللمزيد من التوضيحات الرجوع إلى [4].
فكتفي بذكر الأدوات المستعملة مع وصف مختصر.

- تيمستتر إنجنيرنج: نظام إ.د.ح.م.
 - كاد إكسبورت: مجموعة من التطبيقات التي طورناها لاستخراج، والتشكيل وإدارة المستودع المشترك.
 - برومبت: أداة توفيق بين الأنطولوجيات
 - بلم تو هايبرتوبيك: موزع (سيرفر) هايبرتوبيك يحقق الإتصال بالمستودع وترجمة محتواه
 - أقورا: عميل (كلاينت) هايبرتوبيك يسمح باستعراض وتصفح وإثراء المحتوى
 - أرقوس: الموزع الرسمي لهايبرتوبيك والذي يحتفظ بالتعديلات المضافة عبر أقورا.
- وفيما يلي نصف هذه التطبيقات إلى طبقتين الأولى للاستخراج والثانية للتصفح.

7.1 الاستخراج

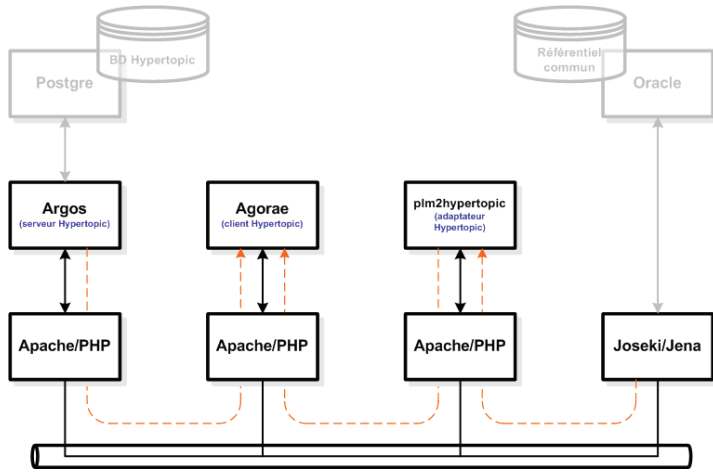
استخراج البيانات، وتشكيلها، ودمجها، وتوفيقيها، وحفظها يتطلب التفاعل بين أربعة أدوات أساسية كما موضح في الشكل 13.



الشكل 13. الهندسة التطبيقية للاستخراج

7.2 التصفح

الأصوات التي تتدخل في عرض، وتصفح، وإثراء محتوى المستودع المشترك مستندة إلى هابير توبيك كمعيار لتمثيل المعارف. مختلف هذه الأدوات يستعرضها الشكل 14.



الشكل 14. الهندسة التطبيقية للتصفح

8 دراسة تجريبية

ندرس في هذا الباب كحالة تجريبية شركة أي بي بي وهي شركة عالمية مصنعة لخطوط تجميع السيارات. لإختصار وقت تنفيذ المشاريع وتقادي الأخطاء المكررة، تود هذه الشركة بتوفير لمهندسيها نظام يسمح بحفظ المعلومات وبسهولة الوصول لها و استعماليتها في مشاريع مستقبلية. فكانت هذه الشركة كإطار تجريبي مثالي للمنهج سبايك و المنصة أي-سمتاك حيث أنهما خضعا لاستعمال تجريبي من طرف مهندسي الشركة المشتغلين بمشاريع إنجاز خطوط التجميع من الطلب إلى التسليم.

قبل بداية استخدام أي-سمتاك، قمنا باستخراج البيانات من نظام إ.د.ح.م. المستعمل في شركة أي بي بي ألا وهو تيمسنتر باستعمال المستخرج الدلالي ووضع هذه البيانات في المستودع المشترك. وبعد ذلك أصبح المحتوى قابل للتصنيف والعرض من طرف المستخدم بواسطة أقورا.

فقد استعملت المنصة في ثلاثة مهمات ألا وهي:

- استجابة للمناقصة
- تحضير عقد الشراء
- اختيار الممون

نتطرق للمهمة الأولى لوضوحها. فلتحضير الإستجابة كان على المهندسين فيما قبل مراجعة المشاريع السابقة لاستعمال أجزاء منها في المناقصة. بعد أن كانت هذه المهمة تنفذ يدويا (أي باستعمال عملية البحث العادية)، أصبح بين يدي المهندسون طريقة للوصول إلى المشاريع التي

تهمهم عن طريق أقورا و ذلك بتصفح وجهات النظر والمواضيع التي يفترض أن تكون قد عرّفت مسبقا باستعمال خصائص مكونات المشاريع.

9 الخاتمة

استعرضنا ضمن هذا البحث عدة مشاكل متعلقة بإدارة دورة حياة المنتج الصناعي ألا وهي اختلاف النظم المستعملة، التطور المستمر لنموذج بيانات هذه الأنظمة، وقصور وظائف البحث. وقد أثبتنا أن كل هذه المشاكل لها صلة مباشرة في عرقلة استعمالية المعارف المتعلقة بالمنتج والمكتسبة خلال تطويره أو خلال أي مرحلة من مراحل حياته.

خلال هذا العمل قمنا بإدراج معايير العنكبوتية الدلالية مع الأنطولوجية من أجل تمثيل المعارف على شكل مشترك وديناميكي. هذا الشكل يحث على دمج وتبادل البيانات بين النظم الغير متجانسة. ولضمان أفضل استعمالية للمعارف برز هايبرتويك كمنهج وثيق لعرض وتصفح المعارف بطريقة متعددة الأوجه.

ثم اقترحنا بعدها المنهج سنبله والذي يتركز على منهجيات هندسة المعارف لتحقيق كل من الاستخراج الدلالي للبيانات، التشكيل والتمثيل الدلاليين للبيانات، والتوفيق والتنسيق بين نماذج البيانات، وتصفح المعارف بسلاسة. ولقد طورنا المنصة أي-سيمنتاك لتحقيق منهج سبايك. كما استعرضنا دراسة تجريبية بحيث أنه اتضح لنا من خلالها مدى استجابة منهج سنبله لمعالجة المشاكل المتعلقة باستعمالية المعرفة المتعلقة بالمنتج.

أدت نتائج هذا العمل إلى انطلاق مشروع آخر في نفس المجال ولكن ليأخذ بعين الاعتبار استعمالية البيانات الهندسية. و كنتيجة أخرى لعملنا هذا قمنا بتطوير برنامج محكم لتبادل البيانات المتعلقة بالمنتج باعتبار نظام (د.ح.م. كوثيقة ل.ت.م. وتنفيذ استعلامات عليها باستعمال لغة الإستعلامات أكس-كوارى. ربما نكلمنا عليه في وقت لاحق.

كما لا نخفي أننا في مشروع حالي وربما في مشاريع لاحقة سنطبق كل من هذه المكتسبات في تطبيقات تخدم اللغة العربية.

10 شكر وتقدير

ساهم في تطوير المنصة أي-سيمنتاك مجموعة من الباحثين وذلك ضمن مشروع بحثي يحمل نفس الاسم ومدعم من طرف الوكالة الوطنية للبحث في فرنسا. حقق هذا المشروع بفضل دوافع والمشاركة النشطة من كل باحث.

المراجع

- [1] Cattan, M. (2001). L'entreprise et ses données techniques, méthodes pour une documentation maîtrisée, AFNOR.
- [2] Ermine, J.-L. (2003). La gestion des connaissances. Eds. Hermès sciences

- publications.
- [3] CIMdata (2003). Product Lifecycle Management "Empowering the Future of Business". CIMdata Inc, Ann Arbor, <http://www.cimdata.com>.
 - [4] Sriti, M.-F. (2008). Démarche et logiciel de gestion des connaissances pour le cycle de vie des produits. Ph.D. thesis in Computer Science, discussed on July 3rd, 2008.
 - [5] Fenves, S.J., Sriram, R.D., Subrahmanian, E., & Rachuri, S. (2005). Product Information Exchange: Practices and Standards. ASME Journal of Computing and Information Science in Engineering 5(3): 238-246
 - [6] Saaksvuori, A. and Immoen, A. (2004). Product Lifecycle Management, Berlin, Springer-Verlag.
 - [7] Eynard, B. (2005). Gestion du cycle de vie des produits et dynamique des connaissances industrielles en conception intégrée. Rapport d'habilitation à diriger des recherches. Université de Technologie de Troyes, France.
 - [8] Bellatreche, L., Xuan, D.N., Pierra, G. & Dehainsala, H. (2006). Contribution of Ontology-based Data Modeling to Automatic Integration of Electronic Catalogues within Engineering Databases, Computers in Industry Journal 57 (8-9): 711-724.
 - [9] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. Scientific American.
 - [10] W3C. (2004). Extensible Markup Language (XML) v1.0. <http://www.w3.org/TR/REC-xml>
 - [11] Rezayat M. (2000). Knowledge-based product development using XML and KCs. Computer-Aided Design 32(5): 299-309.
 - [12] Lubell, J., 2002, "From Model to Markup: XML Representation of Product Data," XML Conference, Baltimore, MD, <http://www.mel.nist.gov/msidlibrary/publications.html>.
 - [13] Mervyn, F., Senthil Kumar, A., Bok, S.H., & Nee, A. (2004). Developing distributed applications for integrated product and process design. Computer-Aided Design 36(8): 679-689.
 - [14] Lee, C., Lau, H., Yu, K.M., & Ip, W.H. (2005). A generic model to support rapid product development: an XML schema approach. International Journal of Product Development 1(3): 323-340.
 - [15] Srinivasan, V. (2005). Open Standards for Product Lifecycle Management, Product Lifecycle Management Emerging solutions and challenges for Global Networked Enterprise, Coordinators: Bouras A. Gurumorthy B., Sudarsan R., Interscience Enterprise Ltd, pp. 475-484.
 - [16] Lassila, O., & Swick R.R. (1999). Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation, February 22nd 1999
 - [17] Terzi, E., Vakali, A., & Hacid, M.S. (2003). Knowledge Representation, Ontologies, and the Semantic Web. In proceedings of Asia Pacific Web Conference, Xi'an, China.
 - [18] Brickley, D. & Guha, R.V. (2004). RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, 10 February 2004.
 - [19] Lin, H.K., & Harding, J.A. (2007). A Manufacturing System Engineering Web Ontology Model on the Semantic Web for Inter-enterprise Collaboration. Computers in Industry 58(5): 428-437.
 - [20] Cahier, J.-P., Zacklad, M. & Monceaux, A. (2004). Une application du Web Socio-Sémantique à la définition d'un annuaire métier en ingénierie. Actes des journées Ingénierie des Connaissances IC2004, mai 2004, Lyon.
 - [21] Zacklad, M. (2005). Introduction aux ontologies sémiotiques dans le Web Socio Sémantique. In: JAULENT, M.-C. 16èmes journées francophones d'Ingénierie des Connaissances, Grenoble: PUG.
 - [22] Charlet, J. (2005). L'ingénierie des connaissances, une science de gestion ? In Teulier R. et Lorino P., coordinateurs, Entre la connaissance et l'organisation, l'activité collective, chapitre 11. La découverte, 2005. Actes du colloque de Cerisy « Activité, connaissance, organisation ».

- [23] Zacklad, M. (2004a). Processus de documentarisation dans les Documents pour l'Action (DOPA) : statut des annotations et technologies de la coopération associées. Le numérique : Impact sur le cycle de vie du document pour une analyse interdisciplinaire, 13-15 Octobre 2004, Montréal.
- [24] Bachimont, B. (2004). Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle. Habilitation à diriger des recherches de l'Université Technologique de Compiègne
- [25] Dieng-Kuntz, R., Corby, O., Gandon, F., Giboin, A., Golebiowska, J., Matta, N. & Ribière, M. (2001). Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire du Knowledge Management. Eds. Dunod.
- [26] Grundstein, M. (2002). GAMETH : un cadre directeur pour reprérer les connaissances cruciales pour l'entreprise. Rapport de recherche N°9. MG Conseil.

المخلص باللغة الانجليزية

Title: SPIKE: Semantic Approach to Improve Knowledge Reusability in Product Lifecycle

Abstract. Product Lifecycle Management is a very vast industrial area and provides new challenges for knowledge engineering. This area faces several problems, especially in sharing, exchange, and reuse of product related knowledge. To solve these problems, it is not enough today to resort to data and information management technology to reach to advanced approaches to identify and provide the useful information to the right person and at the right time. Among these challenges and to solve the mentioned problems the approach SPIKE was born to represent knowledge in a unified and dynamic format (to integrate and exchange data) using the Semantic Web and ontology techniques and to ensure knowledge reuse in a better way (navigate on knowledge using a multi-views manner), by using HyperTopic.

Keywords: Semantic Web, Ontology, Knowledge Engineering, Product Lifecycle Management

المصطلحات

إعادة استعمال	Reuse
استعمالية	Reusability
البيانات الوصفية	Metadata
التصميم بمساعدة الحاسب	Computer-Aided Design
العنكبوتية الدلالية	Semantic Web
إدارة العمليات	Process Management
إدارة الوثائق	Document Management
إدارة دورة حياة المنتج	Product Lifecycle Management

إدارة دورة حياة المنتج (إ.د.ح.م.)	Product Lifecycle Management
إطار توصيف الموارد (إ.ت.م.)	Resource Description Framework (RDF)
أرقوس	Argos
أقورا	Agorae
أكس-كواري	XQuery
أنطولوجية	Ontology
برومبت	PROMPT
بلم تو هايبر توبيك	plm2hypertopic
ترميز	Encoding
تمثيل	Representation
تيمسنتر إنجنيرنج	Teamcenter Engineering
خاصية إ.ت.م.	RDF Property
سلسلة تبادل واستعمالية ومشاركة المعلومات والمعارف المتعلقة بالمنتج (سنبله)	Smooth Product Information and Knowledge Exchange, reuse and sharing (SPIKE)
شكلي، شكلية	Formal
صنف م.إ.ت.م.	Class RDF
عملية	Process
عنصر	Item
قابلة للتشغيل المتبادل	Interoperable
كاد إكسبورت	CADeXPORT
كائن	Object
لغة ترميز موسعة (ل.ت.م.)	eXtended Markup Language (XML)
لغة وجودية العنكبوتية	Web Ontology Language (OWL)
متعدد الأوجه	Multipoint of views
مخطط إطار توصيف الموارد (م.إ.ت.م.)	RDF Schema
مشاركة	Sharing
مورد	Resource
نموذج ما وراء	Metamodel
هندسة المعرفة	Knowledge Engineering



الجلسة العاشرة

تمثيل و تصنيف مطلوبات الجودة على النموذج المستقل عن التكنولوجيا (بي أي إم) في منهج نموذج المعمارية (إم دي إيه)

د. عبدالغفار حامد أحمد¹، البرفسور روبرت كولوم²،
جامعة التكنولوجيا الماليزية، الرمز البريدي ص ب 81310، جوهو بارو، ماليزيا
البريد الإلكتروني: ghaaabdel2@live.utm.my¹، rcolomb@me.com²

الخلاصة. نموذج المعمارية إتجاه جديد في تطوير البرمجيات يقوم على تسير عملية البرمجة بالنماذج. يبني النموذج الأول و يعرف بالمستقل عن التقنية- للعتبير عن المطلوبات (قوانين العمل) بينما يبني النموذج الثاني و الذي يعرف بالمنبني على التقنية- للتعبير عن قضايا التصميم. تقوم بعد ذلك لغات التحويل مثل (كيو في تي) بتحويل النموذج الأول الى النموذج الثاني بصورة آلية. ما زالت هذه العملية بحثية. بما أنه ظهر جليا أثر عدم الاهتمام بجودة البرمجيات في مرحلة مبكرة (النماذج) في عملية التطوير وبما أنه لا يوجد عمل سابق تناول علاقة النموذج الاول(نموذج المطلوبات) بالمطلوبات الكمالية أو مطلوبات الجودة، فان هذه الورقة تسهم في هذا الاطار. تحت هذا الاطار هذه الورقة تمنح تصنيفا جديدا لمطلوبات الجودة ثم صممت لغة خاصة (بروفایل) للتعبير عنها في النموذج المستقل عن التقنية باستخدام اليو ام ال. علاوة على ذلك بهذه الطريقة سنساعد في عملية أتمتة التحويل التي تقوم عليها هندسة نموذج المعمارية.

الكلمات الجوهرية: نموذج المعمارية، النموذج المستقل عن التقنية (بي أي إم)، البروفایل (لغة المظهر الخاص)، اعادة الاستخدام.

1 مقدمة

اليوم غالبية استراتيجيات تطوير البرمجيات تتطلب تعقيد بسبب اعتمادها على منهج المسير-بالكود واليدوي الامر الذي جعل تكلفة المنتجات عالية. المتطلبات المتزايدة والتغيرات من وقت لآخر لتؤدي الى تضخيم الكلفة الإجمالية. وأصبحت حرفة الكود مملة حتى بدأ النظر اليها كأنها عمليات جراحية (إلى حد كبير بسبب الطلب المتزايد على نوعية عالية). من جانب إعتقاد المطورين كليا على الكود بحيث يعبر عن نموذج النظام الذي يبني بلغة البرمجة من الجيل الثالث مثل جافا، # ج ، أو سي ++ (أي حزم، وحدات)، ويتم ترك أي نموذج لتصاميم العمارة منفصل و غير رسمي. من جهة أخرى تغير هذه الحلول يصبح لا مفر منه مثلا فإن التكنولوجيا متقلبة حتى داخل نفس التكنولوجيا تجد عددا من الاصدارات كالجافا تشحن الى المستهلك بسبب تحسين الجودة (الاداء ، السرية ، الخ).

عملية أتمتة صناعة البرمجيات باتت واقعا ملموسا بعد ظهور نموذج المعمارية الذي تبنته أوه إم جي [1] . كانت المشكلة سابقا تكمن في أن المطلوبات توصف باللغة الإنجليزية أو العربية و هما محل إلتباس أو غموض أولا ثم مشكلة التوافق بين النص و النموذج. جاء هذا الإتجاه الحديث لجعل التوصيف رسمي عن طريق إستخدام لغة قياسية بالتالي أصبحت المطلوبات هي النموذج و النموذج هو الكود وذلك عن طريق التحويلات. إن أحد أهم خصائص هذه الطريقة الفصل بين نموذج التصميم و نموذج المطلوبات. مما يتح فرصة إعادة الإستخدام هو الآخر الذي يميز هذه الطريقة. إذ يمكن أن نعيد إستخدام نفس نموذج التوصيف عددا من المرات مع نموذج التصميم (هذه الصورة تشبه التغيير الذي يحدث من تقنية إلى أخرى مثل ماي سكول إلى أوراكل). بالمثل يمكن أن يتغير نموذج المطلوبات مع ثبات نموذج التصميم و ذلك يعني عدد من التطبيقات في مجال معين مثل قواعد البيانات تستخدم نفس نموذج التصميم. نظام سجل الطالب و نظام قبول الطلاب للدراسات العليا يستخدمان النموذج العلائقي كنموذج تصميم واحد. الإستفادة من النموذج في توصيف المطلوبات جعل من الممكن توصيف مطلوبات الجودة و الشروط بصورة ظاهره والتي تعبر عن قوانين العمل. إبتكرت هيئة إدارة الكائن (أوه إم جي) عددا من اللغات لدعم هندسة نموذج المعمارية مثل لغة قيود الكائن (أوه سي إل) التي تستطيع توصيف الشروط و قيود التكامل بصورة دقيقة و منهجية، الموف التي تعبر عن نموذج النموذج (منا متامودل)، ولغة الكيوفيتي [2] للتحويلات من نموذج الى آخر.

لقد جاء الاهتمام مؤخرا بقضية الجودة لتكون في وقت مبكرة من عملية صناعة البرمجيات لانه اكتشف أنها من أقوى العوامل التي تؤثر على تكلفة المشروع اذا لم يكن هناك اهتمام بها .

2 نموذج المعمارية

تبين أن مدى الجودة وطول العمر ، وتكلفة الإنتاج عندما تكون متوازنة في تطوير البرمجيات ذلك يحدد الجدوى الاقتصادية لاستراتيجية تطوير البرمجيات [3]. وذلك يتحقق برفع مستوى التجريد واسلوب إعادة الإستخدام . هذه الملامح هي التي شكلت نموذج المعمارية الذي تبنته أوه إم جي (تضم 800 شركة في تقنية المعلومات) وهي مؤسسة غير ربحية ترفع عدد من المعايير مثل لغة اليو ام ال.

على هذا النحو تعتبر النماذج هي الاله في عملية البرمجة وبالتالي تستخدم لغات عالية المستوى ورسمية للتعبير عن الصياغة والمعنى بطريقة تخلو من اللبس أو الغموض. وهذا بدوره يشجع على الأتمتة لانه يمكن لاله ان تفهم هذه النماذج وبالتالي استخراج الكود سيكون مباشرا اذا توفرت المعلومات الكافية في النموذج. هذا الامر كان منشودا في تاريخ هندسة البرمجة (أدوات الكيس) .

اذا هو اتجاه تطوير جديد يهدف الى جعل من الممكن تطوير التطبيقات في المجالات المحددة دون كتابة رمز البرنامج. تستخدم لغة اليو ام ال و الموف (جزء من اليو ام ال ولكنها أعلى منها) في بناء النماذج بصورة عامة. النموذج الاول وهو محور هذه الورقة يوصف وظائف العمل المعنى بعبارة العمل و بتجريد عالي المستوى. هذه اللغة الموجهة نحو التطبيق تسمى بالنموذج المستقل

عن التقنية.(البي أي ام).لانه لدينا تقنيات (Platform) تعتبر كقواعد لانتاج تطبيقات كثيرة مثل أنظمة قواعد البيانات فانه يبنى نموذج بلغة موف لواجهة تطبيق البرنامج (اه بي أي) لتلك التقنية المعتمدة على الكود.يسمى هذا النموذج بالنموذج المنبني على تقنية. عملية التحويل من اللغة الموجهة للتطبيقات الى النموذج المنبني على الكود يتم تأسيسها بواسطة لغة الاستفسار-مشهد-تحويل (الكيفيتي). هذه اللغة تعبر عن قوانين التحويل من نموذج-الي-نموذج وتشبه لغة برولوج في أنها توصيفية ولايهتم المعماري بشرح كيف- يعمل.نتيجة التحويل هذه تكون عبارة عن تحويل مطلوبات العمل الى مناداة لنموذج الكود والذي يقوم بتنفيذها لانه قام على أساس تقنيه مصنعة وموجوده أصلا. بهذه الطريقة تعتبر اللغات عامل أساسي في نجاح فكرة نموذج المعماريه. ان ما يميز ذلك هو أن اللغة أصبحت ذات وجهين واحد للاله والاخر للبشر [4] .

من هذه النقاط نستنتج أن نظام البرمجيات يصبح بهذا الشكل نظام معلومات في حد ذاته حيث نحتاج الى ادارة عدد من الاصول ،المصنوعات البرمجية ،عملية التحويل ، التحديث بينهم حتي يحدث الاتساق والاكتمال ثم انتاج الكود. كل هذا هو الدور المنوط ببنات نموذج المعمارية أن تقوم بها وبصورة أدق بالادوات المعتمد على موف والتي باتت عديده بعضها مفتوحة المصادر مثل (الاني ام اف) والاخر تجاري .

2.1 النموذج المستقل عن التقنية -البي أي ام

ال او ه ام جي عرفت هرم مستوى اللغات يمكن الرجوع اليه في[2] ولكنه باختصار اذا كان لدينا عدد من الطلاب في عالم ما سنقوم بانشاء الصنف حسب مبادئ منهج الكائن. فالصنف هو نموذج لانواع أو كائنات هي الطلاب نسمي هذه الكائنات المستوى صفر بينما الصنف هو المستوى 1 ولكننا قد نحتاج لتوصيف مستوى الصنف (الطلاب،الماده،الخ) اذا في هذه الحالة نحتاج لمستوى أعلى للتوصيف وهو ما يسمى بالمستوى 2.ولهذا المستوى أيضا مستوى أعلى منه وهو ما يعرف بالمستوى 3 أو مستوى موف والذي يعرف بالمتامودل(وهنا تظهر مفاهيم اللغة مثلا في حالة اليوم ام ال هي العلاقة ،الصنف الخ) وهو الذي يحمل معلومات عن نموذج 2 والذي اتضح أنه الاخير. اذا هو النموذج الفكري للغة النمذجة.تسمى هذه المستويات عدا الأول نموذج-النموذج (المتامودل) حيث أن كل مستوى هو عبارة عن حالة أو كائنات للمستوى الذي أعلى منه.

المتامودل يوفر الية مستقلة عن التقنيات لتوصيف الهيكل ، الصياغة،والمعني للتكنولوجيا[5] .

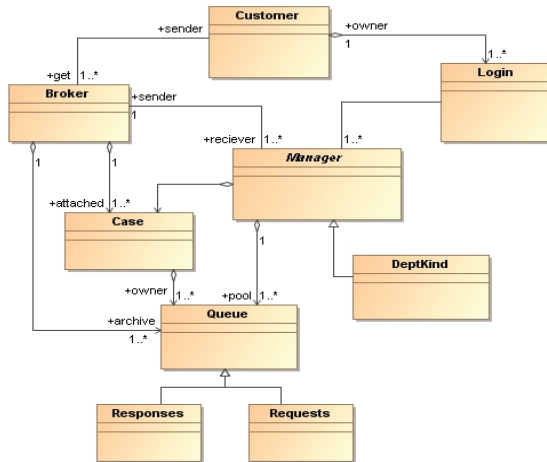
النموذج المستقل عن التقنية عبارة عن نموذج يعبر عنه باستخدام اليوم ال (مستوى 1) أو الموف (مستوى 2).يوصف هذا النموذج قوانين العمل أو التطبيقات بلغة التطبيقات. ويمكن تقسيمه الى نوعين ثابت (مثال نموذج الاصناف)وديناميكي(نموذج النشاطات في اليوم ام ال).هذا النموذج لايتناول أي معلومات تقنيه أو تخص التكنولوجيا. سنأخذ حالة من الحالات لنبين كيف يمكن نمذجة نظام المساعدة الالي الذح أصبح من الأنظمة الحساسة والمهمة في معظم تطبيقات الانترنت و التجارية الالكترونيه.

نظام برمجيات مكتب المساعدة الالي يقوم بخدمة العملاء و الموظف في نفس الوقت. والفكرة هي أن يسمح لمستخدم أن يطرح قضية في بعض الجوانب التي تقدمها الشركة التجارية والتي تحتاج إلى رد من بعض موظفي الشركة عن طريق شبكة الإنترنت. الموظف يطلع من خلال الموقع نفسه على الاسئلة الواردة في شكل حالات استشارة وينبغي أن يجرى تحديث من خلال موقع الويب نفسه على قائمة القضايا المطروحة أو حالة معينة . ثم يتم إرسال التحديث مرة أخرى إلى المستخدم، الذي قد يكون غير موجود عند الاستجابة. وهناك وحدة وسيطة تستخدم لتبادل الرسائل بين النظام والمستخدمين. شرط الجوده أداء النظام وهو أعلى أولوية من شرط الموثوقية. ملاحظة فقط سنتاول من هذا النظام ما يخدم غرض مبحث هذه الورقة .

أمثلة المطلوبات الوظيفية في هذا النظام التي يجب أن يؤديها :

- يسمح المستخدم بالدخول عليه بكلمة المرور واسم المستخدم.
- لا بد أن يكون المستخدم قادر على تسليم حالة للموظف عن طريق النظام بالسماح له بالكتابة النصية
- المستخدم يكون قادر على مشاهدة التحديثات على الحالة المسلمة والتفاصيل.
- المشرف لا بد ان يكون قادر على مشاهدة والتعليق على مجموعة من الحالات أو حالة بعينها. لذلك لا بد من وجود جدولة وترتيب في النظام لهذه الحالات.
- للمشرف القدرة على التعديل في الحالة.

من هذه المطلوبات نستطيع تكوين النموذج الفكري (الب اي ام) كما يظهر في نموذج الاصناف للبر ام ال شكل 1. نجد أنه لم يستخدم مفاهيم تقنية بعينها مثلا الاتصال، الجلسة أو المفتاح الاساسي والاجنبي للجدول لقاعدة البيانات.



الشكل 1. مخطط الاصناف يوصف ويشرح العلاقة بين كائنات نظام مكتب المساعدة الالي

من الشكل يتضح أنه لدينا الكائنات التالية :

broker, case, manager, and Queue customer, login,

أحد مهام المشرف أنه يستطيع ان يدير صف الاستلام والرد والذي يغذى بواسطة الوسيط بالرسائل القادمة من المستخدم لذا كانت هناك علاقة بين كل من *broker, manager- Queue* ، العلاقة *Queue-broker- case* . النظام الوسيط يستطيع ارسال أكثر من رسالة لذلك كانت العلاقة واحد-متعدد كذلك الحال مع الموظف ، المشرف والصف لانه يدير عددا من الحالات بنفس التعددية موجوده بين الحالة و الصف لان الصف يحتوى على عدد منها.

3 متطلبات الجودة (المطلوبات الكمالية)

تعبر المطلوبات الوظيفية عن الاحتياجات التي يجب على البرمجيات المطوره أن تكون قادرة على أدائها ، بينما غير الوظيفية توصف كيفية تحقق المتطلبات الوظيفية في المنتجات البرمجية [5]. هي مثل الكفاءة الزمنية أو الاداء، الاعتمادية، السرية الخ.

هناك مصدران لسمات ومتطلبات الجودة تم التعرف عليهم وهما أ- أصحاب المصلحة (مستفيد بشكل أواخر من المنتج) ب- نطاق التطبيقات مثل أنظمة المعلومات -نظام المرتبات، السجل الأكاديمي [6] .

أصبحت مصدر الاهتمام مؤخرا لانها أصبحت تمثل تكلفة المشروع الحقيقية، الاستقرار، والنضوج [5]. وجود فكرة المعمارية ساعد كثيرا على تيسير دراسة هذا الموضوع والاهتمام به مبكرا لأنها عالية المستوى (يعبر عنها باستخدام اليوم ال). فالبالتي كان هناك اهتمام المطلوبات التي تتعلق بالمعمارية والتي تسمى بخصائص الجودة وهي التي تنشق منها المعمارية. [7] [6]

تمثيل وتصنيف متطلبات الجودة لا تزال قيد البحث لأنها في المقام الأول مسألة ذاتية. كذلك أظهرت نتائج الدراسة التنوع في المصطلحات والتوجه [8]. الايزو على سبيل المثال اقترحت تصنيفان :سمات جودة للمنتج الوسيط (أي حين المنتج في مرحلة الإنتاج) و المنتج المنتهي (في حين أن المنتج جاهز وفي الاستخدام). النوع الاول (الوسيطه) هذه الصفات هي ستة : الوظيفة، والموثوقية ، سهولة الاستخدام والكفاءة والصيانة والنقل .بينما النوع الثاني(المنتهي) صفات الجودة للمنتج في استخدامها هي : الفعالية والإنتاجية، والسلامة.

هناك تصنيف اخر عند باس وآخرون [6] ودوينياس وآخرون [5]، ويقسم سمات الجودة الى : جودة التطوير والتغيير والجودة التنفيذية. الموثوقية والأداء هما من الصفات التنفيذ في وقت

التشغيل، في حين أن التمدد والاندماج لا يمكن أن يكون معالجتهما إلا على في الوقت اللازم لتطوير الهندسة المعمارية.

في مجال هندسة المتطلبات ليس هناك اتفاق على واحد في تعريفها، تصنيفات وتمثيلها. ويمكن لهذه أن يظهر بوضوح في مجهود [9] الذي ألقى تحليل مكثف لهذه القضايا. مشكلة أيت تمثل المتطلبات ذهب فريق إلى تمثيلها في مستند في شكل قالب كقالب فولير أو في قياس جمعية المهندسين الكهربائيين والالكترونيين 1998-830 الذي يفصلها عن المتطلبات الوظيفية في شكل سبع قوالب. أيضا الذين أعتدو على منهج الكائن في التحليل والتصميم قاموا بتمثيلها على نموذج الاستخدام (اليوس كيس) عدا العامة منها (ملحق).

المنهج الذي اتبعه مايينز [9] في التحليل النوعي لها كان على نفس نمط منهج نموذج المعمارية وهو ما ذهبت اليه هذه الورقة في الفصل بيت الاهتمامات حيث أنه فصل بين المتطلبات وخصائصها مثل النوع، الدور، التمثيل.

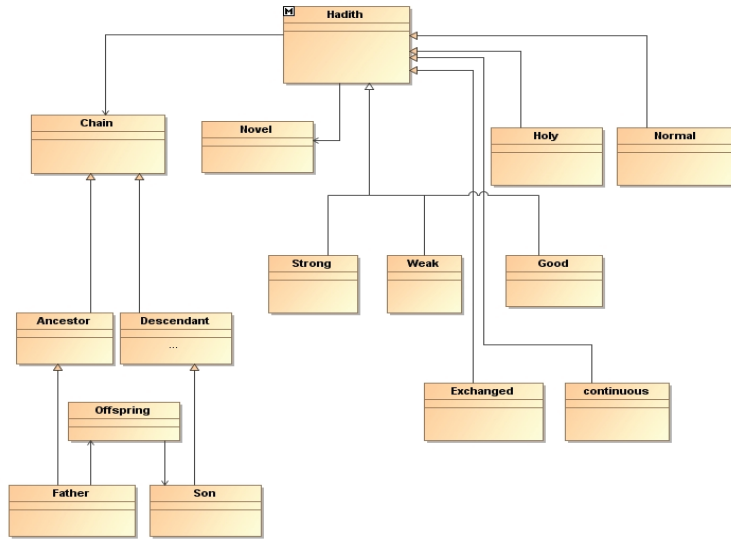
4 التمديد الهامشي و لغة المظهر الخاص

لقد رأينا اتجاه نموذج المعمارية يقوم على مبدأ اللغات في تسير عملية البرمجة. في هذه الفقرة سنتعرض لأحد الأساليب الفعالة في لغة اليو ام ال و المتنبه في منهج نموذج المعمارية والذي سيجعلنا نتبنى لغة خاصة والتي نريدها على النموذج المستقل عن التكنولوجيا.

من المعلوم أن لغة اليو ام ال لغة أغراض عامه . في كثير من الاحيان كما ثبت بالتجربة العمليه يريد المعماري تصميم لغة خاصة مثلا نمذجة مطلوبات الجودة التي تحدثنا عنها في الفصل 3. هذه المعلومات ستجعل من النموذج المستقل عن التكنولوجيا غنيا بالشكل الكافي لمساعدة اتخاذ قرارات التصميم المناسبة أو الاختيار من بينهم .

مصممو لغة اليو ام ام حرصو على تقديم نموذج قوى للتمديد أو الزيادة والذي يحافظ على هيكل اللغة دون تغيير صياغة وهيكل اللغة وذلك يعني دون المساس بنموذج مفاهيم اللغة (المتامودل). هذا الامر يقود الى اعادة استخدام نموذج اليو ام ال هذا وبالتالي نفس ادوات اليو ام ال الواسعة الانتشار مما يذيل عبء التعديل وتكلفته. هذه الطريقة التي تتبنى تعديل هامشي (خفيف) تسمى بالمظهر الخاص أو بالبروفایل وهي عبارة عن نظام توصيف.

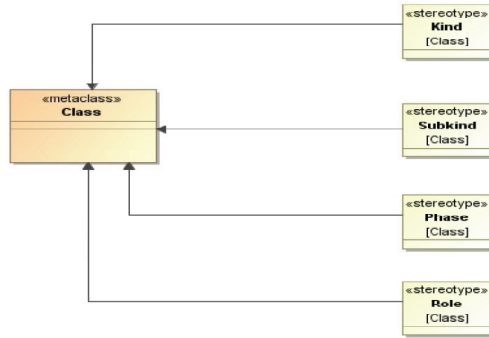
إذا دخلنا على موقع الـ اوام جي نجد عددا من صور البروفایل لأغراض مختلفة مثلا المجهود الكبير لسد الثغرة بين نماذج اليو ام ال و لغة الجافا الذي سمي بال إي جي بي (موقع البروفایل). سنركز في الفقرة التالية في كيفية استخدامه وهو المهم الآن لاننا نريد ان نصمم بروفایل لمطلوبات الجودة وسنترك لمن يريد تفاصيل أكثر أن يرجع الى القياس (سوبر استركتشر). يجب أن نلاحظ أن هناك تمديد جذري يمكن أن يكون وهو اسلوب نموذج مفاهيم اللغة- المتامودل- ونسبة للمساحة يمكن الرجوع له في قياس الموف [2] .



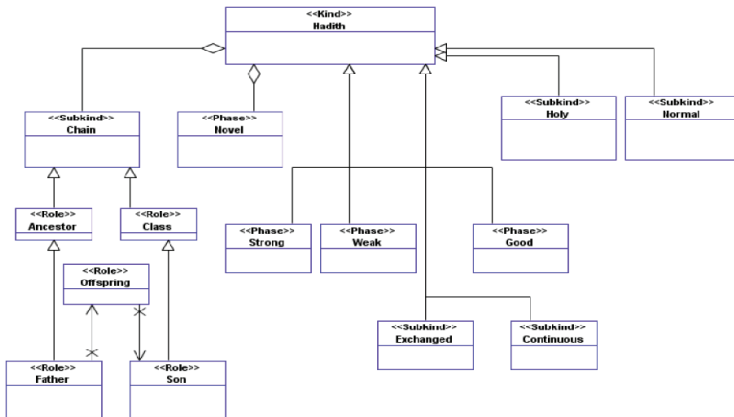
الشكل 2. نموذج الاصناف يوضح مفاهيم من علم الاحاديث الاسلامي

شكل 2 يوضح مفاهيم و مصطلحات في علم الحديث النبوي الشريف حيث يمثل ذلك في شكل عدد من الاصناف مع نظام معقد من السبكلاس بالإضافة لعدد بسيط من العلاقات (الاسوسيشن). من أنواع الحديث القدسي والنبوي وهما متقاطعان من حيث أنها أبناء للاب الحديث (لا يوجد صورة للحديث من النوعين في ان واحد) وكذلك هما مكتملان (كفرنك) (يعنى كل نسخة من كلاس الحديث هي أحد نوعين فقط لا ثالث لهما). المنسوخ والثابت (اكسجنج-كونتينيوس) أيضا لهم نصف الصفات. كل عضو منهم يظل بصورة دائمة نوع من أنواع الحديث ولا يتغير. هناك ثلاث درجات للحديث قوي، حسن وضعيف أيضا لهم خاصية التقاطع والتغطية ولكن دعنا نفترض أن أى عضو منهم يمكن ان يتغير من درجة الى أخرى (مثلا بعض الاحاديث التي لم تصنف أو ان النظام في أوله). أما الخلف والسلف والأبناء هم لتمييز الافراد (الرواة) والعلاقة بينهم. يمكن أيضا أن نقول عن هذا الشكل أنه بمثابة الاونتولوجي لانه يوصف مفاهيم مجال معين والعلاقة بينهم.

يمكن وضع العلاقة المعقدة للسبكلاس في هيئة أفضل وأكثر فهما بتميز أنماط الاشياء المختلفة. يمكن اعتبار الحديث نوع (kind) بينما holy, normal, exchanged, contiouns أبناء أو فرع (subkind) مادامت مصنفات مستقره في دورة حياتها في الواقع. أما نمط الاصناف مثل strong, good, week فهي نظام يعبر عن مراحل (phases) أو رتبه مادام الفرد منهم يمر بمرحلة أو له رتبه. أما البقية فتمثل أدوار (role). يمكن ان نضع هذه اللغة في بروفایل ولكي نحصل على ذلك أجرينا التعديل على عنصر من عناصر لغة اليوم ام وهو الصنف وكان التمديد بالانماط التي ذكرت وهنا تسمى بالاستروتايب (مظهر خاص) كما يظهر في شكل 3 والتي كل منها تم نمذجته كائن للصنف (عصر اليوم ام ال) والذي يسمى في هذه الحالة بالمتى كلاس ونتيجة لذلك نستطيع استخدام هذه اللغة الخاصة في النموذج في شكل 2 ليصبح كما في الشكل 4 والذي أصبح له معنى لان السبكلاس أصبح مرئي.



الشكل 3. مثال للبروفایل في لغة اليو ام ال



الشكل 4. يوضح تعليم البروفایل في شكل 3 لنموذج الحديث الموصف

4.1 منهج التصنيف المتبناه وبروفایل مطلوبات الجودة

في هذا الفصل سننظر كيف أن نموذج المعمارية هو فرصة كبيرة لتخفيف مشكلة تمثيل ونمذجة مطلوبات الجودة. الملمح الوراثي لمنهج هندسة المعمارية الذي تبنته ال ام دي ايه يقود الى تقديم حل عملي لعدد من مشاكل مطلوبات الجودة.

النقطة الاولى المتقدمة التي صنعتها هذه الطريقة أنها جعلت المطلوبات (غير رسميه) هي

النموذج(رسميه) فبالتالي حل لمشكلة التمثيل والذي سيكون مباشرا(البي اي ام) وهذا ما سيشرحه الفصل القادم. ثم انها تقدم هذا التجريد على مستوى العمل (البنزس) وهو مفهوم لدى أصحاب

المصلحة وفي نفس الوقت ممثل بطريقة تساعد على الأتمتة(تفهمه الاله).الميزة الأخرى أن حصر هذا النموذج في نطاق مجال معين دون التعرض لقضايا التكنولوجيا يساعد بصورة أساسية في إعادة الاستخدام لعدد من تطبيقات المجال. مثلا عدد المعاملات في جدول معين مرتفع لايعبر عن أي تقنيه أو تكنولوجيا تطبيق فهو مستقل اذا يمكن يعاد استخدام هذا المعلومه مع عدد من تطبيقات قواعد البيانات.

تقترح هذه الدراسة البحثية تصنيف جديد لمطلوبات الجوده من وجهة نظر منهج المعمارية وهي مطلوبات الحزمة(البكج) و مطلوبات الصنف(الكلاس).الاول يمثل و يعبر عن عمومية المطلوبات بينما الاخر يعبر عن محلية المطلوبات أو انها خاصة.مثلا تمثيل عدد الخصائص غي الهرم الوراثي (الحقول في الاب-الابناء في البو ام ال) في نموذج الاصناف (مخطط الصنف) لا يخص أي عضو من الأعضاء الاب أو الابناء بينما الجميع. يستفاد من هذه المعلومه في اختبار قرار التصميم المناسب والذي يدعم جوده معين(الكفاءة الزمنية). أما معلومة نسبة تكرار عملية تحديث الابن فهي خاصة بصنف معين وليست عامه.

نجد أحيانا أن هناك تصادم بين مطلوبات الجوده (الاداء مقابل الكفاءة) عند تأثيرها على قرارات التصميم كما سنشاهد في الحالة المدروسة في آخر الورقة لذي يجب حل ذلك التصادم.وذلك نسبة لان بعض قرارات التصميم تشترك في عدد من العوامل لسمات جوده مختلفه. سنسميها العوامل المؤثرة.

سنأخذ نفس الحالة التي تم توصيفها في شكل 1 لتوضيح التصنيف المذكور والتعرض لتلك القضايا.ثم سنبين تصميم وتطبيق البروفایل لمطلوبات الجوده فيها. هو برمجيات الرسائل الوسيطه مثل ال جاه ام اس(جافا). نجد أنه لدينا الانواع التالية من العوامل المؤثرة في جودة الكفاءة الزمنية و الاعتماديه.

- 1- Application sensitivity {High, low, medium},
- 2- AppKind {transactional, nontransactional},
- 3- CustomerType {normal, durable},
- 4- AppType {normal, critical},
- 5- and delivery Mode {normal, guaranteed}

هذه السمات(1-5) تلعب دورا في تحديد جودة المنتج مثلا في نظامنا قيد النظر نجد أننا اذا أردنا تحويل الب اي ام الى تصميم سنحتاج لاتخاذ قرار (بعد بناء البي اس ام) بشأن نوع الرسالة (دائمه/مؤقتة)، نوع الاتصال(مباشر/غيرمباشر)و نوع الجلسة (اجرائية/غير اجرائية). فمثلا

الرسالة الدائمة تؤثر على الاداء الزمني سلبا لانه بها عملية تخزين على عكس المؤقتة. نوع الاتصال غير المباشر يتطلب موارد لان الارسال فيه يكون لعدد من المستخدمين في ان واحد. بينما الجلسة الاجرائية تطلب عمل كبير من النظام فيه يحافظ على الاتساق،الاكتمال والضمان بنفس طريقة قواعد البيانات وهذا يؤثر على الزمن بينما يدعم الاعتمادية. أما طريقة التسليم (5) يكون الاتجاه المضمون يتطلب رعاية خاصة للبيانات حتى في حالة قطع الاتصال أو غياب الطرف الآخر لابد من الحفاظ عليها وذلك ينعكس سلبا على جودة الاداء بينما ايجابا لوجودة الاعتمادية. الجدول رقم 1 يظهر بصورة جلية ماهي العوامل المؤثرة أو مطلوبات الجودة في القرارات التصميمية في أول عمود من اليمين. نلاحظ من الجدول أن سمات الجودة قد منحت درجات مثلا عالي/ ومنخفض بالنسبة sensitivity و مضمون/طبيعي في حالة delivery mode. بناء على ذلك يتم اختيار 1.1.1 أو 1.1.2 لتطبيق ما وهلم جرا. نلاحظ كل هذه مصطلحات تصميم لا علاقة لها بمستوى التطبيق (شكل 1) وهو الفصل الذي قصده منهج نموذج المعمارية.

نلاحظ أن مطلوبات الجودة 3،5 تمثل مطلوبات صنف (خاصة) بينما 1،2،4 هي مطلوبات بكج بمعنى أنها على مستوى النموذج ككل.

نخلص الى أن تمثيل مثل تلك المطوبات (1-5 أعلاه) مهم للغاية ليسير عملية التحويل الالي لنموذج المطوبات (البي أي ام) الى نموذج التصميم (البي أي اس ام) في بيئة نموذج المعمارية والذي تقوم به ادوات لغة الكيوفيتي.

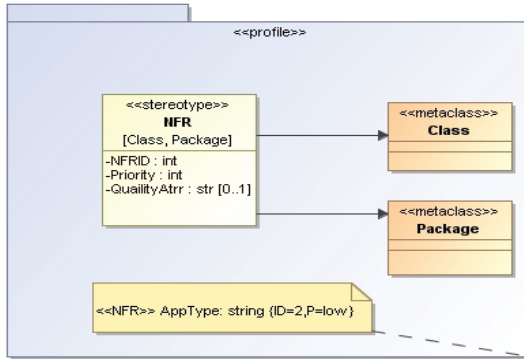
4.1 تصميم بروفایل مطلوبات الجودة وتطبيقه

حتى يصبح من الممكن تمثيل تلك المطلوبات على النموذج المستقل عن التقنيه نحتاج الى صناعة لغة خاصة (مظهر خاص) كما أسلفنا في الفصل 2. سيتم تمديد عنصر واحد من عناصر اليوم ال المتامودم وهو الصنف (متاكلاس) وذلك لانه يمثل نوع لكل من الصنف نفسه و الحزمة (البكج) والذان يمثلان النوعان المراد نمذجتهم. المعلومات المطلوبه في هذا البروفايل هي رقم لتمييز مطلوب الجودة وألوية تحدد بارقام في حالة حدوث التصادم بالاضافة الى التأثير وهو نص يمكن أن يرمز الى التأثير المتوقع. في شكل 5 كل ذلك يظهر كخصائص للصنف متاكلاس والذي تم تمديده بالطفرات (استروتايب). تطبيق هذه اللغة الجديد على نموذج نظام المساعدة الالي في شكل 1 يظهر لدينا في الشكل 6. نلاحظ أننا استطعنا بعد تعريف البروفايل أن نعلم delivery mode بأنها رقم 4 وتمتلك الأولوية P=normal. نلاحظ أن هذه المطلوبات يمكن اعادة استخدامها مع تطبيقات من المجال نفسه مثل الايميل ، الجات،المنتدى الخ. لانها كلها تستخدم نفس قرارات التصميم (نظام الرسائل). ان ميزة اعادة الاستخدام هذه تكمن في تحقيق الاعتمادية(الخلو من الأخطاء) في عملية تطوير البرمجيات اضافة الى صنع استقرار يساعد في الانشغال بالقضايا المهمة. علاوة على ذلك بهذا الطريقة نحصل على توثيق للمطلوبات ممنهج ومتسق يعتبر بمثابة أصول.

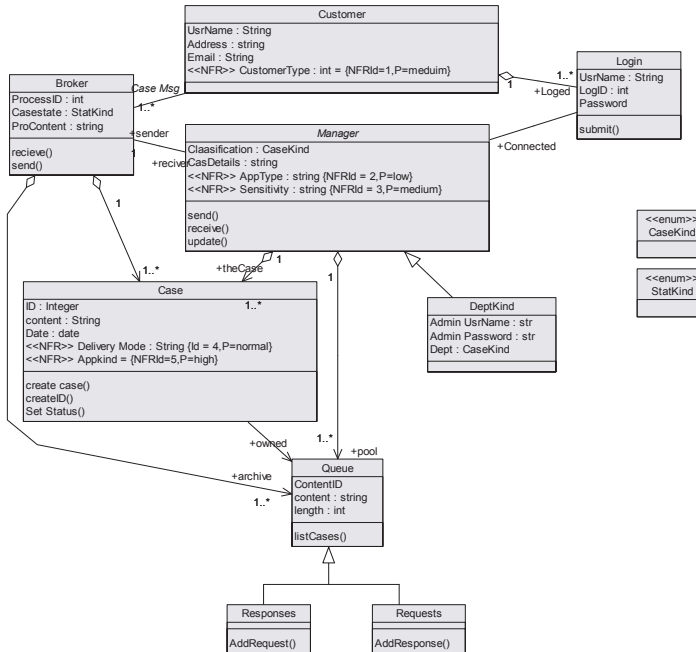
مجال مطلوبات الجودة مازال غير مؤسس كغيره من المجالات مثل المطلوبات الوظيفية ويحتاج الى بحث. هناك عدد من المجهودات مثل باس وآخرون [6] ، زوه ، وإيان [7] ، شنغ [10] ، و قروس وآخرون [11]. المنهج الذي اتبعناه يعتمد على فكرة نموذج المعمارية وبصورة سهلة فمثلا زوه ، وإيان [7] صمم يو ام ال بروفايل عام لمطلوبات الجودة الا أنه لا يصلح للاستخدام هنا لانه بني على اعتبار أن مطلوبات الجودة و قرارات التصميم تكون في مكان واحد. وهو يخالف مبدأ نموذج المعمارية . إطار-السته عناصر (معهد هنسة البرمجيات العالمي) [6] تبني منهج منبني على السنايرو للفصل بين التداخل الذي يحدث بين مطلوبات الجودة. منهجنا تبني طريقة مبسطة تعتمد على الاولوية لحل مشكلة التصادم وبالتالي اختيار قرار التصميم المناسب بصورة تدعم الأتمته ، ولكنه يعتبر منهج مفيد خاصة في مرحلة تحليل المطلوبات الاوليه. الدراسة التي قدمت كانت واسعة وجيده . أغلب الاسهامات لم تركز على مسالة الاتمته وإنما على الجانب البشري فقط كمثال زوه ، وإيان [7] تتبنى support, break, help and hurt في العلاقة بين قرارات التصميم ومطلوبات الجودة.

الجدول 1: يوضح علاقة مطلوبات الجودة (1-5 أعلاه) من مجال برنامج مكتب المساعده اللالي بقرارات التصميم

قرارات التصميم في المعتمد النموذج على التكنولوجيا	سمات الجودة المناسبه	النموذج المستقل عن التقنيه
1.1 Message 1.2 connection	البيانات واجب وجودها مطلوبه بين المستهلك والمنتج. لذا هي متطلب وظيفي لكن في طريقة معالجتها هناك جوده على حسب نوعها وأولويتها.	1-Case
1.1.1 persistence or 1.1.2 nonpersistence	sensitivity {low, high} and delivery mode {normal, guaranteed, ...}	1.1 Data [message]
1.2.1 queue or 1.2.2 topic	CustomerType and AppKind {OLTP, normal}	1.2 Connection
1.2.3 transacted 1.2.4 Nontransacted	Appkind, AppType {critical, normal}	1.2.x.1 session



الشكل 5. بروفایل مطلوبات الجودة



الشكل 6. يوضح تطبيق لغة مطلوبات الجودة على النموذج المستقل عن التقنية في شكل 1

أتجاه نموذج المعمارية هو منحى جديد يعتمد على النماذج في تسير عملية تطوير البرمجيات. بناء هذه النماذج بلغات عالية المستوى مثل الموف واليو ام ال يجعلها غير معتمدة على التكنولوجيا ثم الفصل بيت نموذج المطلوبات ونموذج التقنيه يساعد على اعادة الاستخدام و الذي بدوره يحقق اعتماده في بناء البرمجيات.

الاهتمام بمطلوبات الجودة كان محور هذه الورقة حيث أنه ظهر مؤخرا بظهور المعمارية والتي أتاحت فرصة لدراسة متطلبات الجودة في مرحلة مبكرة وذلك بسبب أنها أصبحت تمثل التكلفة الحقيقية للمشروع والاستقراره. كانت الصعوبات البحثية تتمثل في تعريف وتصنيف وتمثيل مطلوبات الجودة. لم يكن هناك اتفاق بصورة عامة لتعريفها ولا لتصنيفها.

أظهرت هذه الورقة العلاقة بين النموذج المستقل عن التكنولوجيا (بي اي ام) في منهج نموذج المعمارية ومطلوبات الجودة اتضح من خلالها أنه يمكن تمثيل سمات الجودة في شكل عوامل على البي اي ام سمينها العوامل المؤثرة. درسنا حالة من الواقع أظهرت أن هناك عدد من سمات الجودة كانت بالالخص في الاعتماديه والاداء الزمني. ظهر لنا من ذلك تصنيفان جديان وهما مطلوبات صنف ومطلوبات حزمة (بكج).

قمنا أيضا بتصميم لغة خاصة (مظهر خاص) أو بروفابل يسمح للمعماري باستخدامه مع نموذج البي اي ام للتعبير عن سمات الجودة أو العوامل المؤثرة. هذه الطريقة تساعد على الاتمته لأنها مفهومة للاله.

بشكل عام مطلوبات البكج لها فرصة أكبر في اعادة الاستخدام من مطلوبات الصنف لأنها خاصة ولكن تحقق مستوى من اعادة الاستخدام داخل عدد من تطبيقات المجال. بهذه الصورة نجد أن هذا المنهج التوصيفي والمستقل عن التقنيه يقدم الية تزيد من رصيد المصنوعات البرمجية في المكتبة (بنفس طريقة الكمبوننت) وبشكل أفضل. إن نجاح تقنية الساب لم يكن الا في مثل ذلك الاتجاه (رصيد توصيفات).

المراجع

- [1] MDA guide version 1.0.1[online]. Available: <http://www.omg.org/cgi-bin/doc?omg/03-06-01>.
- [2] OMG MOF QVT Final Adopted Specification [online]. Available: <http://www.omg.org/docs/ptc/05-11-01.pdf>, 2005.
- [3] Frankel, S, D., Applying MDA to Enterprise Computing. USA:Wiley Publishing.2003.
- [4] Favre, J.M, Dragan, G, Lammel, R., and Winter,A..Guest Editors' Introduction to the Special Section on Software Language Engineering.Proceding of IEEE Transactions on Software Engineering,

- November/December 2009. 35(6): 737-741
- [5] Dueñas, J. C., de Oliveira, W. de la Puente, J. A software architecture evaluation model. In Proceedings of the second international ESPRIT ARES workshop . Development and Evolution of Software Architectures for Product Families..Las Palmas de Gran Canaria, Spain: Springer-Verlag.148-157; 1998.
- [6] Bass, L., Clements, P., and Kazman, R. Software Architecture in Practice.2ed edition. Reading, Massachusetts:Addison-Wesley.2003
- [7] Zhu,L. and Ian, G. UML Profiles for Design Decisions and Non-Functional Requirements. Processdings of the Second Workshop on SHaring and Reusing architectural Knowledge Architecture, Rationale, and Design Intent (SHARK-ADI'07). May2007. Washington, DC, USA:IEEE.2007.8.
- [8] Matinlassi, M. *Quality-driven Software Architecture Model Transformation*.PhD Dissertation. University of Oulu; 2006.
- [9] Glinz, M. On Non-Functional Requirements. In: 15th IEEE International Requirements Engineering Conference (RE 2007). 15-19 Oct., 2007. Delhi :IEEE. 21–26;2007
- [10] Chung, L., Nixon, B. A., Yu, E., Mylopoulos, J. *Non-functional Requirements in Software Engineering*. Dordrecht: Kluwer Academic Publishers.2000
- [11] Gross D., and E. Yu, From Non-functional Requirements to Design through Patterns.*Requirements Engineering*. 2001. 6(1):18-36.

المخلص باللغة الانجليزية

Title: The nonfunctional requirements classification and representation in PIM o MDA

Abstract. Model-driven architecture (MDA) is a new trand in engineering software whereby software development is driven by models. Platform independent model (PIM) is built to specify application requirements(business rules) using UML-based languages.The platform specific model (PSM)is for specifying implementation aspects.Then transformation language such as QVT used to do the mappings between PIM and PSM automatically.This this process is still under research. Since the effect of not concerning the quality at very early stage in software artifact becomes visible and there is no previous work tackle the relationship between PIM and nonfunctional requirements, here the contribution of this paper comes.To this end two new classification for nonfunctional requirements are found and UML-profile is designed to allow an architect to annotate PIM with quaitly attributes. Moreover, by this way, we will help automate the transformation process upon which MDA engineering is based.

Keywords: Model driven architecture, PIM, Profile, Re-use**المصطلحات**

النموذج	model
مطلوبات الجودة أو سمات الجودة	Nonfunctional requirments
نموذج المعمارية	Model Driven Architecture
النموذج المستقل عن التكنولوجيا	Platform Indpentant model(PIM)
النموذج المعتمد على تقنية	Platform specific model (PSM)
بالاستروايب (مظهر خاص)/ الطفرة	stereotype
بروفایل/لغة مظهر خاص	Profile
أوه سي ال	Object constraint lanaguge(OCL)
متامودل، نموذج-النموذج	Metamodel
هيئة إدارة الكائن (أوه ام جي)	OMG
الاستفسار-مشهد-تحويل/الكيفيتي	Query-view-transform(QVT)
الامتة	autmation
رمو البرنامج/الكود	code
جاه ام اس	Java Messaging system(JMS)
الاني ام اف	elcipse modeling framework (EMF)
الصف/مخطط الصف	Class/Class diagram
السيكلاس/الابن	Subclass
العلاقات	association
التقاطع	disjoint
الاعتماديه	reilability
الكفاءة الزمنية	performance
خصائص	properties
صورة/حالة	instance
الاتصال	connection
التسليم	delivery
اجرائية	transactional
جلسة	session
أصول	core asset
كفر نك/التغطيه	covering

تأثير نموذج العضلة على تصميم استراتيجيات التحكم بقيادة الدراجة عن طريق الإثارة الكهربائية

رشا كاظم مسعود

جامعة دمشق، كلية الهندسة الميكانيكية و الكهربائية، قسم الهندسة الطبية، طريق المطار، دمشق، سورية

البريد الإلكتروني: rqies@yahoo.com

الخلاصة. في هذا البحث تمت دراسة ثلاثة نماذج للعضلات الهيكلية، و كان أول نموذج هو نموذج Ferrarin للعضلة و هو عبارة عن تابع تحويل بين الإثارة الكهربائية وعزم الركبة. بينما يعتمد تصميم النموذجين الآخرين (نموذج Riener ونموذج Virtual Muscle) على فزيولوجية العضلة الهيكلية، تمت نمذجة عضلة Riener باستخدام برنامج Matlab/Simulink، بينما بني نموذج virtual muscle باستخدام برنامج Virtual Muscle 4.0.1 و الذي يمكن تحميله من الشابكة. تمت نمذجة العضلة رباعية الرؤوس الباسطة للركبة باستخدام النماذج الثلاثة، و اختبرت هذه النماذج من حيث الاستجابة للإثارة الكهربائية، ومن ثم استخدمت في استراتيجيات للتحكم تهدف إلى التحكم بسرعة دوران مسنن الدراجة. تمت مناقشة و تقييم أداء النماذج الثلاثة خلال عملية التحكم، و تبين أن نوع نموذج العضلة يؤثر على تصميم استراتيجيات التحكم.

الكلمات الجوهرية: نموذج العضلة، المنطق العائم، العضلة الافتراضية، الإثارة الكهربائية الوظيفية

1 مقدمة

ما زال الباحثون يدرسون نماذج العضلات منذ أكثر من خمسين عاماً [1]، و عادة ما يعتمد نموذج العضلة على التجارب العملية [2] أو فزيولوجية العضلة أو علم حركة العضلة [1]. إن أشهر نماذج العضلات هي نماذج العضلات المعتمدة على نموذج Hill، حيث أن بارامترات هذا النموذج تصف المبادئ المعروفة المتعلقة بتركيب القسم العضلي، و يتألف هذا النموذج من معادلات متعاقبة تعتمد على النتائج التجريبية، و هذا ما جعل النموذج مرغوباً أكثر من تلك النماذج المعتمدة على النظريات كونه أكثر دقة [3].

لقد بذلت جهود حثيثة لتطوير نماذج للجسم البشري من أجل توفير الوقت و التكلفة مقارنة مع التجارب العملية. و قد تم استخدام نماذج الجسم البشري الفعالة (التي تولد الحركة اعتماداً على القوة المولدة من قبل العضلات المحركة) في الكثير من المجالات مثل علم ميكانيك الرياضة (sport's mechanics) و علم حساب الطاقة (ergonomics) و الميكانيك الحيوي (biomechanics) و ذلك لتحسين أداء الإنسان و ضمان السلامة في شروط العمل.

خلال العقود القليلة الماضية أصبحت المحاكاة الحاسوبية لحركات الجسم البشري باستخدام المقاطع المتعددة و نماذج العضلات المتعددة رائجة جداً [4]، لأنها توفر أساساً لتوصيف ميكانيك العضلة، و تصف كيف تولد العضلات الحركة [5].

إن شلل الأطراف السفلية هو عجز وظيفي يحول دون قدرة المريض على الحركة والانتقال من مكان لآخر، وينجم عن مرض أو إصابة في الجهاز العصبي المركزي تحدث توقفاً أو انقطاعاً في مسار الإشارة المرسلة من الجهاز العصبي المركزي إلى العضلات المحركة، و بالرغم من أن العمود الفقري قد حصل فيه أذى إلا أن العضلات مازال سليمة، وبالتالي يتم تمرين العضلات الكبيرة للمشلول من أجل إعادة تأهيله و تحسين صحته. تتم عملية إعادة التأهيل بتطبيق نبضات إثارة كهربائية وظيفية على العضلات المشلولة للأطراف السفلية لقيادة دراجة التمرين، و التي تستخدم للتقليل من فرط التوتر التشنجي في العضلات و تمنع ضمور العضلات، كما تزيد من لياقة الأوعية القلبية وحجم العضلة و تدفق الدم و كثافة العظام الموجودة في الساق، بالإضافة إلى أنها تقلل من فرصة تطور التقرحات الناجمة عن الضغط و الاضطرابات الوعائية القلبية.

في هذا البحث، يتم استخدام كلاً من نماذج الجسم البشري و العضلات من أجل القيام بمحاكاة حاسوبية لعملية التحكم بسرعة دراجة هوائية ثابتة عن طريق التحكم بنبضات الإثارة الكهربائية الوظيفية المطبقة على عضلات إنسان ذو أطراف سفلية مشلولة.

إن نجاح أي نظام للإثارة الكهربائية الوظيفية يعتمد على نظام التحكم المستخدم، و الذي يجب أن يحقق متطلبات مشكلة التحكم المرتبطة بطبيعة الإعاقة. حيث يكون إشراف المستخدم كافياً في بعض التطبيقات، و لكن عادة ما يرغب بمستويات تحكم أخرى.

و من أجل التحكم عن طريق الإثارة الكهربائية الوظيفية، تستخدم نبضات من التيار الكهربائي لإثارة أعصاب العضلات و هذا يؤدي بدوره إلى تقلص العضلات منتجة قوة عضلية تؤدي إلى نشوء عزم حول المفصل مما يسمح بتحريك الجسم. استخدمت الدراسات عدة طرق للتحكم بقيادة دراجة إعادة التأهيل و ذلك باستخدام عدة طرق من الإثارة الكهربائية الوظيفية [6] [7] [8] [9]. كما استخدم معظم الباحثون المحرك ليساعد في التغلب على النقاط الميتة في الحركة أثناء التدوير، و أيضاً كمولد للحركة بالاتجاه المعاكس من أجل مقاومة الحركة [10] [11] [12] [13] [14]. و قد استخدم المرجع [11] تقنية التحكم باستخدام المنطق العائم (fuzzy logic) من أجل التحكم بنبضات الإثارة الكهربائية الوظيفية للحصول على حركة دورية منتظمة لدراجة إعادة التأهيل، و هذا تم عن طريق إثارة عضلات الركبة الباسطة والقابضة.

تضمنت جميع الدراسات السابقة إثارة مجموعتين أو أكثر من العضلات للتحكم بقيادة دراجة إعادة التأهيل عن طريق الإثارة الكهربائية الوظيفية، و لكن هذا البحث ينتهج طريقة باستخدام متحكم المنطق العائم (model free fuzzy logic controller) الذي لا يعتمد على النموذج الرياضي من أجل التحكم بالحركة الدورية للدراجة، و عن طريق إثارة العضلات الباسطة للركبة فقط. وتم تطبيق طريقة التحكم هذه على ثلاثة نماذج مختلفة للعضلات الباسطة للركبة و التي تعتمد إما على نموذج هيل (Hill) أو على النموذج الرياضي.

2 نموذج الجسم البشري

تم تطوير نموذج الجسم البشري باستخدام برنامج Visual Nastran (VN) و تم حساب الأبعاد الفيزيائية لمقاطع الجسم البشري حسب المرجع [13]، حيث تم افتراض أن ارتفاع النموذج 180 cm و وزنه 70Kg. كما أن المفاصل القابلة للتحكم بها هي كل من مفاصل الركبة و الدولاب المسنن في الدراجة (الترس)، أما مفاصل الطرف العلوي فليس لها دور مهم في عملية التدوير. و لجعل التحليل أسهل تم بناء المفاصل بحيث يكون لها درجة حرية وحيدة، أي تسمح لهذه المفاصل أن تتحرك في المستوي السهمي فقط، و يبين الشكل 1 النموذج الكامل لجسم الإنسان مع الدراجة.



الشكل 1 نموذج الجسم البشري مع الدراجة

3 نموذج العضلة

3.1 نموذج العضلة الممثل بتابع تحويل

يقسم نموذج مفصل الركبة إلى النموذج الخامل (غير الفعال) و النموذج الفعال الذي يُمثل بتابع تحويل بين الإثارة الكهربائية المطبقة و عزم الركبة الناتج، و قد تم تعريف هذا التابع في المرجع [14] و الذي استخدم نموذج من نوع ARX و طبق طريقة مربعات الخطأ الصغرى (least square method) للأخطاء المحسوبة بين القيم الحقيقية و خرج النموذج. إن النموذج الديناميكي لمفصل الركبة هو تابع من الدرجة الأولى و يعطى بالعلاقة:

$$H(s) = \frac{G}{1 + s\tau} \quad (1)$$

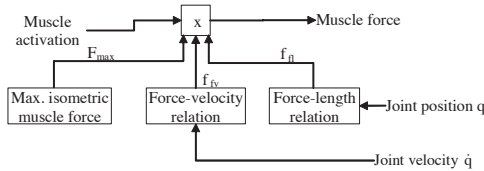
حيث τ هو ثابت زمني و G هو التضخيم، لقد وجد Ferrarin و زملاؤه عام 2000 [14] أن الثابت الزمني τ مستقل عن تردد الإثارة الكهربائية، و لكنه يعتمد بشكل كبير على شكل نبضات الإثارة (خطوية أو صاعدة)، كما أن التضخيم G يعتمد على تردد نبضات الإثارة. كما صرح المرجع [14] بأن نموذجاً أحادي القطب (من الدرجة الأولى) مع مضخم يعتمد على تردد نبضات

الإثارة هو نموذج قادر على تحديد العلاقة بين الإثارة و العزم الفعال للمفصل بشكل دقيق إلى حد ما.

و يعد هذا النموذج مناسباً لمجال حركة الركبة في هذا البحث، حيث أنه قادر على محاكاة خصائص مجموعة العضلات الباسطة للركبة الموجودة في الفخذ، و التي تولد عزم بسط فعال على مفصل الركبة. و سوف يشار إلى هذا النموذج بنموذج Ferrarin أينما ورد في هذا البحث.

3.2 النموذج الفيزيولوجي للعضلة المثارة

في هذا النوع من النماذج ينقسم نموذج مفصل الركبة إلى نموذج مفصل الركبة الخامل و الذي خرجة هو العزم الخامل لمفصل الركبة المحسوب من المرجع [15] ، و نموذج مفصل الركبة الفعال و الذي يمثل خرجة بالعزم الفعال على مفصل الركبة و الذي يمكن حسابه عن طريق ضرب قوة العضلة المثارة الناتجة عن تفعيل العضلة بذراع العزم للعضلة المثارة. و قد تم شرح نموذج تفعيل العضلة بالتفصيل في المرجع [16] حيث أن العناصر الرئيسية لهذا النموذج هي: تفعيل العضلة و العلاقة بين الطول و القوة و العلاقة بين السرعة و القوة. و يظهر الشكل 2 أنه بضرب هذه العناصر مع بعضها يمكن حساب قوة العضلة الناتجة.



الشكل 2 نموذج تقلص العضلة [16]

يمكن تمثيل نموذج العضلة من وجهة نظر هندسية بصندوق كما في الشكل 3، حيث أن مداخله هي إشارة الإثارة الكهربائية و زاوية المفصل الفعال، بينما تمثل قوة العضلة الناتجة خرج هذا الصندوق.



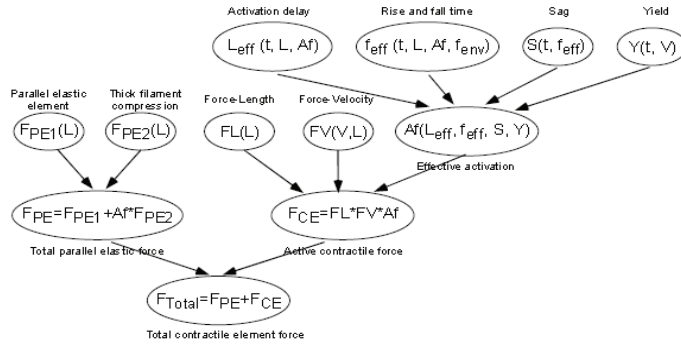
الشكل 3 المخطط الصندوقي لنموذج العضلة

و قد تم بناء أحد هذه النماذج بواسطة الباحث Riener [16] لذلك سوف نطلق على هذا النموذج اسم نموذج Riener أينما ورد ذلك في هذا البحث. يعتمد هذا النموذج على فيزيولوجية العضلة كما أنه يعتمد على زاوية الركبة و سرعتها الزاوية.

3.3 النموذج الذي تم بناؤه باستخدام برنامج Virtual Muscle

تم تصميم برنامج Virtual Muscle (VM) ليعمل في بيئة Matlab/Simulink. و يستخدم البرنامج تابعي Matlab لكل منهما واجهة اتصال مع المستخدم يسميان: بناء نوع الألياف (build fiber type) و بناء العضلات (Build Muscles) يقوم المستخدم بتعريف خصائص كل نوع من الألياف و قياس شكل كل عنصر من أوتار العضلة. و بعد تحديد خصائص كل عنصر وتري عضلي يتم بناء صندوق مستقل للـ Virtual Muscle داخل بيئة الـ Simulink حيث يتم توصيلها مع النموذج الهيكلي للجسم البشري و نظام التحكم [17].

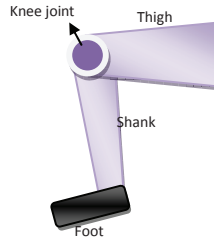
يستخدم برنامج Virtual Muscle لنمذجة الخصائص الميكانيكية المعقدة للعضلات و الأوتار عند شروط توظيف العضلة و أثناء حركتها، و يبين الشكل 4 مخططاً يمثل النموذج العضلي [17].



الشكل 4 مخطط لنموذج Virtual Muscle [17]

4 استجابة نموذج العضلة

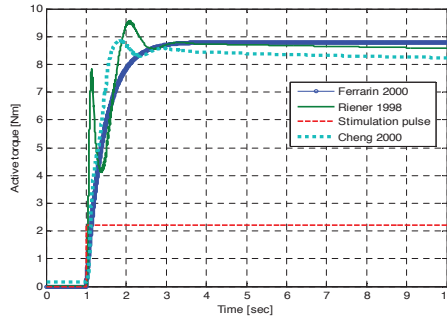
تمت نمذجة الساق الحرة المتأرجحة و المثبتة عند مفصل الورك و الكاحل كما في الشكل 5 و تم التحكم بمفصل الركبة من خلال برنامج Matlab/Simulink عن طريق العزم الفعال الناتج عن نموذج العضلة الباسطة للركبة [15]. و اختبرت نماذج العضلات الثلاثة و سجلت استجاباتها.



الشكل 5 نموذج الساق الحرة المتأرجحة

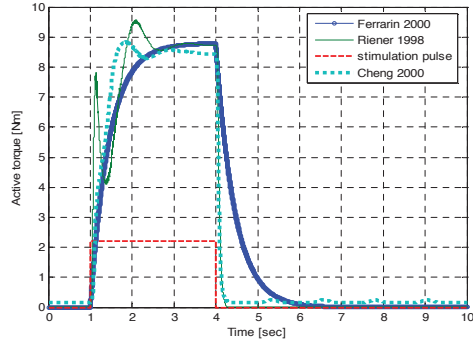
تم تطبيق قطار من نبضات الإثارة على نموذج العضلة و تسجيل العزم الناتج عن كل نموذج (انظر الشكل 6). و من الملاحظ أن زمن الاستقرار هو نفسه لكل النماذج، بينما يؤد كل من نموذجي Riener و VM قمة كبيرة و الكثير من التعرجات عند فترة العبور، و هذا يعود لاعتماد كلا نموذجي العضلة على الزاوية و السرعة الزاوية لمفصل الركبة، إن هذه التعرجات تحدث أيضاً في تابع الاستجابة للعضلة البشرية الحقيقية. و بما أن نموذج Ferrarin للعضلة لا يعتمد على حركة مفصل الركبة لذلك تكون استجابة العبور له أكثر نعومة.

بعد عدة ثوان من بداية الإثارة الكهربائية للعضلة يبدأ العزم الفعال للركبة في نموذج Riener للعضلة بالانخفاض وذلك بسبب نموذج التعب العضلي المتضمن داخل النموذج، كما يُظهر نموذج Virtual muscle نفس الانخفاض، بينما يبقى العزم الناتج عن نموذج Ferrarin ثابتاً، وذلك لأن هذا النموذج لا يأخذ التعب العضلي بعين الاعتبار.



الشكل 6 استجابة العضلات الثلاث لقطار من نبضات الإثارة الكهربائية

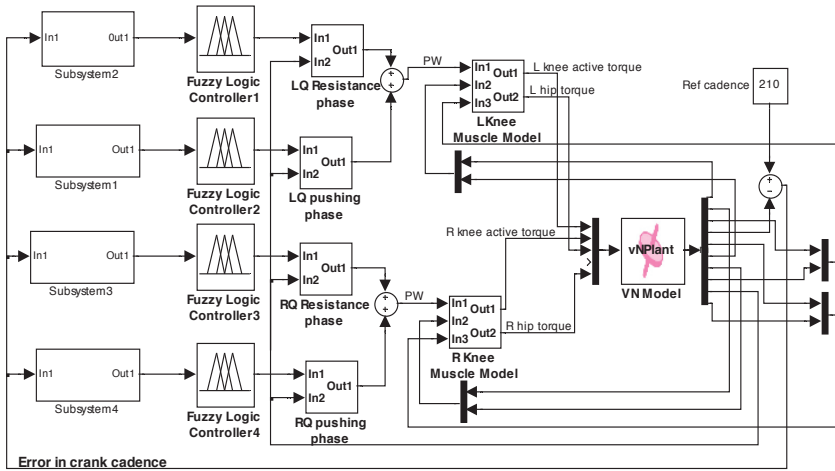
كما يلاحظ اختلاف مهم آخر خلال فترة الهبوط، فعندما تتوقف الإثارة الكهربائية تبدأ منحنيات الاستجابة للنماذج العضلية الثلاثة بالانخفاض في نفس الوقت، ويكون التأخير الزمني في الوصول إلى الصفر في كل من نموذجي Riener و VM حوالي 0.25 sec وذلك لأن حركة شوارد الكالسيوم متضمنة في كلا النموذجين، بينما يكون التأخير الزمني في نموذج Ferrarin حوالي 2 sec و ذلك يعود إلى معادلة تابع التحويل (انظر الشكل 7). هذه الاختلافات في منحنيات تابع الاستجابة تلعب دوراً مهماً في تصميم استراتيجيات التحكم لاستعادة الحركات الوظيفية عن طريق الإثارة الكهربائية للعضلات الموافقة [19].



الشكل 7 استجابة نماذج العضلات لنبضة إثارة كهربائية

5 استراتيجيات التحكم

تم بناء نموذج الجسم البشري و الدراجة باستخدام برنامج Visual Nastran كما ذكر سابقاً، و بناء نظام التحكم باستخدام بيئة Matlab\ Simulink كما في الشكل 8، حيث يتم تطبيق العزم الناتج عن نماذج العضلات الباسطة للركبة اليمنى و اليسرى على مفاصل الركبة الأيمن و الأيسر على الترتيب. إن الدخل الأول لمتحكم المنطق العائم هو الخطأ بين السرعة الحقيقية لمسند الدراجة و السرعة المرجعية، أما الدخل الثاني فهو مشتق ذلك الخطأ. تم استخدام أربعة متحكمات معتمدة على المنطق العائم للتحكم بأربعة أطوار أساسية للتدوير، هذه الأطوار هي: طور الدفع للساق اليمنى و طور المقاومة للساق اليمنى، و طور الدفع للساق اليسرى و طور المقاومة للساق اليسرى. تلعب كل عضلة باسطة للركبة دورين خلال دورة التدوير: (1) دور الدفع لزيادة سرعة الدراجة، و دور مقاوم للسرعة عندما تزيد عن الحد المطلوب.



الشكل 8 المخطط الصندوقي لنظام التحكم الذي تم تصميمه في بيئة Matlab\Simulink

إن العائق الرئيسي في عملية التحكم بقيادة الدراجة هو النقاط الميتة والتي تعرف بأنها موضعي الدواسات الذي يكون عندهما العزم الصافي لدوران مركز المسنن مساوياً إلى الصفر [18]. و عادة ما تحصل هذه النقاط (النقاط الميتة) عند الزوايا 0° و 180° للمسنن.

و من أجل التغلب على النقاط الميتة عند الإنسان الطبيعي، يجب على السائق الموافقة الانبساط والدفع من أجل تحريك المسنن باتجاه حركة الدراجة، بينما تقوم الساق المقابلة بالانقباض لإعطاء عزم محرك كاف و ذلك للتغلب على هذا الوضع. و بعد اجتياز النقطة الميتة تبدأ سرعة الدراجة بالازدياد و تبدأ الساق الموافقة بالانقباض للتغلب على هذه الزيادة و لإزالة أثر الانبساط السابق. هذه الأعمال المهمة لعضلات الركبة القابضة خلال عملية التدويس جعلتها أساسية في هذه العملية، و قد استخدمها الباحثون كعناصر مهمة عند قيادة الدراجة من قبل المشغلين.

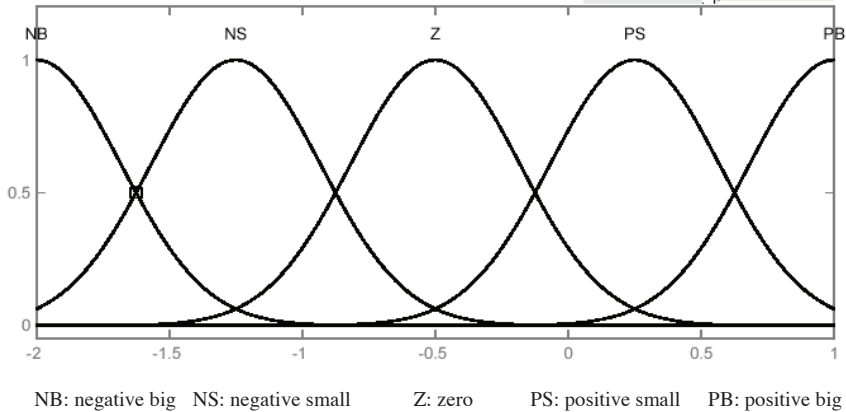
لم تتم إثارة عضلات الركبة القابضة للركبة في هذا البحث، و كانت الوسيلة الوحيدة لإنقاص سرعة التدويس بعد تجاوز النقاط الميتة هي تطبيق عزم معاكس لاتجاه الحركة. و هذا يمكن تحقيقه عند تطبيق عزم بسط على مفصل الركبة للساق المقابلة عوضاً عن قبض مفصل الركبة للساق الموافقة. أي أن الساق المقابلة تقوم بمقاومة الازدياد في سرعة التدويس، وهذا ما يطلق عليه طور المقاومة للساق المقابلة. كما يساعد هذا الانبساط للرجل المقابلة لتحقيق طور الدفع التالي بنجاح، حيث أن الساق التي سميت سابقاً بالمقابلة سوف تصبح الساق الموافقة في النصف التالي من دورة التدويس، و تعوض هذه الآلية عن عدم وجود عضلات قابضة لمفصل الركبة.

يمكن تمثيل أطوار التدويس السابقة الذكر كمخططات للإثارة كما في الشكل 11 و الذي يبين مواضع و أطوار الإثارة لكل من العضلات الباسطة للركبة اليمنى و اليسرى. و تبين الأقواس في المخطط مجالات زوايا المسنن التي تثار عندها العضلات.

5.1 المتحكم المعتمد على المنطق العائم

تعمل كل عضلة باسطة للركبة (اليسرى و اليمنى) خلال طورين (المقاومة و الدفع) و لكل طور متحكم خاص به. تعد عملية اخنيار مداخل و مخارج المتحكم حرجة للغاية، لأنه من الضروري التأكد من أن جميع المعلومات الضرورية حول النظام المتحكم به متوفرة من خلال مداخل المتحكم. لكل متحكم مدخلين هما الخطأ في سرعة دوران المسنن e و مشتق هذا الخطأ Δe ، و اللذين يتم تقييسهما باستخدام معاملي التقييس G_e و $G_{\Delta e}$. و بعد ذلك تتم عملية التحويل إلى متغيرات لغوية، حيث تم تعريف خمس من مجموعات المنطق العائم باستخدام توابع عضوية غوسية، و تم استخدام 25 قاعدة من قواعد المنطق العائم. أما مخرج المتحكم فهو التغير في عرض نبضات الإثارة و التي يتم ضربها بمعامل التقييس G_o

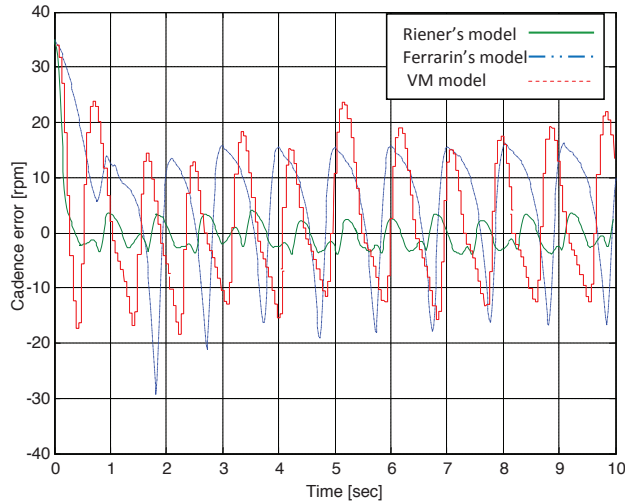
و قد تم توزيع توابع العضوية بشكل متناظر من أجل جميع مداخل و مخارج متحكم المنطق العائم، إلا أن المجال العائم لمخرج متحكم طور المقاومة ليس متناظراً كما في الشكل 9، حيث أن المجال يتراوح بين $[-2, 1]$ و أن تابع العضوية المسمى بـ Z يعطي مخرجاً مقداره -0.5 .



الشكل 9 مجال الخرج غير المتناظر في طور المقاومة [20]

6 النتائج

تم في هذا البحث التحكم بسرعة دراجة إعادة التأهيل (و التي تستخدم الإثارة الكهربائية الوظيفية للعضلات) باستخدام استراتيجية تحكم تعتمد على المنطق العائم و ذلك للمحافظة على سرعة دوران ثابتة لمسنن الدراجة وهذا بدوره يحقق قيادة سلسلة للدراجة. إن سرعة الدوران المرجعية للمسنن في هذا البحث هي 35 rpm و الشكل 10 يبين الخطأ في سرعة دوران المسنن عند استخدام كل نوع من نماذج العضلات المذكورة سابقاً.



الشكل 10 الخطأ في سرعة دوران مسنن الدراجة

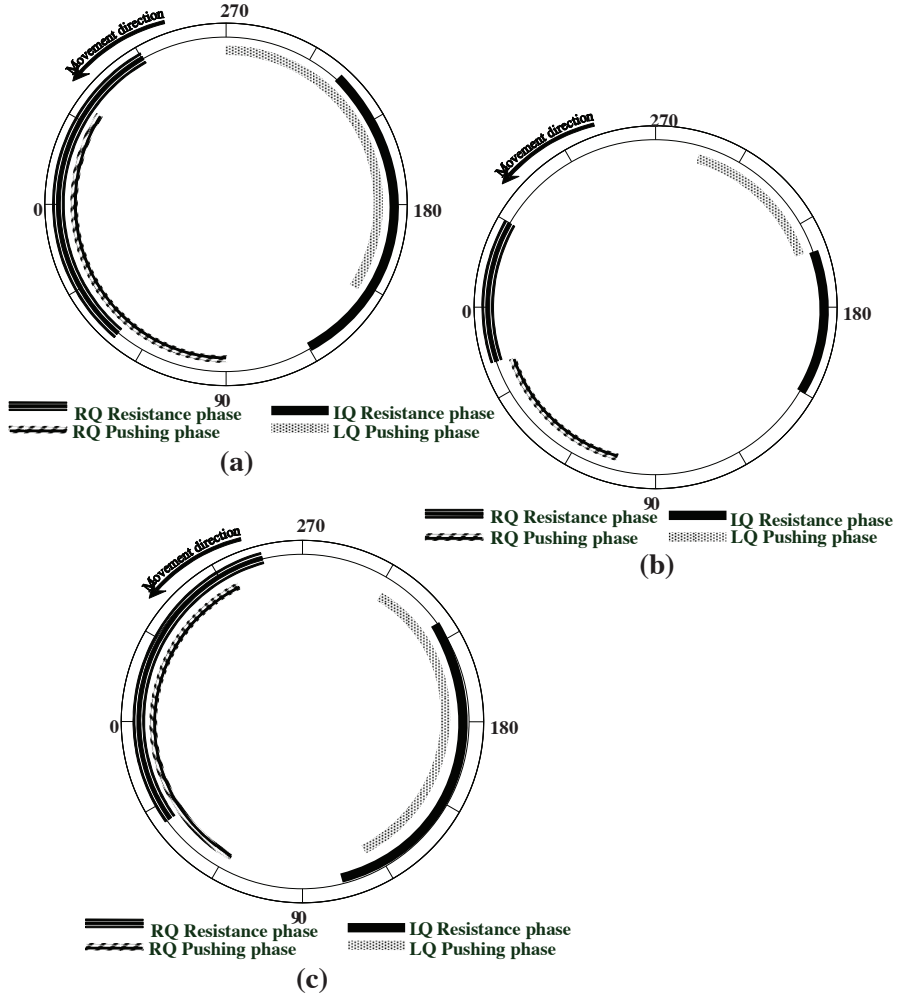
حيث يظهر الشكل 10 أن استخدام نموذج Riener قد حسن أداء النظام ككل و قلل من الخطأ في سرعة دوران مسنن الدراجة، كما يظهر أفضلية هذا النموذج بالنسبة للنموذجين الباقيين.

و مع أن استراتيجية التحكم المستخدمة نفسها قد طبقت على جميع العضلات إلا أنه تمت إعادة ضبط بارامترات المتحكمات (مضخمات المداخل و المخارج) بالإضافة إلى بداية و نهاية فترات كل طور من أطوار دورة التدوير لكل نموذج عضلة على حدا. حيث يبين الشكل 11 نماذج الإثارة لكل من العضلتين الباسطتين لمفصلي الركبة الأيمن و الأيسر، وذلك لكل من النماذج الثلاث. و تظهر الأقواس مجالات زوايا مسنن الدراجة التي تتم عندها إثارة العضلات الباسطة للركبة.

نلاحظ من الشكل 11 أن أقصر مجال إثارة يحدث عند استخدام نموذج Ferrarin للعضلة و أن هذا المجال أقصر بـ 40° عن المجال الناتج عند استخدام نموذج Riener للعضلة أو نموذج VM، كما أن مجال الإثارة الذي يحدث عند نموذج VM مزاح بمقدار 30° عن مجال نموذج Riener.

7 الخاتمة

و على الرغم من أن النتائج أظهرت تشابهاً كبيراً في استجابة نماذج العضلات الثلاثة لقطار من نبضات الإثارة الكهربائية، إلا أنه عند توظيف هذه النماذج في أنظمة التحكم (التحكم بقيادة الدراجة عن طريق تطبيق نبضات الإثارة الكهربائية الوظيفية) تبين أن أداء كل من هذه النماذج يختلف إلى حد ما، فقد أعطى نموذج Riener أداءً أفضل حيث أنه قلل من الخطأ في سرعة دوران مسنن الدراجة و بالتالي أدى إلى قيادة أكثر انسيابية للدراجة و أقرب إلى قيادة الإنسان السليم. كما تبين أن زمن إثارة العضلات يختلف حسب نوع العضلة المطبق، فالنماذج ذات الأساس الفيزيولوجي



الشكل 11 نماذج الإثارة : (a) نموذج Riener (b) نموذج Ferrarin (c) نموذج VM (نموذجي Riener و VM) أظهرت تشابهاً من حيث زمن إثارة العضلة مع وجود بعض الإزاحة، بينما اختلف هذا الزمن عند استخدام النماذج ذات الأساس الرياضي (نموذج Ferrarin).

و من الضروري استخدام البحث الآلي عن بارامترات التحكم التي تعطي الأداء الأفضل من أجل كل نموذج، و بالتالي هذا يجعل المقارنة أكثر وثوقية، كما أنه من الضروري إجراء دراسة عميقة لكل خاصية من نمذجة من خواص العضلة في كل نموذج و البحث عن التشابه و الاختلاف بين النماذج من أجل بناء نموذج دقيق وذكي للعضلات يعتمد على العمل التجريبي و النماذج السابقة في آن واحد.

- [1] A. F. Huxley, "Muscle structure and theories of contraction", Progress in Biophysics and chemistry, vol. 7, pp 257-318, 1957.
- [2] A. V. Hill, "The heat of shortening and the dynamic constants of muscle". Proceedings of the Royal Society of London, B 126 (843), pp. 136-195, 1938.
- [3] Y. C. Fung, "Biomechanics, mechanical properties of living tissues", Springer-verlag, New York, 1981.
- [4] A. J. van Soest, and M. F. Bobbert, "The contribution of muscle properties in the control of explosive movements", Biological Cybernetics, vol. 69, pp. 195–204, 1993,.
- [5] S.L. Delp, F.C. Anderson, A.S. Arnold, P. Loan, A. Habib, C.T. John, E. Guendelman, D.G Thelen, "OpenSim: Open-source software to create and analyze dynamic simulations of movement", IEEE Transactions on Biomedical Engineering, vol. 54, Nov. 2007, pp.1940 – 1950.
- [6] D. J. Pons, C. L. Vaughan and G. G. Jaros, "Cycling device powered by the electrically stimulated muscles of paraplegics". *Medical & Biological Engineering & Computing*, vol. 27, pp. 1-7, 1989.
- [7] K. J. Hunt, B. Stone, et al. "Control strategies for integration of electric motor assist and functional electrical stimulation in paraplegic cycling: utility for exercise testing and mobile cycling". *IEEE Transactions on Neural systems and rehabilitation engineering*, vol. 12, pp.89-101, 2004.
- [8] C.Fornusek, G. M. Davis, P. J Sinclair., and B.Milthorpe, "Development of an isokinetic functional electrical stimulation cycle ergometre". *Neuromodulation*, vol. 7, no.1, pp. 56-64, 2004.
- [9] T. A. Perkins, N. Donaldson, N. A. C. Hatcher, I. D. Swain, and D. E. Wood, "Control of leg-powered paraplegic cycling using stimulation of the lumbo-sacral anterior spinal nerve roots". *IEEE Transactions on Neural systems and rehabilitation engineering*, vol. 10, no. 3, pp. 158-164, 2002.
- [10] K. F. Eichhorn, W. Schubert, and E. David, "Maintenance, training and functional use of denervated muscles". *Journal of Biomedical Engineering*, vol. 6, pp. 205-211, 1984.
- [11] J. J. Chen, N. Y Yu., D. G Huang., B. T. Ann, and G. C. Chang, "Applying fuzzy logic to control cycling movement induced by functional electrical stimulation". *IEEE Transactions on Rehabilitation Engineering*, vol. 5, no. 2, PP. 158-169, 1997.
- [12] J. S. Petrofsky, J. Smith, "Three-wheel cycle ergometer for use by men and women with paralysis". *Medical & Biological Engineering & Computing*, vol. 30, pp. 364-369, 1992.
- [13] D. A. Winter, "Biomechanics and motor control of human movement". Wiley_Interscience New York, 1990.
- [14] M. Ferrarin, and A. Pedotti, "The relationship between electrical stimulus and joint torque: a dynamic model", IEEE Transactions on Rehabilitation Engineering, vol. 8 (3), 2000, pp. 342-352.
- [15] T. Edrich, R. Riener, and J. Quintern, "Analysis of passive elastic joint moments in paraplegics", IEEE Transactions on Biomedical Engineering, vol. (8), 2000, pp. 1058-1064.
- [16] R. Riener, and T. Fuhr, "Patient-driven control of FES-supported standing up: A simulation study", IEEE Transactions on Rehabilitation Engineering, vol. 6, 1998 , pp. 113-124.
- [17] E. J. Cheng, I. E. Brown, and G. E. Loeb, "Virtual muscle: a computational approach to understanding the effects of muscle properties on motor control", Journal of Neuroscience Methods, vol. 101, 2000, pp.117-130.
- [18] J. J. Chen, N. Y Yu., D. G. Huang, B. T. Ann, and G. C. Chang, "Applying fuzzy logic to control cycling movement induced by functional electrical stimulation", *IEEE Trans Rehabil Eng*, 5 (2), 1997, pp.158-69.
- [19] R. Massoud, "Comparative Study of Three Human Muscle Models", UKSim 2010, Cambridge, 2010, UK, 24-26 March.

- [20] R. Massoud, "Intelligent control techniques for spring assisted FES-cycling", PhD Thesis. The University of Sheffield, Sheffield, UK, 2007.

الملخص باللغة الانجليزية

Title: The influence of muscle model on the design of FES-cycling control strategy

Abstract. In this paper, three different muscle models have been investigated. The first model named Ferrarin's muscle model is a transfer function between electrical stimulation and the resultant knee torque. The other two muscles are physiologically based models (Riener's muscle model and Virtual Muscle). Riener's muscle is modelled in this paper by using Matlab/Simulink, while Virtual Muscle model has been built using Virtual Muscle software (Virtual Muscle 4.0.1), which can be downloaded from the internet. A quadriceps is modelled using each of the three models. The three models are tested in terms of their responses to activation and then they are implemented in a control strategy which aims to control the cycling cadence. The performance of the three models during control has been discussed and evaluated, it appears that the type of the muscle model has an influence on the control strategy design.

Keywords: Muscle model, fuzzy logic, Virtual Muscle, FES.

المصطلحات

اسم باحث	Ferrarin
اسم باحث	Riener
عضلة افتراضية	Virtual Muscle
الإثارة الكهربائية الوظيفية	FES (Functional Electrical Stimulation)
برنامج هندسي شهير	Matlab/Simulink
اسم باحث	Hill
علم ميكانيك الرياضة	sport's mechanics
علم حساب الطاقة	ergonomics
علم الميكانيك الحيوي	biomechanics
المنطق العائم	fuzzy logic
متحكم المنطق العائم الذي لا يعتمد على النموذج	model free fuzzy logic controller
برنامج محاكاة	Visual Nastran (VN)
طريقة مربعات الخطأ الصغرى	least square method

قوة العضلة	muscle force
تفعيل العضلة	muscle activation
موضع المفصل	Joint position
سرعة المفصل	Joint velocity
قوة العضلة اللا تقاصرية	Isometric muscle force
العلاقة بين الطول و القوة	Force-velocity relation
العلاقة بين السرعة و القوة	Force-length relation
نوع الألياف	fiber type

هيكلية العنكبوتية بمجموعة أشجار لتبسيط خوارزميات التكرار أخذًا بعين الاعتبار كفاءة وصلات شبكة توزيع المحتوى بالشابكة

فتحي التنزختي

شعبة علم الحاسب ، جامعة نزوى ، سلطنة عمان

البريد الإلكتروني: tenzakhti@yahoo.com

الخلاصة. تهتم هذه الدراسة بالعنكبوتية وهي أهم خدمة تقدمها الشابكة. نحاول من خلال هذه الدراسة أن نقتع بإمكانية هيكلية العنكبوتية بمجموعة أشجار كل شجرة تمثل موقعاً من مواقع العنكبوتية. يكون على رأس هذه الشجرة الموقع، وتتكون فروع هذه الشجرة من الموجهين الذين يقومون بتمرير حزم البيانات عبر المسار الملائم باتجاه المستخدم الهدف. أما أوراق هذه الشجرة فهي الأنظمة التي تتعامل مع الموقع والتي نسميها عملاء الموقع. تمكن هذه الهيكلية من الاستفادة من الكم الكبير من المفاهيم والخوارزميات الجاهزة التي تتعامل مع الشجرات. لبيان أهمية هذه الهيكلية نوضح ضرورتها في بعض آليات التوجيه التي تعمل مع نسخ متعددة للموقع (المفوضين) والتي تطلب شفافية عند تلبية طلب ما يمكن أن يلبيه المفوض أو الموقع الأصلي. نقوم كذلك بصياغة حل رياضي لكيفية تطوير خوارزمية تكرار لتحديد المواقع المثالية للمفوضين في شجرة العنكبوتية أخذًا بعين الاعتبار كفاءة وصلات الشبكة.

الكلمات الحوهرية: التكرار، شجرة العنكبوتية، المفوض، تحسين الأداء، العنكبوتية، كفاءة وصلات الشبكة

1 المقدمة

العنكبوتية هي أهم خدمة توفرها الشابكة. وهي مصدر هام جداً للمعلومات وهي التكنولوجية التي ستضمن التواصل الإلكتروني في الميادين التجارية، التعليمية، الترفيهية و الترفيهية [3,4]. يشككي نظام العنكبوتية حالياً من بعض النقائص نذكر من أهمها بطء الخدمة خاصة هنا في المنطقة العربية وعدم توفرها الدائم حيناً وعدم استعداديتها لتوفير الخدمة حيناً آخر. ولحل هذه المشاكل بطريقة سهلة و ممتعة، لابد من فهم هيكلية هذا النظام وكيفية مرور البيانات فيه من المزود إلى المستخدمين.

في هذه الدراسة نحاول بيان إمكانية هيكلية العنكبوتية بمجموعة أشجار كل شجرة تمثل موقعاً من مواقع العنكبوتية. يكون رأس هذه الشجرة المزود المقيم بالموقع وفروعها الموجهون الذين يقومون بتمرير حزم البيانات عبر المسار الملائم باتجاه المستخدم الهدف والأنظمة التي تتعامل مع هذا المزود والتي نسميها عملاء العنكبوتية. تمكن هذه الهيكلية من الاستفادة من بعض الخوارزميات التي صممت للأشجار للمساعدة على ضمان استجابة جيدة واستعدادية عالية في نظام العنكبوتية. و لبيان أهمية هذه الهيكلية نوضح ضرورتها في بعض آليات التوجيه التي تعمل مع نسخ متعددة للموقع (مفوض) والتي تطلب شفافية عند تلبية طلب ما يمكن أن يلبيه المفوض أو الموقع الأصلي.

وللتدليل على هذه الهيكلية، نستخدم الفكرة التي تقول أنه خلال فترة حددت بست ساعات تقريبا [1,3,5]، تكون طرق الشبكية التي تمر عبرها المعلومات مستقرة بحيث تمر المعلومات التي تنتقل من المزود إلى عملائه بنفس الطرق. وكنتيجة لذلك، يكون خلال هذا الزمن كل مزود على رأس شجرة أوراقها عملاؤه وفروعها الموجهون. تمر المعلومات من المزود عبر الموجهين المقيمين بفروع هذه الشجرة إلى العملاء. وبما أن العنكبوتية هي مجموعة من موزعي المواقع، فيمكن اعتبار نظام العنكبوتية متكونا من مجموعة من الشجرات على رأس كل واحدة منها مزود للموقع.

تحتوي هذه الدراسة على 7 فقرات. الفقرة الثانية تذكر بعض الدراسات المرتبطة بعملنا هذا و تبرز وجوه تطابقها واختلافها مع ما نقوله هنا. الفقرة الثالثة تتحدث عن مكونات نظام العنكبوتية و تركيبته. الفقرة الرابعة تحاول الإقناع بإمكانية هيكلية نظام العنكبوتية بمجموعة أشجار على رأس كل شجرة مزود الموقع. تمر المعلومات من المزود إلى عملائه عبر فروع هذه الشجرة. الفقرة الخامسة تبين أهمية هذه الهيكلية في بعض آليات التوجيه الشفافة التي تعمل مع نسخ متعددة للموقع (المفوضين). الفقرة السادسة تعرض خوارزميات تكرر لتحديد المواقع المثالية للمفوض في شجرة العنكبوتية اعتمادا على الهيكلية التي نقترحها. الفقرة السابعة تختتم الدراسة.

مرفقة بهذه الدراسة جدول لأهم المصطلحات المستخدمة ومرادفها باللغة الإنجليزية. وقد تعمدنا في هذه الترجمة أن نستخدم مفهوم المصطلح ووظيفته بدلا عن مرادفه الذي نرى أنه قد لا يؤدي المعنى الصحيح.

2 أعمال مرتبطة بهذا العمل

أول دراسة اعتمدت فكرة الشجرة كهيكلي لنظام العنكبوتية هي الدراسة المذكورة ب[8]. حيث افترض الباحثون في هذه الدراسة نظام العنكبوتية كمجموعة شجرات على رأس كل واحد منها مزود و تمر المعلومات من المزود إلى العملاء عبر فروع هذه الشجرة. استخدم الباحثون هذه الهيكلية لإيجاد خوارزميات تساعد على وضع نسخة من هذا المزود في بعض عقد هذه الشجرة لتحسين استجابته لطلبات عملائه. ولقد قام الباحثون في هذه الدراسة بتصميم خوارزمية تحدد أفضل عقد الشجرة لاحتواء نسخ من موقع العنكبوتية لضمان أفضل استجابة لطلبات العملاء. وقد اعتمدت هذه الخوارزمية على البرمجة الحركية وتحتاج إلى وقت تنفيذ مرتبط بعدد النسخ وعدد عقد الشجرة التي تسمح باحتواء نسخة الموقع. ضعف هذه الدراسة يتمثل في افتراض الهيكلية الشجرية للواب دون التدليل عليها.

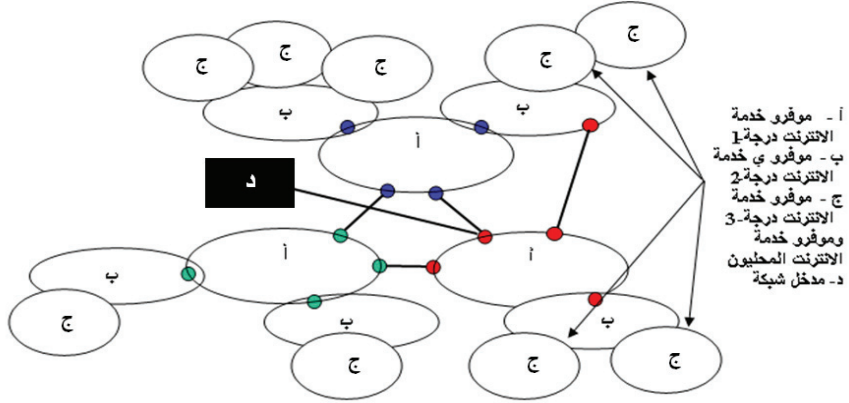
في دراسة ثانية [7] افترض الباحثون أن المسالك التي تمر بها المعلومات من الموقع إلى المستخدم تشكل شجرة رأسها موقع العنكبوتية وقاموا بتصميم خوارزمية لإيجاد مجموعة الإقامة الأقل تكلفة التي تضمن استجابة في أقصر مدة زمنية لطلبات المستخدمين معتبرين أن العنكبوتية يسمح بالقراءة دون الكتابة و أخذين بعين الاعتبار طاقة المعالجة للمزود المستخدم بالموقع. ضعف هذه الدراسة يتمثل في افتراض الهيكلية الشجرية للواب دون التدليل عليها.

3 مكونات نظام العنكبوتية و تركيبته

تتكون العنكبوتية من نظام لمجموعة من خدام الشبكة تحوي ملفات منسقة بطريقة خاصة تعتمد لغة HTML تسمى صفحات العنكبوتية. تتميز لغة HTML بقدرتها على ربط صفحات العنكبوتية بحيث يمكن الانتقال من صفحة إلى أخرى بمجرد النقر على كائن في الصفحة وقد اصطلح على تسميت هذه الارتباطات بالارتباطات التشعبية.

تجدر الملاحظة هنا إلى أن ليس كل مزودي الشبكة جزء من العنكبوتية مما يعني أن العنكبوتية والشبكة مختلفان عن بعضيهما. فالشبكة هي شبكة الشبكات وهي تمثل بنية تحتية تمر من خلالها المعلومات إلى المستخدم وتتكون من أجهزة حاسوب، أجهزة شبكات تتكون من موجهين وأسلاك بالإضافة إلى برمجيات و مراسم تضمن انتقال المعلومة بدقة، سرية، و دون أخطاء.

لفهم تركيبة العنكبوتية لابد أولاً من الحديث عن الهيكل العامة للشبكة حيث تجدر الملاحظة أن الشبكة بصفة عامة تتكون من طرفيات تشمل الحواسيب الشخصية، مزودي العنكبوتية و هي التي تحوي مواقع العنكبوتية، مزودي البريد الإلكتروني و هي التي تقدم خدمة البريد الإلكتروني للمستخدمين وغيرها. ترتبط هذه الطرفيات بالشبكة عبر مداخل الشبكة التي يمكن أن تكون شبكات محلية سلكية أو لاسلكية، أو موفري خدمة الشبكة التي يصلها المستخدم عبر شبكة التلغون العادية أو عبر خطوط تلفونية خاصة [6]. مداخل الشبكة الواقعة على طرف الشبكة ترتبط بالشبكة عبر مجموعة من موفري خدمة الشبكة موزعة حسب سلم هرمي ثلاثي كما في الشكل 1.



شكل 1. هيكلية شبكة الشبكة

في أعلى السلم الهرمي نجد مجموعة صغيرة من موفري خدمة الشبكة تعرف بموفري خدمة الشبكة درجة أولى. تتميز هذه المجموعة بسرعة عالية في نقل البيانات حيث ترتبط هذه المجموعة مع شبكات ثانية عبر أسلاك و موجهين يصل سرعة نقل البيانات فيها إلى 10 مليون بايت في الثانية. يرتبط موفري خدمة الشبكة درجة أولى ببعضها عبر مداخل تربية خاصة و كثيرا ما يرتبط بعضها البعض عبر مداخل الشبكة. يسمح موفري خدمة الشبكة درجة أولى بتغطية عالمية للمستخدمين ويعتبر العمود

الفقري للشابكة. يرتبط موفرو خدمة الشابكة درجة أولى بموفري خدمة الشابكة درجة ثانية كما في الشكل واحد.

موفري خدمة الشابكة درجة ثانية يسمح بتغطية إقليمية أو على مستوى بلد معين وعادة ما يكون مرتبطا بعدد قليل من موفري خدمة الشابكة درجة أولى حتى يتمكن من التواصل مع جزء كبير من المستخدمين الدوليين للشابكة.

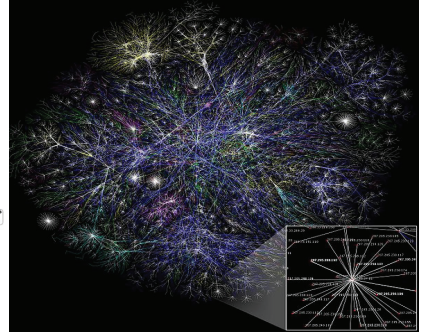
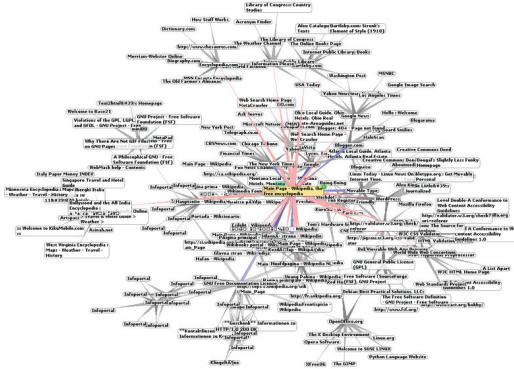
في أسفل السلم الهرمي الثلاثي يقيم موفري خدمة الشابكة درجة ثالثة وهي نفس الدرجة التي يقيم بها الموفرين المحليين لخدمة الشابكة.

كخلاصة نود الإشارة إلى أن هيكل الشابكة يتكون من عشرات موفري خدمة الشابكة درجة أولى ودرجة ثانية وآلاف موفري خدمة الشابكة درجة ثالثة التي تشمل الموفرين المحليين لخدمة الشابكة. تختلف درجة التغطية التي يوفرها الموفرين حسب النوعية حيث البعض يغطي عدة قارات و محيطات بينما يكتفي البعض الآخر بتغطية مساحات ضيقة من العالم. مع الإشارة إلى أن موفرو الدرجة الثالثة و الموفرون المحليون يحتاجون إلى موفري الدرجات العليا من الهرم الثلاثي ليتمكنوا من الاتصال بباقي الشابكة.

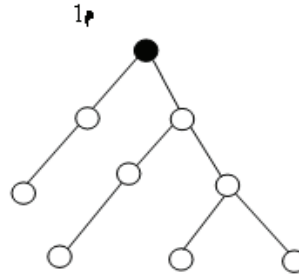
4 هيكل نظام العنكبوتية بمجموعة شجرات

تعتمد نظرية الهيكل الشجري لنظام العنكبوتية الذي نقترحه على حقيقة تقول أنه خلال فترة حددت بست ساعات تقريبا [10] تكون مسارات الشابكة التي تمر عبرها المعلومات مستقرة بحيث تمر المعلومات التي تنتقل من موقع ما بالشبكة العالمية إلى عملائه بنفس المسارات. وكنتيجه لهذا للاستقرار الضرفي، تكون المسارات التي تأخذها المعلومات لتنتقل من الموقع إلى عملائه شجرة أقصر المسارات يكون جذعها الموقع الموفر للمعلومات (انظر الصورة 1). وللتدليل على صحة الافتراض القائل باستقرار مسارات العنكبوتية، قامت بعض الدراسات [9] بتحليل مسارات العنكبوتية واستخلصت إلى أن 80% من المسارات تتغير بنسبة أقل من مرة يوميا. ففي هذه الدراسة قام الباحث كريسشن بتتبع مسارات معلومات انتقلت من مجموعة 15,533 مزود لشبكة المعلومات التابعة لشركة 'Bell Labs' إلى عملائها و توصل إلى أن 93% من المسارات كانت مستقرة خلال التجربة. ولهذا فإن افتراضية استقرار مسارات العنكبوتية هي افتراضية واقعية مثبتة ببعض الدراسات الجادة.

باستخدام نظرية الاستقرار هذه، نلاحظ أن كاننا ما موجود بموقع م و طلبه العميل ع، يمر عند تنقله من م إلى ع عبر مسار ثابت م ← 1 ← 2 ← 3 ← ... ع نسميه بالمسار المختار و نستخدم الرموز مخ(م،ر) للتدليل عليه. يتكون المسار المختار من مجموعة من الموجهين 1ر، 2ر، 3ر، ... والعقد التي تحويها. وكنتيجه لذلك تكون المسارات المتجهة من الموقع م إلى عملائه شجرة جذعها الموقع م، وأوراقها العملاء و عقدها الداخلية متكونة من الموجهين أو من عملاء في مسار عملاء طرفيين. وبما أن العنكبوتية هو مجموعة من المواقع فإننا نستنتج بأنه يمكن هيكلية العنكبوتية بمجموعة شجرات. لكل موقع م شجرة نسميها ش (شكل 2) تمر عبرها الكائنات من الموقع إلى عملائه. وتكون هذه الشجرة اتحادا للمسارات المختارة.



الصورة 1: المسارات المتبعة لنقل البيانات في الواب تكون مجموعة من الشجرات



شكل 2. شجرة ش₁ للخاص م₁

5 تأثير الهيكلية على آليات التوجيه

نظرا لكون موجهي الشبكة يكونون نسبة كبيرة جدا من مكونات الشبكة المادية، فإن الحديث عن آلية توجيه الطلبات في المسار الصحيح هام و ضروري عند الحديث عن أي تغيير أو تحسين نود أجراه على العنكبوتية أو أي خدمة من خدمات الشبكة. في هذه الفقرة، سنتحدث عن فائدة هيكلية العنكبوتية كمجموعة من الشجرات من حيث تسهيل آليات توجيه طلب عميل ما إلى نسخة من نسخ الموقع مفترضين أن الموقع الذي نهتم بدراسته منسوخ في مجموعة من عقد الشجرة التي يرأسها المزود المقيم بهذا الموقع. أهمية نسخ الموقع في أماكن عديدة يسرع خدمة الشبكة ويضمن أداء مقبولا واستعدادية جيدة.

يقع نسخ المزود في ما اصطلح على تعرفته بالمفوض و هو نظام (معدات و برمجيات) يتوسط المزود الأصلي والعميل. يشتغل المفوض كوكيل عن المزود الأصلي حيث يقوم بتوفير كل الخدمات التي يوفرها

المزود الأصلي من بيانات و معلومات و صور و خدمات تفاعلية و غيرها. يقع هذا المفوض بين المزود و العميل و لا يراه العميل و نسميه "مفوض مسيار شفاف".

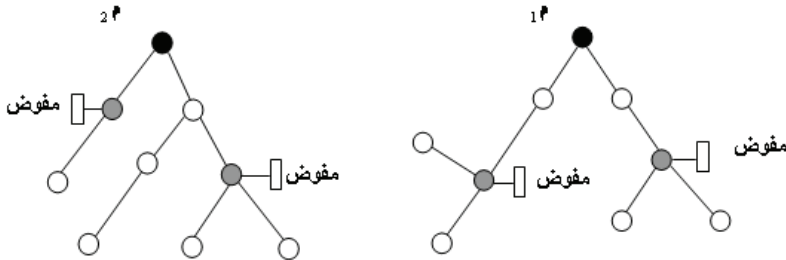
نبدأ بالحديث عن نوعين من الطلبات التي يصدرها عميل ما في اتجاه موقع من مواقع العنكبوتية للحصول على معلومة. النوع الأول يعرف بالطلبات المعروفة لدى الموجه و النوع الثاني يعرف بالطلبات العادية [11]. في الطلبات المعروفة لدى الموجه يحتاج الموجه الى أن يستقبل الطلبات لمعرفة محتواها. أما الطلبات العادية فإنها تمر عبر الموجه دون أن يستقبلها لمعرفة محتواها.

عند عمل مفوض لموقع ما، نرغب في أن تكون عملية توجيه طلب يصدرها عميل للحصول على معلومة من المفوض شفافة بحيث لا يحتاج العميل إلى ذكر عنوان المزود الذي يحوي النسخة ولكن يكفي باستخدام عنوان الموقع الأصلي. تعرف هذه الآلية بشفافية الوصول إلى الكائن. هنا تكمن أهمية هيكله العنكبوتية بمجموعة أشجار حيث يمكن تنفيذ آلية شفافية الوصول إلى كائن عن طريق إحدى النماذج التالية:

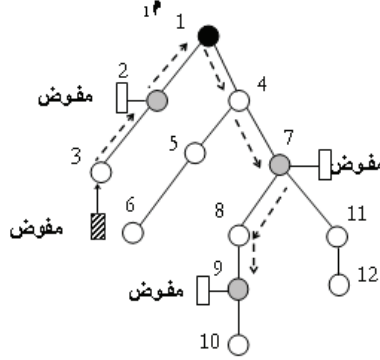
مفوض مسيار : كل طلب يصدره عميل العنكبوتية هو من النوع المعروف لدى الموجه. يقوم موجه مقيم بأحدى العقد بالمسار بين العميل و الموقع الأصلي باستقبال هذا الطلب. إذا كان الكائن المطلوب موجودا بالمفوض المقيم بالعقدة، يقوم المفوض بتلبية طلب العميل وإرسال الكائن له. وفي حال عدم وجود الكائن بالعقدة ، يقوم الموجه بإرسال الطلب إلى الموجه الموالي.

مفوضات هرمية : تكون المفوضات في هذا النموذج مرتبة ترتيبا هرميا بحيث يكون الموقع الأصلي على قمة الهرم و كل مفوض يعرف موقع المفوض الأب من الشجرة. في هذا النموذج كذلك يظهر جليا أهمية هيكله العنكبوتية بمجموعة شجرات إذ عند صدور طلب من عميل إلى كائن بموقع ما، نوجه الطلب عبر المسار المختار بالشجرة إلى أقرب موجه. إذا كانت العقدة التي يقيم بها الموجه قادرة على تلبية طلب العميل فإن مزود العقدة يرسل الإجابة إلى العميل. أما إذا لم تقدر العقدة على تلبية الطلب فإنه يتم إرسال الطلب إلى المفوض الأب. وهكذا إلى أن يلبي الطلب عن طريق أحد المفوضين أو يصل الطلب إلى الموقع الأصلي.

مفوضات هجينة: في هذا النموذج تكون الطلبات معروفة لدى الموجه و تكون المفوضات مرتبة ترتيبا هرميا. في هذا النموذج تكمن أهمية الهيكل الشجرية في أن الطلب لا يحتاج إلى معرفة عنوان المفوض وإنما يقوم الطلب بإتباع المسار المختار بالشجرة في اتجاهه إلى الموقع الأصلي إلى أن يصل إلى موجه. عندها يقوم الموجه باستقبال الطلب. إذا كانت العقدة التي تؤوي الموجه بها مفوض قادر على تلبية الطلب، فإن الطلب يلبي و إلا فإن المفوض يقوم بإرسال الطلب إلى المفوض الأب. وهكذا تستمر العملية إلى أن يلبي الطلب إما بالوصول إلى مفوض به الكائن المطلوب أو الوصول إلى الموقع الأصلي. (الشكل 3 و الشكل 4)



الشكل 3. شجرتان لموقعين مع بعض النسخ المقيمة بعقد الشجرة.



الشكل 4. استجاب طلب للكتابة في الفسخ الهجينة

6 خوارزميات تكرار لتحديد المواقع المثالية للمفوض في شجرة العنكبوتية

سنحدث الآن عن أهمية هذه الهيكلية في تحديد مثالي و بسيط لمواقع المفوضين بشجرة العنكبوتية في عملية تكرار. وسنبدا الحديث بتعريف تقنية التكرار.

التكرار: هي تقنية حفظ نسخ لموقع عنكبوتية في مفوض. يكون هذا المفوض في مكان من الشبكة بحيث يسهل للعميل الوصول إليه.

يستخدم التكرار عادة لحل مشكلة المواقع التي عليها طلب كبير. وبصفة عامة يفيد التكرار في تحسين استجابة المواقع للمعاملات واستفسارات العملاء حيث يقوم المفوض القريب من العميل بالرد عن الاستفسارات عوضا عن المزود الأصلي. كما يفيد التكرار في تحمل العيوب والاستعدادية. ولذا فإن تحديد مواقع المفوضين في الشجرة هام جدا في تحسين أداء العنكبوتية.

لتعريف رسمي لمسألة تحديد مواقع المفوض بشجرة أخذنا بعين الاعتبار كفاءة خطوط الاتصال، ندرج هذه التعريفات:

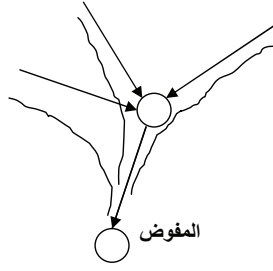
- مجموعة الإقامة R في شجرة T_s جذعها العقدة s هي مجموعة عقد تحوي نسخا للمزود الأصلي الموجود ب s . مجموعة الإقامة الأصغر هي مجموعة الإقامة الأقل تكلفة ضمن مجموعات الإقامة الممكنة بالشجرة.
- المسافة بين عقدتين u و v بشجرة هي $d(u,v)$ وهي طول المسار بين u و v . يعرف المسار بين u و v $\pi(u,v)$. رياضيا، نقول أن:

$$(1) \quad d(u,v) = \sum_{(x,y) \in \pi(u,v)} d(x,y)$$

- أول مفوض يعترض طريق طلب للعميل v من المزود s في الشجرة T_s هو المفوض $p(v,s)$ ، يسمى هذا المفوض بالمفوض الأفضل. يمكن أن يكون هذا المفوض العميل v نفسه إذا كان العميل v مفوض للمزود s أو المزود نفسه إذا خلى المسار بين $\pi(v,s)$ من مفوضين.
 - f هو تردد المعاملات من العميل v للمزود s خلال الفترة الزمنية الفاصلة بين استدعائين لبرنامج تحديد المفوضين.
 - $\beta(v)$ هو العيب الذي تفرضه العقدة v على الوصل الأبوي للمفوض $p(v,s)$.
 - k_u هي كفاءة الوصلة u .
 - P_u هي مجموعة المفوضين ضمن الشجرة T_s التي مفوضها u . $P_u = \{x \in T_s, u = p(x,s)\}$. مجموع المسافات للوصول إلى المجموعة P التي تضم جميع المفوضين بالشجرة المرتبطة بمجموعة الاقامة
- $$(2) \quad d(P) = \sum_{v \in T_s} d(v, p(v,s))$$
- مجموع تكلفة الوصول الى الكوائن بالمفوض المثالي يمكن تمثيلها بالمعادلة:
- $$(3) \quad C(T_s, P) = \sum_{v \in T_s} f(v) d(v, p(v,s))$$

1.6 القيد على قدرة الوصلات

- كل عميل v يتعامل مع المزود s مفوضه الأفضل $u = p(v,s)$ يمارس عيأ على الوصل الأب للمفوض. نسمي هذا الوصل الأب w . إذا اعتبرنا أن كفاءة الوصلات بالشجرة يمكن تخزينها في المتجه $\vec{K}$ ، وأن k_w هي كفاءة الوصلة الأب w فان القيد على كفاءة وصلات الشبكة يحتم على مجموعة العقد التي مفوضها المفضل u ألا تقرض عيأ أكبر من الكفاءة k_w للوصلة الأب للمفوض بمعنى أنه لا بد من أن
- $$(4) \quad \sum_{x \in P_u} \beta_x \leq K_u$$



الشكل 5. القيد على قدرة وصل المفوض

2.6 صياغة رياضية للمسألة

كنتيجة لذلك فإننا إذا اعتبرنا أن المجموعة P_u المتكونة من كل العملاء مفوضهم المفضل u ، فإن

$$\sum_{x \in P_u} \beta_x \leq K_u \quad \text{يجب أن تكون صحيحة.}$$

الآن إذا اعتبرنا أن عدد المفوضين هو n وهو أكبر من واحد، فإن المسألة تصبح مسألة إيجاد مجموعة الإقامة R التي تقلل تكلفة الوصول إلى مفوضي الشبكة في شجرة العنكبوتية أخذاً بعين الاعتبار قدرة الوصلات الرابطة لعقد الشجرة. ومن هنا يمكن اختزال المشكلة في إيجاد التكلفة الأقل للوصول إلى جميع مفوضي الشبكة بالشجرة. رياضياً، تصبح المسألة تتمثل بالمعادلة التالية:

$$(5) \quad C(T_s, n, \kappa) = \min_{P \subseteq T_s, |R|=n} \{C(T_s, R, \kappa)\} \quad \text{أوجد التكلفة الأقل}$$

$$(6) \quad \sum_{p_u} \beta(x) \leq \kappa_u, \forall u \in P \quad \text{أخذاً بعين الاعتبار المعادلة:}$$

الدراسة [4] توضح كيف يمكن حل هذه المسألة ونقترح خوارزمية لبرمجتها. يقع تنفيذ هذه الخوارزمية بالموقع الأصلي الذي يفترض أن يكون عارفاً بنفاصيل الشبكة التي يرأسها. يقوم الموقع الأصلي بترتيب المفوضين الجدد كل ست ساعات واعطاء الأوامر بنقل بعضهم من عقدة إلى أخرى.

7 الخاتمة

حاولنا من خلال هذه الدراسة المختصرة إظهار إمكانية هيكلية العنكبوتية كمجموعة من الشجرات. كل شجرة تخص موقعا معينا و تتكون من المسارات المختارة التي تمر عبرها الطلبات من العميل إلى الموقع و الإجابات من الموقع إلى العميل. ثم ذكرنا أهمية هذه الهيكلية في إنشاء خوارزميات للمساعدة على ضمان استجابة جيدة واستعدادية عالية في نظام العنكبوتية. و لبيان أهمية هذه الهيكلية تحدثنا عن ضرورتها في بعض آليات التوجيه التي تعمل مع نسخ متعددة للموقع والتي تطلب شفافية عند تلبية طلب ما يمكن أن تلبية النسخ أو الموقع الأصلي. وأخيرا تحدثنا عن صياغة رياضية لتحديد المواقع المثالية لمفوضي مزود ما للوصول الى افضل استجابة أخضا بعين الاعتبار كفاءة وصلات الشبكة.

المصادر والمراجع

- [1] فتحى التنتختي، هل يمكن هيكلية العنكبوتية بمجموعة شجرات؟، المؤتمر الدولي الرابع لممارسات علوم الحاسب باللغة العربية، 1-4 أبريل، الدوحة، قطر. 2008
- [2] Anne Benoit, Veronika Rehn-Sonigo, Yves Robert, "Replica Placement and Access Policies in Tree Networks," IEEE Transactions on Parallel and Distributed Systems, vol. 19, no. 12, pp. 1614-1627, Feb. 2008.
- [3] Benoit, V. Rehn, and Y. Robert, "Impact of QoS on Replica Placement in Tree Networks," Research Report 2006- 48, LIP, ENS Lyon, Dec. 2006.
- [4] C.-M. Wang, C.-C. Hsu, P. Liu, H.-M. Chen, and J.-J. Wu, "Optimizing Server Placement in Hierarchical Grid Environments," Proc. First Int'l Conf. Grid and Pervasive Computing (GPC '07), pp. 1-11, 2007.
- [5] Fathi Tenzakhti, *Optimal Placement of Web Proxies in the Internet with Link capacity constraints*, Journal of Digital Information Management, Vol.4. No. 4, 2006.
- [6] L. Qiu, V.N. Padmanabhan, and G.M. Voelker, "On the Placement of Web Server Replicas," Proc. INFOCOM '01, pp.1587-1596, 2001.[
- [7] Li, M.J. Golin, F. Italiano, X. Deng, K. Sohaby,., *On the Optimal Placement of Web Proxies in the Internet*, Proc. IEEE INFOCOM , pp. 1282-1290, 1999.
- [8] James F. Kurose, Keith W. Ross, Computer network, a top-down approach featuring the Internet, third edition., Pearson, Addison Wesley, 2005.
- [9] P. Krishnan, D. Raz, and Y. Shavitt. (2000). The cache Location Problem, IEEE/ACM Transactions on Networking, vol. 8, no. 5, pp. 568-582.
- [10] V. Paxson, *End-to-End Routing Behavior in the Internet*, IEEE/ACM Transactions Networking, vol. 5, no. 5, pp. 601-615, 1997.
- [11] Xu, Bo Li, and D.L. Lee, *Placement Problems for Transparent Data Replication Proxy Services*, IEEE Journal on Selected Areas in Communications, vol. 20, no. 7, pp. 1383-1398, 2002.

Modeling the Web as a Forest of Trees for Simpler Replica Placement Algorithms in a Content Delivery Network Considering Link Capacity

Fathi Tenzakhti

Department of Computer Science, Univesity of Nizwa, Oman

tenzakhti@yahoo.com

Abstract. This study tries to demonstrate that the World Wide Web (the Web for short) could be modeled as a forest of trees. Each Web site has its own tree for which it is the root. The study then tries to show that such web topology allows for the design of simple replica placement algorithms in a content delivery network CDN that optimizes the placement of web proxies in the Internet taking into consideration the network link capacity.

Keywords: Replication, proxy, forest of trees, link capacity, performance

عربي	انجليزي
الاستعدادية	Availability
شبكة توزيع محتوى	Content Distribution Network
الطرفيات	End systems
تكرار المسبار	En-route Replication
تكرار هرمي	Hierarchical Replication
تكرار هجني	Hybrid Replication
موفر خدمة الشبكة	Internet Service Provider (ISP)
الموفرون المحليون لخدمة الشبكة	Local Internet Service Provider
مداخل الشبكة	Network Access Points
علاقة تربية	Peer-to-peer
المسار المختار	Preference Path
نقاط تربية خاصة	Private peering points
الطلبات	Requests
الموجه	Router
الطلبات المعروفة لدى الموجه	Router Aware Requests
آليات التوجيه	Routing mechanisms
شجرة أقصر المسارات	Shortest Path Tree
موفرو خدمة الشبكة درجة أولى	Tier-1 ISPs
موفرو خدمة الشبكة درجة ثانية	Tier-2 ISPs
موفرو خدمة الشبكة درجة ثالثة	Tier-3 ISPs
سلم هرمي ثلاثي	Tiered hierarchy
شفافية الوصول إلى كائن	Transparent object access
الشبكة العالمية للمعلومات (العنكبوتية)	Web
مستخدم الشبكة العالمية للمعلومات (مستخدم العنكبوتية)	Web Client
مزود الشبكة العالمية للمعلومات (مزود العنكبوتية)	Web Server
موقع بالشبكة العالمية للمعلومات (موقع العنكبوتية)	Web Site

Link Capacity	كفاءة شبكة الصلات
Proxy	المفوض
Replication	تكرار
Client	المستخدم
Server	المزود
Transparent Proxy Server	مفوض مسيار شفاف
Fault Tolerance	تحمل العيوب
Availability	الاستعدادية
Parent Link	خط الاتصال الأب

مشكل الاختيار الجماعي بالاستناد إلى صانعي قرار

سوسن كريشان^{1,2} ، سهام بن جويده²

¹ العنوان البريدي: جامعة جندوبة ، كلية العلوم القانونية و الاقتصادية و التصرف ،

ص.ب. 8189 ، الرمز البريدي 8100 ، جندوبة ، تونس

² العنوان البريدي: جامعة تونس ، المعهد العالي للتصرف ، ص.ب. 41 ، الرمز البريدي 2000 ،

باردو ، تونس

البريد الالكتروني: saoussen.krichen@isg.rnu.tn

benjouida_sihem@hotmail.fr

الخلاصة في هذا العمل نقترح تعميما لمشكل التوقف المثالي يتميز بوجود صانعي قرار يطمحان للاتفاق حول اختيار عرض واحد من جملة عروض يتم تلقيها في شكل مجموعات. في كل مرحلة من مراحل العملية الديناميكية يتم تفحص مجموعة عروض و إذا افترضنا انه قد تم رفضها فلا يمكن أن يعاد النظر فيها في المراحل القادمة. وفي هذا السياق، استدعاء العروض يكون مسموحا فقط داخل المجموعة الواحدة. لحل هذا المشكل نقترح إتباع منهج البرمجة الديناميكية وكتابة معادلات رياضية لمختلف أدلة صانعي القرار و نوضح دراستنا من خلال التحقيق التجريبي الذي يستند على الدالة المتوقعة.

الكلمات الجوهرية: الإعلامية و شبكات الاتصال، التوقف المثالي ، البرمجة الديناميكية، نظرية الألعاب، نظرية القرارات الجماعية

1 مقدمة

التوقف المثالي يعرف بأنه العلم الذي يبحث في تحديد أفضل عرض من جملة عروض يتم تلقيها بصفة مسترسلة زمنيا. و لهذا المشكل تطبيقات عديدة لعل أهمها الإعلامية و شبكات الاتصال و الاستثمار في المشاريع و اختيار المرشحين و مشكل شراء الأصول في البورصات.

إن منهج التوقف المثالي يعتبر من المناهج الحديثة فقد تم التطرق إليه من قبل "والد" منذ عام 1945 وذلك في إطار نظرية التحليل المسترسل ثم ذكر مع " غاردنر " (1960) و " لندي " (1961) و " دانكان " (1963).

وبالأساس فالمشكل يفترض وجود صانع قرار وحيد يقوم بتقييم مجموعة عروض عرضا تلو الآخر دون أي معلومات مسبقة ، و ذلك من خلال مقارنة مرتبة العرض الحالي بمجموعة العروض التي فحصها. مع العلم أنه إذا كان العرض i الحالي في المرتبة k وذلك مقارنة بعدد i من العروض التي تم تقييمها سابقا فان هذه المرتبة يمكن لها إما تتقدم في مراحل لاحقة من العملية الديناميكية أو أن تحافظ على ذاتها.

و للتوضيح نقدم هذا المثال الذي يفترض أن صانع القرار يستقبل عددا من العروض الذي يصل عددها إلى $n=5$ والهدف من هذا المثال هو شرح لفظين أساسيين لتحليل أسس التوقف المثالي و هما:

الرتبة المطلقة وهي رتبة العرض الحالي مقارنة بالعدد الجملي للعروض.

الرتبة النسبية وهي رتبة العرض الحالي مقارنة بعدد العروض التي تم تلقيها وصولا إلى المرحلة i .

علما أن العروض تأتي متتالية الواحد تلو الآخر فصانع القرار مدعو إلى اختيار العرض الأفضل (ذو الرتبة المطلقة 1) .

نقدم في ما يلي الرتب المطلقة التي يمكن أن تقترح إلى صانع القرار:

O_1	O_2	O_3	O_4	O_5
3	5	2	1	4

الشكل 1. إمكانية تسلسل الرتب المطلقة ($n=5$)

الشكل 2 يبين إمكانيات يمكن أن تتوافق معها الرتب النسبية للعروض الخمسة وصولا إلى ترتيبها المطلق.

$i = 1$	1				
$i = 2$	1	2			
$i = 3$	2	3	1		
$i = 4$	3	4	2	1	
$i = 5$	3	5	2	1	4

الشكل 2. إمكانيات لتسلسل الرتب النسبية $i=1...5$

وبما أن الهدف الرئيسي لصانع القرار هو تحديد العرض الذي يطمح أن يكون الأفضل بين مجموع العروض فإن ذلك يكون من خلال رفض (i^*-1) العروض الأولى و قبول العرض i^* مع رتبة نسبية تكون الأولى.

من ذلك فإن الدالة العامة لصانع القرار يتم تعريفها من خلال الرتب المطلقة k_a و قد أمكن لها أن تأخذ الشكل التالي إتباعا إلى ليندلى 1961 :

$$v^p(r_p) = \begin{cases} 1 & \text{if } r_p = 1 \\ 0 & \text{otherwise} \end{cases}$$

وفي سنة 1994 تطرق تشان إلى تعميم مشكل التوقف المثالي يقتضي وجود صانع قرار وحيد يهتم بمراقبة سلسلة من العروض تصل في شكل مجموعات. ومن ثم، في كل فترة زمنية يتم تقييم مجموعة من العروض ولصانع القرار الاختيار بين قبول العرض الأفضل داخل المجموعة أو رفضه كاختيار بديل. في حالة الرفض لمجموعة العروض فإنه لا يمكن إعادة النظر فيها في مرحلة لاحقة.

نقترح في هذا العمل تعميما للوضعية تعتبر وجود صانعي قرار يتشاركان لقبول عرض واحد من بين سلسلة عروض تصل في مجموعات الشيء الذي يؤدي إلى نشأة بعض الصراعات :

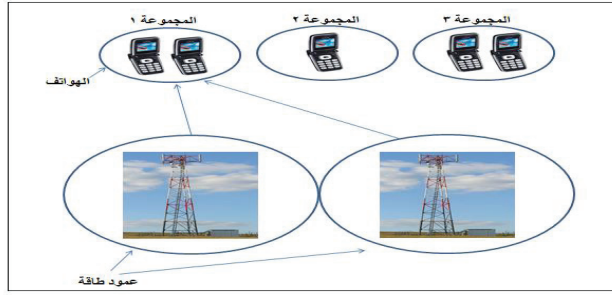
داخل المجموعة الواحدة صانعي القرار مطالبان بالاتفاق حول عرض مشترك حتى وإن كانت الاختيارات الفردية مختلفة.

بين المجموعات العرض المتفق عليه من قبل صانعي القرار يتم إما قبوله أو رفضه وبذلك فإن القرارات الفردية يمكن أن تختلف فيما بينها مما يقتضي تحديد قاعدة توقف تهدف إلى معالجة هذا الاختلاف في القرارات.

في كل مرحلة، و بعد رصد مجموعة من العروض، جملة من القرارات يجب أن تؤخذ إما بقبول العرض الأفضل بالنسبة للمجموعة الجاري تقييمها ومن ثم إيقاف عملية البحث أو برفضها و مواصلة استقبال العروض. و بذلك فإن القرارات الفردية يمكن أن تكون مختلفة و لحل هذا الصراع نقترح في هذا العمل

قاعدة توقف: في سياق المجموعة و خطوة أولى كل صانع قرار مطالب بتحديد إستراتيجيته الفردية مما ينتج عنه جملة من القرارات الفردية. إذا تطابقت هذه القرارات الفردية مع قبول العرض المتفق حوله من قبل صانعي القرار فسيتم إيقاف عملية البحث و كوضع مقابل سيتواصل البحث في حال كان القرار الجماعي يقضي برفض العرض. إذا كان هناك تباين (قبول و رفض) فسيكون حله من خلال تحديد حد أقصى لا ينبغي، للفرق بين قيمة الدالة المتوقعة عند قبول العرض الجاري تقييمه وبين قيمة الدالة المتوقعة في حال رفضه، أن يتجاوزه.

لمزيد التوضيح لنفترض المثال التالي:



الشكل 3. مثال في مجال شبكات الاتصال

* العروض: الهواتف (الاتصالات الواردة)

* صانعي القرار: عمودي الطاقة

* الاختيار: تتوافد مجموعات الاتصالات الواحدة تلو الأخرى. بالنسبة إلى كل مجموعة إذا تم الاتفاق حول تلقي أفضل إرسال من قبل أحد العمودين يكون بذلك قبول العرض أما في حالة إلغاء الاتصال يتم الانتقال إلى المجموعة الموالية

وسيكون تسلسل الفقرات في هذا العمل كالتالي: الفقرة 2 تعرض مشكل التوقف المثالي في حالة توافد العروض في شكل مجموعات و بالإستناد إلى صانعي قرار، الفقرة 3 تقدم الصياغة الديناميكية للمشكل،

الفقرة 4 تصف المنهج الديناميكي و في الفقرة الأخيرة نقدم التحقيق التجريبي للمشكل .

2 بيانات المشكل

بتمحور المشكل في سلسلة من العروض تتوافد زمنيا في شكل مجموعات المجموعه تلو الأخرى و ذلك تبعا لترتيب عشوائي ومن دون أي معلومات مسبقة عنها. كل مجموعة يتم فحصها من قبل صناعي القرار بمختلف التقييمات و التصنيف بهدف الاتفاق حول تحديد أفضل عرض في الإجمال. وبمجرد اختيار عرض عام عن المجموعة الحالية ينبغي لقرارات ثنائية في قبوله أو رفضه أن تتخذ و على كل صانع قرار أن يقرر استنادا إلى تصنيفه النسبي للعرض الجاري تقييمه. في ما يلي نقدم رسم توضيحي لمشكل التوقف المثالي أين يتم تلقي 5 عروض في 3 مجموعات تحتوى كل من المجموعتين الأولى و الثالثة على عرضين في حين تحتوي المجموعة الثانية على عرض واحد.



الشكل 4. 5 عروض في شكل 3 مجموعات

يمكن تحديد ملامح المشكل كالتالي

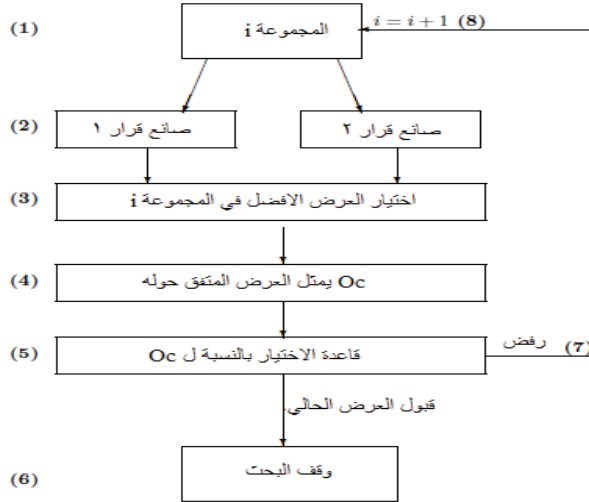
- 1- صناعي قرار مع أدلة مختلفة يشارك في منهج الاختيار.
 - 2- كل صانع قرار يستند إلى دالة فردية يتم تعريفها من خلال الترتيب المطلق للعروض.
 - 3- يتم تحديد العدد الإجمالي للعروض بصفة مسبقة.
 - 4- تتوافد مجموعات العروض بصفة عشوائية.
 - 5- إعادة استدعاء العروض يكون مسموح داخل المجموعة الواحدة و ليس بين المجموعات
 - 6- يتم إنهاء عملية البحث بمجرد قبول عرض و ذلك وفقا لقاعدة التوقف.
 - 7- الوصول إلى المجموعة الأخيرة يقتضي حتما قبول أفضل عروضها.
- كما يمكن أن نصف عملية البحث المبينة في الرسم التوضيحي 5 عدد من خلال المراحل التالية:
- * يقوم صانعا القرار بتقييم مجموعة من العروض (1)

* كل صانع قرار يقيم العروض داخل المجموعة اعتمادا إلى تصنيفها نسبيا و إلى مختلف الدالة المتوقعة (2)

* كل صانع قرار يحدد أفضل عرض بالنسبة إليه استنادا إلى دالته الفردية (3)

* يتم اللجوء إلى منهج التجميع بين القرارات بهدف الاتفاق حول عرض واحد (4)

* تنتهي عملية البحث (6) إذا أدت المفاوضات بين صانعي القرار (5) بقبول العرض المتفق حوله اما في حالة الرفض يتواصل البحث من خلال تفحص المجموعة الموالية (7) و (8).



الشكل 5. مراحل وضعية التوقف المثالي

3 الصياغة الديناميكية

في ما يلي نقوم بتحديد جملة رموز لحل المشكل ديناميكيا. يقتضي هذا الحل اقتراح رموز سيتم استعمالها لاحقا في كتابة المعادلات الرياضية التي تسمح باتخاذ قرار عقلائي للاختيار بين قبول كل عرض أو رفضه.

r_p	الرتبة المطلقة للعرض الجاري تقييمه استنادا إلى صانع القرار p ($p = 1, 2$)
k_p	الرتبة النسبية للعرض الجاري تقييمه استنادا إلى صانع القرار p ($p = 1, 2$)
$V^p(k_p)$	الدالة المطلقة لصانع القرار p
$Ev_s^p(i, k_p)$	الدالة المتوقعة لصانع القرار p ، التابعة لاختياره بإيقاف عملية البحث خلال المرحلة i
$Ev_c^p(i)$	الدالة المتوقعة لصانع القرار p ، التابعة لاختياره بمواصلة عملية البحث خلال المرحلة i
n	العدد الجملي للعروض.
M_i	عدد العروض التي تم تقييمها وصولا إلى المرحلة i
m_i	عدد العروض الموجودة داخل المجموعة i
d_p	القرار الفردي لصانع القرار p ويمكن أن يكون قبول أو رفض
D_p	القرار الجماعي خلال المرحلة i
t_p	حد أقصى لصانع القرار p

بما أن كل صانع قرار يهتم بتحديد العرض الأفضل، فإن ذلك يتحقق بمقارنة المعادلتين التاليتين :

<u>الدالة الفردية المتوقعة إذا تم مواصلة عملية الاختيار</u>	<u>الدالة الفردية المتوقعة إذا تم وقف عملية الاختيار</u>
$Ev_c^p(i) = \sum_{a=1}^{M_{i+1}} Ev^p(a, i+1) \frac{\binom{M_{i+1}-a}{m_{i+1}-1}}{\binom{M_{i+1}}{m_{i+1}}}$	$Ev_s^p(i, k_p) = \sum_{a=k_p}^{n-M_i+k_p} U(a) \frac{\binom{a-1}{k_p-1} \binom{n-a}{M_i-k_p}}{\binom{n}{M_i}}$

وبذلك فإن المرافق المتوقعة لكل صانع قرار عند إتباع إستراتيجية مثلى تتشكل استنادا إلى المعادلتين عند اختيار مواصلة أو إيقاف عملية البحث.

$$Ev^p(i, k_p) = \begin{cases} Ev_s^p(i, k_p)\chi_E + Ev_c^p(i)(1 - \chi_E) & \text{if } i < n \\ v^p(k_p) & \text{if } i = n \end{cases}$$

مع العلم أن

$$X_E = \begin{cases} 1 & \text{if } Ev_s^p(i, k_p) \geq Ev_c^p(i) \\ & \text{and } Ev_c^{p'}(i) - Ev_s^{p'}(i, k_{p'}) \leq t_p \\ 0 & \text{otherwise} \end{cases}$$

4 المنهج الديناميكي

الجدول الموالي يحدد الحالات التي تمكن صانعا القرار من الاتفاق حول رأي مشترك بقبول العرض الجاري فحصه أو رفضه. فقبول العرض يكون إذا كانت الدالة المتوقعة عند توقف عملية البحث لكلا صانعي القرار تفوق الدالة المتوقعة عند مواصلة البحث و يكون الاتفاق علي الرفض في الحالة المخالفة.

قبول العرض الحالي	رفض العرض الحالي
$\begin{cases} Ev_s^1(i, k_1) \geq Ev_c^1(i) \\ Ev_s^2(i, k_2) \geq Ev_c^2(i) \end{cases}$	$\begin{cases} Ev_s^1(i, k_1) < Ev_c^1(i) \\ Ev_s^2(i, k_2) < Ev_c^2(i) \end{cases}$

في حالة اختلاف الآراء بين صانعا القرار أي إذا كان صانع القرار p يقضي بقبول العرض الحالي و صانع القرار p' يطمح إلي رفضه فيكون حل الصراع كما يلي :

إيقاف عملية البحث	مواصلة عملية البحث
$\begin{cases} Ev_s^p(i, k_p) \geq Ev_c^p(i) \\ Ev_c^{p'}(i) - Ev_s^{p'}(i, k_{p'}) \leq t_p \end{cases}$	$\begin{cases} Ev_s^p(i, k_p) < Ev_c^p(i) \\ Ev_c^{p'}(i) - Ev_s^{p'}(i, k_{p'}) > t_p \end{cases}$

و سيتم تنفيذ هذه القاعدة للتوقف في الخوارزمية التالية.

الخوارزمية للتوقف المثالي لصانعي القرار

* التهيئة

1- المعلومات n ، $k_p(p=1, 2)$ ، $t_p(p=1, 2)$.

2- حساب الأدلة المتوقعة لكل صانع قرار.

* العملية التفاعلية

1- بالنسبة إلى كل صانع قرار p إدخال $Ev_s^p(i, k_p)$ و $Ev_c^p(i)$ إذا $Ev_s^p(i, k_p) \geq Ev_c^p(i)$ إذن $d_p \rightarrow$ وقوف

آخر

 $d_p \rightarrow$ مواصلة

نهاية إذا

2- إذا $d_p = d_p$ إذن $d_p \rightarrow D_p$

نهاية إذا

إذا مواصلة $d_p =$ وقوف d_p إذنإذا $t_p - Ev_s^p(i, k_p) \leq Ev_c^p(i)$ إذن $D_p \rightarrow$ وقوف

آخر

 $D_p \rightarrow$ مواصلة

نهاية إذا

نهاية إذا

نفذنا المعادلات الديناميكية التي تم تقييمها في الجزء السابق من العمل استنادا إلى لغة البرمجة جافا. في هذا الجزء سنقوم بتقديم جملة من المشاكل التي تقتضي التوقف المثالي مع اختلاف الأدلة الفردية لصانعي القرار وذلك باعتبار عدد من العروض يتم تلقيه في شكل مجموعات. يتراوح هذا العدد بين 50 و 150 عرضا. لتحقيق نجاعة المنهج الديناميكي الذي تطرقنا له سابقا سنقوم بتغيير الأدلة الفردية لصانعي القرار وسنقوم في مرحلة ثانية تغيير الحد الأقصى لكل مشكل.

1.5 سلوك صانعي القرار مع اختلاف الأدلة المطلقة

في هذا القسم تعميم للمنهج الذي تم التطرق إليه سابقا وذلك بالاستناد إلى الأدلة التالية:

تعريف الدالة	لا شيء ولكن الأفضل	مشكل الرتبة الأدنى	دالة الرتب المعكوسة
الباحثين	* ليندلي 1961 * دانكان 1963 * جيلبرت و مستلي 1966	* تشو و آخرون 1994 * تشان 1994	* تشان 1994 * تشان 1999
الدالة المطلقة	$v^p(r_p) = \begin{cases} 1 & \text{if } r_p = 1 \\ 0 & \text{otherwise} \end{cases}$	$v^p(r_p) = r_p$	$v^p(r_p) = \frac{1}{r_p}$
الدالة المتوقعة إذا تم وقف عملية الاختيار	$Ev_s^p(i, k_p) = \begin{cases} \frac{M_i}{n} & \text{if } k_p = 1 \\ 0 & \text{otherwise} \end{cases}$	$Ev_p^p(i, k_p) = \sum_{a=0}^{n-M_i+k_p} a \times \frac{\binom{a-1}{k_p-1} \binom{n-a}{M_i-k_p}}{\binom{n}{M_i}} = \frac{n+1}{M_i+1}$	$Ev_p^p(i, k_p) = \sum_{a=0}^{n-M_i+k_p} \frac{1}{a} \times \frac{\binom{a-1}{k_p-1} \binom{n-a}{M_i-k_p}}{\binom{n}{M_i}}$
الدالة المتوقعة إذا تم مواصلة عملية الاختيار	$Ev_c^p(i) = \frac{m_{i+1}}{M_{i+1}} \times Ev^p(i+1, 1) + \frac{M_i}{M_{i+1}} \times Ev_c^p(i+1)$	$Ev_c^p(i) = \sum_{a=1}^{M_{i+1}} Ev^p(a, i+1) \times \frac{\binom{M_{i+1}-a}{m_{i+1}-1}}{\binom{M_{i+1}}{m_{i+1}}}$	$Ev_c^p(i) = \sum_{a=1}^{M_{i+1}} Ev^p(a, i+1) \times \frac{\binom{M_{i+1}-a}{m_{i+1}-1}}{\binom{M_{i+1}}{m_{i+1}}}$

النتائج داخل الجدول رقم 1 توافق مع الدالة المتوقعة أثناء المرحلة الأولى من العملية الديناميكية، و

استنادا على هذه النتائج استنتجنا ما يلي:

* احتمال الفوز، بالنسبة للدالة لا شيء ولكن الأفضل و دالة الرتب المعكوسة، يتطور بصفة معكوسة مع تزايد عدد المجموعات.

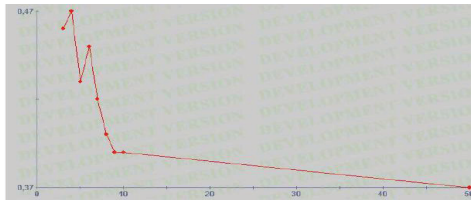
* مع كل عدد من العروض نلاحظ أن قيمة الدالة المتوقعة في حالة الرتبة الأدنى أكبر من قيمة الدالة المتوقعة بالنسبة للرتب المعكوسة، هذه الأخيرة التي هي بدورها أكبر من التي في حالة الأشياء و لكن الأفضل.

* إذا كان اختيار صانعي القرار يتركز علي الدالة نفسها، تصبح الأدلة المتوقعة متطابقة. من ثم فالغموض الرئيسي للمشكل هو عند اختيار أدلة مختلفة من قبل صانعي القرار و بذلك تكون القرارات الفردية متباينة.

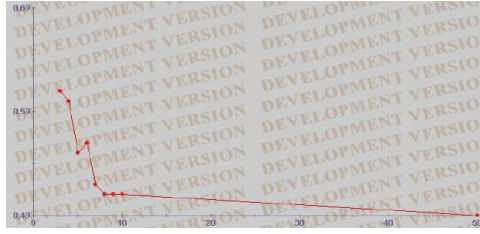
هذه النتائج مبرهنة من خلال الرسمين 6 و 7. الدالتان علي أن احتمال الفوز ينقص مع تزايد عدد المجموعات.

n	$\left\{ \begin{smallmatrix} 1 \\ 0 \end{smallmatrix} \right\}$	$\left\{ \begin{smallmatrix} 1 \\ 0 \end{smallmatrix} \right\}$	$\frac{1}{k}$	$\left\{ \begin{smallmatrix} 1 \\ 0 \end{smallmatrix} \right\}$	$\left\{ \begin{smallmatrix} 1 \\ 0 \end{smallmatrix} \right\}$	k
10 (5,5)	0.5	0.5	0.7	0.5	0.5	1.83
10 (3,2,5)	0.5	0.5	0.58	0.5	0.5	2.74
10 (2,3,3,2)	0.47	0.47	0.6	0.47	0.47	3.66
10 (2,3,2,2,1)	0.45	0.45	0.6	0.45	0.45	3.66
50 (20,10,20)	0.46	0.46	0.61	0.46	0.46	2.42
50 (10,15,15,10)	0.47	0.47	0.6	0.47	0.47	4.63
50 (10,10,10,10,10)	0.43	0.43	0.55	0.43	0.43	4.63
50 (5,5,10,15,5,10)	0.45	0.45	0.56	0.45	0.45	8.5
50 (5,10,5,10,5,5,10)	0.42	0.42	0.52	0.42	0.42	8.5
50 (10,5,5,5,5,5,5,10)	0.4	0.4	0.51	0.4	0.4	4.63
50 (10,5,5,5,5,5,5,5,5)	0.39	0.39	0.51	0.39	0.39	4.63
50 (5,5,5,5,5,5,5,5,5,5)	0.39	0.39	0.51	0.39	0.39	8.5
100 (50,50)	0.5	0.5	0.69	0.5	0.5	1.98
100 (20,40,40)	0.53	0.53	0.6	0.53	0.53	4.8
100 (25,35,10,30)	0.49	0.49	0.6	0.49	0.49	3.88
100 (20,20,20,20,20)	0.43	0.43	0.55	0.43	0.43	4.8
100 (20,20,10,10,30,10)	0.42	0.42	0.55	0.42	0.42	4.8
100 (15,15,15,15,5,25,10)	0.42	0.42	0.53	0.42	0.42	6.31
100 (5,25,5,5,5,15,25,5,10)	0.41	0.41	0.51	0.41	0.41	16.8
150 (75,75)	0.5	0.5	0.69	0.5	0.5	1.98
150 (50,50,50)	0.5	0.5	0.57	0.5	0.5	2.96
150 (50,35,15,50)	0.45	0.45	0.55	0.45	0.45	2.96
150 (25,25,35,15,50)	0.45	0.45	0.54	0.45	0.45	5.8

الجدول 1. النتائج التجريبية للمشكل التوقف المثالي



الشكل 6. اختلاف الدالة المتوقعة مع اختلاف عدد المجموعات ($V(k)=(1,0)$, $n=50$)



الشكل 7. اختلاف الدالة المتوقعة مع اختلاف عدد المجموعات ($V(k)=(1/k)$, $n=50$)

2.5 تحليل الحساسية بالنسبة إلى t_p

لنفترض في هذه الفقرة أن صانعي القرار يطمحان إلى تحديد أفضل عرض لكل منهم. و على هذا الأساس تم إجراء تجارب على عدد من العروض يتراوح بين 50 و 150 تتوافد في شكل مجموعات بأعداد مختلفة. و قد قمنا بتقديم المنهج التفاعلي لكل سلسلة عروض و ذلك مع اختلاف الحد الأقصى

$$t_p (t_p = 0.3, 0.5, 0.8)$$

لعدد n من العروض توجد n^2 إمكانية لتصنيف الرتب النسبية، هذه الرتب التي ستحدد بمقتضاها

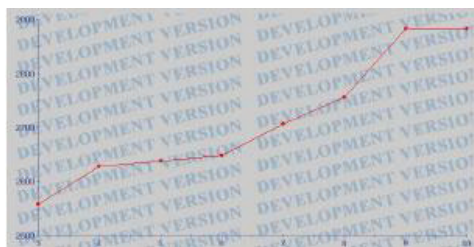
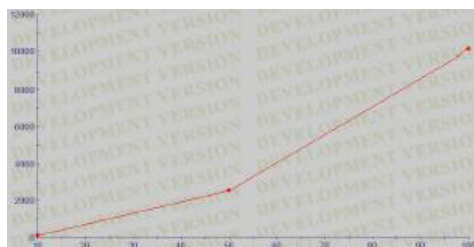
القرارات الفردية بقبول العرض الجاري تقييمه أو برفضه. و من هنا فبمجرد اتخاذ القرار بالقبول يدعى العرض المتفق عليه "أفضل اتفاق".

الجدول الثاني يبين عدد العروض المسماة "أفضل اتفاق" و ذلك اعتبارا لكل الرتب النسبية مع اختلاف العروض و اختلاف مجموعاتها. و انطلاقا من الجدول رقم 2 نتبين أن عدد العروض المتفق حولها

يتزايد مع تزايد عدد العروض و عدد المجموعات و الحد الأقصى t_p و هو مبين في الرسم 8 و 9.

n	الدالة المتوقعة f	أفضل دالة		
		$t_p = 0.8$	$t_p = 0.5$	$t_p = 0.3$
10 (5,5)	0.5	109	101	101
10 (2,3,5)	0.5	109	105	101
10 (2,3,3,2)	0.47	124	113	109
10 (2,3,2,2,1)	0.45	139	112	108
50 (20,10,20)	0.46	2559	2530	2501
50 (10,15,15,10)	0.47	2628	2565	2541
50 (10,10,10,10,10)	0.43	2638	2530	2501
50 (5,5,10,15,5,10)	0.45	2648	2540	2540
50 (5,10,5,10,5,5,10)	0.42	2707	2575	2575
50 (10,5,5,5,5,5,5,10)	0.4	2756	2565	2536
50 (10,5,5,5,5,5,5,5)	0.39	2884	2585	2537
50 (5,5,5,5,5,5,5,5,5)	0.39	2884	2555	2502
100 (50,50)	0.5	10099	10001	10001
100 (20,40,40)	0.53	10119	10060	10001
100 (25,35,10,30)	0.49	10258	10130	10071
100 (20,20,20,20,20)	0.43	10278	10060	10001
100 (20,20,10,10,30,10)	0.42	10397	10140	10091
100 (15,15,15,15,5,25,10)	0.42	10516	10200	10092
100 (5,25,5,5,5,15,25,5,10)	0.41	10556	10105	10002
150 (75,75)	0.5	22649	22501	22501
150 (50,50,50)	0.5	22699	22600	22501
150 (50,35,15,50)	0.45	22868	22685	22502
150 (25,25,35,15,50)	0.45	22868	22585	22501
150 (25,25,35,15,25,25)	0.44	23117	22710	22626

الجدول 2. النتائج التجريبية للمنهج التفاعلي للمشكل

الشكل 8. اختلاف عدد العروض المتفق حولها مع اختلاف عدد المجموعات ($n=100$, $t_p=0.8$)

الشكل 9. اختلاف عدد العروض المتفق حولها مع العدد الجملي للعروض

6 الخاتمة

اقترحنا، في هذه الورقة، صياغة جديدة لمشكلة التوقف المثالي حيث أن صانعي قرار يطمحان إلى الاتفاق حول عرض مشترك. تتوافد العروض في شكل مجموعات وفي ترتيب عشوائي. كتبنا المعادلات الديناميكية و المنهج التفاعلي الديناميكي لحل المشكل و توفير التحقيق التجريبي و ذلك بالاعتماد علي أعداد مختلفة من العروض. تبين التجارب إن احتمال الوقوع علي أفضل اختيار من قبل كل صانع القرار يقل مع تزايد عدد العروض. و من خلال اختلاف عدد العروض قمنا بتقييم نهجنا الديناميكي و أظهرت النتائج أن عدد أفضل العروض يتطور مع عدد العروض و عدد الجماعات.

وكمشاريع مستقبلية يمكن لقاعدة التوقف التي قمنا بالاستناد إليها لحل مشكل التوقف المثالي أن يكون لها تطبيقات في مجال الإعلامية و شبكات الاتصال.

المراجع

- [1] Ben Abdelaziz F and Krichen S (2005). An interactive method for the optimal selection problem with two decision makers. European Journal of Operational Research 162: 602-609.
- [2] Ben Abdelaziz F and Krichen S (2007). An Optimal Stopping Problem with two Decision Makers. Sequential Analysis 26: 467-480.
- [3] Ben Abdelaziz F and Krichen S (2009). An Experimental Investigation of the Optimal Selection Problem with Two Decision Makers. Economics and Mathematical Systems 26: 175-185.
- [4] Chun Y H, Moskowitz H and Plante R D (1994). Dynamic programming formulation of the group interview problem with a general utility function. European Journal of Operational Research 78: 81-92.
- [5] Chun Y H (2004). Generalized best choice problem based on the information economics approach. J Opl Res Soc 55: 988-999.
- [6] Gardner M (1960). Mathematical Games. Scienti_c American 3: 150-156.
- [7] Dynkin E B (1963). Optimal Choice of the stopping moment of a markov process, Dokl. Akad. Nauk. SSSR 150: 238-240.
- [8] Gilbert J and Mosteller F (1966). Recognizing the maximum of a sequence. American statistical Association Journal 61: 35-73.
- [9] Gusein-Zade S M (1966). The problem of choice and the optimal stopping rule for a sequence of independent trials. Theory of Probability and its Applications 11: 472-476.
- [10] Lindley D V (1961). Dynamic programming and decision theory. Applied Statistics 10: 39-51.
- [11] Sakaguchi M (1984). Bilateral sequential games related to the no-information secretary problem. Mathematica Japonica 29: 961-973.

[12] Sakaguchi M and Mazalov V (2004). A non-zero-sum no-information best-choice game. Mathematical Methods of Operations Research, 60; 437-451.

الملخص باللغة الانجليزية

Title: Group Selection Problems by Two Decision Makers

Abstract. The present paper proposes a generalization of the optimal stopping problem characterized by two decision makers that are about selecting jointly a single offer among a stream of offers arriving by groups. At each stage of the dynamic process, a group of offers is examined. We assume that if a group of offers is declined all offers of this group cannot be re-examined in later stages. However, inside one group the recall of offers is allowed. We propose a dynamic programming approach and write the backward recursive equations for various utilities of the DMs. We then illustrate our study by an experimental investigation showing the impacts in terms of expected utility.

Keywords: Computer Science, Telecommunication Networks, Optimal stopping, Dynamic programming, Game theory, Group decisions

المصطلحات

التوقف المثالي	Optimal Stopping
صانعي قرار	Decision Makers
العملية الديناميكية	Dynamic proces
عروض	Offers
البرمجة الديناميكية	Dynamic programming
معادلات	Equation
نظرية الألعاب	Game theory
نظرية القرارات الجماعية	Group Decision
الرتب المطلقة	Absolute ranks
الرتب النسبية	Relative ranks
الدالة العامة	Utility function
الصراع	Conflict
الدالة المتوقعة	Expected utility
الدالة المتوقعة عند قبول العرض	Expected utility when the offer is selected

الدالة المتوقعة عند رفض العرض	Expected utility when the offer is discarded
منهج الاختيار	Selection process
التصنيف	Ranking
قاعدة التوقف	Stopping rule
عملية البحث	Searching process
حد أقصى	threshold
أفضل اتفاق	Best compromise
المنهج الديناميكي	Dynamic approach

كوك شيك ونج

مصطفى صفاء عبدالكريم

العنوان البريدي: Department of Computer System and Technology, Faculty of Computer Science & Information Technology University of Malaya, 50603 Kuala Lumpur, MALAYSIA

البريد الإلكتروني: koksheik@um.edu.my , mustafa.alwahaib@siswa.um.edu.my

نستعرض في هذا البحث خوارزمية جديدة عن طريق ترميز العناصر المتعاقبة المتشابهة (. . .) في الصورة بتقنية مائعة للتضخم الذي يؤدي إلى زيادة حجم العناصر نتيجة الترميز التقليدي. تستند الخوارزمية ترميز إنتروبي مقترح كلمات ترميزية لترميز التدرجات الرمادية بصورة وفقا للإحتمال الرياضي لظهور كل هيستوجرام. ترميزي باحتوائها على إشارة علمية مدمجة معها ، تستخدم للتفريق بين قيمة في الدفق الثنائي للصورة المرمزة بتقنية ت. . . .

الحاجة لحجز كلمات ترميزية خاصة أو ثنائية تستخدم كإشار علمية كما هو الحال في الطرق التقليدية. لدينا من مصادر ، تعد الخوارزمية المقترحة الأولى من نوعها التي ترمز عنصرين متتاليين من نفس القيمة بكلمة ترميزية واحدة ، وقادرة على ترميز عدد لانها .

تظهر أن الخوارزمية المقترحة تحل مشكلة التضخم في تقنيات ت. . . .

ليدية	65%	يحقق الترميز المقترح	أعلى من ترميز
هوفمان	80%		

الكلمات الجوهرية: الضغط المحافظ ، ترميز إنتروبي ، ترميز هوفمان ، ترميز العناصر المتعاقبة المتشابهة

1.

أصبحت الصور الرقمية وسطا رئيسيا للمعلومات خلال السنوات الأخيرة بفضل التطور الكبير بكفاءة أجهزة التقاط الصور وعرضها ، ويعزز ذلك انخفاض أسعار تلك الأجهزة أيضا. من الناحية المبدئية ، تحفظ الصورة الرقمية () على شكل مصفوفات من العناصر الرقمية () كبير بالعادة؛ يتناسب حجمه مع عمق الألوان وعدد قنواتها في الصورة. ومن ثم ، تخضع الصورة عادة لعملية ضغط تؤدي لتصغير حجمها ، لغرض حفظها وأرشفتها أو نقلها عبر قنوات الاتصال بكفاءة.

يرتبط نوع الضغط المستخدم بنوع الصورة المزمع تصغير حجمها. في الصور العادية ، مثل تلك الملتقطة لأغراض ترفيهية أو تجارية ، يمكن السماح بتغيير بعض معلومات الصورة الأصلية نتيجة لضغطها ، بحيث تكون الصورة المستعادة بعد إزالة الضغط مقاربة أو مشابهة لحد كبير للصورة الأصلية ، وربما لا تلحظ العين البشرية تلك التغييرات الطفيفة فيها. هذا النوع من الضغط نسميه: الضغط الغير محافظ ، وفي هذا النوع من الضغط ، يستحيل استعادة الصورة الأصلية بمعلوماتها الأصلية التي كانت قبل تصغير حجمها بتقنية ضغط غير محافظ.

في الصور الطبية والعسكرية وتلك الملتقطة من قبل الأقمار الصناعية ، ينبغي تجنب أي خسارة (ولو جزئية) لأي معلومة في الصورة نتيجة الضغط ، بمعنى ، يجب استعادة الصورة كاملة ، على هيئتها قبل ضغطها ، بعد هذا النوع من ضغط الصور نطلق عليه: ضغط محافظ ، لأنه يحفظ الصورة ويستعيد معلوماتها

الأصلية بعد إزالة الضغط. في الحقيقة، هناك العديد من تقنيات الضغط المحافظ في الأدبيات المنشورة، وتعد تقنية "ترميز العناصر المتعاقبة المتشابهة" والمسماة أحيانا "ترميز أطوال الاستمرار" [1] أحد أهم وأبسط تلك التقنيات وهي جزء من تقنية ضغط الصور المعياري "جي بيج" [2] المعيار الخاص باتحاد الاتصالات العالمي [3]، وأيضا ضمن العديد من تقنيات ضغط الصور [4]. من هنا وصاعداً، سنشير لتقنية ترميز العناصر المتعاقبة المتشابهة بالأحرف ".....".

تقوم تقنية بترميز مجموعة من العناصر المتجاورة والتي تحمل نفس القيمة العددية من خلال تسجيل قيمتها العددية، متبوعة بعدد تكرارها. على سبيل المثال، فإن العناصر التالية: (555555999999) تُرمز بتقنية : (7 5 6 9)، حيث أن الـ 6 7 يشيران لعدد مرات تكرار العناصر 9 5. تُعد تقنية ذات كفاءة عالية بترميز الصور المحتوية على عناصر متشابهة تتجاور مع بعضها. من العناصر المتكررة، مثل الصور المحتوية على لونين فقط: أسود وأبيض، لكن تقل كفاءة تلك التقنية حينما تستخدم لترميز صور المحتوية على العديد من العناصر ذات الألوان (أو القيم) في الصورة، بشكل يمنع من ظهور سلاسل طويلة من صر متجاورة ذات قيم متكررة [5]. التباين العالي بين قيم عناصرها، يؤدي ترميزها بتقنية الـ إلى زيادة (القيمة، المرزمة، بدلا من ضغطها. على سبيل المثال، لو فرضنا وجود سلسلة من العناصر الغير متشابهة ذات القيم: (5 4 3 2 1) : (1 5 1 4 1 3 1 2 1 1). وحيث أن إضافة قيمة (1) قياسا إلى حجمها قبل الترميز). هذا السلوك مع العناصر الغير متشابهة نطلق عليه:

في هذه الورقة، نستعرض حلا مقترحا لمشكلة التضخم المذكورة أعلاه من خلال خوارزمية بتقنية قائمة على استخدام ترميز إنتروبي جديد تتم توليد كلماته الترميزية وفقا لقواعد معينة؛ حيث بإمكان الترميز الإنتروبي المقترح التمييز بين العناصر المفردة أو المكررة مرتين فقط، ومن ثم لا يقوم بإضافة عدد التكرار بعدها، وبين تلك المكررة ثلاث مرات أو أكثر، ومن ثم تكون القيمة العددية للعناصر المكررة متبوعة بكلمة ترميزية تشير إلى عدد تكرارها. تتم عملية التمييز من خلال إضافة "إشارة علمية" ملحقة بكل كلمة ترميزية وهذا ينفي الحاجة لحجز خانة ثنائية لكل كلمة ترميزية - حجز كلمات ترميزية - علمية - كما هو الحال في التقنيات الحالية - يد نوع الكلمة الترميزية القادمة، من حيث كونها قيمة للعنصر. تمتاز التقنية المقترحة بقدرتها على ترميز عدد لانهاية من التكرارات ضمن كلمة ترميزية واحدة، في مقابل محدودية هذا العدد في التقنيات المستخدمة حاليا، وبحسب ما توافر لدينا من مصادر، تعد التقنية المقترحة هي الأولى من نوعها التي بإمكانها ترميز عنصرين متجاورين متشابهين بكلمة ترميزية واحدة، ومن ثم ضغطهم إلى عنصر واحد ممثلا بكلمة ترميزية مفردة. ملية خوارزمية

2. التقنيات التقليدية

في الأدبيات المشار إليها في [2] [5] [6]، أقرح استخدام تقنية الـ فقط حينما يكون العنصر مكررا ثلاث مرات أو أكثر، بينما لا يتم ترميز أي عنصر مكرر لمرة واحدة (بمعنى أنه يحمل قيمة تختلف عن العنصرين المجاورين له من اليمين واليسار)، كما لا يتم ترميز أولئك العناصر المكررة لمرتين فقط ولكي يتم، يتم استخدام إشارة علمية تمثيل ثنائي مميز (بمعنى أنه يمكن تمييزه عن غيره من الكلمات الترميزية) للتمييز بين الكلمات الترميزية التي تشير إلى "قيمة عنصر" وتلك التي تشير إلى "التنائي الذي يشكل ترميز الصورة."

التقنيات التي تقوم بقراءة الصورة وترميزها صفا صفا، من اليسار لليمين، هناك تقنيتان رئيسيتان لتنفيذ الإشارة العلمية الطريقة الأولى قائمة على حجز كلمة ترميزية كاملة لاستخدامها كإشارة علمية، بحيث تسبق الإشارة العلمية في الدفق الثنائي يعني أن الترميزية القادمة تشير إلى قيم ويتبع ذلك كلمة أخرى تشير إلى عدد تكراره. [6] يتم حجز الكلمة الترميزية لأي لون غير موجود في هيستوجرام الصورة لغرض استخدامها كإشارة علمية. من الواضح أن هذهريقة لا يمكن تطبيقها في حال كانت جميع الألوان (على سبيل المثال جميع الـ 256 لونا في الصور الرمادية حيث يتم استخدام 8 ثنائيات لترميز كل لون) موجودة في هيستوجرام الصورة. نفس التقنية تم استخدامها في الأدبيات المشار إليها في [7] [8]. من هنا وصاعداً، سنطلق على هذه التقنية : التقليدية الأولى.

3. الترميز الإلكتروني

في طبقة الأغصان ، وهي الطبقة التي تعلق طبقة الصبغ العامة في المخطط الشجري بالشكل رقم 1 ، يتم تعيين تبدليتين ثنائيتين من ل لكل من " " " " في اي صبغة عامة ، وفقا للقاعدتين الثالثة والرابعة ، واللذان توضحان التبديلات الثنائية التي يمكن منحها لـ " " " " " " . " " " " " " فتظل بدون تبدلية ثنائية في طبقة الأغصان ، لسبب الموضوع في الفقرة التالية.



إذن يمكن توليد كلمة ترميزية من الفئة ب من كل كلمة ترميزية من الفئة أ ، وبشكل عام ، لحساب عدد الكلمات الترميزية من الفئة ب التي يمكن توليدها من أي صيغة عامة ، فإن عدد التباديل التي يمكن تعيينها لـ " " هو 2 ، وعدد التباديل التي يمكن تعيينها لـ " " هو 2 ، وذلك وفقا للقاعدة الرابعة ، بينما يكون عدد التباديل التي يمكن تعيينها لـ " " هو 2 .
فإن العدد الكلي للكلمات الترميزية من الفئة ب التي يمكن اشتقاقها من أي صيغة عامة هو:

$$2^2 = 4$$

وبوضح الجدول رقم: 2 ، وتحديدًا تحت العمود اليسار كلمات الترميزية .
وجميع الكلمات الترميزية من الفئتين: أ و ب تمتاز بخاصية القابلة للكشف بشكل فريد. وهذه الخاصية تجعل كل كلمة ترميزية قابلة للقراءة والإكتشاف بشكل فريد دون أن تتداخل مع الكلمة المجاورة ضمن الدفق الثنائي لترميز . وجود هذه الخاصية في الكلمات الترميزية في الفئتين المشار إليهما يعود لتمييز كل كلمة ترميزية بوحدة على الأقل من الخصائص التالية:

1. الصيغة العامة المستخدمة لاشتقاق الكلمة الترميزية.
2. عدد تكرار ثنائيات البداية " " الكلمة الترميزية.
3. موقع الثنائيات الوسطى " " بين ثنائيات البداية " " في الكلمة الترميزية.

2 : الكلمات الترميزية من الفئة أ والفئة ب

الصيغة العامة =	
01 00	01 01 00
10 00	10 01 00
11 00	11 01 00
00 01	01 10 00
10 01	10 10 00
11 01	11 10 00
.	.
.	.
.	.

1: أمثلة للصيغ العام

1	.
2	.
3	.
4	.
5	.
6	.
7	.
8	.
.	.
.	.
.	.

4. خوارزمية الضغط المحافظ باستخدام ت . . .

4.1 الترميز

نستعرض في هذه الـ خوارزمية مقترحة لضغط الصور بشكل محافظ باستخدام الكلمات الترميزية من الفئة أ ونشير ابتداءً إلى أن الغرض من تحديد تباديل ثنائية " " في الفئة ب كان لجعل الكلمة الترميزية قابلة للتكيف مع متطلبات الترميز ، والتي يتم تحديدها أثناء عملية الترميز ، ويتم استيفاءها من خلال منح " " تبديلة من المجموعة ل خلال مرحلة الترميز. ونستعرض فيما يلي خوارزمية ضغط عناصر الصورة مستند . . . مع طريقة كفوة في تحديد الإشارة العلمية.

ترميز الصورة هو عملية استبدال عناصر الصورة بكلمات ترميزية. تبدأ مرحلة الترميز بتعيين كلمات ترميزية من الفئة ب للترددات الرمادية في الصورة (في حال كان الألوان في الصورة ممثلة بثمانية ثنائيات). تتم عملية التعيين وفقا للإحتمال الرياضي لظهور كل تدرج رمادي في هيستوجرام الصورة ، بحيث تعين كلمات ترميزية ذات طول صغير للترددات الرمادية ذات احتمال الأكبر للظهور في الصورة، والعكس بالعكس. ويتم حفظ أبعاد الصورة والتي تحدد عدد الصفوف والأعمدة فيها ، وذلك لتحديد النقطة التي يتوقف عندها فاتح الترميز عند إزالة ضغط

الصورة لاستعدادتها ، ويتم أيضاً حفظ عدد التدرجات الرمادية الموجودة في هيستوجرام الصورة ، متبوعاً بترتيبها حسب الاحتمال الرياضي لكل منها ، وكل ما سبق من المعلومات تحفظ في مقدمة الملف الذي تخزن الصورة الرمزة فيه.

يلي ذلك قراءة قيمة عناصر الصورة من اليسار إلى اليمين ، ويعين لكل عنصر كلمة ترميزية من الفئة ب. " " غير محددة التباديل في هذه الفئة ، يتم استغلال " " لتكون إشارة علمية ، بحيث تحدد حالة الكلمة الترميزية القادمة من ناحية كونها "قيمة لتدرج رمادي" "شابهة المتعاقبة.

العلمية ثلاث قيم تبديل ، يمثل كل منها حالة معينة للإشارة العلمية ، وهي $1' \ 2' \ 3'$ ، حيث أن $1' > 2' > 3'$ ، علماً بأن جميع هذه الحالات هي تبديل محققة للشروط الخامس في الشكل رقم: 1.

ومن ثم ، إذا قرأ الرمز عنصراً ضمن سلسلة متعاقبة متشابهة ، جميعها يحمل نفس القيمة ، وتكرر لثلاث مرات أو أكثر ، يتم ترميز هذه السلسلة بكلمة ترميزية واحدة تشير إلى قيمة التدرج الرمادي لتلك السلسلة المتشابهة ، ويتم تعيين " " لتلك الكلمة الترميزية بتبديلة تمثل $3'$ ، وهي التبديلة ذات أعلى قيمة ضمن التبديلات " " . تشير $3'$ في الكلمة الترميزية إلى أن الكلمة التالية لها تشير إلى عدد تكرار العنصر السلسلة ، وليست قيمة تدرج رمادي. وهنا ، يتم تعيين كلمة ترميزية من الفئة أ للإشارة إلى قيمة التكرار ، حيث يمكن للكلمات الترميزية لهذه الفئة ترميز عدد لانتهائي من التكرارات.

مكرراً لمرةٍ فقط (أي أنه يتشابه فقط مع العنصر المجاور له من جهة اليمين)

الرمز يعين كلمة ترميزية واحدة تشير إلى قيمة التدرج الرمادي لهذين العنصرين ، ويتم تعيين تبديلة لـ " "

$2'$ ، والتي تشير إلى أن العنصر الحالي مكرر لمرةٍ فقط ، والكلمة الترميزية التالية هي قيمة التدرج الرمادي للعنصر الثالث الذي يلي العنصرين الحاليين. وبحسب ما توافر لدينا من الخوارزمية المقترحة في هذه الورقة هي الأولى من نوعها التي ترمز عنصريين يحملان نفس القيمة بكلمة ترميزية واحدة ، وهو أمر يساهم في تصغير حجم البيانات التي تمثل عناصر الصورة ، حيث تقوم الطرق التقليدية بترميز العنصرين بكلمتين ، إحداها للإشارة إلى قيمة التدرج الرمادي للعنصرين ، والأخرى لتحديد عدد التكرار (وهو 2) ، كما هو موضح في الأدبيات المشار إليها في [6] [7]. بينما تكفي طرق أخرى بترك العنصرين المتجاورين اللذان يحملان نفس القيمة بدون ترميز ، كما هو الحال في الطرق المشار إليها في الأدبيات [9] [8] [10].

أخيراً ، إذا قرأ الرمز عنصراً مفرداً غير مكرر ، بمعنى أن قيمة تدرجه الرمادي تختلف عن جاره الأيمن ، وجاره الأيسر ، يقوم الرمز بتعيين كلمة ترميزية تشير إلى قيمة التدرج الرمادي للعنصر ، ويعين لـ " " تبديلة $1'$ ، للإشارة إلى أن هذا العنصر غير مكرر ، والكلمة الترميزية القادمة هي قيمة التدرج الرمادي للعنصر المجاور له.

الصورة يتم ترميزها باستخدام الحالات المشار إليها. ويُلخص الجدول رقم: 3 طريقة الترميز ، بينما يوضح الملحق أ كود البرمجة الوهمي لخوارزمية الترميز. ومن الواضح أن الخوارزمية المقترحة تمنع ظهور تضخم في حجم العناصر المرمزة من خلال منع إضافة قيمة للتكرار للعناصر الغير مكررة . وبسبب هذا ، يتم توصيف هذه الخوارزمية بأنها مائعة .

3: ملخص عملية الترميز

عبرة الترميز	
عنصر غير مكرر	(قيمة التدرج الرمادي للعنصر ، $1'$)
عنصر مكرر مرتين	(قيمة التدرج الرمادي للعنصر ، $2'$)
	(قيمة التدرج الرمادي للعنصر ، $3'$)

فك الترميز (حل الترميز [13]) هي عملية قراءة الملف المرمرز للصورة وفك شفرته القيم الأصلية لعناصر الصورة. تبدأ عملية الترميز بقراءة مقدمة ملف الصورة لمعرفة أبعادها التدرجات الرمادية ، والاحتمال الرياضي لظهور كل منها، لمعرفة الكلمات الترميزية التي مُنحت لكل لون من ألوان التدرج الرمادي ، يلي ذلك قراءة الدفق الثنائي الذي يشكل ترميز الصورة المرمرزة ، حيث يتم تمييز كل كلمة ترميزية من خلال تطبيق القاعدة الأولى بالشكل رقم: 1. حينما يتم تمييز كلمة ترميزية ، يتم قراءة الأجزاء التي " " " " لمعرفة قيمة التدرج الرمادي للعنصر الذي تمثله الكلمة الترميزية ، ويتم قراءة الجزء الأخير من الكلمة الذي يمثل " " لمعرفة نوع الكلمة الترميزية التالية و عدد مرات تكرار العنصر الذي تشير إليه الكلمة الترميزية الحالية.

$= 1$ فهذا يعني أن العنصر الذي تشير إليه الكلمة الترميزية الحالية غير مكرر، بمعنى أنه عنصر وحيد ، والكلمة الترميزية التالية تشير إلى لون التدرج الرمادي للعنصر المجاور له.

$= 2$ ، فإن هذا يشير إلى أن العنصر الذي تشير إليه الكلمة الترميزية الحالية مكرر مرتين ، فتم استعادة عنصرين من عناصر الصورة ، وتكون الكلمة الترميزية المجاور مشيرة الرمادي للعنصر الثالث من جهة اليمين.

$= 3$ ، فإن هذا يعني أن العنصر الذي تشير إليه الكلمة الترميزية الحالية مكرر لثلاث مرات أو أكثر ، وأن الكلمة الترميزية المجاورة تشير إلى عدد التكرار ، وهي من فئة: عناصر المتعاقبة المتشابهة بالمقدار الذي تحدده الكلمة من الفئة أ ، وهكذا تستمر قراءة الدفق الثنائي حتى تتم استعادة جميع العناصر المشار إليها بقيمة أبعاد الصورة المحفوظة في مقدمة الملف.

5. نتائج الاختبارات العملية

تمت إجراء الاختبارات العملية لقياس أداء تقنية ت. . . " " (7.10.1) ، حيث تمت الاختبارات على مجموعة من الصور القياسية وعددها 20 [12] ، جميعها بالأبعاد 512 512 () ، حيث تم تحويلها إلى صيغة BMP بمعنى أن عدد الألوان في كل صورة هو 256 . ويوضح الشكل رقم: 3 التقنية المقترحة مع أداء كل من ت. . . التقليدية الأولى والتقليدية الثانية ، كما تمت المقارنة مع ترميز هوفمان حل تقييم أداء الترميز الإنتروبي ، حيث يعتبر ترميز هوفمان من أفضل أنواع الترميز الإنتروبي الذي يحقق ضغطاً للبيانات مقارباً للحالة المثالية () ، باستخدام كلمات ترميزية ، وهو جزء من الضغط المعياري جي بيج [3] (مزيد من التفاصيل حول ترميز هوفمان يرجى الرجوع للأدبيات [1-3]).

نظراً لأن الهدف من تقنية ت. . . المانعة للتضخم هو ضغط البيانات أولاً ، ثم منع تضخمها (أي زيادة حجمها) نتيجة لاستخدام تقنية ت. . . العادية ، فقد تمت مقارنة الأداء من هذه النواحي على الصور الاختبارية. بخصوص ضغط البيانات تم قياس () صورة ، ويتم احتسابها بقسمة حجم الصورة الأصلية (ضغطها) على حجم الصورة بعد ضغطها. أما التضخم ، فيتم احتسابه من خلال مراقبة نسبة الإنضغاط ، فإن كانت أكبر من واحد فهذا يدل على أن حجم الصورة الم أصغر من حجمها قبل ضغطها فهذا يعني أن حجم الصورة المرمرزة ازداد نتيجة لتطبيق تقنية الت. . . () حالة الحصول على نسبة إنضغاط مساوية للواحد ، فهذا يدل على أن حجم الصورة الأصلية مساوي لحجم الصورة المضغوطة ، وبالتالي لا يوجد تخفيض لحجم الصورة الأصلية باستخدام تقنية الت. . . بالكيفية المشار إليها 2 4 :

يلخص الجدول رقم: 4 للتقنية المقترحة مع التقنيات المشار إليها. بشكل عام ، يتفوق المعدل العام لنسبة الإنضغاط للخوارزمية المقترحة على جميع التقنيات التقليدية ، حيث تزيد المعدل العام 68% . . . التقليدية الأولى ، 100% . . . التقليدية الثانية ، و 79% الإنضغاط التي يحققها ترميز هوفمان.

تقسيلاً ، 1 13 (65% من الصور الاختبارية) . . . التقليدية الأولى والثانية ، من حيث أداء نسبة ضغط أعلى من واحد ، بمعنى عدم وجود بالتقنيتين التقليديتين إلى زيادة حجم الصور (نسبة إنضغاطها أقل من واحد) وظهور تضخم نتيجة الترميز ، وهذا بشكل ناجحاً بمنع التضخم باستخدام الخوارزمية المشار إليها ب 4:

14 20 (35% من الصور الاختيارية) فلم تنتج الخوارزمية المقترحة بإزالة التضخم ، لسببين ، الأول: يعود لطبيعة تلك الصور ، حيث تحتوي على تباين عالي بين قيم التدرجات الرمادية بين عناصرها مما يمنع ظهور سلاسل طويلة من العناصر المتعاقبة المتشابهة ، ويمكن ملاحظة ذلك التباين على سبيل المثال في الصور () () 3: مقارنة مع تباين أقل ووجود مساحات ناعمة أكثر في الصورة () .

: استخدام كلمات ترميزية طويلة نسبيا ، تتراوح بين 16 18 ثنائية ، لترميز العناصر القليلة الظهور وفقا للإحتمال الرياضي في هيستوجرام تلك الصور ، حيث يوضح الجدول رقم: 5 العدد الأقصى للثنائيات الذي تم استخدامه لكل صور ، ويمكن ملاحظة أن الصور ذوات الأرقام من 14 20 ، تحت عمود الطريقة المقترحة تراوحت أطوالها القصوى بين 16 18 ثنائية ، بين يقل العدد في الصور ذوات الأرقام من 1 13 ، لتتراوح بين 6 14 ثنائية فقط () 6 8 9 حيث أنه رغم استخدام الطول 16 ثنائية كحد كثيرة () . وهذا الطول الكبير

الترميزية المستخدمة لترميز الصور التي عانت من مشكلة تضخم يعود إلى عدم استخدام نمذجة تسبق ضغط تقوم بتصغير عدد الثنائيات المطلوبة لترميز عناصر .

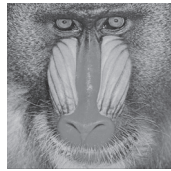
فيما يتعلق بـ فائدة الترميز المقترح مع ترميز هوفمان ، يوضح الجدول رقم: 4 أن الترميز المقترح يتفوق على ترميز هوفمان بتحقيق نسب إنضغاط أعلى من تلك التي يحققها ترميز هوفمان ولجميع الصور 80% الصور الاختيارية () 15 17 ، وأيضا 19 ، حيث تنخفض نسبة الانضغاط للطريقة المقترحة 3% عن تلك التي يحققها ترميز هوفمان لأغلب الصور الأربعة المشار إليها ، باستثناء الصورة رقم: 17 حيث 13% () . ويمكن تفسير ذلك من خلال ملاحظة العدد الأقصى للثنائيات المستخدمة لكلا الترميزين ، والتي يوضحها الجدول رقم: 5 ، حيث يقل هذا العدد لأغلب الصور باستخدام الترميز المقترح بترميز هوفمان. وبشكل عام ، فإن المعدل العام للطول الأقصى للثنائيات في ترميز هوفمان هو 15.75 ثنائية ، بينما ينخفض هذا المعدل عند استخدام الترميز المقترح إلى 14.5 ثنائية ، وهذا يعادل تخفيضا في حجم الكلمات الترميزية بنسبة 8% عن ترميز هوفمان ، مما يدل على فوق الترميز المقترح على ترميز هوفمان من ناحية الطول الأقصى للكلمات الترميزية المستخدمة ، وهو ما ينعكس على نسبة الإنضغاط.

4: نسب إنضغاط الصور بالخوارزمية المقترحة مقارنة مع الطرق التقليدية

الخوارزمية	ترميز هوفمان	الثنائية التقليدية	التقليدية	[12]	
1.13	0.75	0.98	0.96	Truck	1
1.24	0.69	0.98	0.96	tank1	2
1.09	0.75	0.98	0.96	tank2	3
1.07	0.77	0.98	0.96	tank3	4
1.24	0.71	0.98	1.00	Stream and bridge	5
1.02	0.89	0.98	0.96	Splash	6
8.53	0.12	0.98	4.26	Ruler	7
42.67	0.55	0.98	0.98	gray21	8
1.00	0.86	0.98	0.96	f_16	9
1.13	0.76	0.98	0.96	Car and APCs1	10
1.15	0.74	0.98	0.96	Car and APCs2	11
1.33	0.63	0.98	0.96	APC	12
1.89	0.50	0.98	1.04	airplane1	13
0.94	0.87	0.98	0.96	aerial1	14
0.88	0.92	0.98	0.96	Lena	15
0.84	0.92	0.98	0.96	Lake and boat	16
0.89	0.92	0.98	0.98	house1	17
0.90	0.90	0.98	0.96	Boat	18
0.80	0.92	0.98	0.96	Baboon	19
0.94	0.86	0.98	0.98	Tiffany	20
3.534	0.7515	0.98	1.134	Average	

5 : العدد الأقصى للثنائيات في الترميز المقترح مقارنة مع ترميز هوفمان

		ترميز هوفمان (ثنائية)	الترميز المقترح (ثنائية)
1	Truck	18	14
2	tank1	18	14
3	tank2	18	14
4	tank3	18	14
5	Stream and bridge	10	10
6	Splash	13	16
7	Ruler	1	6
8	gray21	5	16
9	f_16	17	16
10	Car and APCs1	18	14
11	Car and APCs2	18	14
12	APC	18	14
13	airplane1	19	14
14	aerial1	17	16
15	Lena	18	16
16	Lake and boat	18	16
17	house1	17	16
18	Boat	17	18
19	baboon	19	16
20	Tiffany	18	16
Average		15.75	14.5



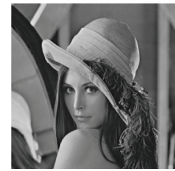
Baboon ()



APC ()



Boat ()



Lena ()

3: نماذج للصور المستخدمة في الإختبارات العملية

6.

يمتاز الترميز الإنتروبي المقترح بالقدرة على تجاوز مشكلة التضخم في الترميز التقليدي بتقنية ت. . .
 3.5% في الصور التي تم إجراء اختبارات عليها ، ودون أي نمذجة مسبقة تؤدي لتخفيض عدد الثنائيات اللازمة لترميز ألوان الصور ، مع تسجيل 35%
 الاختبارية لأسباب تعود لطبيعة تلك الصور. كما يقدم الترميز الإنتروبي حلا لمشكلة الإشارة العَلَمِيّة من خلال دمج الإشارة مع الكلمة الترميزية ، وهو حل يتفوق على طريقة حجز كلمة ترميزية أو حجز ثنائية معينة في الدفق استخدامها كإشارة علمية كما هو الحال في الطرق التقليدية. وهذه الميزة في الترميز المقترح تجعل خوارزمية الضغط قادرة على ترميز عنصرين متتاليين من نفس القيمة بكلمة ترميزية واحدة.
 التكرارات التي يمكن ترميزها في كلمة ترميزية واحدة لانهائي ، على عكس المحدودية في الطرق التقليدية كما هو 2: كما تظهر النتائج العملية تفوق الترميز المقترح على ترميز هوفمان بنسبة 80%
 الصور الإختبارية ، مع تخفيض في العدد الأقصى للثنائيات المستخدمة في الكلمات الترميزية بنسبة 8%. ولتحسين نسبة الإنضغاط بما يحقق انعدام التضخم لجميع الصور ، سنقوم في أبحاثنا القادمة بإدراج نمذجة ما للصورة قبل ضغطها بالخوارزمية المقترحة ، لتحقيق أداء أفضل.

- [1] 1994 "معالجة الصور الطبية"
- [2] Acharya, T. and Tsai, "JPEG2000 Standard for Image Compression", *Wiley-Interscience*.2005
- [3] Roberto Togneri, Christopher J. S. DeSilva, "Fundamentals of information theory and coding design", *CRC Press*, p. 174, 2002.
- [4] Marv Luse, "Bitmapped Graphics Programming in C++", Pp. 664-689, *Addison-Wesley*, 1993.
- [5] David Salomon, "Data compression: the complete reference", 3rd Edition, pp.28,29. *Springer*, 2004.
- [6] Glassner, S. Andrew, "Adaptive Run-Length Encoding", *Graphics Gems II*, Academic Press Inc., 1991
- [7] Run Length Coding, Leadtools Imaging Developer SDKs. [On-Line]: www.pdsite.com/LEAD/SDK/Raster/raster-compression.htm
- [8] Wenfa Zhan El-Maleh, "A new collaborative scheme of test vector compression based on equal-run-length coding". *International Conference on Computer Supported Cooperative Work in Design*, pp. 1155 - 1158 Vol.2 .2009.
- [9] Michael Dipperstein, "Run Length Encoding Discussion and Implementation". 2008. [On-Line]: <http://michael.dipperstein.com/rle/index.html>
- [10] James D. Foley, "Computer graphics: principles and practice", 2nd Edition, P. 845, *Addison-Wesley*, 1995.
- [11] Intergraph Corporation, "Intergraph Raster File Format Guide", DRA 220700, Version 3.2.0. 1994.
- [12] USC-SIPI Image Database, *Signal & Image Processing Institute*, School of Engineering, University of Southern California. [On-Line]: <http://sipi.usc.edu/database/>
- [14] قاموس مؤسسة فيليبس للنشر ، العنوان الإلكتروني:
<http://www.phillips-publishing.com/files/englex.doc>

```

1      Let  $i = 0$ ;
2      Let  $count = 0$ ;
3      Let  $i = i + 1$ ;
4      Read Pixel( $i$ ).
5      If (Pixel( $i$ ) = Pixel( $i-1$ ))
6          {
7              count=count+1;
8              goto 3;
9          }
10     else
11         {
12             If ( $count \geq 3$ )
13                 {
14                     Encode pixels with (A-Class) codeword and set  $y$  to  $S_3$ 
15                     Encode count with (B-Class) codeword
16                 }
17             If ( $count = 2$ )
18                 Encode pixels with (A-Class) codeword and set  $y$  to  $S_2$ 
19             If ( $count = 1$ )
20                 Encode pixels with (A-Class) codeword and set  $y$  to  $S_1$ 
21         }
22     Goto 2 Until scanning all images pixels

```

المخلص باللغة الإنجليزية

Title: A Lossless Image Compression Algorithm Based on an Excess-Free Entropy Code

Abstract: A novel lossless compression algorithm based on an excess-free Run Length Coding (RLC) is presented. A novel entropy coding that encodes intensity levels according to their probability of occurrences is proposed. The codewords of the proposed entropy code are uniquely decodable of variable lengths; and have consolidated flags utilized to distinguish between intensity levels and counts in the bit stream of RLC. Thus, there is no need to dedicate a codeword or a bit (such the MSB) in order to be used as a flag, as in the traditional RLC methods. Consequently, and to the best of our knowledge, our proposed method is the first RLC method that encodes two adjacent pixels that have similar values into a single codeword, with the ability to encode infinite number of count into a single codeword. Results show that the proposed algorithm overcomes the excess problem of the traditional RLC methods in 65% of test images; and outperform Huffman coding in term of the compression ratio in 80% of test images.

Keywords: Lossless compression, entropy code, Huffman code, Run-Length coding

binary permutation	تبديلة ثنائية
bit	ثنائية
bit stream	
codeword	كلمة ترميزية
compression ratio	
decoding	فك الترميز
encoding	ترميز
excess	
flag	إشارة علمية
gray level	
histogram	هستوجرام
image dimension	
JPEG	جي بيج
lossless compression	
lossy compression	ضغط غير محافظ
modeling	
permutation	تبديلة
pixel	
probability of occurrence	الاحتمال الرياضي للظهور
pseudo-code	كود برمجة وهمي
run-length coding	ترميز العناصر المتعاقبة المتشابهة
smooth area	
uniquely decodable	القابلية للكشف بشكل فريد

تقنيات التعليم الإلكتروني وأدواته في خدمة القرآن الكريم بين النظرية والتطبيق

جميل أحمد إطميزي
جامعة فلسطين الأهلية، قسم تقنية المعلومات
ص.ب. 1041، بيت لحم، فلسطين

البريد الإلكتروني: j.itmazi@paluniv.edu.ps

الخلاصة. إن التقدم الكبير في تقنية المعلومات فتح المجال للمؤسسات التعليمية ومنها المؤسسات المعنية بخدمة القرآن الكريم بتوسيع خدماتها إلى العالم. وسمحت تقنيات التعليم الإلكتروني بتلبية الحاجة لوجود بيئات تعليمية تعلم القرآن وعلومه، وتحاكي عمل مراكز تحفيظ القرآن الكريم التقليدية لخدمة المسلمين عرب وغير عرب، في أوطانهم أو خارجها.

ونظرا لحدائث الموضوع، وبعد استقراء تقنيات المواقع التي تقدم هذه الخدمة، لوحظ الضبابية والارتجالية في عملها، وظهرت الحاجة لأبحاث وصفية وتطبيقية حول هذا الأمر.

في هذا البحث تم توضيح ماهية التعليم الإلكتروني، واستعراض مجالات التعليم والتدريب الخاصة بالقرآن الكريم، وتبيان أسباب استخدام تقنيات التعليم الإلكتروني في هذه المجالات وفوائدها، كما عُرض ملخص للتقنيات المستخدمة في معظم المواقع ذات الصلة. وبناءً على أبحاثنا ودراساتنا وخبرتنا، تم اقتراح إطار عمل مرّن يوفر لهذه المراكز إمكانية استعماله حسب خصوصيتها لتقديم خدمة تعليمية وتدريبية متكاملة، حيث وضحت طبيعة أدوات التعليم الإلكتروني وتقنياته اللازمة، تبع ذلك تقديم موجز عن أشهر الأدوات المتوفرة حالياً وأفضلها لتقديم خدمات تعليم القرآن الكريم، وانتهى البحث بالخلاصة.

الكلمات الجوهرية: تقنيات التعليم الإلكتروني، أدوات التعليم الإلكتروني، تعليم القرآن عن بعد، تعليم القرآن عبر الإنترنت.

1 مقدمة

لا يخفى على ذي بصيرة التطور الكبير والتسارع المتلاحق في التقنيات الإلكترونية، وما وصلت إليه من تقدّم، وبخاصة في مجال الحاسوب والاتصالات، وقد استفادت المؤسسات التعليمية من هذه التقنيات. وأهل القرآن مدعوون ليس للاستفادة من هذه التقنيات فحسب بل أيضاً لحل مشكلة الملايين من المسلمين حول العالم المتعطشين لحفظ القرآن وتعلمه وفهم تفسيره، ولا يستطيعون لأسباب عديدة سنذكرها لاحقاً.

فقد أتم الله سبحانه وتعالى نعمته علينا بأن هيا لنا أسباباً، وسخر لنا أشياء، وعلمنا ما لم نعلم، قال تعالى: {عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ (5)} / العلق، ومن هذه الأسباب التي هياها لنا سبحانه العلوم العظيمة التي علمنا إياها؛ التقنيات الحديثة المعاصرة بصورها المختلفة التي أفادت في مجال خدمة القرآن الكريم. وقد تضافرت الأدلة الفقهية من المجتهدين على إباحة استعمال هذه الأدوات والتقنيات في خدمة القرآن وعلومه، بل هي باب خير فتح على مصراعيه لأهل القرآن.

وغالباً ما تتمثل معيقات تطويع هذه الوسائل والتقنيات في قلة المعرفة بإمكاناتها، أما بقية المعوقات ومنها المالية فأزعم أنها أسهل لتوفر أهل الخير في طول الأرض وعرضها، وقد جاء البحث ليجيب على سؤال أساس، كيف يمكننا الارتقاء بخدمة القرآن الكريم عبر تقنيات التعليم الإلكتروني؟ ومجال البحث يدور حول هذه التقنيات والأساليب والوسائل التي يمكنها تقديم خدمة تعليمية متكاملة تحاكي عمل مراكز القرآن، بل وتقدم خدمات متنوعة قد يصعب توفيرها بالطرق التقليدية.

2 القرآن الكريم والتقنيات الحديثة

2.1 فضل تعليم القرآن وتعلمه

يعلم كل مسلم عظم فضل خدمة أجل كتاب في الكون ألا وهو القرآن الكريم، فقد أنعم الله تعالى علينا أنه أنزله وجعله هادياً ومبشراً ومنذراً ومذكراً، وتكفل بحفظه منذ نزوله إلى يوم الدين، فابتداءً جعله سبحانه في اللوح المحفوظ: {بَلْ هُوَ قُرْآنٌ مَّجِيدٌ (21) فِي لَوْحٍ مَحْفُوظٍ (22)} / البروج، وتكفل سبحانه بحفظه إلى يوم الدين: {إِنَّا نَحْنُ نَزَّلْنَا الذِّكْرَ وَإِنَّا لَهُ لَحَافِظُونَ (9)} / الحجر.

وقد بين النبي عليه السلام علو منزلة قارئ القرآن في الجنة، حيث قال: (يقال لقارئ القرآن يوم القيامة اقرأ وارق ورتل كما كنت ترتل في الدنيا، فإن منزلتك عند آخر آية تقرؤها) [1]، وقال: (مثل الذي يقرأ القرآن وهو حافظ له مع السفرة الكرام البررة، ومثل الذي يقرأ القرآن وهو يتعاهده، وهو عليه شديد، فله أجران) [2]. وحث صلى الله عليه وسلم على تعلم القرآن وعلى تعليمه حين قال: (خيركم من تعلم القرآن وعلمه) [3]، وقال في أهل القرآن: (إِنَّ لِلَّهِ أَهْلِينَ مِنَ النَّاسِ. قَالُوا: يَا رَسُولَ اللَّهِ مَنْ هُمْ؟ قَالَ: هُمُ أَهْلُ الْقُرْآنِ، أَهْلُ اللَّهِ وَخَاصَّتُهُ) [4].

2.2 مجالات الخدمة

قبل اقتراح إطار عمل لطبيعة أدوات التعليم الإلكتروني التي ترتقي من شأن تدريس القرآن الكريم وتحفيظه وتقديم علومه، علينا استقراء مجالات التعليم والتدريب الخاصة بالقرآن. أن أبرز صور الخدمة هي التقنيات الحديثة المعاصرة بصورها المختلفة تتمثل في ثلاثة أمور [5]:

أولاً: حفظ القرآن وترتيبه: تعددت الوسائل التي تعين المرء على تلاوة القرآن في أي وقت شاء ليلاً أو نهاراً، وفي أي مكان كان مسكناً أو مركباً أو غيره، وعلى أية حال كان قائماً أو قاعداً.

ثانياً: تجويد القرآن وترتيبه: لقد أفادت هذه الوسائل في جانب تجويد القرآن وترتيبه إفادة عظيمة؛ عن طريق نشر القراءة الصحيحة السليمة، والمجودة المرتلة، وبث الدورات التجويدية، وإمكانية الالتحاق بها، وأخذ الإجازات القرآنية عن طريق التعليم عن بعد بكل سهولة ويسر.

ثالثاً: تفسير القرآن وتدبره: أثمرت هذه التقنيات بفضل الله علماً واسعاً في تفسير القرآن وتدبره، ونشره، مما أفاد وغير كثيراً في جوانب حياة فئات المجتمع في شتى أنحاء العالم.

رابعاً: تعليم العربية للناطقين بغيرها: ويعود ذلك إلى أهمية تعلم العربية حيث أشارت إحصائية أن 86% من الأسباب الرئيسة وراء تعلم العربية هي الرغبة في قراءة القرآن الكريم وتعلمه [6].

2.3 الأحكام الشرعية في استخدام التقنيات المعاصرة

نكتفي ببيان أن الحكم الشرعي العام في استخدام تقنية المعلومات في تيسير تعلم القرآن الكريم وتعليمه هو الجواز، حيث القاعدة الشرعية تنص على أن "الوسائل لها أحكام المقاصد" [7، 8].

2.4 دوافع استخدام التعليم الإلكتروني في خدمة القرآن

يتشوق أي مسلم لحفظ القرآن الكريم، وإتقان أحكام تجويده وترتيبه، ودراسة علومه، ولكن واقع الحال إن هناك معوقات قد تحول دون القيام بذلك في المسجد أو في مقر مركز تحفيظ القرآن الكريم. إن الطريقة التقليدية - في تعليم القرآن - لم تشمل "أولئك الذين منعهم المرض أو كبر السن من الحركة والتنقل، ولم تشمل من حال بينهم وبين هدفهم بعد مسافة أو قلة ذات اليد. لكن ومع التطور الهائل في مجال التعليم الإلكتروني، أصبح من السهولة تخطي تلك الحواجز، وكسر تلك العقبات، فأوجدت بيئة تعليمية متكاملة عبر مواقع عربية عدة ساهمت في تعليم القرآن الكريم" [9].

ويمكننا استقراء الأوضاع التي يستطيع فيها التعليم الإلكتروني خدمة القرآن الكريم:

- (1) المرض وكبر السن ممن يتعذر معه الحركة والتنقل من مكان إلى مكان.
- (2) بعد مسافة بين المتعلم ومراكز التعليم.
- (3) قلة ذات اليد والتي تحول دون الانتقال إلى المركز التعليمي أو العيش بقربه.
- (4) الأماكن التي لا يتوفر فيها مثل هذه المؤسسات مثل بعض البلاد سواء بلاد عربية أو عجمية.
- (5) تعارض أوقات طلب الرزق أو الدراسة مع أوقات المؤسسات ذات العلاقة.

3 التعليم الإلكتروني

3.1 تعريف التعليم الإلكتروني

عُرِّف بأنه "استخدام تقنيات الوسائط المتعددة الجديدة، والإنترنت لتحسين جودة التعليم عن طريق تسهيل الوصول للمصادر، والخدمات إضافة إلى التعاون والتبادل عن بعد" [10]، وعُرِّف أيضاً بأنه "أي استخدام لتقنية الويب، والإنترنت لإحداث التعلم" [11].

وعموماً نميل إلى اعتباره مجموعة من الأنشطة التي تغطي التعلم والتعليم التي يمكن أن تتم في أي زمان ومكان، على جهاز حاسوب موصول عموماً بشبكة.

3.2 أساليب تقديم التعليم الإلكتروني

- 1) التعليم الإلكتروني المتزامن: يصف النشاط التعليمي الذي يتم في الوقت الحقيقي، تحت قيادة المدرب حيث يتواجد هو وجميع الدارسين في نفس الوقت ويتواصلون مباشرة مع بعضهم البعض، ولكن ليس تواجداً فيزيائياً بنفس المكان [12]. وهذا الأسلوب يضمن المرونة بالمكان وليس بالزمان مثل المؤتمرات المرئية، والردشة، واللوحة الإلكترونية.
 - 2) التعليم الإلكتروني غير المتزامن: حيث لا يشترط تواجد المدرب والطلبة في الوقت نفسه أو المكان نفسه. ومن الأمثلة عليه: استخدام البريد الإلكتروني، ومنتديات النقاش، ووضع المحتويات التعليمية والمحاضرات على موقع التعليم الإلكتروني.
- وبالعموم، فإن كلا أسلوب التعليم الإلكتروني، له حسناته وسيئاته، وتقديم خدمة تحفيظ القرآن وتدريب علومه تقتضي الاستعانة بكليهما.

3.3 مزايا التعليم الإلكتروني

تمثل فوائد التعليم الإلكتروني المتزامن وغير المتزامن مقارنة بأسلوب التعليم التقليدي في مجال خدمة القرآن الكريم تحديداً بالآتي:

■ المرونة والملائمة:

1. سرعة الوصول في أي وقت وأي مكان.
3. تغذية راجعة فورية عند استخدام واجبات وامتحانات وتمارين مباشرة على الإنترنت.
4. سهولة الوصول إلى المكونات التعليمية، وسرعة مراجعتها وتحديثها وتحريرها وتوزيعها.
5. متكيف حسب الدارس، فيمكن مثلاً له أن يدرس حسب قدرته (بسرعة أو ببطء).
6. يقدم تسهيلات وأساليب تعليمية متنوعة تمنع الملل.
8. سهولة وصول الآلاف للمصادر الإلكترونية في الوقت نفسه بخلاف المصادر الورقية.

■ الوقت والمال:

1. توفير الوقت وحسن استغلاله لأنه مثلاً يتيح للدارس القفز عن مواد ونشاطات يعرفها.
2. تنظيم الوقت بحيث يجدول الدارس دروسه حسب عمله وعائلته.
3. تخفيض تكلفة السفر والتنقل والمعيشة.
4. تخفيض تكلفة الإنتاج والتوزيع للمواد التعليمية.
5. تخفيض تكلفة المكاتب والمدرسين.

■ الاتصالات والتفاعل:

1. إمكانية الاتصال والتفاعل بين الطلبة والمدرّب عبر دروس مباشرة على الإنترنت.
2. إمكانية الدراسة في أي مكان يتوافر فيه جهاز حاسوب وإنترنت.
4. يتوفر وسائل تعليمية عديدة تخدم التعليم التعاوني.

4.1 استقرار الواقع الحالي

قام الباحث بعملية بحث شاملة عبر الشبكة العنكبوتية عن الوسائل التي تستخدمها مواقع تعليم القرآن الكريم، حيث تم حصر قائمة من 147 موقعا كما في الملحق 1، مع مراعاة الآتي:

- تم إدراج المواقع التي تتضمن نوعا من التعليم عبر الإنترنت، وليس مواقع خدمة القرآن الخاصة بقراءته والبحث فيه.
- المواقع المشمولة هي مواقع تحتوي - ولو جزئيا- على خدمة التعليم المتزامن أي التعليم الذي يلقي فيه دارسون مع معلمهم بشكل مباشر وفي الوقت الحقيقي عبر الإنترنت.
- الباحث لم يقم بفحص صدقية هذه المواقع أو جودتها، فهذا يحتاج إلى دراسة أخرى.
- لوظ تشابه وتكرار في بعض المواقع، مما قد يوشر على أن بعضها قد يكون لأغراض تجارية بحتة وربما يكون عدد منها مواقع مخادعة.
- لا يستطيع الباحث التأكيد أن هذه القائمة هي كل مواقع تعليم القرآن الكريم التي تقدم تعليما تزامنيا، رغم بذل الوسع في تحصيلها حتى منتصف شهر شباط / فبراير للعام 2011م.

4.2 الأدوات البرمجية المستخدمة حاليا

تم استقراء الأدوات التي استخدمت في المواقع الـ 147، والجدول رقم 1 يعرض نسبة استخدام كل أداة، وتم إدراجها حسب أكثرها استخداماً.

جدول 1: نسب استخدام أدوات التواصل في مواقع تعليم القرآن الكريم

اسم الأداة - عربي	اسم الأداة - انجليزي	النسبة المئوية	عدد المواقع التي تستخدمها
سكاى بى	Skype	79%	116
جو تو ميٲينج	Gotomeeting	34%	50
ميكاجو	Mikogo	16%	24
ٲيم فيور	Teamviewer	14%	20
ياهو ماسنجر	Yahoo Messenger	4%	6
مسن/ليف ماسنجر	MSN/Hotmail Messenger	4%	6
جوجل ٲوك	Google Talk	2%	3
ألفور شات	All2chat	2%	3
يوجو	Yuuguu	1%	1
ويندوز نت ميٲينج	Windows Netmeetings	1%	1
الشبكة الافتراضية الخاصة	Virtual Private Network (VPN)	1%	1
البالتوك	Paltalk	1%	1
إنسبيك	Inspeak Messenger	1%	1

علماء أن 23 موقعاً منها (16%) لم تحدد الوسائل التزمانية المستخدمة.

4.3 رُخص الأدوات البرمجية

أغلب البرمجيات على نوعين من الرخص، وهما:

- البرمجيات المفتوحة المصدر، وهي تسمح للمستخدمين باستعمالها وتوزيعها، كما وتسمح برؤية الكود البرمجي (البرنامج الأساس) وتعديله، وتوزيعه، كل ذلك مقابل الحفاظ على حقوق الاسم والماركة [13].
- البرمجيات التجارية أو الاحتكارية، وهي برمجيات تضع شروطاً على الاستخدام، ولا تسمح بتوزيع البرمجية أو تعديلها، بل في الغالب لا تسمح برؤية الكود البرمجي.
- وقد غدت برمجيات التعليم مفتوحة المصدر هي مجال صاعد واعد، وملائم، ومفضل للتعليم.

4.4 نبذة عن الأدوات المستخدمة

أولاً برمجية سكاى.بى (www.skype.com): برنامج اتصال يوفر نسخاً مجانية للمهام الأساسية، ويقدم الخدمات الآتية:

- اتصالاً صوتياً واتصالاً مرئياً بين المشاركين فيه عبر الإنترنت.
- رسائل فورية مع إمكانية إرسال صور وملفات متنوعة أثناء المحادثة.
- مشاركة الشاشة.

ثانياً برمجية جوتوميٲينج (www.gotomeeting.com): برنامج يتطلب الدفع حسب عدد المشتركين، ويسمح بتجربته لمدة شهر مجاناً، ويقدم خدمة اللقاء المرئي وخدمات اللقاءات الإلكترونية المباشرة، ومن ميزاته:

- إمكانية تسجيل الجلسة صوتاً وصورة، بحيث يمكنك الإطلاع عليها لاحقاً.
- نافذة دردشة كتابية مع إمكانية المشاركة بالصوت.

- المدرب يمكنه عرض شرائح من جهازه يستطيع أن يراها المشاركون من أجهزتهم.

ثالثاً برمجية ميكاجو (www.mikogo.com): يقدم هذا البرنامج المجاني مجموعة من الخدمات لعمل الاجتماعات على الإنترنت ومنها الاجتماعات المرئية وكذلك عمل عروض ومحاضرات، والمناقشة وتبادل الصور، وتشمل جلسات هذا البرنامج الآتي:

- تقاسم الشاشة مع عدة مشاركين.

- تحكم بلوحة المفاتيح والفأرة عن بعد.

- تنظيم الاجتماعات عبر الإنترنت.

- خدمة نقل الملفات والصور.

رابعاً برمجية تيم.فيور (www.teamviewer.com): هو برنامج يوفر نسخاً مجانية للمهام الأساسية ويقدم خدمة الاتصال بأي حاسوب حول العالم بحيث يتسنى التحكم عن بعد في حاسوب الطرف الآخر، وعموماً يمكنه تقديم الخدمات الآتية:

- الدخول عن بعد، ونقل الملفات

- الاجتماعات، والأعمال الجماعية، والتدريب، والعروض عبر الإنترنت.

- إجراء محادثة كتابية مع الطرف الآخر.

خامسا) برمجيات المسنجرات: وهي برامج للتواصل مشهورة ومعروفة ومجانية، وهذه الأدوات تتشابه وتشتمل في حدها الأدنى ميزات مثل الدردشة الكتابية، واللقاء الصوتي والمرئي، ونقل الملفات والصور، ومنها:

- برمجية ياهو ماسنجر (<http://messenger.yahoo.com>):

- برمجية مسن/ليف ماسنجر (<http://explore.live.com>):

- برمجية جوجل توك (www.google.com/talk):

- برمجية إنسيك (<http://messenger.yahoo.com>):

سادسا) برمجية ألفورشات (<http://all2chat.com>): برنامج متخصص بالدردشة يتطلب الدفع حسب عدد المشتركين، ويسمح بتجربته لمدة شهر مجانا، ويقدم خدمة اللقاء المرئي والصوتي المباشرة، ومن ميزاته:

- سهولة الاتصال، والسماح للمشاركة بالدخول لأكثر من غرفة بنفس الوقت من نفس الحاسوب.

- الدردشة صوتا وصورة.

- مستويات تحكم متعددة.

سابعا) برمجية يوجو (www.yuuguu.com): برنامج متخصص بالمحادثة يتطلب الدفع حسب عدد المشتركين، ويسمح بتجربته لمدة شهر مجانا، ويقدم خدمة اللقاء المرئي وخدمات اللقاءات الإلكترونية المباشرة، ومن ميزاته:

- خدمات اللقاءات الإلكترونية المباشرة ومنها خدمة اللقاء المرئي.

- مشاركة الشاشة.

- عروض وتدريب مباشر.

- خدمة الدخول إلى أجهزة الحاسوب عن بعد.

ثامنا) ويندوز نت.ميتينج (www.microsoft.com): برنامج اتصال وتواصل مجاني باستثناء خدمة الاتصال الهاتفي، وهو خاص بنظام ويندوز، ولكنه لم يعد يدعم بعد نظام التشغيل ويندوز اكس بي، ومن ميزاته:

- خدمات اللقاءات الإلكترونية المباشرة المرئية والصوتية، والدردشة الكتابية.

- مشاركة الشاشة ومشاركة الملفات.

- يمكن لكافة المشاركين في الاجتماع العمل على مستند في وقت واحد.

- تسهيل إجراء المكالمات باستخدام دفاتر العناوين من على حاسوب المشارك.

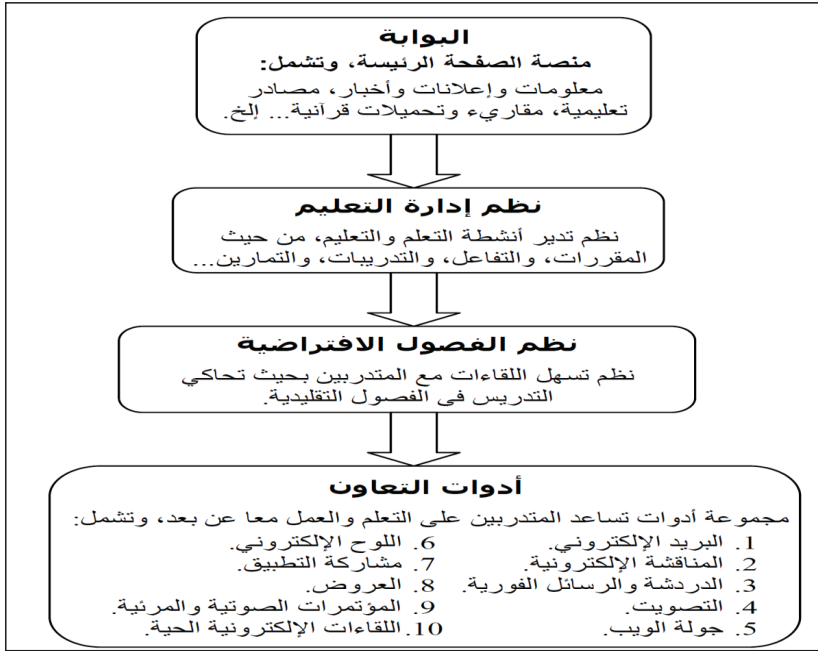
تاسعا) خدمة البالتوك (www.paltalk.com): هو برنامج اتصال يقدم خدمة المحادثة الفورية على الإنترنت، بالصوت والصورة والكتابة مع مستعملين آخرين لنفس البرنامج، حيث تُقسم الأماكن إلى غرف، ويوجد منه نسختان نسخة مجانية ونسخة مدفوعة القيمة، ويستخدمه البعض في التعليم والتدريس.

عاشرا) أداة الشبكة الافتراضية الخاصة: هي شبكة ظاهرية بين المشارك وبين مقدم الخدمة (بوابة اتصال)، ورغم أنها افتراضية لكنها تؤدي واجبها على أكمل وجه كأكثر أنواع الشبكات أمانا وأكثرها شيوعا، وتتم عبر استخدام الشبكة العنكبوتية بحيث يمكن تصور الشبكات الافتراضية وكأنها نفسها الشبكة العنكبوتية لكن تم توظيف خصائصها لتلائم سرية نقل البيانات والحفاظ على أمن المعلومات، ومن ميزاتها توفير الكثير من المال خاصة في تكاليف الأجهزة [14].

5.1 تقديم

إيجاد مؤسسة أو مركز يقدم خدمات تعليمية وتدريبية عبر الإنترنت باستخدام حلول التعليم الإلكتروني خاصة بتدريس القرآن الكريم وتحفيظه وتقديم علومه بحيث تحاكي بل وتتفوق على تقديم هذه الخدمات بشكلها التقليدي، يقتضي استخدام أحدث ما هو متوفر من التقنيات البرمجية التي تشمل الأدوات التزامنية والأدوات غير التزامنية.

وسنقدم إطار عمل عام (Framework) يوضح طبيعة الأدوات البرمجية اللازمة.



الشكل 1: إطار عمل لموقع ويب يقدم خدمة متكاملة في تعليم القرآن الكريم

5.2 أقسام الإطار

الإطار المقترح في الشكل 1 ينقسم إلى الطبقات الآتية:

- الطبقة الأولى: البوابة (منصة الصفحة الرئيسية).
- الطبقة الثانية: نظم إدارة التعليم.
- الطبقة الثالثة: نظم الفصول الافتراضية.
- الطبقة الرابعة: أدوات التعاون.

5.3 الطبقة الأولى: البوابة

وهي منصة الصفحة الرئيسية، والتي تشمل التعامل مع جميع ما ترغب المؤسسات بعرضه كصفحة رئيسة أو بوابة مثل إدارة شؤون المسجلين على عدة مستويات، وإدارة شؤون الإعلانات والأخبار، ونشر المعلومات، والتعامل مع الملفات النصية والرسومية والوسائط المتعددة ... وغيرها الكثير. ويمكن للمؤسسة برمجة هذه الصفحة الرئيسية أو استخدام إحدى برمجيات إدارة المحتويات الجاهزة (Content Management System).

5.4 الطبقة الثانية: نظم إدارة التعليم

أنظمة إدارة التعليم هي برمجيات توفرت إدارة نشاطات التعلم والتعليم، من حيث المقررات، والتفاعل، والتدريبات، والتمارين... الخ. [15]، وتستخدم من قبل المؤسسات التعليمية والتدريبية كأحد أهم حلول التعليم الإلكتروني الأساسية. وعموما توجد مفاهيم قريبة لهذا المفهوم مع اختلافات يسيرة، ومنها: أنظمة إدارة المقررات، وأنظمة إدارة التعليم، وأنظمة إدارة محتويات التعليم، ومنصات التعليم الإلكتروني. ومن أهم مهمات هذه الأنظمة الآتي:

- إدارة المقررات والفصول والدورات.
 - إدارة تسجيل المستخدمين واتصالهما، ومتابعة دخول الطلبة ونشاطاتهم ونتائج امتحاناتهم وتمارينهم.
 - أدوات اتصال وتواصل مثل: المنتديات، والدردشة، والاقتراحات.
 - تقارير متنوعة للإدارة.
 - أدوات تأليف المحتوى، وأدوات إضافة الأنشطة والمصادر وإدارتها.
- ولا ينصح ببرمجة أنظمة جديدة لكلفتها العالية، لأنه يتوفر في الوقت الحاضر بدائل جاهزة وبجودة عالية.

5.5 الطبقة الثالثة: نظم الفصول الافتراضية

وهي أنظمة تسمح بالتعليم الإلكتروني الميسر والموجه بالمعلم، وهي تجمع بين قدرات إدارة التعليم، وإدارة المحتويات، ونظم التعاون. ونظم الفصول الافتراضية تميل إلى أن تكون موجهة إلى المقرر في حين إن أدوات اللقاءات الإلكترونية تميل إلى أن تكون مجموعة من أدوات التعاون المعزز للتعليم الإلكتروني، على الرغم من أن نظم الفصول الافتراضية يمكنها إجراء لقاءات حية [16].

وهذه النظم توفر بيئة تعليمية افتراضية تحاكي الفصول التقليدية فيما قد يسمى الحلقات الإلكترونية حيث توفر وسائل تتيح للمتدرب سرعة التواصل والتفاعل ومتابعة الحفظ مع المدرب، ويلتقي الدارسون في الزمن الحقيقي مع مدربيهم عن بعد، مستعينين بأدوات تواصل كتابية وصوتية، وأحيانا باللقاء المرئي لكل مشارك، وتوفر أدوات عديدة تسهل مهمة عرض الدروس والشرحات

مثل السبورة الإلكترونية، كما وتسهل كذلك التفاعل والمشاركة المتنوعة، حيث تشمل العديد من أدوات التعاون التي سيأتي شرحا لاحقا.

وهذه النظم غالبا ما توضع وصلاتها ضمن مقرر موجود على نظم إدارة التعليم السابق شرحها، بما في ذلك مواعيد الجلسات وأي أخبار ذات العلاقة.

5.6 الطبقة الرابعة: أدوات التعاون

هي مجموعة من الأدوات التي تساعد الدارسين على التعلم والعمل معا عن بعد، حيث تسمح لهم بتبادل الأفكار ومشاركتها، وكذلك الاقتراحات وحتى الشكاوي والهموم. وهذه الفئة تغطي مجموعة واسعة من الأدوات، من نص بسيط في البريد الإلكتروني إلى أدوات التواصل الإلكترونية الحية المعقدة، وتوفير بيئة تعاونية كاملة قد يحتاج إلى الجمع بين العديد من الأدوات والتقنيات المنفصلة.

ومعظم أدوات التعاون تعمل بنفس الآلية حيث يكتب شخص ما رسالة في أداة تعاون، ومن ثم ترسل إلى مستخدمين آخرين، والرسالة قد تكون أي شيء تقريبا، مثلا قد تكون نص بريد إلكتروني، أو ردا على دردشة، أو رسماً وكتابة على اللوح الإلكتروني، أو جملة في مؤتمر صوتي، أو مقطع فيديو في مؤتمر مرئي.

أدوات التعاون الموجودة في الطبقة الثانية والثالثة

الحقيقة إن النظم الجيدة من نظم إدارة التعليم ونظم الفصول الافتراضية تتضمن الكثير من هذه الأدوات، ولكنها تتفاوت من برمجية إلى أخرى، ولهذا وضعنا الطبعة الرابعة والتي يمكن للمؤسسة أن تشمل فيها بقية أدوات التعاون غير المتوفرة في هاتين الطبقتين.

قائمة بأشهر أنواع أدوات التعاون

- **البريد الإلكتروني:** هو أقدم نوع من أدوات تعاون، وبالنسبة لكثير من المهام، لا يزال الأكثر فعالية، وهي خدمة بسيطة، وموثوقة ومألوفة وغير مكلفة، ومنتشرة في كل مكان، وأي شخص يمكنه استخدام تقنية الحاسوب، يمكنه استخدام البريد الإلكتروني.
- **المناقشة الإلكترونية:** وتسمى أيضا منتديات النقاش، وهي تجتذب الأفراد المتباعدين ذوي الاهتمامات المشتركة، وذلك لتبادل الأفكار وتركز معظم منتديات النقاش على كتابة مواضيع ومشاركات وردود بحيث يمكن عرضها في شكل يسهل اتباعها. وكثيرا ما يؤكد الدارسون أن المناقشة هي أفضل جزء في التدريب، وليس غريبا بقاء المناقشات أشهراً بعد نهاية الدورة.
- **الدردشة:** هي رسائل نصية فورية تقدم وسيلة فورية وعفوية لتبادل الرسائل، ويمكن أن تستخدم في النقاشات التي تجري في الوقت الحقيقي من حول موضوع التدريب.
- **التصويت:** وهي مجموعة برمجيات تقدم خدمات التقييم أو الاقتراح أو التصنيف أو المسح ومنها الاستبانات الإلكترونية، وهو يتضمن وضع أسئلة ومن ثم مشاهدة ردود المشاركين

عليها. وآلية عمله بسيطة جداً، فأولاً، يوجد أمام الدارسين أسئلة مع قائمة من الإجابات لكل سؤال، فيقوم الدارسون بالتصويت (الاختيار)، ومن ثم يتم جدولة الإجابات وتعرض الإحصائيات بشكل بياني أو رقمي، ويمكن تخزينها في قاعدة بيانات، لتحليلها في وقت لاحق.

- **جولة الويب:** وهي نتيج للمدرب الإبحار في شبكة الويب بينما المشاركون يتابعونه من على متصفحات الويب، وتدعى هذه الميزة التصفح المتزامن للويب أو مشاركة تصفح الويب.

- **اللوحة الإلكترونية:** هو أداة تعاون تحاكي الاتصال الذي يحدث عندما يرسم المدرب على لوح أبيض معلق على الجدار، ثم يدعو أحد الدارسين للمساهمة في الرسم، فإذا رفع المدرب رسماً للنقاش، فإنه يبيث لجميع المشاركين بحيث يراه الجميع على نافذة (لوحة أبيض) على حاسوب كل مشارك. وحالما يضيف المشارك أية ملاحظة أو يعدل الرسم فإنها تثبت فوراً إلى بقية المشاركين.

- **مشاركة التطبيق:** وهو يسهل وصول اثنين أو أكثر من المستخدمين إلى تطبيق أو وثيقة مشتركة من أجهزة الحاسوب الخاصة بهم في وقت واحد في الوقت الحقيقي. وهو يتيح للمدرب مشاركة (وتقاسم) البرامج، والنوافذ، أو كامل الشاشة مع المشاركين الذين يرون بالضبط ما يظهر على شاشة المدرب. وغالباً ما يستخدم لتعليم برامج الحاسوب وشرحها.

- **العروض:** أصبحت الشرائح الروائية التي يوجد فيها نصوص وصور وأحياناً صوت وفيديو... إلخ تشكل عنصراً أساسياً في التعليم لدرجة أن معظم الفصول الدراسية التقليدية وأدوات اللقاءات الإلكترونية الحية تشمل سبلاً لنشر العروض الحية عبر الشبكة العالمية.

- **المؤتمرات الصوتية:** تتيح للمشاركين الحديث مع بعضهم البعض، وهي ميزة توجد في حزم برمجيات التعاون، والقدرات السمعية مفيدة لأن الاتصالات اللفظية أو سرعة تبادل الأفكار هي أمر هام عندما لا يكون هناك وقت كاف لكتابة كل شيء، أو كانت هناك أهمية لعفوية التعبير عن الفكر.

- **المؤتمرات المرئية (الفديو كونفرنس):** المؤتمرات المرئية تتيح للمشاركين المشاهدة والاستماع لبعضهم البعض، ويسمح بسماع الصوت، ورؤية إشارات اليد، وتعابير الوجه، والمشاركون يستطيعون رؤية الشخص المتحدث كشخص، ورؤية الأشياء ثلاثية الأبعاد، والحركات والإيماءات. وتستخدم غالباً لتقديم العروض بحيث يكون هنالك قنوات خلفية للدردشة أو للصوت تقوم بنقل أسئلة الدارسين وتعليقاتهم، وبهذه الكيفية فإن المؤتمرات المرئية مفيدة في التعريف بالمدرب، وعرض الأشكال ثلاثية الأبعاد، وإضافة الواقعية والتأثير والتواصل البشري الانفعالي من خلال تعابير الوجه، والإيماءات، ونبرة الصوت... إلخ.

- **أدوات اللقاءات الإلكترونية الحية:** وهي تقدم أنواع متعددة من الوسائل التي تسمح باللقاءات والاجتماعات عن بعد.

6 نموذج تطبيقي

6.1 تقديم

يوفر هذا النموذج التطبيقي نوعاً من الإرشاد للمؤسسات أو المراكز التي ترغب في تقديم خدمات تعليمية وتدريبية خاصة بتدريس القرآن الكريم وتحفيظه وتقديم علومه عبر الإنترنت باستخدام حلول التعليم الإلكتروني حيث قدمنا نبذة عن أحدث ما هو متوفر من التقنيات البرمجية التي تشمل الأدوات التزامنية والأدوات غير التزامنية، ونبذة عن بعض استخداماتها في تدريس القرآن الكريم وتحفيظه وتقديم علومه.

6.2 برمجيات البوابة

يمكن برمجة منصة الصفحة الرئيسية، ولكن يُنصح باستخدام إحدى برمجيات إدارة المحتويات الجاهزة (Content Management System)، حيث يتوفر منها أنظمة مشهورة لنشر المحتويات على الشبكة العنكبوتية العالمية، ويمكن استعمالها كبوابة لأية مؤسسة ترغب في موقع إنترنت متسق ومتحكم فيه، ويضمن قوالب عديدة لملائمة تنوع المؤسسات.

■ وقد حازت بعض أنظمة إدارة المحتويات ذات الرخصة المفتوحة على جوائز كأفضل أنظمة في هذا المجال، ومنها الأنظمة الآتية حسب موقع التصنيف الشهير توب.تن. ريفور (17):

- منصة جملة (www.joomla.org).

- منصة ووردبريس (http://wordpress.org).

- منصة دروبال (www.drupal.org).

وجميع هذه الأنظمة ذات رخصة حرة مفتوحة المصدر بحيث تسمح بإمكانية تعديلها وتكييفها حسب رغبة مدير الموقع، وهي تدعم عدة لغات ومنها العربية.

■ البوابة المستخدمة لتقديم خدمات تعليمية وتدريبية خاصة بتدريس القرآن الكريم وتحفيظه وتقديم علومه قد تشمل الآتي:

- معلومات عن المؤسسة.

- إعلانات وأخبار عامة.

- المصادر التعليمية العامة مثل مصادر للتفسير، وأحكام التجويد ... إلخ.

- المقاريء الإلكترونية، بحيث يتاح لزائر الموقع الاستماع إلى أي مقطع من سور القرآن الكريم، وذلك عن طريق تحديد رقم المقطع واسم القارئ ثم الاستماع.

- تحميلات قرآنية عامة: مثل صور وفلاشات قرآنية... إلخ.

6.3 برمجيات نظم إدارة التعليم

تعتبر هذه الأنظمة العمود الفقري للتعليم الإلكتروني عبر الإنترنت حيث يمكنها إدارة المكونات التعليمية (الأنشطة والمصادر)، وتسجيل الدارسين، وتنسيق المدربين، وتحصيل الرسوم، وإصدار التقارير، كما تشمل متابعة دخول المقررات، والدروس، والمكونات، وتسجيل العلامات، وتخزين الدروس الصوتية المسجلة لمن يرغب بإعادتها، إضافة إلى كونها مرجعاً ومنطلق البرمجيات الأخرى الموجودة في الطبقة الثالثة والرابعة.

ولا ينصح ببرمجة أنظمة جديدة لكلفتها العالية، ولتوفر بدائل جاهزة وبجودة عالية. ففي الوقت الحاضر، هناك عدد كبير من أنظمة إدارة التعليم، حيث يوجد أكثر من 70 مؤسسة تنتج هذه الأنظمة [18]، وهذه الحُزم منها برمجيات تجارية (احتكارية)، ومنها برمجيات وأنظمة مشهورة ذات رخصة حرة مفتوحة المصدر مما يعني أنها مجانية وتسمح بتعديلها وتكييفها حسب رغبة المؤسسة.

■ ومن أشهر الأنظمة الجاهزة:

- نظام مودل مفتوح المصدر (moodle.org).

- نظام بلاكبودر غير مفتوح المصدر (www.blackboard.com).

وهذه الأنظمة تتميز بانتشارها الواسع، وتنظيم تخزين الأنشطة والمصادر التعليمية ضمن مقررات، وبإمكانية تطوير أنشطة تعليمية، وإدارة شؤون الدارسين والامتحانات، وتتضمن بعض الأدوات التعاونية مثل الدردشة ومنتديات نقاش والتصويبات، وهذه الأنظمة تدعم عدة لغات ومنها العربية.

■ عموماً، لا فرق في برمجيات نظم إدارة التعليم بين تقديم خدمات تعليمية وتدريبية في مجال تدريس القرآن الكريم وتحفيظه وتقديم علومه أو في بقية المجالات.

6.4 برمجيات نظم الفصول الافتراضية

■ هذه النظم توفر بيئة تعليمية افتراضية وتحاكي الفصول التقليدية وتوفر لمركز تقديم الخدمات التعليمية والتدريبية الخاصة بتدريس القرآن الكريم القيام بالآتي:

- اللقاء الدارسين في الزمن الحقيقي مع مدربيهم عن بعد في الحلقات الإلكترونية.

- سرعة تواصل الدارسين وتفاعلهم ومتابعة الحفظ مع المدرب.

- الاستعانة بأدوات تواصل كتابية وصوتية، وأحياناً باللقاء المرئي لكل مشارك تفيد مثلاً في إظهار كيفية أداء التجويد أو مخارج الأحرف العربية.

- الاستعانة بأدوات عديدة تسهل مهمة عرض الدروس والشروحات مثل السبورة الإلكترونية لشرح ما أشكل على الدارس فهمه من حكم تجويدي مثلاً أو معنى كلمة معينة وعرض صفحات من القرآن أثناء التلاوة والشرح.

- الاستعانة بأدوات تعاون عديدة تسهل التفاعل والمشاركة المتنوعة.

■ ومن البرمجيات الجاهزة:

- نظام الومنييت (www.illuminate.com): وهي من أشهر الأدوات في هذه الشأن، وهي تحتوي العديد من الأدوات التعاونية مثل اللقاءات المرئية والصوتية والردشة الكتابية واللوحة الإلكترونية ومشاركة الشاشة والعديد من الميزات الرائعة، كما يمكن تخزين كامل الجلسة، ومن ثم يستطيع الدارس أن يتابعها بعد انتهائها.
- يمكن اعتبار برمجيات جوب.تو.ميتينج، وميكاجو، ويوجو، وويندوز نت.ميتينج التي سبق الإشارة إليها من هذا النوع.

6.5 برمجيات أدوات التعاون

- هذه البرمجيات قد لا تتوفر جميعها في حزمة واحدة، ولكن بعض الحزم تشمل الكثير منها، كما أن أغلبها يتوفر بشكل برمجية منفصلة. ويمكننا استعراضها حسب الإطار المذكور كالآتي:
- نظم إدارة التعليم (الطبقة الثانية) قد تتضمن الكثير من هذه الأدوات، ولكنها تتفاوت من برمجية إلى أخرى. فمثلا نظام مودل يشمل أدوات تعاون مثل التعامل مع البريد الإلكتروني، والمناقشة الإلكترونية، والردشة والرسائل الفورية، والتصويت.
 - نظم الفصول الافتراضية (الطبقة الثالثة) قد تتضمن الكثير من هذه الأدوات، ولكنها أيضا تتفاوت من برمجية إلى أخرى. فمثلا نظام المونيت يشمل أدوات تعاون مثل التعامل مع البريد الإلكتروني، والردشة والرسائل الفورية، والتصويت، واللوحة الإلكترونية، ومشاركة التطبيق، والعروض، والمؤتمرات الصوتية، والمؤتمرات المرئية (الفيديو كونفرنس).
 - تقوم المؤسسة بالاستعانة بالبرمجيات التي توفر خدمات تعاون غير المشمولة في البرمجيات التي تبنتها في الطبقات السابقة.

◀ كيفية استخدامها في التعليم الإلكتروني المخصص لخدمة القرآن الكريم:

- برمجيات البريد الإلكتروني: يوجد مئات من مزودي البريد الإلكتروني ولعل أشهرها بريد جوجل (Gmail) وبريد ياهو وهوتميل.
- 📧 وتستخدم رسائل البريد الإلكتروني في التواصل غير المتزامن، وفي تلقي أخبار التدريب وإعلاناته، وكذلك في تبادل الملفات والمصادر التعليمية والتدريبية.
- برمجيات المناقشة الإلكترونية: ويوجد منها مئات، ويمكن للمؤسسة الحصول على البرمجية وتركيبها على موقعها الإلكتروني. ومن أمثلة برمجيات إدارة المنتديات:
- برنامج في.بوليتين (vBulletin) وهو غير مجاني من (www.vbulletin.com).
- برنامج بي.أتش.بي.بي.بي (phpBB) ذو الرخصة الحرة من (www.phpbb.com).

📌 والمناقشات ما زالت أكثر الأدوات استخداما في التعليم والتدريب، بحيث يشارك الدارسون ومدرّبهم في النقاشات، حتى لو لم يكونوا متواجدين في نفس الوقت، وهي تتيح طرح الأسئلة وتلقي الإجابات.

■ برمجيات الدردشة: خدمة الرسائل النصية الفورية متوفرة في أغلب برمجيات الطبقة الثانية والثالثة، كما توجد برمجيات منفصلة متخصصة في تقديم خدمة الدردشة، ومنها:

- برنامج ليف.هلب.نو (LiveHelpNow) وهو غير مجاني من (livehelpnow.net).

- برنامج فيلارو (Velaro) وهو غير مجاني من (www.velaro.com).

📌 والدردشة توفر وسيلة تعليمية وتدريبية تتميز بالفورية والعفوية حول موضوع التدريب، مثلا في أثناء التدريب والتعليم فإن الدردشة وسيلة فعالة للمشاركين لطرح الأسئلة وتوفير التغذية الراجعة للمدرّب.

■ برمجيات التصويت والاختيار: وهي مجموعة برمجيات تسهل عملية التقييم أو الاقتراع أو التصنيف أو المسح. ويمكننا تصنيفها إلى الآتي:

- خدمات المستوى البسيط والمكونة من سؤال مع عدة خيارات متوفرة في أغلب برمجيات الطبقة الثانية والثالثة. كما يمكن الاستعانة ببرمجيات جاهزة لتركيبها على أي برمجية مفتوحة المصدر، ومنها مثلا البرمجية المجانية ستار.رأينج.سيسستم (Star Rating System) من (www.grafxsoftware.com/product.php?id=8).

- الخدمات المسحقة والاستبانات الإلكترونية: وهي برمجيات تقدم خدمة تصميم استبانات متنوعة على الانترنت وإتاحة استخدامها للمعنيين، وتشمل أنواع أسئلة متنوعة بحيث تحاكي عالم الاستبانات الورقية. وتوجد مواقع تقدم هذه الخدمة مجانا أو باشتراك، ومن المواقع التي تقدمها بشكل مجاني موقع كيوك.سيرفيس (KwikSurveys) من (www.kwiksurveys.com) وموقع إي.سيرفيس.برو (eSurveysPro) من (www.esurveyspro.com).

أغلب هذه البرمجيات تقوم بجولة الإجابات وتعرض الإحصائيات بشكل بياني أو رقمي، ويمكن تخزينها في قاعدة بيانات، لتحليلها في وقت لاحق.

📌 ويمكن استخدام هذه البرمجيات في التعليم والتدريب الخاص بالقرآن الكريم لأجل مساعدة المشاركين في تكوين رأي، وفي الكشف عن آرائهم، وقد يستخدمها المدرّب في تحليل خلفية المستخدمين واهتماماتهم، والكشف عن المواقف والمشاعر، وقياس الفهم ورصده خلال التدريب، والإجبار على التفكير والمناقشة إضافة إلى جمع البيانات والمعلومات.

■ برمجيات جولة الويب: وغالبا ما تكون هذه الخدمة ضمن حزم برمجية تقدم خدمات متنوعة، وهي موجودة في برمجيات ذكرت سابقا مثل برمجية المونيت، وجو.تو.مينيتنج، وميكاجو.

المساهمة التعليمية والتدريبية الأساسية لهذه الخدمة هو متابعة المشاركين لتصفح مدربهم مواقع في شبكة الويب. ويمكن استخدامها مثلا في تصفح مواقع لخدمة القرآن وشرح أجزاء تلك المواقع ومفرداتها.

■ برمجيات اللوح الإلكتروني: من ناحية تقنية تعتبر خدمة اللوح الإلكتروني هي خدمة عروض متفاعلة بحيث تسمح للمشاركين والمدرّب بالمشاركة في الكتابة والرسم والتأشير وغيره. وهي غالبا ما تكون ضمن حزم برمجية تقدم خدمات متنوعة، ويمكن الاستعانة ببعض البرمجيات التي تقدم هذه الخدمة مثل:

- وايت بورد (WhiteBoard) من (www.tucows.com/preview/411066).

- نوتيت.أت (NotateIt) من (www.notateit.com).

- كما أن بعض البرمجيات المذكورة سابقا مثل برمجية يوجو وميكاجو تقدم هذه الخدمة.

المساهمة التعليمية والتدريبية الأساسية في مجال الدراسة لهذه الخدمة هي تقديم عروض وشروحات مثل شروحات حية عن اللغة العربية وأحكام التجويد، كما أنها تسمح بالتفاعل بحيث يمكن عرض صفحات من القرآن خالية من علامات التجويد مثلا، ويقوم المشاركون بوضع رموز التجويد وكذا عرض صفحات من النصوص العربية ووضع علامات الترقيم والفتحة والضمة ... إلخ.

■ برمجيات مشاركة التطبيق: وهي عموما تشابه جولة ويب والفرق هو أن المدرّب يعرض برنامجا أو تطبيقا من حاسوبه وليس موقع ويب. وغالبا ما تكون هذه الخدمة ضمن حزم برمجية تقدم خدمات متنوعة، وهي موجودة في برمجيات ذكرت سابقا مثل برمجية المونيت، وجو.تو.ميتينج، وميكاجو.

المساهمة التعليمية والتدريبية الأساسية لهذه الخدمة هو متابعة المشاركين عمل مدربهم على برنامج أو تطبيق حاسوبي، والاستخدامات المتوقعة لهذه الخدمة كثيرة ومنها شرح كيفية استخدام برمجيات ذات علاقة بالقرآن أو باللغة العربية.

■ برمجيات العروض: من الناحية التقنية هي برمجيات تقوم بتقديم خدمة العروض التي يستخدمها المدرّب من جهاز عرض في قاعة تقليدية ولكن على الانترنت. وتوجد مواقع عديدة مشهورة تقدم هذه الخدمة بشكل مجاني، ومنها:

- وثائق جوجل (Google Docs) من (google.com/google-d-s/presentations).

- بري.زنتيت (PreZentit) من (<http://prezentit.com>).

- زو.شو (Zoho Show) من (<http://show.zoho.com>).

المساهمة التعليمية والتدريبية لهذه الخدمة مشابهة لعروض الشرائح الروائية التي يوجد فيها نصوص وصور وأحيانا صوت وفيديو ... إلخ والتي تشكل عنصرا أساسيا في التعليم التقليدي.

- برمجيات المؤتمرات الصوتية والمؤتمرات المرئية: هذه البرمجيات مفيدة في التعليم والتدريب عندما لا يكون هناك وقت كاف لكتابة كل شيء، أو إذا كانت هناك أهمية لعفوية التعبير عن الفكر، ويمكن الاستفادة منها في عقد جلسات أسئلة وأجوبة، ونقاشات جدلية، وعصف ذهني أو إثارة الأفكار، وتكرار الأدوار، وإلقاء محاضرة، أو خطبة، أو تقرير، أو اقتراح، أو تلاوة... الخ. وتتميز المؤتمرات الصوتية بإمكانية تشغيلها على الإنترنت بسرعة ليست عالية، وتتميز والمؤتمرات المرئية بإمكانية رؤية إشارات اليد، وتعبير الوجه، ورؤية الأشياء ثلاثية الأبعاد، والحركات والإيماءات.

ومن البرمجيات التي تقدم هذه الخدمات والتي سبق ذكرها: سكايب، وجو.تو.ميتينج، وميكاجو، وإنسبيك، ومسنجرات ياهو - مسن/ليف - جوجل توك.

ومن البدهي فأنتها في تعليم التحدث باللغة العربية والقراءة الصحيحة للقرآن الكريم.

■ أدوات اللقاءات الإلكترونية الحية: وهي تقدم أنواعا متعددة من الوسائل التي تسمح باللقاءات والاجتماعات عن بعد.

ومن البرمجيات التي تقدم هذه الخدمات والتي سبق ذكرها: برمجية جو.تو.ميتينج، وميكاجو، ويوجو، وويندوز نت. ميتينج والمونيت.

ومن فائدتها التعليمية والتدريبية تتمثل في تسهيل لقاءات المدرب مع طلابه.

7 الخاتمة

لقد زاد الوعي من أهل القرآن الكريم على أهمية استخدام تقنيات التعليم الإلكتروني وتطويرها في سبيل خدمة القرآن على صعيد التحفيظ، والترتيل، وتعليم أحكام التجويد، والتفسير، وتعليم اللغة العربية، ولكنه ما زال بحاجة إلى توجيه وإلى عمل أبحاث لترقي به إلى درجة الاحتراف، عبر الاستفادة من أفضل التقنيات وأحدثها.

وقد جاء هذا البحث ليضع إطارا عاما يوضح الطريقة التي يلزم أن تسلكها مراكز القرآن المعنية بتقديم خدمة تعليم القرآن عبر الإنترنت، حيث بدأ البحث بمقدمة عن التعليم الإلكتروني وأنواعه، ووضح مجالات التعليم والتدريب الخاصة بالقرآن الكريم، كما وبين أسباب استخدام تقنيات التعليم الإلكتروني في هذه المجالات وفوائدها، كما عُرض ملخص للتقنيات المستخدمة في معظم مواقع تعليم القرآن الكريم الحالية، حيث تم استقراء 147 موقعا، ووضع إطار العمل المقترح بشكل مرّن عبر أربع طبقات ليوفر لهذه المراكز إمكانية استعماله حسب خصوصيتها لتقديم خدمة تعليمية وتدريبية متكاملة، وتبعه توضيح بطبيعة أدوات التعليم الإلكتروني وتقنياته اللازمة لذلك وضمنت كل طبقه فيه بعض أشهر الأدوات المتوفرة حاليا وأفضلها لتقديم خدمات تعليم القرآن الكريم.

- [1] أخرجه الترمذي في سننه (36-35/5)، برقم: 2914.
- [2] متفق عليه، أخرجه البخاري في صحيحه برقم: (4937)، ومسلم في صحيحه برقم: (798).
- [3] رواه البخاري في صحيحه (74/9) -مع الفتح- برقم: 5027.
- [4] رواه النسائي في السنن الكبرى برقم: (7977) وقد صححه الشيخ الألباني في صحيح الجامع الصغير برقم: (2165)، وفي صحيح الترغيب والترهيب (210/2).
- [5] ابتسام الجابري، استخدام التقنيات الحديثة في خدمة القرآن، منشورة على موقع ندوة "القرآن الكريم والتقنيات المعاصرة" المنعقدة في مجمع الملك فهد لطباعة المصحف الشريف بالمدينة المنورة في الفترة من 13-2009/10/15م. شوهدت في 2011/3/15م، الوصلة: www.quranit.org/remis/cms/?la=&p=38
- [6] محمد الخطيب، التوظيف التقني للقرآن الكريم في تعليم العربية للناطقين بغيرها، بحث منشور في ندوة "القرآن الكريم والتقنيات المعاصرة" المنعقدة في مجمع الملك فهد لطباعة المصحف الشريف بالمدينة المنورة في الفترة من 13-2009/10/15م. الوصلة: www.quranit.org
- [7] يوسف القرضاوي، الإسلام واستعمالات الإنترنت، موقع فضائية الجزيرة، 2004/11/20م، شوهدت في 2011/3/15م، الوصلة: www.aljazeera.net/channel/archive/archive?ArchiveId=102557
- [8] محمد القبي، حكم إدخال الإنترنت إلى البيت، موقع الشيخ الرسمي، شوهدت في 2011/3/15م، الوصلة: www.ferkous.com/rep/Bq62.php
- [9] نهى البداح، طريق الحقيقة: تعليم القرآن وعلومه باستخدام تقنيات ويب 2.0، المؤتمر الأول لتقنيات واتجاهات الويب الحديثة، جامعة الملك سعود، 2009/1/28م. شوهدت في 2011/3/15م، <http://webference.wordpress.com>
- [10] Commission of the European Communities. The eLearning Action Plan: Designing tomorrow's education, COM, 172 final, Brussels, March 28, 2001, online available, viewed on April 1st, 2010: http://ec.europa.eu/education/archive/elearning/annex_en.pdf
- [11] Horton, W. and Horton, K. E-learning tools and technologies: A consumer's guide for trainers, teachers, educators, and instructional designers. Indianapolis, Indiana, Wiley Publishing Inc. pp. 591. ISBN: 0471444588.
- [12] Weingardt, K. R., The Role of Instructional Design and Technology in the Dissemination of Empirically Supported, Manual-Based Therapies, Clinical Psychology: Science and Practice, 11(3), p.313-331, published by Oxford University Press, August 2004.
- [13] OpenSource, The Open Source Initiative, Web site of OS Initiative. online available, viewed on April 1st, 2010: <http://opensource.org>
- [14] Virtual Private Networks, C. Scott, P. Wolfe and M. Erwin, 2nd edition, Publisher: O'Reilly, ISBN: 1-56592-529-7, 255 PP, January 1999.
- [15] Janet Clarey, E-Learning 101: An Introduction to E-Learning, Learning Tools, and Technologies, Brandon Hall Research, Sunnyvale, CA 94087. April 2007.
- [16] جميل إطميزي، "نظم التعليم الإلكتروني وأدواته"، ردمك: 844966-557-0-978، الطبعة الأولى، الناشر: فيليبس للنشر، 2010/4/30م.
- [17] [toptenreviews.com](http://cms-software-review.toptenreviews.com), <http://cms-software-review.toptenreviews.com>, on April 25 2011.
- [18] Howard, C., Learning Content Management Systems: What Works, Industry Trends, Best Practices, and Vendor Profiles, Bersin & Associates, 130 pp, October, 2005.

Title: eLearning Tools and Technologies at Holy Quran Services between Theory and Practice.

Abstract. The significant progress in IT opens the way for educational institutions, including institutions concerned with serving the Quran to expand its services to the world, and allowed eLearning technologies to meet the need for learning environments to learn the Quran and its Sciences and simulate the work of the real Holy Quran Memorization Centers, to serve the Muslim Arabs and non-Arabs, in home or abroad.

Upon the novelty of the subject and after extrapolation techniques sites that offer this service, the lack of clarity and improvised work were observed, which indicate the need for more related descriptive and practical research.

This research, provides an introduction to eLearning, reviews the education and training areas of the Holy Quran, explains the reasons for using eLearning Technologies in these areas as well as their benefits. Moreover, it presents a summary of the techniques used in the most relevant sites. Upon our research, studying and experience a flexible framework is proposed to help these centers use eLearning technologies in adaptive way to provide educational and training integrated service. The paper will conclude with a summary of the best and the most currently available tools to provide education services with the holy Quran, at last, an abstract will be provided.

Keywords: eLearning technologies, eLearning tools, distance teaching of Quran, teaching of Quran via Internet.

المصطلحات

المناقشة الإلكترونية	Online discussion
في التصويت الإلكتروني	Online voting
في جولة الويب	Web tour
اللوحة الإلكترونية	Whiteboard
أدوات اللقاءات الإلكترونية الحية	Online Meeting Tools
المتزامن - غير المتزامن	Synchronous - Asynchronous
والبرمجية المفتوحة المصدر	Open Source Software
الشبكة الافتراضية الخاصة	virtual private network - VPN
أنظمة إدارة المقررات	CMS -Course Management System
أنظمة إدارة التعليم	LMS – Learning Management System
أنظمة إدارة محتويات التعليم	LCMS - Learning Content Management System
منصات التعليم الإلكتروني	eLearning Platform
نظم أنظمة إدارة المحتويات	Content Management System

عناوين المواقع التي تم زيارتها على الشبكة العنكبوتية

#	الوصلة	#	الوصلة	#	الوصلة
1	aquranonline.com	51	khuddamulquran.com	101	onlinequran.wordpress.com
2	www.alquranacademy.com	52	www.kidslearnquran.com	102	www.quranonlinelearning.com
3	www.alquranclasses.com	53	www.kidsquran.com	103	www.quranreciting.org
4	www.alhudapk.com	54	www.learnquran.com	104	www.quranrecitingacademy.com
5	alquraneducators.com	55	www.learnonlinequran.com	105	quranrecitingonline.com
6	www.alquraninfo.com	56	www.learn-quran-tutor.com	106	quranshine.com
7	www.alquranonlinecourse.com	57	learnquranlive.com	107	www.quranstudyonline.com
8	www.alquranstudy.com	58	readquranonline.weebly.com	108	www.qurantaafheem.com
9	alqurannow.com	59	www.learningquranonline.com	109	www.quranteaching.com
10	www.alquranulkareem.com	60	www.quranequran.com	110	qurantutorlive.com
11	www.alquranwithtajweed.com	61	www.read-quranonline.com	111	www.qurantutoring.com
12	www.altajweedulquran.com	62	learningquran.net	112	www.qurantutoringacademy.com
13	azeem-ul-quran.com	63	www.learn-quran.net	113	www.qurantutorsonline.com
14	bait-ul-arqam.com	64	learnquranonline.net	114	quranathome.org
15	www.bestquranreading.com	65	www.livequranread.com	115	www.quranicinstitute.com
16	betulquran.com	66	www.livequranstudy.com	116	www.quranmaster.com
17	www.click4quran.com	67	www.mahadalquran.com	117	quranmate.com
18	e-quranlearning.com	68	modernquranteaching.com	118	www.quranondesk.com
19	ellearnquran.net	69	mualimulquran.com	119	quranonline.com.pk
20	elearning.minhaj.org	70	muslimcityschool.com	120	www.quranpreaching.com
21	www.e-learningquran.com	71	www.myequran.com	121	www.quran-read.com
22	www.equrancollege.com	72	www.nooralquran.com	122	www.quranreading.com
23	www.equraninstitute.com	73	learningequran.com	123	www.quranreadingonphone.com
24	equranlearn.com	74	muslimeducators.org	124	www.quranreadings.com
25	www.equranschool.com	75	onlinequranacademy.co.uk/	125	www.quran-teaching.com
26	www.globalizingquran.com	76	www.onlinequranclasses.com	126	www.qurantrainer.com
27	equrantajweed.com	77	www.onlinequraneducation.com	127	qurranonline.com
28	www.faizequran.com	78	www.onlinequranhome.com	128	www.readalquranonline.com
29	https://fehmulquran.com	79	www.onlinequraninstitute.com	129	www.readquranbook.com
30	www.1stquranread.com	80	www.institutequranlearning.com	130	www.readquranonline.net
31	holyyquranacademy.coml	81	alquranicfoundation.webs.com	131	www.recitequran.net
32	www.homequran.com	82	quranonlinetutor.com	132	soatulquran.com
33	www.homequranteaching.com	83	www.siratulnoor.com	133	tahseenulquran.com
34	www.homequranschool.com	84	www.onlinequranstudy.com	134	www.tajveedequran.com
35	www.iquranschool.com	85	onlinequranuv.com	135	taleemeislam.com
36	www.ilmulquran.net	86	www.onlineteachingquran.com	136	www.tarteelquran.com
37	quranlearningonline.org	87	www.qarisl.com	137	www.tarteelulquran.com
38	www.quranreadingonline.com	88	www.alquran4all.com	138	www.teachingquran.com
39	www.qurantutoronline.com	89	www.quranalfurqan.com	139	learn.alquraan.net
40	quranlearnonline.com	90	www.qurancampus.com	140	zain-ul-quran.com
41	iqraquran.net	91	www.qeah.co.uk	141	zaoulquran.com
42	iqraquranacademy.com	92	quranes.com	142	www.tajac.org
43	www.iqraquranlive.com	93	www.quranfocus.com	143	www.alwhyyn.net
44	www.iqraulquran.com	94	quranforummah.com	144	www.ta7feez.com
45	www.iquranreading.com	95	qurankream.com	145	www.islam-sister.com
46	www.ireadquran.com	96	www.tajweedacademy.com	146	www.factway.net
47	www.islamiceducationsystem.com	97	www.quranlearningacademy.com	147	malahsan.net/forum
48	www.islamicnet.com	98	www.quranlearningonline.com		
49	istudyquran.com	99	www.quranlearningschool.com		
50	www.joinquran.com	100	www.quranlearningtajweed.com		



الجلسة الحادي عشرة

المسافة الصرفية بين الكلمات العربية

أحمد خرصي

كلية علوم الحاسب والمعلومات

قسم علوم الحاسب

جامعة الإمام محمد بن سعود الإسلامية

الرياض، المملكة العربية السعودية

amakhorsi@imamu.edu.sa

ملخص

إن من أكثر تطبيقات التحليل الصرفي وتقسيم الكلمة شيوعاً توحيد شكل متغيرات الكلمة الواحدة في مرحلة تسويق استخراج المعلومات. والفكرة أن التحليل الصرفي قد يسمح بتحديد الأصل الصرفي لكلمتين سواء جزأ أو جذعا ثم مقارنة أصلي الكلمتين، فإن كان الأصل واحداً اعتبرت الكلمتان متشابهتين معنا وإلا فلا. في هذا البحث نناقش إمكانية المقارنة المباشرة بين الكلمتين لقياس التشابه الصرفي بينهما دون الحاجة لاستخراج أصليهما. هدفنا هنا الاستغناء عن استعمال المعلومات اللغوية كقواعد بيانات السوابق والوفاق والجذور. واختصار المعالجة على مرحلة واحدة هي المقارنة المباشرة. أما كون مقياس المسافة الصرفية لا يعتمد على معلومات لغوية فهذا له فائدتان مهمتان: 1. إمكانية توسيع استعماله للغات غير اللغة العربية خاصة تلك التي لها تركيب كلمات مشابهة للغة العربية دون الحاجة إلى تغيير كبير في طريقة الحساب. 2. عدم الحاجة لبناء ذخيرة معرفية قد لا تكون شاملة وأو خالية من الأخطاء. وأما كون المعالجة تقتصر على المقارنة المباشرة فهذا يعني مبدياً أنها أسرع من المقارنة ذات المراحل العدة كالتحليل الصرفي.

الكلمات الجوهرية: التحليل الصرفي، معالجة اللغة البشرية، استخراج المعلومات، البحث عن المعلومات، التعلم الآلي

1. مقدمة

في تطبيقات استخراج المعلومات كمحركات البحث، تسبق عملية البحث واستخراج المعلومات مرحلة تنقيح جملة الاستعلام. خلالها تمر الجملة التي أدخلها المستخدم (طالب المعلومة) بمعالجة لكلماتها وجملها. من هذه المعالجات اشتقاق كلمات ذات صلة صرفية بالكلمات المدخلة، كاشتقاق صيغة الجمع من صيغة المفرد وصيغة التأنيث من المذكر وهكذا. من بين التقنيات المستعملة لهذا الغرض، التحليل الصرفي والتقسيم الصرفي للكلمات. مبدؤها إرجاع الكلمات إلى أصل صرفي مشترك أكبر قدر من متغيرات الكلمة كالجزر. ثم البحث عن الكلمات التي لها نفس الأصل الصرفي تماماً. أما التحليل الصرفي لكلمات اللغة العربية فقد حظي باهتمام في السنوات الأخيرة. ولعل أشهر محلل صرفي هو الذي صممه بك وولتر [بك وولتر، 2002]

ثم الذي صممه بيزلي [بيزلي، 1996]. وغيرهما [عبد الرحيم ورقيق، 2008] [عواجان، 2003] الغرض هنا ليس عرض كل المحلات الصرفية العربية فقد سبق إلى مثل هذه الدراسة غيري [الصغير والخراسي، 2004] وكذلك الحال لتقسيم الكلمات حيث الهدف ليس إرجاع الكلمة إلى أصلها الصرفي الحقيقي وإنما إيجاد جزء الكلمة المشترك بين الكلمة وصيغها الأخرى. وعادة يسمى هذا الجزء المشترك جذعا. ونذكر منها على سبيل التمثيل لا الحصر [لاركي ومن معه، 2002] [تاغفا ومن معه، 2005] [إروقاتي ومن معه، 2003] [الجيل وفريزر، 2002] وغيرها. الحاصل أن الكل يهدف إلى توحيد الأشكال المختلفة لكلمات مشتقة من أصل صرفي واحد ليتمكن نظام استخراج المعلومات فيما بعد من العثور على كل الوثائق التي تحتوي تلك الأشكال المختلفة للكلمة. المقاربة البديلة التي نطرحها في هذا البحث، مقارنة الأشكال المختلفة بشكل ينتج قياس مدى التشابه الصرفي بينها، وحينها كلمتان بقيمة تشابه صرفي عالية تعني أن الكلمتين من أصل صرفي واحد بقيمة متدنية تعني أنهما من أصلين مختلفين دون الحاجة لمعرفة الأصل الصرفي. في الفقرة التالية نناقش طريقة القياس التي نقترحها. في الفقرة الثالثة نعرض آلية التجريب التي اتبعناها ونتبعها بفقرة نتايج ثم نختم بخلاصة البحث.

2. المسافة

ليكن كلمتان عربيتان

كلمة₁ = ح ١ ح 2... ح ن

و

كلمة₂ = خ ١ خ 2... خ م

حيث أن ح ≠ خ؛ حروف.

ليكن تردد نص (ح) نسبة تكرار الحرف في النص نص. فيما يلي سندرج على عدم كتابة نص حيث أن النص (الذخيرة النصية) الذي نستخرج منه ترددات الأحرف لا يتغير.

سنسمي أطول سلسلة مشتركة بين كلمتين أكبر عدد من الأحرف المشتركة بين كلمتين في نفس الترتيب وسنرمز لها بـ لسم.

2.1. مثال

لسم (الكتاب، تكرب) = كب

ا ل ك ت ا ب

ت ك ر ب

لسم (أقلام، قلم) = قلم

أ ق ل ا م

ق ل م

نسمي مكمل أطول سلسلة مشتركة بين كلمتين كل الأحرف من الكلمتين التي لا تشارك في لسم

مثال:

لسم (الكتاب، تكرب) = التتار

و

لسم (أقلام، قلم) = ا ا

نرمز لحرف من كلمة ما كالتالي

كلمة [رقم الحرف]

مثال

كلمة₁ [2] = ح₂

حساب المسافة الصرفية بين الكلمتين يتم إذن بالمعادلة التالية:

$$\text{ما يعني أن المسافة الصرفية} = \left\| \text{مجم}_{1=\varnothing}^{(2\varnothing, 1\varnothing)} \right\| \text{لو} \left(\frac{1}{\text{تردد}(\text{لسم}(\varnothing, 1, 2\varnothing) \mid \text{س})} \right) - \left\| \text{مجم}_{1=\varnothing}^{(2\varnothing, 1\varnothing)} \right\| \text{لو} \left(\frac{1}{\text{تردد}(\text{لسم}(\varnothing, 1, 2\varnothing) \mid \text{س})} \right)$$

هي مجموع لوغاريتيمات مقلوب ترددات أحرف أطول سلسلة مشتركة مطروح منه مجموع ترددات أحرف مكمل أطول سلسلة مشتركة.

أساس هذه المعادلة فكرة أن تردد أحرف مادة الكلمة أقل بكثير من تردد الأحرف الزائدة. فمثلاً ترددات أحرف كلمة "كتاب" في الذخيرة التي استعملناها في استخراج الترددات :

تردد(ا) = 9.32 %

تردد(ب) = 3.01 %

تردد(ت) = 3.00 %

تردد(ك) = 1.37 %

واضح أن تردد الحرف الزائد أكبر بكثير من أحرف مادة الكلمة الأصلية. أي أن الاختلاف في أحرف المادة مع التشابه في الأحرف المزيّدة يعني تدني التشابه الصرفي المحسوب بالمعادلة السابقة، في حي أن التشابه في الأحرف الأصلية يرفع قيمة التشابه الصرفي وإن وجدت اختلافات في الأحرف المزيّدة.

3. التجارب

للتثبت من فاعلية طريقة الحساب المقترحة أجرينا تجارب على ذخيرة لغوية.

3.1. الذخيرة

قمنا باستخراج ترددات الأحرف من ذخيرة تضم أكثر من 500 ألف كلمة عربية. وباستخدام محلل صرفي أنشأنا ذخيرة من أكثر من 200 ألف كلمة كل منها مرفقة بجذرها. فكان عدد الجذور يتجاوز 4000 جذر. من الـ 200 ألف كلمة اخترنا ما يتجاوز 80000 كلمة بشكل عشوائي. لكل كلمة من الـ 80000 استخرجنا عدد الكلمات ع المشتقة فعلاً من نفس جذر الكلمة، ثم قمنا بحساب التشابه الصرفي بينها وبين كل من الـ 200 ألف كلمة ثم استقينا فقط ع كلمة التي لها أعلى تشابه.

وبوصف أدق:

لنكن

ذ₁ : ذخيرة من 200 ألف كلمة حيث

ذ₁ = { <كلمة₁₁₂₂

ولنكن مجموعة الـ 80000 كلمة ذ₂ جزء من ذ₁

وليكن جذر (كلمة) هو مجموعة جذور كلمة حسب الذخيرة ذ₁

أي

جذر (كلمة) = { جذر₁ : <كلمة₁₁> ∪ ... }

ليكن م ج (كلمة) مجموعة الكلمات التي تجتمع مع كلمة في جذر على الأقل أي:

م ج (كلمة) = { كلمة₁ : جذر (كلمة) ∩ جذر (كلمة₁) ≠ ∅ }

و #م ج (كلمة) عدد عناصر م ج (كلمة) أي عدد الكلمات التي تشترك مع كلمة في جذر على الأقل.

ليكن م مسـمـن (كلمة) : ن كلمة من ذ₁ ذات أعلى قيم التشابه الصرفي مع كلمة

لقياس فاعلية التشابه الصرفي المقترح استعملنا معيارين : الأول المكسب التراكمي المخفض [سقاي، 2007] NDCG والثاني نسبة الدقة [أسلم ويلمز، 2005] [بولكلي وفوريز، 2000] r-precision

لحساب المكسب التراكمي لابد من تحديد كيفية حساب الصلة.

سنقول أن كلمة كلمة₁ ذات صلة بالكلمة كلمة₂ إن وفقط إن اشتركت الكلمتان في جذر على الأقل. أي:

$$\text{صلة (كلمة}_1\text{ ، كلمة}_2\text{)} = \left\{ \begin{array}{l} 1 \text{ إذا كان كلمة}_2 \ni \text{ كلمة}_1 \\ 0 \text{ إن لم يكن} \end{array} \right.$$

بحسب المكسب التراكمي المخفض كالتالي:

$$\frac{\text{المكسب الفعلي}}{\text{المكسب المثالي}} = \text{مكسب التراكمي المخفض}$$

حيث

$$\frac{1}{\text{مجم}_{2=0}^{(كلمة)_{\text{ع}} \neq \#}} + 1 = \text{المكسب المثالي}$$

و

$$\frac{\text{صلة (كلمة، كلمة، كلمة)_{\text{ع}}}{\text{مجم}_{2=0}^{(كلمة)_{\text{ع}} \neq \#}} + 1 = \text{المكسب الفعلي}$$

أما

$$\frac{\text{صلة (كلمة، كلمة، كلمة)_{\text{ع}}}{\text{مجم}_{1=0}^{(كلمة)_{\text{ع}} \neq \#}} = \text{نسبة الدقة}$$

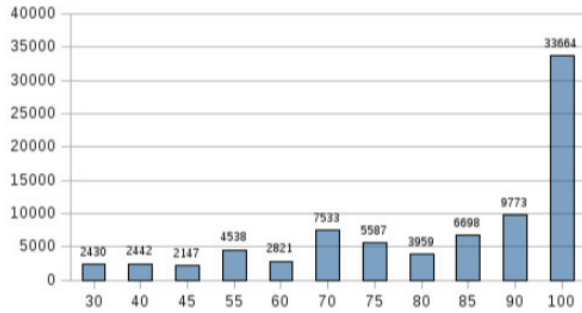
4. النتائج

فيما يلي نناقش فاعلية مقياس التشابه الصرفي في استخراج عدد ثابت من متغيرات الكلمة العربية الواحدة. ونعني بمتغيرات : الكلمات ذات المعنى والشكل المقاربين. مثلاً صيغة الجمع متغير لصيغة المفرد وكذلك صيغة المؤنث للمذكر وهكذا...

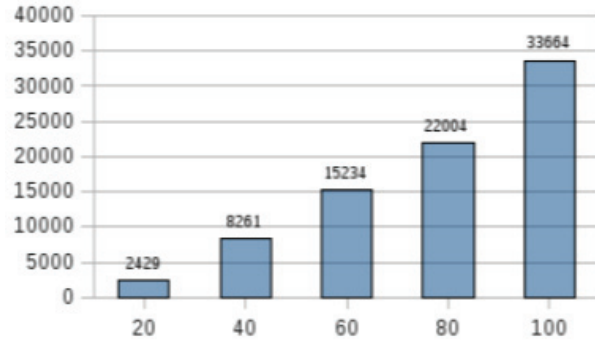
أثبتت التجارب أن استعمال المقياس من أي معالجة أولية خاصة عدا اعتبار الأشكال المختلفة لبعض الحروف كان كفيلاً باستخراج خمس متغيرات لكل من الـ 80000 كلمة من الذخيرة سابقة الذكر بمكسب تراكمي مخفض 82.76% ونسبة دقة 78.68%. الرسم 1 يوضح توزيع عدد الكلمات على نسب المكسب التراكمي المخفض. أما الرسم 2 فتوزيع عدد الكلمات على نسب الدقة.

الرسمتان تظهران أن أكثر من 33000 كلمة تم استخراج متغيراتها الخمسة بدقة مقاربة لـ 100%. وأكثر من 20000 كلمة تم استخراج متغيراتها الخمسة بنسبة مقاربة لـ 80%.

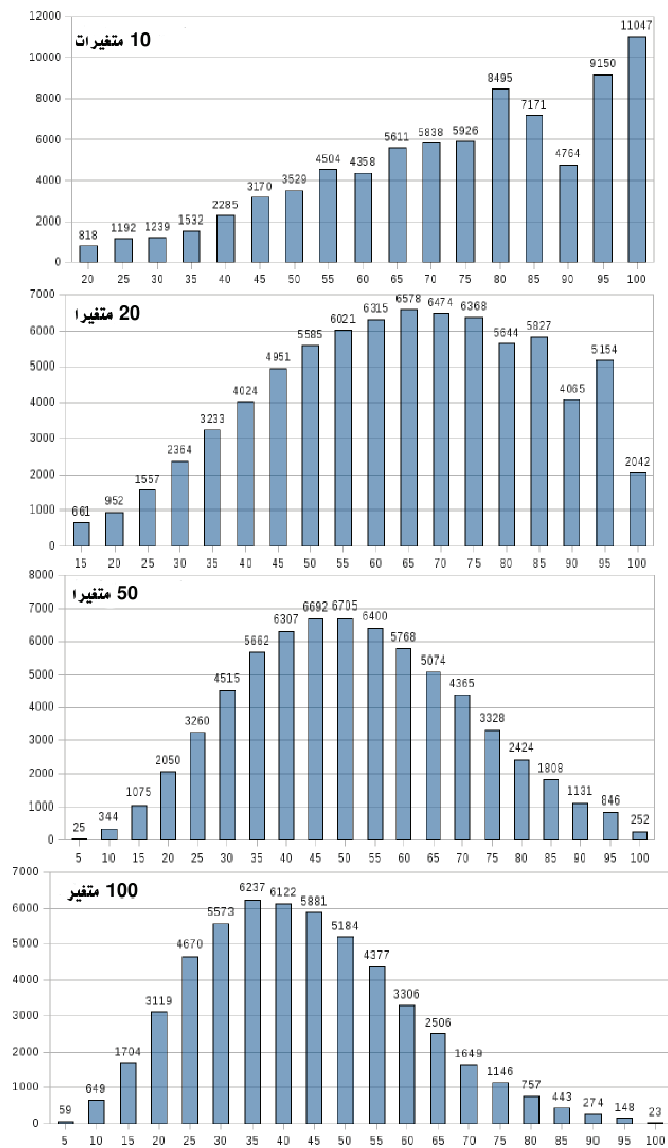
لألا نحصر تحليلنا للنتائج في عدد معين من المتغيرات جربنا النظام في استخراج عدد متصاعد من المتغيرات لنرى أداءه مع زيادة عدد المتغيرات المطلوب استخراجها. الرسم 3 يوضح توزيعات المكاسب التراكمية على أعداد الكلمات لأعداد مختلفة من المتغيرات. أما الرسم 4 فتوزيع عدد الكلمات على نسب الدقة.



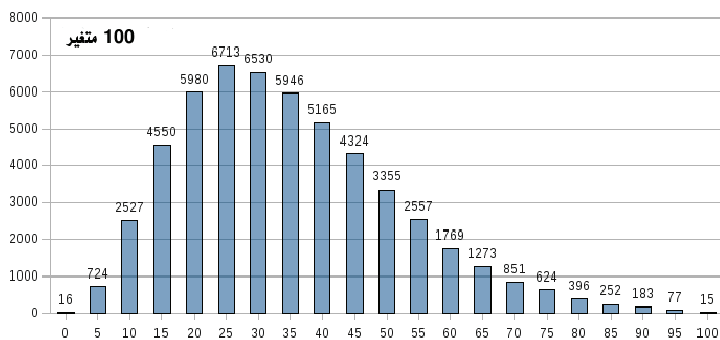
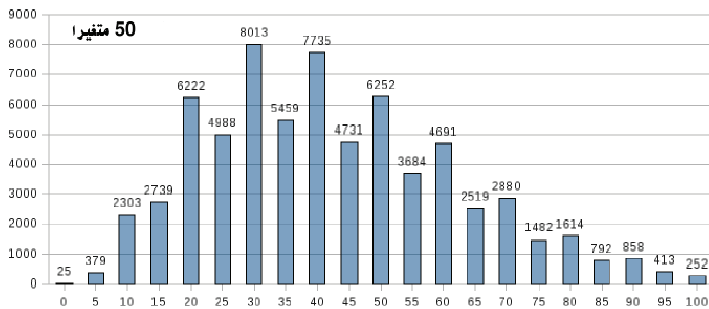
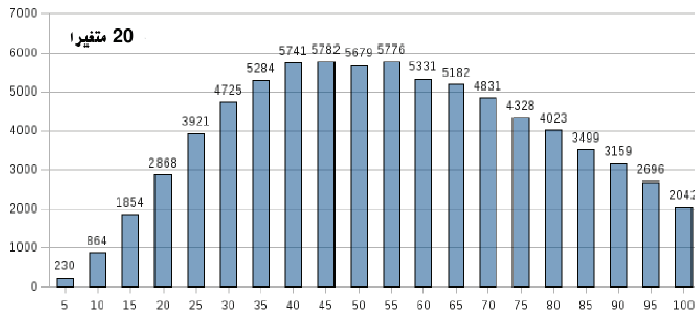
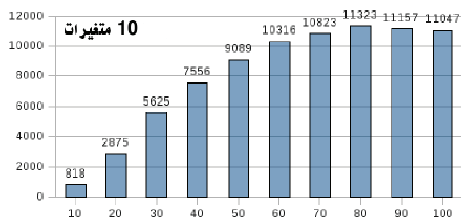
رسم 1: توزيع عدد الكلمات على المكاسب التراكمية
المخفضة في استخراج 5 متغيرات



رسم 2: توزيع عدد الكلمات على نسبة الدقة في استخراج
متغيرات



رسم 3: توزيع المكاسب التراكمية المخفضة لأعداد متغيرات مختلفة



رسم 4: توزيع نسبة الدقة لأعداد متغيرات مختلفة

واضح أن النتائج تضعف كلما زاد عدد المتغيرات المطلوب استخراجها. السبب الأقرب في نظرنا هو أن عدد الأحرف المختلة بين الكلمة ومتغيراتها يزيد كلما زاد عدد المتغيرات. فمثلاً "كتيب" و "وللكتيبات" كلتاهما متغير للكلمة "كتاب" غير أن عدد الأحرف المختلفة بين "كتيب" و "كتاب" لا يزيد عن "يا" أما بين "كتاب" و "وللكتيبات" فـ "ولليات" مايعني أن قيمة التشابه الصرفي كما اقترحنا حسابه لابد وأن يكون أقل بكثير مما هو بين الكلمتين "كتاب" و "كتيب". الحل الأقرب هو استعمال التجذيع [خرصي، 2012] للتخلص من السوابق والواحق قبل حساب التشابه الصرفي بين الكلمتين. مشكل آخر يكمن في التغيرات التي تطرأ على أحرف العلة في المتغيرات. فمثلاً الكلمة "نل" وكلمة "نيل" تشتركان في الجذر "ن و ل" غير أن التغير الذي دخل على الحرف العلة جعل مقياس التشابه الصرفي يعتبر الحرفين "اي" أحرف اختلاف مع أنهما إنما من أصل واحد. الحل الأقرب لهذا استعمال قواعد مقارنة الأحرف مختلفة تأخذ بعين الاعتبار هذا النوع من التغيير. إضافة لهذا بما أن النص العرب غالباً ما يكتب عارياً عن التشكيل والشدة فإن الأحرف المشددة التي تضاعف في المتغيرات تعتبر أحرف اختلاف. مثلاً كلمة "عد" و "عدد" كلاهما من جذر واحد غير أن الطريقة الحالية في حساب التشابه سيعتبر الحرف "د" حرف اختلاف.

5. خلاصة

ناقش هذا البحث استعمال طريقة جديدة لقياس التشابه الصرفي لغرض استخراج الكلمات المشتقة من أصل صرفي واحد (في هذا البحث اعتبرنا الأصل الصرفي هو الجذر) دون اللجوء للتحليل الصرفي، بل ودون الحاجة لأي معرفة لغوية خاصة وبالاتحاد على إحصائيات مستخرجة من ذخائر نصية بسيطة. النتائج الأولية للمقياس المقترح تظهر نتائج جيدة في استخراج خمس متغيرات لكلمة. لا شك أن هذا العدد من المتغيرات مقنع في أنظمة استخراج والبحث عن المعلومة كمحركات البحث. أظهرت النتائج ضعف النتائج في حال زاد عدد المتغيرات المراد استخراجها. وقد ناقشنا الأسباب المحتملة لذلك. أهم تطوير يمكن استهدافه هو تطبيق التجذيع قبل حساب التشابه الصرفي لاجتناب تدخل السوابق والواحق في قيمة التشابه الصرفي. كما أن تعديل قواعد المقارنة بين الأحرف العربية قد يجنب الأخطاء الناتجة عن اعتبار الأشكال المختلفة للحرف حرفاً مختلفاً.

6. المراجع

[أسلم ويلماز، 2005]

J.A. Aslam and E. Yilmaz, "A geometric interpretation and analysis of R-precision," *Proceedings of the 14th ACM international conference on Information and knowledge management*, New York, NY, USA: ACM, 2005, pp. 664–671.

[بك وولتر، 2002]

T. Buckwalter, *Buckwalter {Arabic} Morphological Analyzer Version 1.0*, 2002.

[بولكلي وفور هيز، 2000]

C. Buckley and E.M. Voorhees, "Evaluating evaluation measure stability," *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA: ACM, 2000, pp. 33–40.

[بيزلي، 1996]

K.R. Beesley, "Arabic Finite-State Morphological Analysis and Generation," *COLING*, 1996, pp. 89–94.

[تاغفا ومن معه، 2005]

K. Taghva, R. Elkhoury, and J.S. Coombs, "Arabic Stemming Without A Root Dictionary," *ITCC (1)*, IEEE Computer Society, 2005, pp. 152–157.

[الجيل وفريدر، 2002]

M. Aljlal and O. Frieder, "On arabic search: improving the retrieval effectiveness via a light stemming approach," *CIKM*, ACM, 2002, pp. 340–347.

[خرصي، 2012]

A. Khorsi, "Effective Unsupervised Arabic Word Stemming: Towards an Unsupervised Radicals Extraction," *IAJIT*, vol. 9.

[روقاتي ومن معه، 2003]

M. Rogati, J.S. McCarley, and Y. Yang, "Unsupervised Learning of Arabic Stemming Using a Parallel Corpus," *ACL*, 2003, pp. 391–398.

[الصغير والخراشي، 2004]

I.A. Al-Sughaiyer and I.A. Al-Kharashi, "Arabic morphological analysis techniques: A comprehensive survey," *JASIST*, vol. 55, 2004, pp. 189–213.

[عبد الرحيم ورقيق، 2008]

M.E.A. Abderrahim and F.B. Reguig, "A Morphological Analyzer for Vocalized or Not Vocalized Arabic Language," 2008.

[عواجان، 2003]

A. Awajan, "A Rule-Based Morphological Analyzer of Arabic Words," Malta, Sept. 1-3, 2003: WSEAS, 2003, p. 9.

[سقاي، 2007]

T. Sakai, "On the reliability of information retrieval metrics based on graded relevance," *Information Processing and Management: an International Journal*, vol. 43, Mar. 2007, pp. 531–548.

[لاركي ومن معه، 2002]

Larkey, L. S, L. Ballesteros, Connell, and M. E, "Improving stemming for Arabic information retrieval: light stemming and co-occurrence analysis," *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2002, pp. 275–282.

نواة نظام تعلّم الأنطولوجي من النصوص العربية

م.فاضل الحسن، د.غيداء ريداوي، د.بسام الكردي

العنوان البريدي:المعهد العالي للعلوم التطبيقية والتكنولوجيا، قسم المعلومات

ص.ب 31983، دمشق، سوريا

البريد الإلكتروني: fadel.alhassan@hiast.edu.sy

ghaida.rebdawi@hiast.edu.sy, bassam.kurdy@hiast.edu.sy

الخلاصة. تحوي النصوص أو المعطيات النصية التي تصف مجالاً معيناً توصيفاً كاملاً أو جزئياً للمجال الذي تتحدث عنه وذلك في مختلف اللغات. وبمعنى آخر يمكننا القول إن النصوص ما هي إلا تعبير الكاتب عن الأنطولوجي التي كونها عن العالم أو عن المجال أو التطبيق الذي يحاول أن يكتب عنه وذلك باستخدام اللغة أو اللغات التي يتقنها. فإذا أمكننا إيجاد طريقة أو برنامج يقوم بعكس هذه العملية، أي استنتاج أو تعلّم الأنطولوجي من النصوص؛ نكون قد قطعنا شوطاً كبيراً في تسريع عملية بناء الأنطولوجي، سنقوم في هذا التقرير بإعطاء صورة مختصرة للأنطولوجي وأهميتها ومكانها في الوب الدلالي، بعدها سنقدم عرض مبسط لأجزاء نظام تعلّم الأنطولوجي من النصوص العربية والتي تتضمن نظام استخراج التعابير ونظام استخراج الهرمية ونظام استخراج العلاقات غير الهرمية. أخيراً سوف نشير إلى أهم الأفكار والنتائج التي توصلنا إليها في نظام تعلّم الأنطولوجي من النصوص العربية.

كلمات جوهرية: الأنطولوجي، الوب الدلالي، تعلّم الأنطولوجي، معالجة اللغات الطبيعية، الذكاء الصناعي، إدارة المعرفة، التنقيب بالنصوص.

1. الوب الدلالي والأنطولوجي:

يعتبر الوب في هذه الأيام المصدر الأوسع للمعلومات في جميع المجالات والمواضيع، وهو في توسع دائم سواء في حجم المعلومات أو في حجم المستخدمين. إن هذا الحجم الكبير من المعلومات والتوسع السريع في الوب يجعلان إدارة المعلومات والوصول إليها أمراً في غاية التعقيد؛ وقد عانى معظمنا من العدد الكبير من الوثائق والروابط التي يعيدها أي محرك بحث رداً على أية عملية استعلام، هذا عدا عن عدم توافق نتائج البحث مع طلب المستخدم!

يعود هذا التعقيد إلى سببين أساسيين: الأول، البنية المستخدمة لتخزين الوثائق، فمعظم الوثائق في الوب تُخزّن بشكل معطيات نصية يقوم الإنسان بتصنيفها تصنيفاً يدوياً دون أي إمكان للتعبير عن محتوى هذه الوثائق تعبيراً يسمح للآلة أن تتعامل معه، والثاني، طرائق التصنيف والبحث الآلي

المستخدمة حالياً، فجميعها طرائق إحصائية لا تتعدى مستوى التحليل الصرفي للنصوص المكتوبة (البحث من خلال الكلمات المفتاحية).

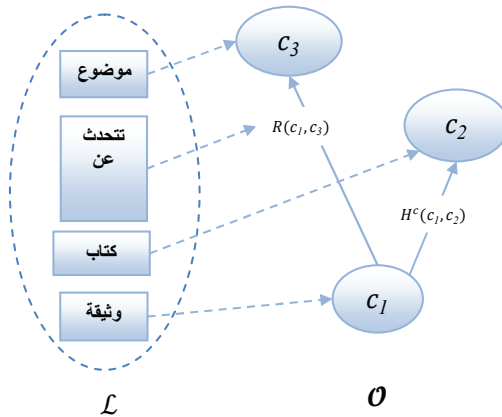
إذن، نحتاج إلى تعديل بنية الوب الحالي تعديلاً جذرياً للوصول إلى بنية أكثر ملاءمة للتعبير عن الوثائق والاستعلام عنها.

نتيجة لعدد من المؤتمرات والأبحاث، جرى طرح ما يعرف اليوم بالوب الدلالي كتوسعة للبنية الحالية للوب، تُعرّف الموارد فيه باستخدام لغة توصيفية تسمح بالتعبير عن المحتوى الدلالي للمورد بطريقة يمكن للألة تفسيرها. وقد جرى اقتراح عدد من اللغات والآليات للتعبير عن المحتوى الدلالي للمورد، واستخدامه لاحقاً في عمليات التصنيف والبحث، غير أن كل آلية من هذه الآليات تهتم بعدد من الجوانب التطبيقية وتغفل جوانب أخرى. لذا، جرى توحيد جميع هذه اللغات والآليات في بنية موحدة أكثر شمولية وديناميكية هي الأنتولوجي، تتيح التعبير عن المحتوى الدلالي للموارد تعبيراً فعالاً يفهمه الإنسان (باستخدام اللغات الطبيعية) والآلة على حد سواء، إضافة إلى الإمكانيات الواسعة التي تقدمها هذه البنية للاستعلام عن الموارد.

2. الأنتولوجي:

تعرّف الأنتولوجي الخاصة بمجال معين بأنها فهم لهذا المجال، يجري الاتفاق عليه بين مجموعة من المختصين والخبراء في المجال. يُعبر عن هذا الفهم بعدد من المفاهيم التي تخص المجال، ومجموعة من العلاقات التي تربط هذه المفاهيم، ويجري تغليف هذه البنية بمجموعة من التعابير اللغوية بلغة خاصة. تشمل هذه التعابير المفاهيم وبعض العلاقات. سنوضح بنية مبسطة للأنتولوجي، وآلية استخدامها في الوب الدلالي بالمثال التالي (الشكل (1)):

لنأخذ مجال الوثائق. يوجد عدد من المفاهيم في هذا المجال، ويمكننا التعبير عن كل منها بتعبير لغوي واحد أو أكثر، كمفهوم الوثيقة أو المورد أو مفهوم المؤلف أو الكاتب.



شكل 1 أنتولوجي مصفوفة لمجال الوثائق

ترتبط هذه المفاهيم بعدد من العلاقات، منها علاقات هرمية، كعلاقة الكتاب بالوثيقة وعلاقات غير هرمية كعلاقة الوثيقة بالموضوع الذي تحدث عنه.

يمكن إغناء هذه الأنتولوجي بعدد من الإضافات، مثل ربط كل مورد بعنوان وتاريخ... إلخ.

كما يمكن إغناؤها بعدد من القواعد التي تجعل منها قاعدة معرفة يمكن استخدامها في محركات الاستدلال^١، مثل: { يمكن أن يكون للوثيقة عدد من

^١ يقصد بمحرك استدلال: برنامج أو نواة برنامج، تستخدم مجموعة من القواعد المعرفة مسبقاً، ومجموعة من الحقائق والمعطيات المتغيرة لاستنتاج معلومات أو قرارات معينة وفقاً لهذه المعطيات. مثال على هذه المحركات Prolog.

المؤلفين}، {قد لا يكون للوثيقة ناشر}، {كل مؤلف يجب أن يكون قد شارك بتأليف كتاب واحد على الأقل}.

بعد استقرار البنية المتفق عليها للأنطولوجي، تُستخدم هذه الأنطولوجي لتخزين المعلومات في تطبيق معين يتعامل مع مفاهيم هذا المجال.

فمثلاً يمكن استخدام البنية السابقة للتعبير عن مجموعة الوثائق التي يحتويها موقع معين. وتشكل كل وثيقة غرضاً من الصف C_1 .

بعد هذا التخزين يمكننا الاستعلام مثلاً عن جميع الوثائق التي قام مؤلف معين بتأليفها، أو الاستعلام عن جميع الوثائق التي تتحدث عن موضوع معين أو.... الخ.

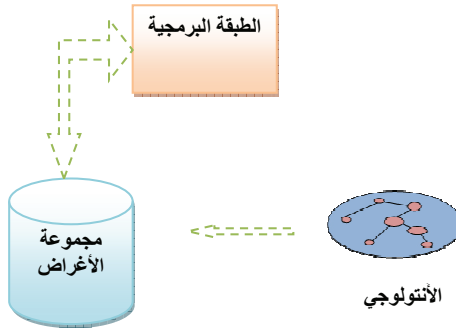
ولمزيد من الفعالية يمكن أن يكون مفهوم المجال أنطولوجي كاملة أو مجموعة من الأنطولوجي، ينطبق الشيء نفسه بالنسبة لمفهوم المؤلف أو الناشر.

تعتبر الأنطولوجي المعبرة عن مجال معين واحدة في جميع اللغات، ويكون الاختلاف في التعابير التي ترتبط بكل مفهوم، لذلك من الممكن أن نعرف مفهوم C_1 مكافئاً للمفهوم C_1 يرتبط بمجموعة التعابير المعبرة عن الوثيقة في اللغة الإنكليزية مثلاً؛ وبذلك يمكننا استخدام الأنطولوجي في تطبيقات أخرى مثل الترجمة الآلية (من حيث المبدأ).

3. البنية العامة لتطبيقات الويب الدلالي:

تتألف البنية العامة لأي تطبيق في الويب الدلالي من ثلاثة مكونات أساسية: هي الأنطولوجي، ومجموعة الأغراض، والطبقة البرمجية. وسنتناول هنا هذه المكونات بشيء من التفصيل.

- الأنطولوجي: هي مجموعة الأنطولوجي التي تصف مجال التطبيق أو مجالاته. فمثلاً، في تطبيق إدارة الوثائق تكون الأنطولوجي التي أشرنا إليها في الفقرة السابقة أحد الأنطولوجي المستخدمة في هذا التطبيق، إضافة إلى عدد من الأنطولوجي الأخرى مثل الأنطولوجي الخاصة بالأشخاص والأنطولوجي المتعلقة بالمواضيع التي تدور الوثائق حولها.



شكل 2 البنية العامة لتطبيقات الويب الدلالي

- مجموعة الأغراض: هي قاعدة معطيات، تعبر عن مجموعة الأغراض المنشأة من صفوف الأنطولوجي السابقة. ففي تطبيق إدارة الوثائق تكون هذه الأغراض هي الوثائق المراد إدارتها، ويعبر عن كل وثيقة بغرض من الصف وثيقة، يحمل جميع معلومات الوثيقة، ويجري تخزين هذه الوثائق في قاعدة معطيات أو في أي بنية تخزين أخرى.

- **الطبقة البرمجية:** مجموعة التوابع والخوارزميات التي تستخدم مجموعة الأغراض السابقة لتقديم عدد من الوظائف التي بُني التطبيق لأجلها، كالبحث في الوثائق أو مقارنة الوثائق أو ترجمتها... إلخ. يمكن أن تكون الطبقة البرمجية محرك استدلال يعتمد مجموعة الأغراض كحقائق، ومجموعة القواعد التي جرى تعريفها في الأنتولوجي كقواعد استدلال كما يمكن أن تكون تطبيقاً برمجياً يستخدم خوارزميات البحث والتصنيف والمقارنة المعروفة للوصول إلى النتائج المرجوة. تختلف الطبقة البرمجية من تطبيق لآخر، ولا يمكن بناء طبقات برمجية عامة تؤمن جميع الوظائف المرجوة من الأنتولوجي.

يمكن أن تبني مجموعة الأغراض بناءً يدوياً، كأن يجري إدخال جميع معلومات الوثائق في مكتبة إلكترونية معينة يدوياً. إلا أن هذا العمل مرهق ويستغرق وقتاً طويلاً. وقد ظهر مؤخراً عدد من التطبيقات القادرة على استخراج الأغراض وفقاً لأنتولوجي معينة، من مجموعة من الوثائق الإلكترونية، وربط هذه الوثائق بالأغراض.

أما بناء الأنتولوجي الخاصة بكل مجال فهو يجري إلى الآن يدوياً، ويستغرق الكثير من الجهد والوقت؛ وذلك لصعوبة الاتفاق على بنية موحدة لكل مجال، إضافة إلى صعوبة تحديد حدود كل مجال من المجالات، وكثرة المجالات التي تشمل مختلف جوانب الحياة.

لذلك، كان لابد من إيجاد طريقة تمكننا من بناء الأنتولوجي الخاصة بمجال معين بناءً آلياً وسريعاً.

وهنا اقترح عدد من الاتجاهات، ورافق كل اتجاه عدد من الأدوات الخاصة بلغة معينة ومنها العامة.

ولعل أهم طريقة اقترحت لبناء الأنتولوجي آلياً هي تعلّم الأنتولوجي من النصوص التي تصف مجالاً معيناً. وقد شهد هذا الاتجاه الكثير من الأبحاث نتج عنها عدد من الأدوات الخاصة بلغات محددة، تقوم هذه الأدوات باستخراج الأنتولوجي من مجموعة من الوثائق تتحدث عن مجال واحد، وتنتج بعد ذلك لخبراء هذا المجال إمكان التعديل على هذه الأنتولوجي والاتفاق على بنية أكثر ملاءمة، كما تنتج عدداً من الوظائف الهامة التي تقيد في تخزين واستخدام الأنتولوجي المستخرجة في التطبيقات المختلفة.

4. تعلّم الأنتولوجي من النصوص:

عندما يقوم مؤلف بكتابة كتاب أو مقال في مجال معين، فإنه يحاول إبراز وجهة نظره التي كونها خلال حياته عن هذا المجال. تتكون وجهة النظر هذه من مجموعة من الأفكار، يمكن التعبير عن هذه الأفكار بمجموعة من العلاقات بين عدد من المفاهيم التابعة لهذا المجال. للتعبير هذه الأفكار نصياً، يستخدم المؤلف التعابير التي تعبر عن المفاهيم أو عن العلاقات بين المفاهيم فيصاغ هذه الأفكار في نصوص مكتوبة بإحدى اللغات الطبيعية.

إذا عكسنا العملية السابقة، فإننا سوف نصل نظرياً إلى الأنتولوجي التي اعتمد عليها المؤلف في كتابته.

إذن، فالفكرة بشكلها المبسط السابق ممكنة وقابلة للتحقيق. ولكن الأمر أعقد من ذلك لعدة أسباب مثل استخدام أساليب لغوية تجعل النصوص غامضة وتحتل عدة أوجه، أو لأن أي نص مهما كان محدداً

لا يمكن أن يعبر عن أنتولوجي واحدة وإنما هو نتاج لكل ما مرَّ به الكاتب في حياته، وغيرها من الأسباب.

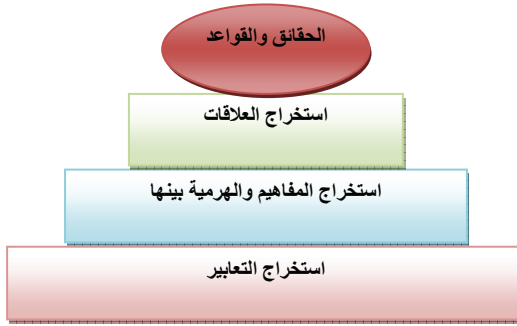
غير أننا إذا قصرنا عملنا على نصوص محددة ومكتوبة بلغة مبسطة نسبياً (الصحف اليومية، النصوص العلمية) يمكننا الوصول إلى نتائج مقبولة ومفيدة في العديد من التطبيقات. وتدرجياً، يمكننا زيادة مستوى التعقيد في الآليات المستخدمة في استخراج الأنتولوجي من النصوص حسب التطبيق وحسب النصوص المستخدمة.

يمكننا تحديد المراحل أو الطبقات الرئيسية في إجراء تعلم الأنتولوجي من النصوص كما يلي:

المرحلة الأولى: يجري استخراج التعابير اللغوية التي يحتويها النص، ويعرّف التعبير اللغوي بأنه عبارة لغوية تدل على مفهوم واحد مثل معهد أو مثل الحرب العالمية الثانية.

تعتبر عملية استخراج التعابير اللغوية من أهم المراحل في تعلم الأنتولوجي لأنها سوف تؤثر في جميع المراحل اللاحقة. يمكن أن يدل التعبير اللغوي على مفهوم معين أو على علاقة بين مفهومين أو أن يكون غرضاً من مفهوم أو من علاقة. سوف ينعكس غنى هذه المرحلة ودقتها على غنى ودقة الأنتولوجي المستخرجة.

المرحلة الثانية: بعد استخراج التعابير اللغوية، يجري استخراج المفاهيم والعلاقة الهرمية (التراتبية) بينها. يمكن أن تكون عملية استخراج المفاهيم منفصلة عن استخراج الهرمية، وقد تجريان في مرحلة واحدة حسب الآلية المستخدمة.



شكل 3 مراحل تعلم الأنتولوجي من النصوص

يتألف كل مفهوم من مجموعة من التعابير المستخرجة في المرحلة السابقة. تشترك هذه التعابير بالصفات والخصائص التي يجري استخراجها من النصوص. أما فيما يتعلق بالتراتبية أو ما نسميه الهرمية، فنقصد بذلك علاقة أب-ابن بين المفاهيم. ويعتبر المفهوم ابناً لمفهوم

آخر إذا كانت مجموعة خصائصه المستخرجة تحوي مجموعة خصائص المفهوم الأب.

المرحلة الثالثة: بعد استخراج المفاهيم والهرمية، يجري استخراج العلاقات غير الهرمية مثل علاقة الكل مع الجزء أو علاقة السببية أو العلاقات الناتجة عن الأفعال. لكل نوع من العلاقات السابقة طرائق خاصة لاستخراجها من النصوص آلياً، حتى أننا يمكن أن نعتبر العلاقات الهرمية نوع من العلاقات بين المفاهيم ونستخرجها في هذه المرحلة، غير أن معظم آليات استخراج العلاقات غير الهرمية تعتمد على العلاقات الهرمية بين المفاهيم وتعتبرها دخلاً أولاًً تعتمد عليه للحصول على المزيد من الدقة.

المرحلة الرابعة: أما موضوع استخراج الحقائق والقواعد، فهو لا يزال موضوعاً بحثياً ولا توجد حتى الآن آليات معتمدة لاستخراج الحقائق أو القواعد من النصوص بشكل مفيد وفعال، وذلك لعدة

أسباب منها: أن الحقائق قد لا تذكر أصلاً في هذه النصوص، وإن ذكرت فإنها تذكر وفق أساليب لغوية مختلفة لا يمكن تحديدها. كما أن القواعد لا تذكر بشكل واضح في النصوص، بل إن الإنسان نفسه قد يعجز عن تلمسها من كامل النص. لذلك، فإن معظم البيانات التي تعمل على تعلّم الأنطولوجي من النصوص تقوم ببتاحة إدخال الحقائق والقواعد يدوياً.

إن تعلّم الأنطولوجي من النصوص آلياً مهما بلغ من الدقة والفعالية، لن يعطي نتائج يمكن اعتمادها اعتماداً نهائياً؛ وإنما تحتاج هذه النتائج إلى العديد من المراحل اللاحقة للوصول إلى نموذج نهائي للأنطولوجي في مجال معين. بعض من هذه المراحل قد يكون آلياً مثل مقارنة الأنطولوجي وتقييمها ودمج الأنطولوجي، وبعضها يجري يدوياً. تنتج بيانات تعلّم الأنطولوجي من النصوص التدخل اليدوي في جميع المراحل، كما توفر بعض البيانات آليات تهدف إلى عرض الأنطولوجي المستخرجة على عدد من خبراء المجال لإبداء الملاحظات والتعديلات المقترحة. تُعتمد هذه التعديلات بعد الاتفاق عليها ويُعاد عرض الأنطولوجي مرة أخرى على الخبراء وهكذا إلى أن نصل إلى البنية التي يتفق عليها جميع الخبراء. غير أن معظم هذه البيانات هي بيانات خاصة بلغة معينة، وهي تعتمد بشكل كبير على خصوصية اللغة التي بنيت هذه الأداة لها لاستخراج مكونات الأنطولوجي. لذلك، كان من الضروري بناء أدوات مماثلة لاستخراج الأنطولوجي من النصوص العربية، لمواكبة التطورات الحاصلة في هذا المجال، ولبناء تطبيقات الويب الدلالي المختلفة والموجهة للغة العربية.

في إطار بناء نواة نظام تعلّم الأنطولوجي من النصوص العربية، قمنا بوضع الأرضية النظرية والبحثية والتصميم العام لنواة بيئة متكاملة، بهدف تعلّم الأنطولوجي من النصوص العربية، ومن ثم تعديل واستكمال هذه الأنطولوجي من قبل الخبراء، وتصدير الأنطولوجي المستخرجة إلى مختلف لغات توصيف الأنطولوجي المستخدمة في تطبيقات الويب الدلالي.

تتكون هذه النواة بشكل أساسي من أربع أنظمة جزئية هي: نظام إعداد الوثائق، ونظام استخراج التعابير، ونظام استخراج الهرمية، ونظام استخراج العلاقات غير الهرمية. تتكامل هذه الأجزاء الأربعة في نظام واحد لتحقيق الوظيفة الأساسية وهي تعلّم الأنطولوجي.

سنقوم في ما يلي بشرح مبسط وعرض موجز لمختلف الآليات التي اعتمدناها لاستخراج مختلف مكونات الأنطولوجي في النواة، وأهم النتائج التي وصلنا إليها.

5. استخراج التعابير اللغوية:

التعبير لغوي هو كل عبارة تشير إلى كائن مستقل بدلالته. قد يتألف التعبير من عدة كلمات، وهي الحالة الأعم في جميع اللغات وفي اللغة العربية بشكل خاص، مثل: "علم الاقتصاد" أو "سرطان الغدد للمفاوية"، وقد يتكون التعبير في بعض الحالات من كلمة واحدة مثل: "الاقتصاد" أو "المرض".

تستخدم التعابير في توصيف المكونات اللغوية التي تشكل المادة الرئيسة لتوصيف مجال معين، فمثلاً نجد لكل اختصاص تعبيره الخاصة ولكل مهنة تعبيرها الخاصة التي تميزها عن المهن الأخرى.

خلال مراحل بناء نظام استخراج التعابير بشكل آلي، قمنا باختبار عدد من الطرائق المعتمدة في لغات أخرى لاستخراج التعابير اللغوية. بعض هذه الطرائق لغوية، تعتمد على وضع نماذج لغوية

لاستخراج التعابير من النصوص بعد تحليلها صرفياً أو نحوياً؛ وبعضها إحصائية، تعتمد على الخصائص الإحصائية للتعابير اللغوية مثل تردد التعبير في النص وطول التعبير. ولكن جميع الطرائق السابقة غير كافية لبناء نظام آلي دقيق لاستخراج التعابير من النصوص. ويعود ذلك إلى سببين أساسيين، الأول: عدم وجود مقاييس إحصائية أو طرائق لغوية تمكنا من تقويم مدى ارتباط التعابير المستخرجة بسباق النص؛ فليس من المهم مثلاً استخراج تعبير من مجال الاقتصاد، إذا كان النص يتحدث عن مرض معين. السبب الثاني: هو أن كل طريقة من الطرائق السابقة -لغوية كانت أم إحصائية- تعبر عن عدد من خصائص التعبير اللغوي وتهمل باقي الخصائص، ولا توجد طريقة أو مقياس يمكنه التعبير عن كامل الخصائص.

1.5 الآلية المقترحة لاستخراج التعابير:

لبناء نظام آلي لاستخراج التعابير، قمنا بتقسيم عملية استخراج التعابير إلى ثلاث مراحل:

ترشيح العبارات:

يجري في هذه المرحلة استخراج العبارات المرشحة كتعابير لغوية. في هذه المرحلة، يجب أن نضمن ترشيح جميع التعابير الموجودة في النص. لتحقيق ذلك اعتمدنا نموذجاً لغوياً شديداً العمومية للتعبير اللغوي، وهو أن كل عبارة اسمية مرشحة لأن تكون تعبيراً لغوياً.

تقويم العبارات:

تهدف هذه المرحلة إلى تحديد دقة كل عبارة مستخرجة من المرحلة السابقة.

استخدمنا في هذه المرحلة مقاييس إحصائية: الأول تردد العبارة في جميع الوثائق التي وردت بها، والثاني تجانس التعبير. وقد جرى تطوير المقياس الأخير في سياق تطوير نواة النظام، وهو يعبر عن عدد من الخصائص الإحصائية للتعبير، كما يقدم حلاً للمشكلة الأولى التي أوردناها في الفقرة السابقة، وهي عدم وجود طريقة لقياس مدى ارتباط التعبير المستخرج بالسباق العام للنص.

المفاضلة بين العبارات:

للاستفادة من جميع التقويمات التي جرى تحديدها في المرحلة السابقة، قمنا ببناء نظام خبير يعتمد المنطق الترجيحي للمفاضلة بين العبارات المستخرجة. يعتمد هذا النظام على عدد من القواعد المبنية على المقاييس السابقة وتقويماتها، جرى استخلاص هذه القواعد تدريجياً أثناء مراحل تطوير النظام.

2.5 اختبار نظام استخراج التعابير:

جرى اختبار النظام على عينتين متوسطتي الحجم من الوثائق، نتحدث المجموعة الأولى عن علم الاقتصاد ونتحدث المجموعة الثانية عن مرض السرطان وأنواعه.

الوثائق التي نتحدث عن مرض السرطان هي الصفحات الناتجة عن عملية بحث عن مرض السرطان باستخدام محرك البحث غوغل، ويبين الجدول التالي التعابير التي استخرجها النظام من هذه الوثائق.

الطول	الدرجة	التجانس	التردد	التعبير
2	0.8375	0.4	57	سرطان الرئة
2	0.8375	0.4	94	سرطان البنكرياس
2	0.8375	0.4	75	سرطان الكبد
2	0.8375	0.4	58	بسرطان القولون
2	0.8375	0.4	51	للإصابة بسرطان
2	0.8375	0.4	65	مرض سرطان
2	0.8375	0.4	73	خلايا سرطانية
2	0.8375	0.4	49	علاج سرطان
2	0.81554878	0.224390244	68	والعلاج الكيميائي
1	0.8125	0.2	1025	سرطان
1	0.751307726	0.112195122	575	علاج
2	0.75	0.224390244	30	العلاج الإشعاعي
2	0.75	0.224390244	13	علاج المرض
2	0.75	0.4	15	الأورام السرطانية
2	0.75	0.4	15	أمراض سرطان

جدول 1 بعض التعبيرات المستخرجة من وثائق مرض السرطان

يعبر حقل التجانس في الجدول عن قيمة المقياس الثاني للتعبير، أما الدرجة فهي نتيجة نظام التقويم الترجيحي؛ نلاحظ أن التعبيرات التي تم استخراجها من الوثائق هي تعابير وفقاً للتعريف الذي أوردناه في البداية، ولكن لا نستطيع أن نتأكد من أن جميع التعبيرات التي وردت في الوثائق قد تم استخراجها. للتحقق من ذلك، لابد من استعراض الوثائق يدوياً من قبل عدد من الخبراء ومن ثم مقارنة النتائج مع نتائج النظام.

تمثل النتائج السابقة الخرج المباشر للنظام دون أي تعديل أو تدخل يدوي، لذلك نلاحظ أن العديد من التعبيرات المستخرجة تحتاج إلى إعادة صياغة وحذف بعض الأحرف الزائدة.

يمكن الاطلاع على نتائج النظام على المجموعة الثانية من الوثائق، وعلى عدد من الملاحظات والطرائق المقترحة لتقويم النظام بشكل دقيق في تقرير مشروع بناء نواة النظام [افضل الحسن، 2009].

6. استخراج المفاهيم الهرمية:

تعتبر الهرمية بين المفاهيم في كثير من المراجع التي تتحدث عن الأنطولوجي البنية الأساسية للأنطولوجي، وباقي العلاقات هي علاقات إضافية تتعلق بالتطبيق الذي تستخدم الأنطولوجي فيه. كما أن معظم تطبيقات استخراج العلاقات غير الهرمية تعتمد العلاقات الهرمية كدخل أولي لها.

عند استخراج الهرمية بين المفاهيم نكون ضمناً قد استخرجنا المفاهيم، ويتم الفصل بين المرحلتين فقط في الطرائق التي تعتمد المفاهيم كدخل لإيجاد الهرمية بينها وهي قليلة نسبياً.

الهرمية:

نقول عن مفهوم A أنه ابن أو توسعة لمفهوم B إذا كان A يمتلك جميع خصائص B. أو بطريقة أخرى إذا كانت جميع الأغراض من المفهوم A هي أغراض من المفهوم B.

1.6 آلية استخراج الهرمية من النصوص:

يوجد عدد من الطرائق المستخدمة في استخراج الهرمية من النصوص، منها لغوية ومنها إحصائية أو رياضية ولكل منها ميزاته ومشاكله، ولا يمكننا الاعتماد على إحدى هذه الطرائق فقط للحصول على هرمية كاملة.

الطرائق اللغوية:

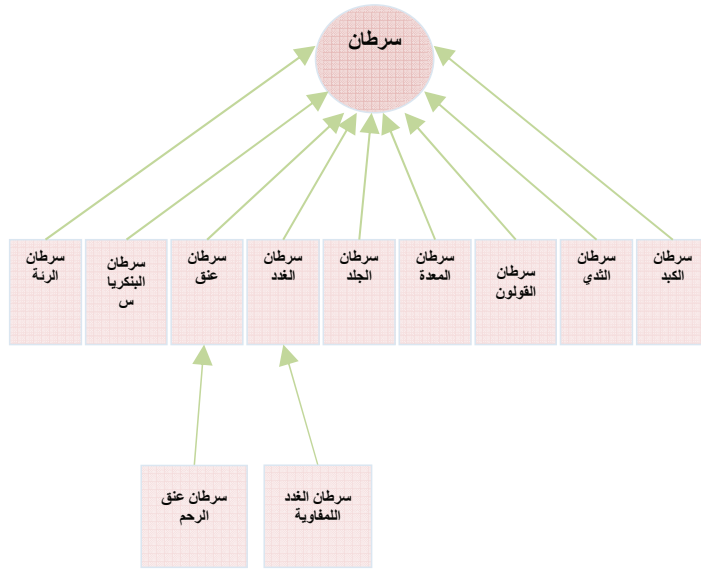
تعتبر النماذج اللغوية من أهم الطرائق اللغوية المستخدمة في استخراج الهرمية من النصوص، حيث يقوم الخبراء اللغويون بتعريف عدد من النماذج اللغوية للهرمية بين مفهومين في لغة معينة، وبعد تحليل النصوص نحويًا تستخرج أجزاء الهرمية الواردة في النصوص وفقاً للنماذج السابقة.

تتميز هذه الطريقة بصحة النتائج التي تعطيها، ولكنها غير كافية. كما أنها تحتاج إلى تحليل النصوص نحويًا. هذا عدا عن أن هذه النماذج اللغوية لا تستخدم إلا في نصوص تلتزم مستوى لغويًا معينًا، كالمعاجم والتعاريف.

لذلك، قمنا باستخدام طريقة لغوية بديلة للنماذج اللغوية، تسمى هذه الطريقة التجريبيات اللغوية. تتميز هذه التجريبيات بأنها مستمدة من نموذج لغوي معين ولكنها تكتب بأسلوب لغوي مبسط يستخدم بكثرة في النصوص اليومية، ولا يحتاج إلى التحليل النحوي لاستخراجه. نورد فيما يلي مثالين يوضحان الفرق بين النماذج اللغوية والتجريبيات اللغوية. من النماذج اللغوية، لدينا النموذج: {جملة اسمية} {هو، هي} {جملة اسمية}، ويعتبر المفهوم الذي تعبر عنه الجملة الأولى ابناً للمفهوم الذي تعبر عنه الجملة الاسمية الثانية.

أما من التجريبيات اللغوية فلدينا التجريبية: علاقة الأجزاء بين التعابير "مرض" و "مرض السرطان"، حيث يعتبر مرض السرطان ابناً للمرض. نلاحظ في التجريبية السابقة أننا لا نحتاج إلى التحليل النحوي للنص وإنما يكفي رسم هذا النص بالتعابير المستخرجة.

طبقتنا طريقة التجريبيات اللغوية على مجموعة من الوثائق نتناول موضوع مرض السرطان بعد أن استخرجنا التعابير باستخدام نظام استخراج التعابير الذي أشرنا إليه في الفقرة السابقة، فأعطى النظام مجموعة من الهرميات المهمة نعرض في (الشكل 4) جزء منها. نلاحظ أن جزء الهرمية الظاهر في الشكل يعبر عن أنواع مرض السرطان التي وردت في النصوص التي تمت معالجتها.



شكل 4 جزء من هرمية السرطان

جرى استخدام التعابير التي أعطاها نظام استخراج التعابير كدخلٍ أولي لنظام استخراج الهرمية دون أي تعديل على هذه التعابير، وذلك بهدف اختبار التكامل بين النظامين. أعطت أيضاً طريقة التجريبيات اللغوية عدد من الهرميات الأخرى كهرمية أنواع العلاج الواردة في الوثائق، كما أتاح استخدام عدد من التجريبيات المقترحة في سياق النظام الحصول على نتائج مفيدة مثل الهرمية بين المرض والسرطان. ولكن تبقى الطرائق اللغوية غير كافية لاستخراج كامل الهرمية الواردة في النصوص.

الطرائق الإحصائية:

معظم الطرائق الإحصائية المستخدمة في استخراج المفاهيم والهرمية بينها من النصوص تعتمد على ما يعرف بالصحة اللغوية للتعبير أو للمفهوم تماشياً مع القول الشائع: "قل لي من تصاحب أقل لك من أنت". فمن معرفة الكلمات المصاحبة للتعبير يمكن معرفة الكثير من خصائص هذا التعبير ومعرفة التعابير المشابهة أو المعممة له أو المخصصة له.

إذن، تعتمد الطرائق الإحصائية بدايةً على استخراج الكلمات المصاحبة لكل تعبير، نقوم بعدها بتطبيق إحدى الخوارزميات لتجميع هذه التعابير في مجموعات واستخلاص علاقات الهرمية بين هذه المجموعات.

من أهم الخوارزميات المستخدمة في بناء الهرمية بين مجموعة من التعابير، لكل منها عدد من الكلمات المصاحبة التي ترد معها بتردد معين، خوارزمية العقدة الهرمية. تبدأ هذه الخوارزمية باعتبار كل تعبير مفهوماً مستقلاً، ثم يجري دمج أقرب مفهومين في مفهوم أو عنقود أب اعتماداً على

صحة هذه المفاهيم وباستخدام مقياس للتشابه بين المفاهيم²؛ وهكذا حتى الوصول إلى مفهوم واحد يحوي جميع التعابير. لدى اختبار هذه الطريقة على مجموعة من الوثائق، وجدنا أن هذه الطريقة لا تعطي هرمية مفصلة وإنما تعطي نموذجاً عاماً للهرمية. فمثلاً إذا اعتمدنا هذه الطريقة في استخراج الهرمية من مجموعة وثائق تتحدث عن أصناف الحيوانات، فإننا نجد في أحد مستويات الخرج مفهومين يحوي الأول منهما، الحيوانات الفقارية، والثاني يحوي الحيوانات اللافقارية؛ وأما باقي المستويات فهي تفصيلات، غير دقيقة غالباً، في أصناف كل من المفهومين.

الطرائق الرياضية:

توجد العديد من الأبحاث الرياضية المتعلقة بإيجاد ما يعرف بشبكة المفاهيم أي الهرمية، بين مجموعة من الأغراض يمتلك كل منها عدد من الصفات. من أهم هذه الأبحاث الخوارزمية التي قدمها العالم الفرنسي بوردا عام 1986 لإيجاد شبكة المفاهيم. أهم ما تتميز به هذه الخوارزمية هو قدرتها على إيجاد المفاهيم والهرمية بينها في نفس المرحلة. ولكن هذه الخوارزمية ذات تعقيد أسي وقد يكون استخدامها مستحيلاً في بعض التطبيقات التي تمتلك فيها الأغراض عدداً كبيراً نسبياً من الصفات. قمنا في إطار البحث بتقديم دراسة مفصلة لتسريع عمل هذه الخوارزمية. ولدى اختبارنا أداء هذه الخوارزمية على مجموعة من الوثائق توصلنا إلى الاستنتاجات التالية:

إن دقة النتائج التي تعطيها الطرائق الرياضية متعلق إلى حد كبير بدقة الطريقة المستخدمة لاستخراج الصفات التي تصف الأغراض (التعابير)، و تصل إلى نتائج مثالية إذا كانت هذه الصفات موضوعاً بشكل دقيق. ولكن في تطبيق يقوم باستخراج هذه الصفات من النصوص وجدنا أن هذه الطرائق لا تعطي النتائج المرجوة منها، وذلك لأن الصفات التي ترد في نص معين وتصف تعبيراً معيناً، تعاني من بعض النواقص التي تعيق عملية بناء الهرمية. من أهم هذه النواقص عدم ورود جميع الصفات التي تصف تعبيراً محدداً في النص، بالإضافة إلى أن عدد الصفات التي ترد في نص ما وتصف مفهوماً أباً لمفهوم آخر، يكون غالباً أكبر من عدد صفات الابن؛ وهذا يعاكس المبدأ الذي تعتمد هذه الطرائق في بناء الهرمية، وهو اعتبار مفهوم ما ابناً لمفهوم آخر، إذا كان يمتلك جميع صفاته. لذلك، لكي نستخدم هذه الطرائق استخداماً فعالاً، لا يكفي أن نعتمد على المعطيات النصية في تحديد الصفات، وإنما نحتاج إلى المعاجم مثلاً أو حتى إلى أنطولوجي قديمة.

في النهاية، لا بد للوصول على نتائج مقبولة في عملية استخراج الهرمية من النصوص من دمج جميع النتائج السابقة كي تكمل كل طريقة من الطرائق السابقة ما تغفله الطرائق الأخرى. يمكن الاطلاع على التوابع والخوارزمية المقترحة لدمج الهرميات في [فاضل الحسن، نظام استخراج الهرمية، 2009]. كما يمكن الاطلاع على الدراسة المقدمة لتسريع خوارزمية بوردا في نهاية الفصل للمرجع ذاته.

7. استخراج العلاقات غير الهرمية:

تختلف العلاقات غير الهرمية عن العلاقات الهرمية بكثرتها واختلافها، مما يجعل عملية استخراجها من النصوص الجزء الأكثر تعقيداً في عملية تعلم الأنطولوجي [Maedche, 2002]. تعتبر عملية استخراج العلاقات غير الهرمية بين مجموعة من التعابير أو الأغراض عملية هامة لعدد من

² اعتمدنا في نظامنا تابع تجيب الزاوية بين الشعاعين الممثلين لكل تعبير، والذي يمكن حسابه ببساطة بدلالة الجداء السلمي للشعاعين .

المجالات الأخرى (الهندسة الجينية، مجالات الأعمال والتسويق،)، وقد جرى بناء عدد من الطرائق والخوارزميات لاستخراج أنواع العلاقات المختلفة من مجموعة من المعطيات. يمكن تصنيف العلاقات بين المفاهيم في عدة مجموعات، من أشهر هذه التصنيفات {علاقة الجزء بالكل، علاقة السببية، العلاقات الناتجة عن الأفعال}. كما توجد بعض التصنيفات ضمن كل مجموعة من المجموعات الثلاث السابقة.

1.7 آلية استخراج العلاقات غير الهرمية:

كما هو الحال في عملية استخراج الهرمية، يوجد اتجاهان أساسيان في عملية استخراج العلاقات غير الهرمية من النصوص: اتجاه لغوي واتجاه إحصائي.

الطرائق اللغوية:

يجري تحديد عدد من النماذج اللغوية لصياغة نوع معين من العلاقات غير الهرمية. وبعد تحليل النصوص تحليلاً نحوياً، تستخرج العلاقات الواردة في النص والتي تحقق أحد النماذج. تتميز هذه الطريقة بدقتها، بالإضافة إلى إمكان تحديد اسم أو نوع العلاقة المستخرجة، غير أنها محدودة بأنواع معينة من العلاقات وبنصوص معينة، كما أنها تحتاج إلى تحليل النص نحوياً. لذلك قمنا باستخدام طريقة التجريبيات اللغوية التي استخدمناها في استخراج العلاقات الهرمية، حيث قمنا بتحديد عدد من التجريبيات اللغوية، لنوع معين من العلاقات والتي يمكن استنتاجها من أحد النماذج اللغوية. كمثال على هذه التجريبيات التجريبية **إيعاني (فعل) مريض السرطان (تعبير) من (حرف جر) نقص في الوزن (تعبير)**. تستخرج هذه التجريبيات من النصوص البسيطة دون الحاجة إلى تحليل النص تحليلاً نحوياً.

الطرائق الإحصائية:

تختلف الطرائق الإحصائية عن الطرائق اللغوية بعموميتها حيث تستخرج جميع أنواع العلاقات الموجودة في النصوص، ولكنها غالباً لا تحدد اسماً لهذه العلاقات. هناك العديد من الطرائق الإحصائية المستخدمة لاستخراج العلاقات غير الهرمية بين مجموعة من التعابير أو المفاهيم؛ والموجودة في مجموعة من المعطيات. تختلف هذه الطرائق حسب نوع المعطيات المستخدمة (قواعد معطيات، نصوص، تقارير) ولكن يمكننا في معظم هذه الطرائق فصل الجزء الخاص بنمط المعطيات وتطبيق هذه الطريقة على مختلف أنواع المعطيات. تراعي كل طريقة من الطرائق السابقة عدداً من الجوانب مثل زمن التنفيذ ودقة النتائج. ولكننا قمنا خلال بناء نواة النظام بتطوير آلية خاصة بنا. تتألف هذه الآلية من مرحلتين:

مرحلة إحصائية: يجري فيها استخراج جميع العلاقات الواردة في النصوص والتي تتجاوز عتبة تردد معين (نقصد بتردد العلاقة عدد مرات ورود مجموعة التعابير المكونة للعلاقة مع بعضها البعض في النصوص)، وعتبة ثقة معينة (نقصد بالثقة الاحتمال الشرطي لورود الطرف الثاني من العلاقة عند ورود الطرف الأول). استخدمنا في هذه المرحلة إحدى الخوارزميات الإحصائية المستخدمة في مجال التنقيب بالمعطيات والتي تقوم باستخراج العلاقات المحققة لشرطي التردد والثقة السابقين دون أن تحتاج إلى المرور على جميع العلاقات الممكنة [Han, et al., 2005].

مرحلة تصحيح بالهرمية: من المهم في عملية استخراج العلاقات من النصوص اعتبار الهرمية بين المفاهيم التي تملك هذه العلاقات، وذلك لاستخراج علاقات أكثر عمومية، حيث يجري إعادة صياغة ودمج العلاقات الخاصة مع العلاقات الأكثر عمومية اعتماداً على الهرمية بين المفاهيم. كما يفيد

استخدام الهرمية قبل التقويم باستخدام الشرطين الإحصائيين في المرحلة السابقة، في استخراج علاقات مهمة في النصوص وردت بأشكال متعددة وفقاً لهرمية معينة، كأن يرد مفهوم معين مع كل ابن من أبناء مفهوم آخر عندها يكون تردد كل علاقة من العلاقات السابقة أقل من العتبة المحددة ولا تُستخرج تلك العلاقة مطلقاً [Maedche, 2002].

عند اختبار الطريقة الإحصائية السابقة على مجموعة الوثائق التي تتحدث عن مرض السرطان واستخدام التعابير المستخرجة باستخدام نظام استخراج التعابير المبني ضمن النظام، كان من نتائج المرحلة الأولى العلاقات التالية:

{السرطان} ← {العلاج، علاج السرطان}

{السرطان} ← {أعراض السرطان}

{العلاج} ← {طرائق العلاج}

{السرطان} ← {تشخيص المرض}

{السرطان} ← {الخلايا، الخلايا السرطانية}

{السرطان} ← {أنواع السرطان}

كما جرى استخراج عدد من العلاقات غير المهمة.

تحتاج النتائج السابقة إلى مرحلة التصحيح بالهرمية عندها سوف نجد نتائج أكثر عمومية وأكثر أهمية كما يمكن أن يظهر عدد من العلاقات الجديدة المهمة. ولكننا لم نقم في هذه النسخة من النواة ببناء هذه المرحلة بعد، لعدد من الأسباب من أهمها عدم استقرار نظام استخراج الهرمية.

8. الختام:

يعتبر تعلم الأنتولوجي من النصوص الحل الأبرز لمشكلة بناء الأنتولوجي التي يحتاجها الوب الدلالي في تطبيقاته المختلفة، لكن عملية التعلم بشكل آلي مهما بلغت دقتها فإنها لن تعطي نتائج نهائية كاملة وصحيحة، لذا لابد من مرحلة لاحقة لتعديل هذه النتائج واستكمالها والتوافق عليها من قبل الخبراء. قدمنا في هذا التقرير لمحة سريعة عن البنية العامة لنواة نظام تعلم الأنتولوجي من النصوص العربية، وأهم ما أنجز من هذه النواة، كما عرضنا بعض نتائج اختبار النواة في المراحل المختلفة للتعلم.

اعتمدنا في المراحل المختلفة لبناء النواة، على ما تم الوصول إليه من نتائج لغوية في لغات أخرى مثل: اللغة الألمانية والإنكليزية والصينية، وعلى الخبرة الشخصية التي تكون معظمها في مراحل العمل ونتيجة لعدد من الدراسات السريعة والمبسطة، لذلك لابد لنا لبناء تطبيقات لغوية دلالية للمحتوى العربي من توفر الخبرة اللغوية اللازمة، وقد حصرنا هذه الخبرة في ثلاث نقاط رئيسية هي: تحديد النماذج اللغوية المعتمدة في اللغة العربية لبناء المكونات والعلاقات الدلالية المختلفة. وضع التصنيفات الدلالية المختلفة للأفعال والأسماء والتي تستخدم في تطبيقات الوب الدلالي. الدراسات الإحصائية لمختلف المكونات الدلالية مثل خصائص التعبير الإحصائية في اللغة العربية أو خصائص الكلمات التي تصف مفهوم معين الخ.

استغنيا في هذه النسخة من النواة عن التحليل النحوي للنصوص اللغوية، وقدمنا طريقة بديلة هي التجريبيات اللغوية والتي تعطي حلاً فعال ومفيد في العديد من الحالات، ولكن لابد مستقبلاً من الاعتماد على التحليل النحوي للوصول إلى نتائج أكثر دقة وغنى.

تعتبر النتائج التي تعطيها النواة كافية لبعض تطبيقات الويب الدلالي مثل العنقدة واستخراج المعلومات، ولكن هذه النواة تعتبر مرحلة من مراحل مشروع متكامل لبناء تطبيقات الويب الدلالي للمحتوى العربي. لذا لابد مستقبلاً من استكمال بناء النواة واختبارها، ولابد أيضاً من بناء باقي الأجزاء الأخرى التي تحتاجها تطبيقات الويب الدلالي مثل أدوات تخزين الأنطولوجي والبحث فيها، وأدوات رسم الوثائق بالأنطولوجي وتصنيف هذه الوثائق اعتماداً على هذا الوسم، أدوات عرض الأنطولوجي، الخ.

1. **فاضل الحسن.** *نواة نظام تعلم الأنطولوجي من النصوص العربية*. دمشق، 2009.
2. **فاضل الحسن.** *نظام استخراج الهرمية: نواة نظام تعلم الأنطولوجي من النصوص العربية*. دمشق، 2009.
3. **Maedche, Alexander.** *Ontology Learning For The Semantic Web*. London : Kluwer Academic Publishers, 2002.
4. **Han, Jiawei and Kamber, Micheline.** 2005. *Data Mining: Concepts and Techniques*. s.l. : Morgan Kaufmann Publishers, 2005.
5. *arabency.net*. [Online]
6. **Driankov, D, Hellendroon, H and Reinfrank, M.** *An Introduction to Fuzzy Control*. s.l. : Springer-Verlag, 1993.
7. **Buitelaar, Paul, et al.** *Ontology Learning from Text*. Porto, Portugal : s.n., 2005. Tutorial at ECML/PKDD 2005.
8. **Davies, John, Fensel, Dieter and Harmelen, Frank van.** *TOWARDS THE SEMANTIC WEB*. West Sussex, England : JOHN WILEY & SONS, LTD, 2003.

المصطلحات:

العنقدة	Clustering
استدلال	Inference
التنقيب بالمعطيات	Data mining
العنقدة الهرمية	Hierarchical clustering
أنطولوجي	Ontology
تجريبيات لغوية	Linguistic heuristic
تعابير	Terms
عامل الثقة	Certainty factor
قاعدة معرفة	Knowledge base
هرمية	Taxonomy

Ontology learning system from Arabic texts

Fadel Al Hassan, Ghaida Rebdawi, Bassam Kurdy

Higher Institute of Applied Sciences and Technology

p.o.Box 31983, Damascus, Syria

Email: fadel.alhassan@hiast.edu.sy

ghaida.rebdawi@hiast.edu.sy, bassam.kurdy@hiast.edu.sy

Textual documents about certain domain contain full or partial specification of that domain. In other words, the author uses texts to describe his ontology about the world or about a specific domain, using a natural language. If we can reverse writing process (e.g. from text to the ontology), we will go a long way in ontology building. We present in this paper a brief definition to the ontology and its role in semantic web, with an introduction to the Arabic ontology learning system and its parts which include Terms Extraction, Taxonomy Extraction and Non-taxonomic Relations Extraction. We conclude with the most important results that we obtained from our ontology learning system from Arabic documents.

من التحليل الصرفي إلى إدراك وإصلاح أخطاء التطابق في نصوص عربية غير مشكولة

مكرم بوجلبان، نهى شعبان كمون، شفيق علولو، لمياء هدرش بلغيث
فريق بحث إن.إل.ب.^١، مخبر ميراكل، جامعة صفاقس، كلية العلوم الاقتصادية والتصرف
بصفاقس، ص.ب. 1088، الرمز البريدي 3018، صفاقس، تونس.
الهاتف: 74 278 777 (216)، الفاكس: 74 279 139 (216)

boujelben_makram@yahoo.fr , nouha.chaaben@gmail.com,
Chafik.aloulou@fsegs.rnu.tn, l.belguith@fsegs.rnu.tn

الخلاصة: تقدّم هذه الورقة منظومة "مورف2" للتحليل الصرفي ومنظومة "ديكور2" لكشف وإصلاح أخطاء التطابق في نصوص عربية غير مشكولة والتي تتمثل في أخطاء في الجنس والعدد والضمير والتعريف والإعراب وصفة العاقل. وتتكوّن منظومة "ديكور2" من عدّة برمجيات متكاملة وهي كالآتي: برنامج المحلل الصرفي مورف2، برنامج "مصرف الأفعال" العربية غير المشكولة، برنامج "مصنغ الكلمات" الذي يحتوي على معجم الكلمات المفردة مرفقة بصيغتها في الجمع، برنامج لكشف الأخطاء الذي يستعمل نتائج مورف2 لتقسيم الجملة إلى عدّة مركبات ثمّ يجمع المركبات التي توجد بينها علاقة تطابق ومن ثمّ يدقق في خصائصها الصرفية لإدراك الأخطاء، برنامج لإصلاح الأخطاء الذي يعتمد على نتائج كشف الأخطاء ليقوم بإصلاحها مستعينا ببرنامج مصنغ الكلمات الذي يمكن من تصريف فعل حسب الضمير والزمان المناسبان وبرنامج مصنغ الكلمات الذي يمكن من صياغة كلمة حسب الجنس و العدد المناسبان.

الكلمات الجوهرية: التحليل الصرفي، التحليل التركيبي، أخطاء التطابق، تصريف الأفعال، صياغة الكلمات.

1. المقدمة

تعتبر منظومة التحليل الصرفي ذات أهميّة بالغة لا يمكن أن تستغني عنها منظومات أخرى مثل منظومة كشف وإصلاح الأخطاء اللغوية والإملائية، ويمكن إدماجها في عدّة برمجيات أخرى مثل منظومات معالجة النصوص والترجمة الآليّة والتعليم بمساعدة الحاسوب والتلخيص الآلي الخ.... وفي وقتنا الحاضر لا توجد منظومة إصلاح جيّدة لهذه الأنواع من الأخطاء في اللغة العربية نظرا لوجود مشاكل معقّدة ناتجة عن ثراء هذه اللغة تجعل من الصّعب حلّها بطريقة آليّة مثل رسم الهمزة وغياب الشكل في معظم النصوص و ترتيب الكلمات في الجملة العربية والكلمات المركّبة إضافة إلى مشاكل التحليل الصرفي. كلّ هذا من شأنه أن يعيق المعالجة الآليّة

¹ Arabic Natural Language Processing Research Group (ANLP-RG): <http://sites.google.com/site/anlprg/>

للغة العربية على العديد من المستويات سواء كان على المستوى الصّرفي أو اللفظي أو النحوي أو الدلالي [1].

نقدّم في هذه الورقة في مرحلة أولى المحلل الصّرفي مورف2 الذي يمثل جزءاً مهماً من مراحل منظومة ديكور2. ثمّ نقدّم بعد ذلك منظومة ديكور2 لكشف وإصلاح أخطاء التّطابق في النّصوص العربيّة غير المشكولة والتي تتمثل في أخطاء في الجنس والعدد والضمير والتّعريف والإعراب وصفة العقل. وتجدر الإشارة أنّ منظومة ديكور2، والتي تمثّل نسخة مطوّرة لمنظومة ديكور1، تقترح العديد من احتمالات الإصلاح عند الكشف عن خطأ في التّطابق وتنتج للمستعمل اختيار الاحتمال المناسب [2]. في ختام هذه الورقة، نقدّم النتائج التي توصّلنا إليها على إثر تقييم ديكور2.

2. تعريف أخطاء التّطابق والطرق المعتمدة لمعالجتها

بدأت البحوث العلميّة في مجال كشف وإصلاح الأخطاء في أواخر الخمسينات، ونذكر من بينها أعمال قلانتز [3]، بلار [4] و داميرو [5] وقام هؤلاء بعدّة محاولات لإدراك وإصلاح الأخطاء الإملائية في نصوص مكتوبة باللغة الإنجليزيّة. ولم تنطلق البحوث في مجال معالجة أخطاء التّطابق إلا في الثمانينات، نذكر على سبيل المثال أعمال هيدورن [6] بالنسبة للغة الإنجليزيّة وأعمال لابلالم وريتشارد [7] بالنسبة للغة الفرنسيّة وأعمال بلغيث [8] في بالنسبة للغة العربيّة. ويبدو لنا أنّ جُلّ منظومات الإصلاح الأخرى أقدر على معرفة الأخطاء الإملائية عن غيرها من أخطاء التّطابق نظراً لصعوبة كشفها. [7].

وقد قسم شومسكي أخطاء التّطابق، والتي تنتج في العديد من الحالات عن عدم احترام القواعد اللغويّة، إلى قسمين: الأخطاء الكفائية التي تنتج عن سهو أو عدم إتباع القواعد اللغويّة، والأخطاء الأدائية التي تنتج عن أسباب تقنيّة مثل أخطاء الرّقن وعدم التركيز. هذه المصطلحات استعملها عديد الباحثين من بينهم فيرونيس [9] و فوكوري [10] و جونسبال [11] و بلغيث [12].

أمّا عن الطرق المعتمدة لإدراك الأخطاء فيمكن تقسيمها إلى نوعين: الأولى تعتمد على تحليل لفظي معمّق اعتماداً على قواعد نحويّة معمّقة أمّا الثانية فتعتمد على تحليل جزئي للجملة قصد التعرف على تركيبها اعتماداً على بعض الكلمات الخاصّة والتي أطلق عليها "كلمات مفاتيح". وإصلاح هذه الأخطاء يمكننا إتباع العديد من الطرق من بينها الطريقة اللسانية، الطريقة الحاسوبية أو الطريقة الهجينة التي تجمع بين كلتا الطريقتين. تجدر الإشارة أنّ الطريقة الأولى تعتمد على بعض الكلمات المسيطرة التي تفرض خصائصها الصرفية على بقية كلمات الجملة. أمّا الطريقة الثانية فإنّها تسعى إلى تقليل عدد التّغييرات الممكنة على مستوى الكلمات في الجملة المراد إصلاحها.

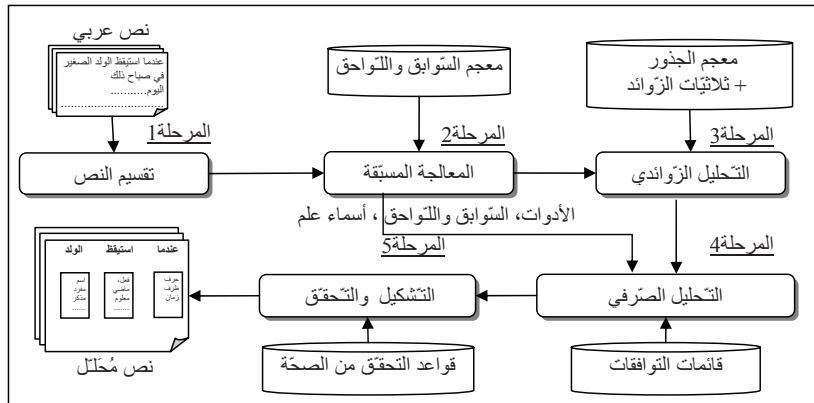
في هذا الإطار، اعتمدنا على طريقة التحليل الجزئي لكشف الأخطاء باعتبارها الأسرع وعلى الطريقة الهجينة لإصلاح الأخطاء نظراً لنجاحاتها في عملية ترتيب الإصلاحات من الأكثر إلى الأقل احتمالاً.

3. المحلل الصرفي مورف2

3.1. طريقة التحليل الصرفي المعتمدة

يُعتبر التحليل الصرفي أحد أهم مراحل المعالجة الآلية للغة الطبيعية ويتمثل دور هذه المرحلة في التعرف على المكونات الأساسية للكلمة واستخراج الخصائص الصرفية والنحوية المناسبة والتي تُعدُّ مهمة بالنسبة للمراحل الأخرى من المعالجة الآلية للغة. ومن هذا المنطلق تثبت أهمية انجاز محلل صرفي يتميز بالشمولية وسهولة الاستعمال والإدماج في أنظمة المعالجة الآلية للغة الطبيعية ذات النطاق الواسع نذكر منها نظم التحليل الآلي للنصوص ونظم الإصلاح الآلي للأخطاء ونظم الترجمة الآلية ونظم التلخيص الآلي. أما طريقة التحليل الصرفي المعتمدة [13] فتتضمّن إلى الطرق اللغوية التي تستطيع بالاعتماد على (1) معجم يحتوي على أقلّ ما يمكن من المعلومات (جذور، أسماء غير مشتقة، الخ) و(2) أيضاً على بعض القواعد الصرفية من استنتاج كلّ الخصائص الصرفية للمصطلحات العربية. وتتكوّن هذه الطريقة من خمسة مراحل (انظر الشكل1):

- تقسيم النص : يتمثل دور هذه المرحلة في تقسيم النص إلى فقرات وجمل ومن ثمّ يقع تقسيم كل جملة إلى كلمات. يعتمد تقسيم النص إلى فقرات و جمل على الاستكشافات السياقية لعلامات التنقيط ولأدوات الربط بين الجمل بالإضافة إلى بعض الأدوات الأخرى [14]. أمّا تقسيم الجملة إلى كلمات فيقتصر على تحديد الكلمات الموجودة بالجملة بالاعتماد على الفراغات بين الكلمات وعلامات التنقيط وبعض الحروف. كلّ كلمة مستخرجة في هذه المرحلة تخضع للمراحل الموالية من التحليل الصرفي.



الشكل 1. طريقة التحليل الصرفي المعتمدة

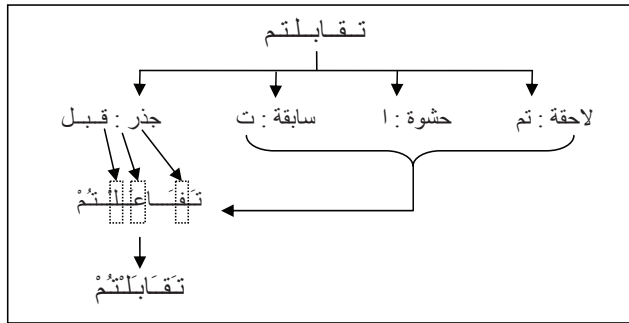
- المعالجة المسبقة : تمكّنتنا هذه المرحلة من التعرف على السوابق والتواحق الملصقة بالكلمة بالاعتماد على قوائم السوابق والتواحق المخزّنة ذلك أنّها تمثّل وحدات لغوية لها خصائصها الصرفية الخاصة بها. وتتألف قائمة السوابق من بعض حروف الجر وحروف العطف وأدوات الاستفهام وغيرها من الأدوات التي لا يمكن أن توجد إلا ملتصقة بكلمة أخرى في حين تتكوّن

قائمة اللّواحق من الضّمائر المتّصلة. بعد ذلك نقوم بعملية بحث عن الكلمة المتبقية في قائمات الأدوات وأسماء العلم والأعداد والتواريخ. في صورة عدم تواجد الكلمة في إحدى هاته القائمات نمر إلى مرحلة التحليل الزواني.

● التحليل الزواني : يهدف التحليل الزواني إلى التعرّف على المكونات الأساسية لتركيبية الكلمة ألا وهي ثلاثية الزوائد (سابقة، حشوة، لاحقة) والجذر. وينقسم التحليل الزواني إلى عدة مراحل أهمّها : التعرّف على الترابطات (سابقة، لاحقة)، التعرّف على ثلاثيات الزوائد الممكنة، التّصفية المعجمية، التحقق من الصّلات "جذر" و (سابقة، حشوة، لاحقة) والتعرّف على التحوّلات الممكنة في الكلمة. كما تقوم عمليات تصفية بحذف التحليلات الطفيلية المتعرّف عليها من مرحلة إلى أخرى [15].

● التحليل الصّرفي : تقوم هذه المرحلة بالاعتماد على الأنموذج (جذر، سابقة، حشوة، لاحقة) الذي وقع استخراجه في المرحلة السابقة لاستخراج كلّ الخصائص الصّرفية النّحوية الممكنة (الصّفة النّحوية، الجنس، العدد، الضّمير، الصّيغة، إلخ). وتنقسم هذه المرحلة بدورها إلى ثلاثة مراحل [2]. يتمثّل دور المرحلة الأولى في تحديد الصّفة النّحوية للكلمة (اسما أو فعلا أو أداة). أمّا المرحلة الثانية فتهمّ بتحديد بقية الخصائص الصّرفية النّحوية لكلّ صفة نحوية مستخرجة للكلمة. بعد ذلك تقوم المرحلة الثالثة بعملية تصفية لقائمات الخصائص المستخرجة قصد حذف القائمات غير الصّحيحة.

● التّشكيل والتّحقق : تقوم هذه المرحلة بعملية التّشكيل اعتمادا على الخصائص الصّرفية النّحوية والأنموذج (جذر، سابقة، حشوة، لاحقة) المستخرجين في المراحل السابقة. فتقع إضافة الحركات لكلّ تحليل ممكن للكلمة وفق خصائصها الصّرفية النّحوية. وللقيام بعملية التّشكيل يقع دمج الجذر المُستخرج في مرحلة التحليل الزواني بتفعيل الكلمة المشكولة كليّا كما سنوضحه في المثال التالي.



الشكل 2. عملية تشكيل كلمة "تقابلتُم"

مثال : اثر مرور كلمة "تقابلتُم" بمرحلة التحليل الزواني نتحصّل على الأنموذج (جذر، سابقة، حشوة، لاحقة) الخاص بها والذي يتمثّل في (قبل، ت، ا، تم). بعد ذلك يقع استخراج التفعيل المشكولة كليّا والموافقة لثلاثية الزوائد (ت، ا، تم) لنحصّل على "تَقَاعَلْتُمْ". ثمّ يقع دمج الجذر المستخرج "قبل" مع التفعيل المشكولة المستخرجة "تَقَاعَلْتُمْ" لنحصّل على الكلمة المشكولة "تَقَابَلْتُمْ" (انظر الشكل 2).

إلا أنّ عمليّة التشكيل يمكن أن تضفي إلى حلول غير صحيحة وذلك يعود إلى طبيعة التصريف في اللّغة العربيّة الذي يتضمّن تحوّل وحذف وإدغام بعض أحرف الكلمة في عديد الأحيان خاصّة عندما يحتوي جذر الكلمة على حروف العلة أو الهمزة. وحتىّ نأخذ بعين الاعتبار كلّ هذه التغيّرات نقوم في مرحلة التحقق من الصّحة بالتّثبت من صحّة الكلمة المشكولة وذلك بالاعتماد على قواعد التحول وقواعد الحذف وقواعد الإدغام المخزّنة للحصول على شكل صحيح للكلمة المحلّلة.

3.2. المحلّل الصّرفي مورف2

يعتمد المحلّل الصّرفي مورف2 على طريقة التّحليل الصّرفي المفصّلة سابقاً لتحليل الكلمات العربيّة. وقد وقعت برمجة هذا المحلّل بالاعتماد على لغة البرمجة "جافا" كما يعتمد المحلّل على معجم مختصر إلى جانب مجموعة من القواعد اللّغويّة.



الشكل 3. نافذة نتيجة تحليل "التحقق بالمدرسة"

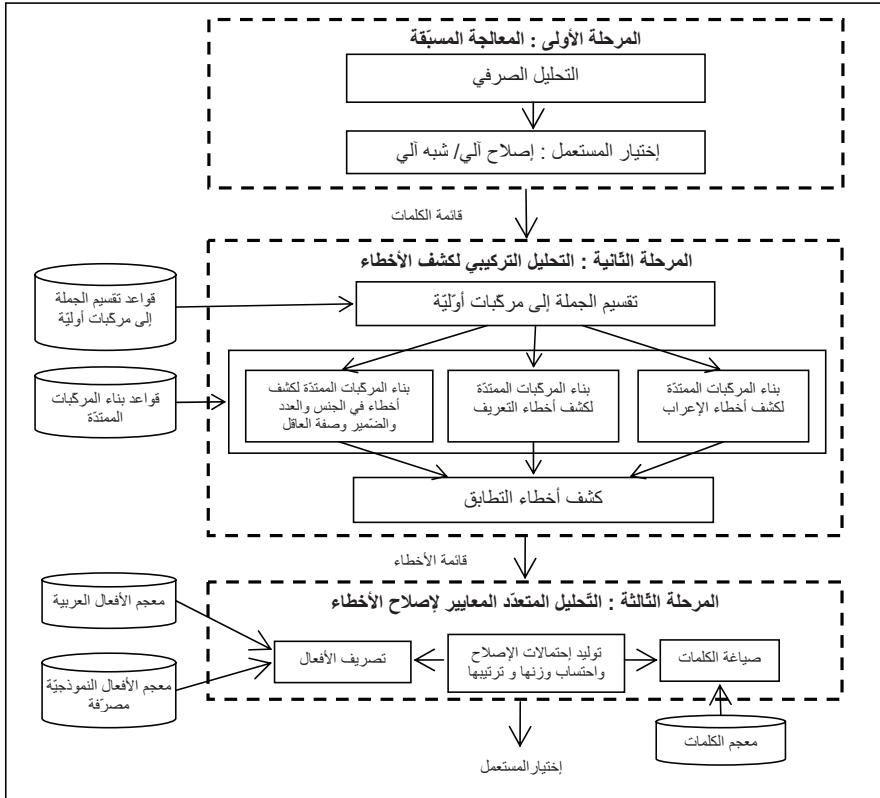
يبين الشكل 3 نافذة تشغيل مورف2 من خلال مثال تحليل "التحقق بالمدرسة" إذ تمكّن من التّعرّف على كلّ الخصائص الصرفية النحوية الممكنة لكلّ أجزاء الكلمات الموجودة في الجملة.

4. طريقة معالجة أخطاء التطابق

تمرّ عملية معالجة أخطاء التطابق بثلاث مراحل وهي على التوالي : مرحلة المعالجة المسبقة، مرحلة التحليل التركيبي لكشف الأخطاء ومرحلة التحليل المتعدد المعايير لإصلاح الأخطاء [19]. تعتمد كلّ مرحلة على نتائج المرحلة التي تسبقها ويبيّن الشكل 5 كيفية تسلسل هذه المراحل.

4.1. المرحلة الأولى: المعالجة المسبقة

تنقسم المعالجة المسبقة إلى مرحلتين: التحليل الصرفي واختيار المستعمل. يتم التحليل الصرفي وفق الطريقة المفصلة في الفقرة 3 إذ يهتم بإسناد الخصائص الصرفيّة والنحويّة لكلّ كلمة يراد إصلاحها. وتجدر الإشارة إلى أنّ هذه الخصائص (التي تعرف بخطوط التطابق) ستستعمل لاحقاً في عمليات كشف وإصلاح الأخطاء.



الشكل 5. طريقة معالجة أخطاء التطابق

بعد ذلك يمكن للمستعمل أن يختار كيفية الإصلاح، بمعنى أنه يمكن أن يختار بطريقة يدوية الخصائص المناسبة لكل كلمة في الجملة من خلال النتائج المستخرجة في مرحلة التحليل الصرفي. كما يمكن أن يتم الإصلاح بطريقة آلية ويكون ذلك باختيار الاحتمال الأول لكل كلمة. إن اختيار الإصلاح شبه الآلي من شأنه أن يبسط عملية الإصلاح ولكنه يحسن النتائج النهائية بشكل ملحوظ خاصة في حالات اللبس. أما بالنسبة للإصلاح الآلي، فإنه أكثر سرعة لكن ذلك يكون على حساب دقة النتائج.

4.2. المرحلة الثانية : التحليل التركيبي لكشف الأخطاء

تنقسم هذه المرحلة إلى ثلاث مراحل وهي تعمل بطريقة متسلسلة كالآتي: مرحلة تقسيم الجملة إلى مركبات أولية، مرحلة تجميع المركبات الأولية إلى مركبات ممتدة حسب نوع الخطأ المراد إصلاحه ومرحلة كشف الأخطاء.

4.2.1. مرحلة تقسيم الجملة إلى مركبات أولية

تهتم هذه المرحلة بتقسيم الجملة إلى عدة مركبات أولية اعتماداً على مجموعة من الكلمات "الأدوات" التي نسميها "الحدود التركيبية" ونذكر منها: حروف الجر، حروف العطف، أدوات النفي، أدوات الاستفهام، أدوات النصب، أدوات الجزم، أدوات الاستثناء، ظروف المكان، ظروف الزمان، كان وأخواتها، إن وأخواتها، كاد وأخواتها، أفعال الظن، أفعال الرجاء، أفعال المقاربة والشرع، الأسماء الموصولة، أسماء الإشارة، الضمائر المتصلة، الضمائر المنفصلة، إلخ... وتعتمد عملية التقسيم على مجموعة من القواعد بحيث نتحصل في النهاية على مجموعة من المركبات الأولية جاهزة للتجميع.

4.2.2. مرحلة تجميع المركبات الأولية إلى مركبات ممتدة

في هذه المرحلة يتم تجميع المركبات الأولية التي توجد بينها علاقة تطابق، إلى مركبات ممتدة حسب نوع الأخطاء المراد إصلاحها. ويقوم نظام ديكور21 بثلاث عمليات تجميع اعتماداً على مجموعة من القواعد التي تم استخلاصها من معاجم قواعد اللغة العربية ([20]، [21]) :

- التجميع الأول: وهو بناء المركبات الممتدة لكشف أخطاء الإعراب.
- التجميع الثاني: وهو بناء المركبات الممتدة لكشف أخطاء التعريف.
- التجميع الثالث: وهو بناء المركبات الممتدة لكشف أخطاء في الجنس والعدد والضمير وصفة العاقل.

4.2.3. مرحلة كشف الأخطاء

إن بناء المركبات الممتدة، تتم عملية التدقيق ويكون ذلك استناداً إلى الخصائص الصرفية للكلمات التي تكون المركب الممتد واعتماداً على تطبيق قواعد اللغة العربية بالنسبة للتجميع الأول وللتجميع الثاني (ما هي الخصائص؟ وكيف يجب أن تكون؟) وعملية توحيد الخصائص بالنسبة إلى التجميع الثالث (يجب أن تكون خصائص الكلمات متساوية). وفي حالة وجود خلل في إحدى خطوط التطابق (أي الخصائص)، بمعنى أن الخاصية الموجودة مختلفة عما يجب أن تكون

عليه حسب القاعدة المطبقة وذلك بالنسبة للمركبين الممتدين الأول والثاني أو وجود اختلاف في الخصائص بالنسبة إلى المركب الممتد الثالث، يقوم برنامج كشف الأخطاء بتمريرها إلى برنامج الإصلاح. أما في حالة عدم وجود خلل، يقع محو المركب الممتد نهائياً لخلوه من الأخطاء.

4.3. المرحلة الثالثة: التحليل المتعدد المعايير لإصلاح الأخطاء

تقوم هذه المرحلة بتوليد الجملة الصحيحة المناسبة (احتمال واحد) بالنسبة إلى أخطاء الإعراب وتوليد جميع الاحتمالات الممكنة بالنسبة إلى بقية الأخطاء. عند توليد احتمال، يقوم البرنامج باستبدال بعض الخصائص الصرفية للكلمات التي تخلف بقواعد التطابق ويمكن لهذه الكلمات أن تكون اسماً أو صفة أو فعلاً. ففي حالة استبدال الخصائص الصرفية لاسم أو صفة، يرسل الاسم أو الصفة في صيغة المفرد إلى برنامج مصيغ الكلمات صحبة الخصائص التي يجب أن تكون عليها أي الجنس والعدد المناسبين. أما في حالة استبدال الخصائص الصرفية لفعل، يرسل أصل الفعل إلى برنامج مصرف الأفعال مع الخصائص التي يجب أن يكون عليها أي حالة الفعل (معلوم أو مجهول) والضمير والزمان المناسبين. وسنتوسع لاحقاً في وصف هذين البرنامجين المهمين في عمليات توليد الاحتمالات. وتجدر الإشارة أنه في حالة وجود العديد من الاحتمالات المتنافسة، يقع احتساب وزن كل احتمال بالتوازي مع عملية توليده وذلك قصد ترتيبها ترتيباً تنازلياً قصد مساعدة المستخدم على اختيار الاحتمال الأكثر صواباً. هذا ويتم احتساب وزن كل تغيير طرأ على الكلمات التي تكون الجملة قصد إصلاحها وفقاً لثلاثة معايير (تم اقتراح هذه المعايير في أطروحة الدكتوراه التي قدمتها د. لمياء هديش بلغيث سنة 1999) [2]:

- المعيار الأول وهو المعيار الترددي: يقترح الاحتمال الذي يغير أقل ما يمكن من الخصائص الصرفية للكلمات بمعنى أنه يحافظ على الخصائص الأكثر تردداً. ويمثل عدد الخصائص العدد المسند لهذا المعيار. وهو معيار للتعظيم بمعنى أن قيمته سوف تضرب بالعدد 1.
- المعيار الثاني وهو المعيار الرقني: يهتم بالأخطاء اللفظية الناتجة عن نقص أو زيادة حرف من الحروف أو استبدال حرف بأخر أو قلب مواقع حرفين متتاليين وينتج عن هذا الخطأ كلمة صحيحة لكنها تخرق قاعدة من قواعد التطابق. ويسند لكل نوع من الأخطاء الرقنية عدداً بين 0 و 1. ويكون هذا المعيار للتقليل بمعنى أن قيمته سوف تضرب بالعدد (-1).
- المعيار الثالث وهو المعيار الصرفي: يشجع على تغيير الخصائص الصرفية للكلمة الخاطئة دون استبدالها بكلمة أخرى لا تشاركها نفس الجذر (مثال يستبعد هذا المعيار استبدال كلمة "ولد" التي تمثل صيغة المذكر بكلمة "بنت" التي تمثل صيغة المؤنث). ويسند لهذا المعيار العدد 1 إذا تغير أصل الكلمة و 0 في الحالة الأخرى. ويستخدم هذا المعيار للتقليل أيضاً.

في النهاية يحتسب الوزن الجملي لكل احتمال وترتب الاحتمالات المقترحة تنازلياً حسب أوزانها الجمالية.

5. مثال لتشغيل نظام ديكور 2

أنجز ديكور 1 سنة 1999 في نسخته الأولى بلغة البرمجة س++ [2]، وتعالج هذه المنظومة الأخطاء في الجنس والعدد والضمير فقط، ثم وقع تطويرها بإضافة أخطاء التعريف والإعراب

وصفة العاقل. وتمت إعادة برمجتها بلغة "جافا". ويتم تبادل المعلومات بين مختلف البرامج المدمجة و تسجيل البيانات بواسطة ملفات "إكس أم أل".

يبين الشكل 6 نتائج كشف الأخطاء بالنسبة للجملة التالية " الرجل مسرور يهتف وتصفق ":

الخطأ الأول : خطأ في التعريف بين كلمة "رجل" و كلمة "مسرور"

اقتراح : يجب أن تكون كلمة "مسرور" معرفة

الخطأ الثاني : خطأ في الجنس بين كلمة "رجل" و كلمة "مسرور" و كلمة "يهتف" و كلمة "تصفق"

كما تبين هذه النافذة احتمالات إصلاح الجملة مرتبة حسب أوزانها وهي:

الاحتمال الأول: " الرجل المسرور يهتف ويصفق " بوزن يساوي 0,141

الاحتمال الثاني: " المرأة المسرورة تهتف وتصفق " بوزن يساوي 0,69 -



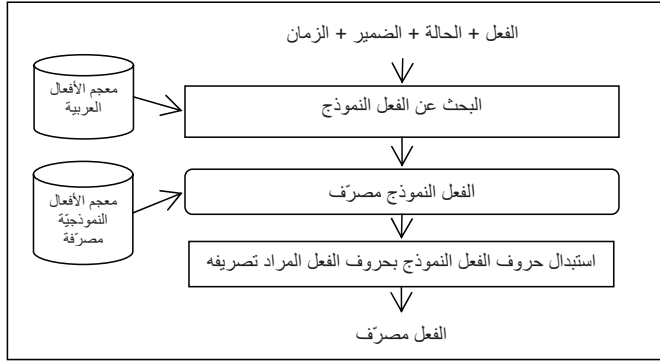
الشكل 6. نافذة عرض لبرنامج ديكورا 2

وتجدر الإشارة أنه قد اخترنا في هذا المثال طريقة الإصلاح الآلي، أما إذا أردنا أن نختار بأنفسنا الخصائص الصرفية لكل كلمة فيتحتّم علينا اختيار الإصلاح شبه الآلي، حيث يعرض لنا ديكورا 2 نتائج المحلل الصرفي مورف 2. ويتم اختيار الخصائص بطريقة يدوية من طرف المستعمل. نقدّم في ما يلي برنامجي مصرف الأفعال ومصيغ الكلمات المعتمدين في عمليات توليد الاحتمالات.

6. برنامج مصرّف الأفعال

يعتبر هذا البرنامج بالغ الأهمية لمنظومة ديكور2 حيث توكل إليه مهمة تصريف الأفعال ويستعمل من طرف برنامج إصلاح الأخطاء الذي يقوم بإرسال الفعل صحة الحالة التي يجب أن يكون عليها والضمير والزمان المطلوبين فيتسلمها هذا البرنامج ويقوم بتصريف الفعل ومن ثم إرجاعه مصرفاً.

لقد قمنا بعملية مسح في معاجم تصريف الأفعال العربية ([20]، [22]) - وهي تتناول الأفعال العربية مشكولة - قمنا بمقارنة أنماط تصريف الأفعال النموذجية فوجدنا أن كثيراً منها تتشابه إذا رفعنا عنها الشكل. هكذا استطعنا حصر مختلف أنماط الفعل العربي غير المشكول وتحصلنا على 216 فعلاً "نموذج" تجسّد أنماط تصريف 14000 فعلاً ، وقمنا بإدخالها جميعاً في قاعدتي البيانات المستخدمة لتحقيق صفة التكمال والشمول في هذا البرنامج. ويبين الشكل7. كيفية تصريف فعل.



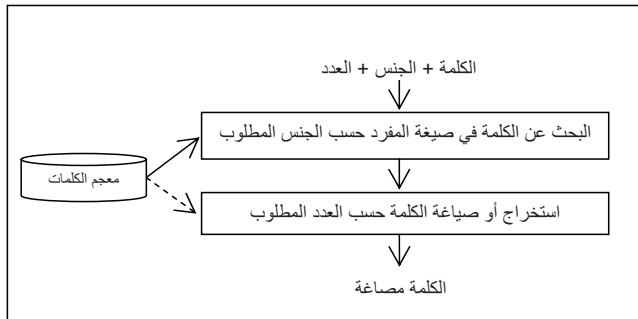
الشكل 7. مراحل تصريف فعل

وتتمّ عملية تصريف فعل بالبحث عن فعله "النموذج" في قاعدة معجم الأفعال العربية ومن ثمّ إيجاد الفعل "النموذج" مصرفاً في قاعدة "معجم الأفعال النموذجية المصرفة" ويتمّ بعد ذلك استبدال حروف الفعل "النموذج" بحروف الفعل المراد تصريفه. و يعدّ برنامج "مصرف الأفعال" منظومة مستقلة يمكن إدماجها في منظومات أخرى.

7. برنامج مصيغ الكلمات

يعتبر هذا البرنامج مهماً أيضاً، فهو الذي توكل إليه مهمة صياغة الأسماء والصفات، ويستعمل من طرف برنامج إصلاح الأخطاء الذي يقوم بإرسال الكلمة صحة الجنس والعدد المناسبان ثمّ يتمّ بعد ذلك صياغتها. تكون صياغة الكلمة سهلة في حالتها المثني والجمع السالم ولكنها ليست كذلك في حالة جمع التكسير، فاللغة العربية ليس فيها قواعد ثابتة وتعتمد على السماع أو ما قاله العرب. ولاجتناب هذه المسألة الصعبة، قمنا بتخزين الكلمات في صيغة المفرد صحة صيغتها في

الجمع وصيغتها في المؤنث (أو المذكر) إذا كانت الكلمة في صيغة المذكر ومرادفتها في صيغة المؤنث ليستا مشتقتان من نفس الجذر. ويبين الشكل 8. كيفية صياغة كلمة.



الشكل 8. مراحل صياغة كلمة

تتم عملية صياغة الكلمة بالبحث عن الكلمة المناسبة حسب الجنس المطلوب وذلك في قاعدة معجم الكلمات ومن ثم النظر إلى العدد المطلوب. إذا كان العدد في المفرد، تؤخذ الكلمة كما هي. إذا كان العدد في المثني تكون صياغتها سهلة حيث يكفي إضافة الحرفين "ان" إلى الكلمة. أما إذا كان العدد في الجمع، يقع الرجوع إلى المعجم لاستخراج الكلمة في صيغة الجمع. عندما تكتمل صياغة الكلمة يقع إرجاعها إلى برنامج إصلاح الأخطاء الذي يتسلم الكلمة مرفوعة ثم يجزمها أو ينصبها حسب حالة إعرابها في الجملة.

8. تقييم نظام ديكورا2

لتقييم نظام ديكورا2، قمنا باختباره على 300 جملة مستخرجة من مواضيع إنشاء لتلامذة التعليم الأساسي والثانوي ومن مطالب بعض المواطنين لدى البلديات ولدى شركة اتصالات هاتفية وسعينا إلى تنوع المستوى الثقافي للأشخاص. وقد قسمنا الأخطاء إلى ثلاثة أقسام: الأخطاء الحقيقية المعالجة ببرنامج ديكورا2 (ونرمز إليها بحرف "ق") والأخطاء الحقيقية التي لم يستطع ديكورا2 إدراكها (نرمز إليها بحرف "ر") والأخطاء غير الحقيقية المضافة من قبل ديكورا2 (نرمز إليها بحرف "ب"). ويكون الخطأ حقيقياً عندما يتعرف عليه أخصائي في اللغة العربية. وتحسب معايير التقييم (وهي الشمولية والتدقيق) على النحو التالي:

$$\text{الشمولية} = \text{ق} / (\text{ق} + \text{ر})$$

$$\text{التدقيق} = \text{ق} / (\text{ق} + \text{ب})$$

وقد توصلنا إلى النتائج التالية : $\text{ق} = 18$ ؛ $\text{ر} = 15$ و $\text{ب} = 3$

وبالتالي تكون المعايير: الشمولية = 54,5% والتدقيق = 85,7%

ويمكن القول بأن النتائج المحققة مشجعة. وتفسر بعض النقصان بوجود حالات تتجاوز مستوى التحليل الصرفي إلى مستوى التحليل التركيبي.

9. الخاتمة و الأعمال المستقبلية

قدّمنا في هذه الورقة منظومة "مورف2" للتحليل الصرفي ومنظومة "ديكور2" لكشف وإصلاح أخطاء التطابق في نصوص عربية غير مشكولة. ولقد توصلنا إلى نتائج مشجعة ونطمح إلى إدخال المزيد من التحسينات على "ديكور2" تمكّنا من معالجة وإصلاح نصوص كاملة عوضاً عن معالجة جمل معزولة ولا يكون ذلك ممكناً إلا بتحسين نظام "مورف2" بإدماج آليات ناجعة لتقسيم النص إلى جمل تكون قادرة على معالجة حالات اللبس في التحليل الآلي والنتيجة عن ثراء لغتنا العربية.

المراجع

- [1] Boujelben M., Belguith Hadrich L. and Aloulou C., « Détection des erreurs d'accord dans des textes arabes non voyellés », Cinquièmes journées scientifiques des jeunes chercheurs en Génie Electrique et Informatique : GEI'2005, Sousse, Tunisie, 2005.
- [2] Belguith Hadrich L., « Traitement des erreurs d'accord de l'arabe basé sur une analyse syntagmatique étendue pour la vérification et une analyse multicritère pour la correction », Thèse de doctorat en informatique, Faculté des sciences de Tunis, février 1999.
- [3] Glantz H. T., « On the recognition of information with a digital computer », ACM, vol. 4, n°2, p. 178-188, 1957.
- [4] Blair C., « A program for correcting spelling errors », Information and control, vol. 3, n° 1, p. 60-67, 1960.
- [5] Damerau F.J., « A technique for computer detection and correction of spelling errors », *Communications of the ACM*, vol. 7, n° 3, p. 171-176, mars 1964.
- [6] Heidorn G. E., Jensen K., Miller L. A., R. J. Byrd, M. S. Chodorow, « The EPISTLE text-critiquing system », IBM System Journal, vol. 21, n°3, 1982.
- [7] Lapalme G., Lacourte R., « Une implantation informatique du français fondamental », *Technique et Science Informatique*, vol. 7, n° 5, 1988.
- [8] Belguith L., Ben Hamadou A., « Un vérificateur et correcteur des accords pour l'arabe non voyellé », *Actes de la 3^e conférence annuelle sur le Traitement Automatique du Langage Naturel (TALN'96)*, 22-24 mai 1996, Marseille, France.
- [9] Véronis J., Contribution à l'étude de l'erreur dans le dialogue homme-machine en langage naturel, Thèse de Doctorat, Université de Droit d'Economie et des Sciences d'Aix-Marseille, 1988.
- [10] Fouqueré C., Systèmes d'analyse tolérante du Langage Naturel, Thèse de Doctorat, Université Paris-Nord, 1988.
- [11] Genthial D., Contribution à la construction d'un système robuste d'analyse du français, Thèse de Doctorat en informatique, Université Joseph Fourier, Grenoble I, 1991.
- [12] Belguith L., Ben Hamadou A., Aloulou C., « An agreement error correction method based on a muticriteria approach: an application to the Arabic language », 10th International conference : Research On Computational Linguistics (ROCLING-X), août 1997, Academia Sinica, Taipei, Taiwan.

- [13] Kammoun Chaâben N., Belguith Hadrich L. and Ben Hamadou A., 1. «The MORPH2 new version: A robust morphological analyzer for Arabic texts», Actes des 10èmes journées internationales d'analyse statistique des données JADT'2010, 9-11 Juin 2010, Rome, Italy.
- [14] Belguith Hadrich L., Baccour L. and Mourad G. (2005). Segmentation des textes arabes basée sur l'analyse contextuelle des signes de ponctuations et de certaines particules. Actes de la 12ème Conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN 2005), pp. 451-456.
- [15] Ben Hamadou A. (1993). Vérification et correction automatique par analyse affixale des textes écrits en langage naturel : le cas de l'Arabe non voyellé. Thèse d'Etat en informatique, Faculté des sciences de Tunis.
- [16] Dichy J. (1997). Pour une lexicomatique de l'arabe : l'unité lexicale simple et l'inventaire fini des spécificateurs du domaine du mot. Meta 42, printemps 1997, Québec, Presses de l'Université de Montréal: 291-306.
- [17] Zaafrani R. (2004). Un dictionnaire électronique pour apprenant de l'arabe (langue seconde) basé sur corpus. Actes de la 11ème Conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN 2004).
- [18] Abbès R., Dichy J. and Hassoun M. (2004). The Architecture of a Standard Arabic lexical database: some figures, ratios and categories from the DIINAR.1 source program, in Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages- COLING 2004 – University of Geneva, 28th August 2004 : 15-22.
- [19] Boujelben M., Aloulou C. et Belguith Hadrich L., « Toward a robust detection/correction system for the agreement errors in non-voweled Arabic texts», The International Arab Conference on Information Technology, ACIT'2008, December 16th- 18th 2008, Hammamet, Tunisia.
- [20] أنطوان الدحاح، " معجم قواعد اللغة العربية في جداول ولوحات"، مكتبة لبنان، الطبعة الرابعة، بيروت 1989.
- [21] شمس الدين، "مرجع الطلاب في الإعراب"، دار الكتب العلمية، الطبعة الأولى، بيروت 2000.
- [22] يوسف ديشي و سامي عمّار، "الشّامل في تصريف الأفعال العربيّة"، هاتيه (Hatier)، باريس، أكتوبر 1999.

From morphological analysis to agreement error detection and correction for non voweled Arabic texts

Abstract. This paper presents two Arabic natural processing systems: “MOPRH2”, an Arabic morphological analyzer, and “DECORA2”, a system of agreement error detection and correction for Arabic texts. The agreement errors handled by DECORA2 consist of gender errors , number errors, case errors, pronoun errors, determination errors and errors in human/non human feature. DECORA2 system is composed of the following complementary softwares: the Arabic morphological analyzer MORPH2, the “verb conjugator” system for non voweled Arabic verbs, the “inflexion word” system, the “detection error” system and the “correction error” system.

Based on MORPH2 result, the “detection error” system decomposes the sentence into phrases. Then, it extracts related phrases and checks their morphological features to detect the errors. The “correction system” aims to correct the detected errors using the “verb conjugator” and the “word inflexion” systems. The “verb conjugator” system is used to conjugate the verb into the appropriate tense and person. The “inflexion word” system generates the word (i.e. noun or adjective) with the appropriate gender and number.

Keywords: Morphological analysis, Syntagmatic analysis, Agreement error, verb inflexion, word framing.

المصطلحات

عربي	إنجليزي
ميراكل	MIRACL
التحليل الزوائد	Affixal analysis
التحليل الصرفي	Morphological analysis
التحليل اللفظي	Lexical analysis
التحليل النحوي	Syntactic analysis
التحليل الدلالي	Semantic analysis
التحليل التركيبي	Syntagmatic analysis
التحليل المتعدد المعايير	Multi-criteria analysis
خطوط التطابق	agreement features
مركب أولي	Initial syntagm
مركب ممتد	Extended Syntagm
الحدود التركيبية	surface indicators
المعيار الترددي	Frequency criterion
المعيار الرقفي	Typographic criterion
المعيار الصرفي	Morphologic criterion
الشمولية	Recall
الدقيق	Precision
تقسيم النص إلى جمل	Tokenization

نمذجة ظاهرة اندماج الكلمات في الجمل العربية المنطوقة وأثرها على الأداء في أنظمة التعرف الآلي على الكلام

ضياء أبوزينة¹، وصفي الخطيب²، مصطفى الشافعي³

¹ جامعة الملك فهد للبترول والمعادن ، كلية علوم وهندسة الحاسب الآلي ، ص.ب. 5066 ، الظهران ، 31261 ، المملكة العربية السعودية

² جامعة الملك فهد للبترول والمعادن ، كلية علوم وهندسة الحاسب الآلي ، ص.ب. 1127 ، الظهران ، 31261 ، المملكة العربية السعودية

³ جامعة الملك فهد للبترول والمعادن ، كلية علوم وهندسة الحاسب الآلي ، ص.ب. 405 ، الظهران ، 31261 ، المملكة العربية السعودية

البريد الإلكتروني : shafei@mit.edu ، wasfi@kfupm.edu.sa ، abuzeina@kfupm.edu.sa

الخلاصة. تشكل الاختلافات الموجودة بين اللغة العربية المكتوبة واللهجات العربية المحلية المتحدث بها بين عامة الناس واحدة من أهم أسباب انخفاض الاداء في أنظمة التعرف الآلي على الكلام العربي ، حيث تُلغظ بعض الكلمات بصورة مختلفة عن بعضها البعض تبعاً لطواهر التبدل الصوتي واندماج عدد من الكلمات، حيث تعتبر هذه الظواهر مصدراً رئيسياً لانخفاض الاداء في أنظمة التعرف الآلي على الكلام. وقد قمنا في هذه الدراسة باستخدام عدد من قواعد علم الصوتيات الكلامية الخاصة باللغة العربية (Phonology) في نمذجة ظاهرة اندماج الكلمات في اللغة العربية المعاصرة من خلال إضافة جميع الكلمات المتوقع اندماجها إلى القاموس الصوتي وكذلك إلى النموذج اللغوي. تم استخدام مدونة صوتية تم تجميعها من نشرات إخبارية تحتوي على 5.4 ساعة من الكلام العربي (لكلا الجنسين) لإنتاج القاموس الصوتي والنموذج اللغوي. كما تم استخدام نظام سفنكس (Sphinx 3) للتعرف الآلي على الكلام والمطور في جامعة كارنيجي ميلون في تنفيذ وفحص الطريقة المقترحة، وقد أثبتت النتائج تحسن الأداء بنسبة 2.3%.

الكلمات الجوهرية: التعرف على الكلام ، الاختلافات في اللفظ، اندماج الكلمات، القاموس الصوتي، اللغة العربية الفصحى.

1 المقدمة

مع دخول الحاسب الآلي في الكثير من التطبيقات العلمية المعاصرة ومع زيادة الاهتمام في الأبحاث التي تعنى بالمعالجات الصوتية، كان لتقنية التعرف الآلي على الكلام نشاط واضح في عدد من اللغات العالمية وخصوصاً اللغة الإنجليزية. ومع أن البحث العلمي في مجال التعرف الآلي على الكلام العربي قد بدأ متأخراً، إلا أن هناك العديد من التطبيقات والابحاث التي ظهرت في الفترة الأخيرة لخدمة هذا الحقل العلمي الهام. وتجدر الإشارة إلى أن هذا التأخر في الانتاج العلمي والبحثي لا ينفي أصالة علم الصوتيات عند العلماء العرب لقرون عديدة خلت. فقد اهتم العلماء العرب ومنذ القرن الثاني الهجري بالصوتيات نظراً للحاجة إلى معرفة الوجوه الصوتية التي اشتملت عليها القراءات القرآنية والتي أدت إلى ظهور علم التجويد، وهو العلم الذي يحدد الطريقة المثلى والصحيحة لتلاوة القرآن الكريم.

ومع بزوغ شمس القرن العشرين شهد العالم ثورة علمية وتقنية كان لها أثر ملحوظ في شتى مجالات العلم والمعرفة، كما برزت الحاجة جلية إلى الكثير من التطبيقات الصوتية، مما أدى إلى تقدم كبير في علم الصوتيات، وقد استُخدمت هذه التقنية في كثير من التطبيقات المحوسبة وفي مجالات شتى نذكر منها على سبيل المثال لا الحصر:

- التعرف على الكلمة والمتكلم (البصمة الصوتية)
- التعرف على الجمل الاستفهامية
- تحويل النص المكتوب إلى كلام منطوق وبالعكس
- خدمات ذوي الاحتياجات الخاصة
- التعرف على المواضيع
- الأنظمة التفاعلية مع الحاسب
- الخدمات الهاتفية (الرد الآلي)
- أنظمة التحكم

إلا أن هناك العديد من العوامل التي تؤثر سلباً على دقة أنظمة التعرف الآلي على الكلام منها:

- اللهجات المختلفة: إن وجود لهجات مختلفة للغة العربية يؤثر على طريقة نطق بعض الحروف من لهجة إلى أخرى، وعند تدريب النظام على التعرف على اللغة بلهجة معينة تتضح معالم انخفاض الأداء عند محاولة التعرف على كلام قيل بنفس اللغة ولكن لمتكلم بلهجة مختلفة. فهناك - مثلاً - فرق بين اللهجة المصرية واللهجة العراقية، كما أن هناك فرق في طريقة الكلام بين أبناء المدينة وأبناء البادية لنفس البلد. كما يتضح هذا الفرق بتصوير المسألة بمجتمع مثل الولايات المتحدة يتكلم شعبه اللغة الإنجليزية - لغته الأم - بلكنات مختلفة فضلاً عن لكنات المهاجرين إليه من أصول مختلفة مثل بلاد الشرق الأوسط وشرق آسيا ووسط أفريقيا.
- الحالة النفسية للمتكلم: تحمل الإشارة الصوتية الصادرة عن الإنسان عدد من الرسائل التي تعكس حالته النفسية، وتظهر هذه الحالة في كثير من الأحيان على نبرة صوته فيمكن تمييز نبرة صوت الشخص السعيد من الحزين أو الخائف، فتختلف بعض صفات الحروف الصوتية تبعاً لاختلاف الحالة النفسية للمتكلم.
- سرعة الكلام: تختلف صفات الحروف الصوتية من حيث التأثير بسرعة الكلام أو بطئه كما أشار كوابارا في [1] مما يؤدي إلى انخفاض الأداء نتيجة لتدريب النظام على سرعة كلام

معينة ثم اختباره على كلام ذو سرعة مختلفة، وقد وجد مارتينيز ومن معه في [2] أن معدل الخطأ للكلمة يزداد بمقدار 2.8 ضعف مقارنة مع النظام المرجعي عند استخدام كلام ذو سرعة عالية أو منخفضة مقارنة مع سرعة الكلام التي تم تدريب النظام عليها.

- اتصال الكلام: بخلاف الكلام المنفرد كأن ينطق المستخدم رقما واحدا مثل "واحد" أو "اثنين" أو يقول كلمة تحكم مثل "نقذ" للدلالة على طلب تنفيذ أمر ما، تتميز الجمل (الكلام المتصل) باندماج عدد من الكلمات مع بعضها البعض لتشكيل كلمات جديدة غير مدرجة في القاموس الصوتي الذي يعتبر مصدر الكلمات في النظام. ومن الأمثلة على الكلام المتصل كأن يدور حوار بين المستخدم الذي يتحدث بلغته الطبيعية والحاسب الذي يستمع لهذا الكلام من أجل تفسيره ومن ثم الاستجابة. وفي هذه الحالة فإن أي حالة إضغام أو إقلاب بين كلمتين أو في الكلمة نفسها سوف يؤدي إلى صعوبة أكبر في التعرف على الكلام.

وبشكل عام فإن تعزيز قدرة النظام على التعامل إيجاباً مع هذه المؤثرات سيقفل من الأخطاء المحتملة ويجعله أكثر ملاءمة للاستخدام. في هذا البحث سوف نتناول ظاهرة الكلام المتصل كواحدة من أسباب انخفاض الأداء في أنظمة التعرف الآلي على الكلام. فمن المعلوم أن بناء القاموس الصوتي في هذه الأنظمة يحتاج إلى الكلمات الواردة في المدونة فقط، وبالتالي فإن الكلمات الجديدة التي ظهرت بعد عملية الاندماج سوف لن تكون مدرجة في القاموس الصوتي، الأمر الذي يؤدي حتماً إلى ضعف الأداء. وقد تم العمل في هذا البحث على بيانات تخص اللغة العربية الفصحى وهي لغة التخاطب التي تستخدم عادة في الفضائيات ومحطات الإذاعة التي تستهدف المواطن العربي من المحيط إلى الخليج بعيداً عن اللهجات المحلية. ونظراً لوجود درجة عالية من التشابه في طريقة النطق بين اللغة العربية الفصحى ولغة القرآن الكريم فإننا استخدمنا قواعد التجويد في نمذجة عملية الاندماج بهدف التقليل من أثرها السلبي. تم ترتيب هذا البحث على النحو التالي: نستعرض في القسم الثاني مسحا أدبيا لظاهرة الكلام المتصل، ننبه بالتعريف بالحروف الصوتية العربية في القسم الثالث، أما القسم الرابع فيتناول تفاصيل بناء القاموس الصوتي والنموذج اللغوي، ثم نقوم بعرض طريقتنا المقترحة في معالجة ظاهرة الكلام المتصل في القسم الخامس، أما القسم السادس فيستعرض النظام المحسن ونتائج الأداء ثم نختم هذه الدراسة في القسم السابع.

2 المسح الأدبي

لقد تناولت العديد من الأبحاث ظاهرة اتصال الكلام وأثرها السلبي على أداء أنظمة التعرف الآلي على الكلام في اللغات الأخرى، ونذكر هنا عدداً من الأبحاث التي أكدت أن نمذجة تغيرات الكلام قد يؤدي إلى تحسن الأداء، ومنها ما قام به سلوبودا ووايل في [3] من إثبات أن وجود عدد أكبر من التمثيلات الصوتية في القاموس الصوتي لنطق بعض الكلمات يزيد من أداء التعرف الآلي، كما توصل مكلاستر ومن معه في [4] إلى نفس النتيجة، أما فوسلر - لوسبير ومن معه فقد وجدوا في [5] أن الاختلاف بين الحروف الصوتية للكلمة المتعرف عليها بمحall الترميز والترميز الصوتي المخزن في القاموس الصوتي يزيد من معدل الخطأ للكلمة في التعرف على الكلام المتصل،

وبالمقابل استنتج ساركلار ومن مع في [6] أن دقة التعرف الآلي على الكلام تزداد بشكل كبير إذا وجد تشابه كبير بين الحروف الصوتية المتتالية المتعرف عليها باستخدام محلل الترميز والترميز الصوتي المقابل للكلمة والمخزن في القاموس الصوتي. لقد أثبتت الدراسات السابقة أن غنى القاموس الصوتي بأكبر قدر ممكن من المفردات الصوتية يرفع الأداء بشكل مؤكد. هناك طريقتان مختلفتان للوصول الى نمذجة هذه الظاهرة وهما: طريقة القواعد الصوتية وطريقة البيانات الصوتية. تعتمد طريقة القواعد الصوتية على مجموعة من القواعد الصوتية الثابتة والمحددة والتي لا تتبدل ولا تتغير عبر العصور، اذ تم الاتفاق عليها بين علماء اللغة جيلاً بعد جيل ، وباختصار فإن قوانين تبدل الصوت تكون ذات مرجعيات واضحة . أما طريقة البيانات الصوتية فهي قواعد صوتية غير ثابتة يتم استنباطها بصعوبة من مجموعة كبيرة من البيانات الصوتية وذلك للخروج بصيغة موحدة حول التغيرات الطارئة على الكلام ، فهذه القواعد متغيرة حسب الأشخاص واللهجات والزمان والبيئة التي تم فيها جمع هذه البيانات. ومع ذلك، فقد أشار أمدال ومن معه في [7] إلى أن هناك هجر لطريقة قواعد علم الصوتيات لصالح استخدام البيانات الصوتية.

أما بخصوص قواعد علم الصوتيات الخاصة باللغة العربية فهي مستقاة من قواعد تجويد القرآن الكريم ، ويُعرّف علم التجويد على انه تلاوة القرآن وفق ما أنزله الله على قلب رسوله صلى الله عليه وسلم ويتحقق ذلك بإخراج كل حرف من مخرجه وإعطائه حقه ومستحقه [8]. ولقد اخترنا طريقة قواعد علم الصوتيات لنمذجة ظاهرة اندماج الكلمات في الكلام المتصل باستخدام قواعد الادغام والإقلاب المعروفة في علم التجويد من أجل اكتشاف جميع الكلمات المتوقعة اندماجها أثناء الكلام. علماً بأن تطبيق هذه القواعد يتم على جميع بيانات المدونة (بيانات التدريب و بيانات الاختبار) . كما أن الكلمات الجديدة المتولدة سوف تضاف الى القاموس الصوتي بالإضافة الى النموذج اللغوي.

إن هذا البحث هو الأول من نوعه - على حد علم الباحثين - والذي يتناول ظاهرة اندماج الكلام في أنظمة التعرف الآلي على الكلام العربي. حيث نهدف إلى تقديم طريقة جديدة لتعزيز قدرة أنظمة التعرف على الكلام العربي المتصل وذلك بنمذجة ظاهرة تداخل الكلمات المتجاورة مع بعضها البعض باستخدام قواعد علم الصوتيات العربية ، حيث نبدأ بحثنا هذا في القسم التالي بمناقشة الحروف الصوتية العربية.

3 الحروف الصوتية العربية

يُعرّف الحرف الصوتي لكلمة معينة على أنه أصغر وحدة صوتية يتم من خلالها تغيير الكلمة حال استبداله بحرف آخر. ويُعتبر الحرف الصوتي الوحدة الأساسية المستخدمة في عملية التعرف على الكلام. حيث يوضح الجدول 1 الحروف الصوتية المستخدمة في النظام، وقد حصلنا على هذا التصنيف من [9] و [10].

الجدول 1. الحروف الصوتية العربية كاملة مع ترميزها المستخدم في النظام

الحرف الصوتي	الترميز المستخدم في النظام	الحرف الصوتي	الترميز المستخدم في النظام
ر (Reh)	/R/	Fatha - <ب>	/AE/
ز (Zain)	/Z/	<بَاب>	/AE:/
س (Seen)	/S/	فتحة مفخمة - <خ>	/AA/
ش (Sheen)	/SH/	نصف مفخمة - <ذ>	/AH/
ص (Sad)	/SS/	Damma - <بُ>	/UH/
ض (Dad)	/DD/	و - <وُن>	/UW/
ث (Thal)	/DH/	ـ - <غُضن>	/UX/
ر (Reh)	/R/	Kasra - <بنت>	/IH/
ط (Tah)	/TT/	سي - <فيل>	/IY/
ظ (Thah)	/DH2/	ـ - <صيف>	/IX/
ع (Ain)	/AI/	و - <لوم>	/AW/
غ (Ghain)	/GH/	ـي - <ضيف>	/AY/
ف (Feh)	/F/	ء (Hamza)	/E/
ق (Qaf)	/Q/	ب (Beh)	/B/
ك (Kaf)	/K/	ت (The)	/T/
ل (Lam)	/L/	ث (Theh)	/TH/
م (Meem)	/M/	جيم فصحة (Jeem)	/JH/
ن (Noon)	/N/	جيم مصرية	/G/
ه (Heh)	/H/	ح (Hah)	/HH/
و (Waw)	/W/	خ (Khah)	/KH/
ي (Yeh)	/Y/	د (Dal)	/D/
		ذ (Thal)	/DH/

توضيح مسألة استبدال الحروف الصوتية بحسب الموضع

قبل توضيح مسألة اندماج الكلمات المتجاورة سنتعرض إلى مبحث استبدال الحروف الصوتية حسب وقوعها في الجملة ، فقمنا بإجراء فحص بسيط لملف صوتي للتأكد من حصول الاختلاف في

اللفظ بين النطق المفترض (حسب كلمات القاموس الصوتي) والنطق الفعلي (حسب ادراك النظام لها) وذلك من اجل المقارنة بينهما. وقد تم استخدام خدمة التعرف على الحروف الصوتية المتوفرة في نظام سفينكس 3 حيث أظهرت النتائج وجود اختلاف بين حروف القاموس والحروف المنطوقة، حيث تم نطق النون ميماً انسجاماً مع القواعد الصوتية العربية كما هو واضح في الجدول 2:

الجدول 2. مثال يوضح مسألة استبدال الحروف الصوتية

 من بينها سبيل التابغة ...	إشارة ملف صوتي يليه النص الخاص بمحتواه
M IH N B AY N IH H AE: S IH L T IH L E T AE: B IH AI AE H	الحروف الصوتية المقترضة
M IH M B AY N IH H AE: S AE L S TT R IX E L E AE T E AE B IH AI AE:	الحروف الصوتية الفعلية

بناء على الجدول السابق فإن حرف النون قد تم لفظه ميماً، أي تم قلب النون الى ميم حسب قاعدة الانقلاب والتي تفيد بأن النون الساكنة تقلب ميماً إذا جاء بعدها باء. وعليه، فإن هناك امكانية للاستفادة من بعض قواعد التجويد في نمذجة ظاهرة اندماج الكلمات المتجاورة والتي تؤدي إلى ظهور كلمات جديدة ، من وجهة نظر القاموس الصوتي . وكما هو معلوم فإن هناك العديد من قواعد النطق في اللغة العربية يمكن الحصول من خلالها على هذه الكلمات المتوقع ظهورها في سياق الكلام. وبالتالي إضافة هذه الكلمات المتوقعة الى القاموس الصوتي.

4 قاعدة البيانات الصوتية

تشمل قاعدة البيانات الصوتية لأنظمة التعرف الآلي على الكلام ثلاثة أمور رئيسية: النموذج الصوتي ، القاموس الصوتي، والنموذج اللغوي. وسوف نستعرض في الأقسام التالية القاموس الصوتي والنموذج اللغوي فقط، إذ أن الطريقة المقترحة لا تنطبق إلى النموذج الصوتي، ولا تقترح أي تعديلات عليه. وبمعنى آخر فإن النموذج الصوتي لكلا النظامين (المرجعي والمحسن) واحد.

4.1 القاموس الصوتي

يعتبر القاموس الصوتي احد اركان انظمة التعرف الآلي على الكلام ، وذلك لانه يمثل حلقة الوصل بين الكلمات الموجودة في المدونة (بيانات التدريب والفحص) وبين الحروف الصوتية الخاصة بكل كلمة على انفراد. وقد قام محمد علي ومن معه في [10] بتطوير برنامج يقوم بإنتاج القاموس

الصوتي. ونشير هنا الى ان هذا البرنامج ينتج القاموس مع الاخذ بعين الاعتبار عدد من الاختلافات التي قد تظهر في نطق الكلمة. ومن الأمثلة على ذلك تطبيق قاعدة الإقلاب اذا حصل داخل الكلمة، وليس بين كلمات متجاورة. بالاضافة الى ذلك فانه ياخذ بعين الاعتبار همزة الوصل في بداية الكلمة والتاء المربوطة في نهايتها. ومن الأمثلة على ذلك ما هو موضح في الشكل 1:

أدبيرة	E AE D IH M B R AA H
أدبيرة (1)	E AE D IH M B R AA T
أدبيرة (2)	E AE D IH N B R AA H
أدبيرة (3)	E AE D IH N B R AA T

الشكل 1. اختلاف الحروف الصوتية لكلمة أدبيرة

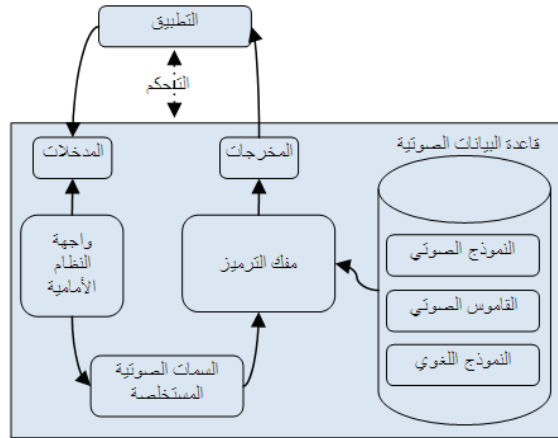
إن الكلمة الاولى في الشكل 1 تشير الى اللفظ الافتراضي لكلمة ادبيرة ، اما الكلمات المرقمة التالية من 1 إلى 3 فهي تشير الى الالفاظ المختلفة التي يمكن حدوثها اثناء نطقها. ولا يعني هنا ان اللفظ الذي يحمل الرقم (1) له افضلية اكثر من غيره وانما العبرة بدرجة التشابه بين الحروف الصوتية التي تم التعرف عليها وتلك المدرجة في القاموس الصوتي. وبذلك فان نظام سفينكس قد اخذ بعين الاعتبار وجود اكثر من لفظ لنفس الكلمة وذلك باستخدام الارقام كما هو موضح اعلاه.

4.2 النموذج اللغوي

يستخدم النموذج اللغوي في كثير من تطبيقات معالجة اللغات الطبيعية وأنظمة التعرف على الكلام، ويعتبر احد المكونات الاساسية في هذه الانظمة. حيث يتم بناء النموذج اللغوي باسناد قيم الاحتمالات التالية (بالاعتماد على المدونة):

- (1) إيجاد احتمالية الكلمات المستعملة
- (2) إيجاد احتمالية كلمتين متتابعين
- (3) إيجاد احتمالية تتابع ثلاث كلمات

ان الهدف الرئيس من وجود النموذج اللغوي هو توقع حدوث كلمة معينة بعد كلمة او كلمتين متتاليتين، وبذلك امكانية تشكيل الجملة التي يراد التعرف عليها. فعند تحديد المقطع الصوتي للتعرف عليه، يبدأ النظام بالتعرف على الكلمة بالاستعانة بالنموذج الصوتي والقاموس الصوتي. بعد تحديد الكلمة الأقرب في القاموس الصوتي والحصول على كلمة جديدة مجاورة، يتم استخدام النموذج اللغوي لربط الكلمتين، ثم لربط ثلاث كلمات ، وهكذا تستمر العملية حتى يتم تشكيل الجملة التي احتوى عليها الملف الصوتي. ويوضح الشكل 2 مكونات نظام سفينكس 3 بحيث يشتمل على النموذج الصوتي، القاموس الصوتي ، و النموذج اللغوي:



الشكل 2. مكونات نظام سفينكس 3

5 الطريقة المقترحة

تم استخدام عدد من قواعد التجويد لنمذجة ظاهرة اندماج الكلمات في اللغة العربية الفصحى، حيث تم اختيار مجموعة محددة من قواعد التجويد وهي القواعد الأكثر ممارسة من قبل الناطقين باللغة العربية. حيث أنه من المعلوم أن المتكلم العادي لا يطبق جميع القواعد التجويدية إلا عند تلاوة القرآن الكريم. ويوضح الجدول 3 هذه القواعد مع ذكر أمثلة عليها، علماً بأنه تم الحصول عليها من [11]:

الجدول 3. قواعد التجويد المستخدمة مع الأمثلة

القاعدة	اسم القاعدة	الحرف الاخير من الكلمة لاولى	الحرف الاول من الكلمة الثانية	حرف الادغام
1	<u>إدغام المتمثلين</u> إذا التقى حرفان متمثلان أولهما ساكن والثاني متحرك أدغم الأول في الثاني. مثل: ما لكم من قرأ ماكمّن. <u>ملاحظة:</u> لا تنطبق هذه القاعدة على الالف، الواو، او الياء.	<u>ساكن</u> بْ، تْ، ثْ، جْ، حْ، خْ، دْ، ذْ، رْ، زْ، سْ، شْ، صْ، ضْ، طْ، ظْ، عْ، غْ، فْ، قْ، كْ، لْ، نْ، هْ	<u>متحرك (ـَ، ـُ، ـِ)</u> بَ، تَ، ثَ، جَ، حَ، خَ، دَ، ذَ، رَ، زَ، سَ، شَ، صَ، ضَ، طَ، ظَ، عَ، غَ، فَ، قَ، كَ، لَ، نَ، هَ	<u>مشدد</u> بَ، تَ، ثَ، جَ، حَ، خَ، دَ، ذَ، رَ، زَ، سَ، شَ، صَ، ضَ، طَ، ظَ، عَ، غَ، فَ، قَ، كَ، لَ، نَ، هَ
2	<u>إدغام النون الساكنة أو التنوين</u> هو دمج النون الساكنة أو التنوين	ـُ، ـِ، ـَ، نْ	يَ، رَ، مَ، لَ، وَ	يَ، رَ، مَ، لَ، وَ

نَ ،	ن ،		بحرف من حروف الإدغام بحيث يصيران حرفاً واحداً مشدداً، هو حرف الإدغام. حروف إدغام النون ستة: تجمعها كلمة (يرملون). مثل: من وآل تقرأ مؤال.	
مَ	ب	نَ ، نْ ، نٍ ، نٌ	الاقلاب هو قلب النون الساكنة أو التنوين ميماً قبل الباء مع مراعاة الغنة. فعند ورود نون ساكنة أو تنوين وبعد هما باء، سواء في كلمة واحدة أو كلمتين، تقرأ النون ميماً. مثل: من بعد تقرأ ممبعد.	3
دَ	د	تَ	ادغام التاء والدال مثل: أجيبت دعوتكما تقرأ: أجيبتُ دعوتكما.	4
طَ	ط	تَ	ادغام التاء والطاء مثل: قالت طائفة تقرأ قالطائفة.	5
تَ	ت	ذَ	ادغام الدال والتاء مثل: قد تبين تقرأ: قُبين.	6
مَ	م	بَ	ادغام الباء والميم مثل: اركب معنا تقرأ اركمنا.	7
ظَ	ظ	ذَ	ادغام الدال والظاء مثل: إذ ظلمتم تقرأ إظلمتم.	8
قَ	ق	كُ	ادغام الكاف والقاف مثل: ألم نخلقكم تقرأ نخلقكم.	9
رَ	ر	لَ	ادغام اللام والراء مثل: قل رب تقرأ قُرب.	10

يبين الجدول 3 وجود عشر قواعد تم اختيارها لاستكشاف حالات الاندماج في النظام. ولتوضيح كيفية تطبيق هذه القواعد، نأخذ المثال التالي: لنفترض وجود جملة تحتوي على المقطع التالي من الكلمات: "... من رفعها...". وبذلك فإن القاموس الصوتي سوف يحتوي على كلمات المدونة بما

فيها "من" و "رفعها". تشير القاعدة رقم 2 (ادغام النون الساكنة او التنوين) ان التقاء نون ساكنة مع حرف الراء يؤدي الى راء مشددة، وبالتالي فان الكلمة تصبح " مَرَفَعُهَا ". وبما اننا نعلم ان بيانات القاموس الصوتي هي بيانات المدونة فقط، فان الكلمة الجديدة المندمجة لن تكون مدرجة في القاموس الصوتي مما يسبب انخفاضاً في الاداء نظراً لاختيار اقرب كلمة مشابهة لها. ويمكن توضيح عملية الاندماج من خلال المثال الموضح في الشكل 3.

حالة اندماج حسب القاعدة الثانية			مثال: الجملة قبل الاندماج		
يلخطر	رفعها	من	واجب	شهر	بعد
الكلمة السادسة	الكلمة الخامسة	الكلمة الرابعة	الكلمة الثالثة	الكلمة الثانية	الكلمة الاولى
يلخطر			واجب	شهر	بعد
الكلمة الجديدة نضاف الى القاموس			الجملة بعد الاندماج حيث نضاف الى المدونة لاغراض النموذج اللغوي		

الشكل 3. مثال يوضح عملية الاندماج

وفيما يلي عدد من حالات الاندماج حسب بيانات المدونة المستخدمة:

(القاعدة الاولى سين مع سين)

إِنَّ هَذَا الْمَجْلِسَ سَيَشْرَفُ عَلَى الثَّرْوَةِ النَّفْطِيَّةِ (بيانات المدونة)

إِنَّ هَذَا الْمَجْلِسَ يَشْرَفُ عَلَى الثَّرْوَةِ النَّفْطِيَّةِ (عملية التحويل)

(القاعدة الثانية تنوين مع ميم)

يَتِمُّ طَرْحُ جُزْءٍ مِنْ رَأْسِ مَالِهَا (بيانات المدونة)

يَتِمُّ طَرْحُ جُزْءٍ مِرَّاسٍ مَالِهَا (عملية التحويل)

(القاعدة الرابع تاء مع دال)

كُشِفَتْ دِرَاسَةٌ حَدِيثَةٌ أَنَّ بَرِيطَانِيَا (بيانات المدونة)

كُشِفَ دِرَاسَةٌ حَدِيثَةٌ أَنَّ بَرِيطَانِيَا (عملية التحويل)

(القاعدة الخامسة تاء مع طاء)

تُعْتَزَمُ شَرَكَةُ طَيَّرَانَ الْإِمَارَاتِ طَلَبَ (بيانات المدونة)

تُعْتَزَمُ شَرَكَةُ طَيَّرَانَ الْإِمَارَاتِ طَلَبَ (عملية التحويل)

(القاعدة السادسة دال مع تاء)

يَقُولُ مُنْتَقِدُوهَا إِنَّهَا قَدْ تَوَجَّجَ التَّضَحُّمُ (بيانات المدونة)

يَقُولُ مُنْتَقِدُوهَا إِنَّهَا قَتْوَجَّجَ التَّضَحُّمُ (عملية التحويل)

الخوارزمية التالية تحدد الخطوات المطلوبة للطريقة المقترحة:

- البحث بشكل متسلسل في بيانات المدونة عن حالات الاندماج حسب القواعد اعلاه.
- انتاج الشكل الجديد للكلمة (الكلمة المندمجة)
- اضافة الكلمة المندمجة الى القاموس الصوتي
- انتاج جملة جديدة مطابقة للجملة التي وجدت بها حالة الاندماج باستثناء كلمتي الاندماج لتحل محلهما الكلمة الجديدة المندمجة.
- تكرار العملية حتى يتم انتاج جميع الكلمات المندمجة حسب القواعد المطبقة
- بناء النموذج اللغوي حسب الاضافات الجديدة.

و توضح الأمثلة التالية شكل النسخ التحريرية (المكتوبة) الجديدة التي سوف تضاف إلى المدونة بعد تطبيق القواعد المقترحة:

بَعْدَ شَهْرٍ وَاحِدٍ مِنْ رَفْعِهَا لِلْحَظَرِ (الجملة التي حصل فيها الاندماج)

بَعْدَ شَهْرٍ وَاحِدٍ مِنْ رَفْعِهَا لِلْحَظَرِ (تضاف كجملة جديدة الى المدونة)

تُجْبِرُهَا عَلَى الْإِبْتَعَادِ عَنْ مَلَاعِبِ الثُّبُسِ (الجملة التي حصل فيها الاندماج)

تُجْبِرُهَا عَلَى الْإِبْتَعَادِ عَنْ مَلَاعِبِ الثُّبُسِ (تضاف كجملة جديدة الى المدونة)

لِلْإِسْتِرَاكِ فِي الْمَزَادِ الْعَالَمِيِّ مِنْ بَيْنِ سَبْعَةٍ (الجملة التي حصل فيها الاندماج)

لِلْإِسْتِرَاكِ فِي الْمَزَادِ الْعَالَمِيِّ مِنْ بَيْنِ سَبْعَةٍ (تضاف كجملة جديدة الى المدونة)

وبعد انتهاء عملية التعرف على الكلام يتم البحث في بيانات النتائج عن جميع الكلمات المندمجة لاعادة استبدالها بالاصل (فك الاندماج). حسب المثال التالي:

مَرْفَعُهَا ◀ مِنْ رَفْعِهَا ◀ عَمَلَايَبِ ◀ عَنْ مَلَاعِبِ ◀ مَمْبِينِ ◀ مِنْ بَيْنِ

6 تقييم الأداء

تم تطبيق الطريقة المقترحة اعلاه على بيانات صوتية (لكلا الجنسين) تم تجميعها من نشرات اخبارية متفرقة ولموضوعات مختلفة، حيث تم الحصول على 249 مقطع اخباري، بحيث تحتوي المقاطع الاخبارية على مضامين مختلفة، فمنها الموضوعات السياسية، والاقتصادية ، والرياضية، الى غير ذلك من الامور المتفرقة التي تشتمل عليها النشرات الاخبارية. تم تقسيم هذه البيانات الصوتية الى 4572 ملف منفصل وبمتوسط فترة (طول الملف) 4.5 ثانية لكل ملف. كما تم تقسيم هذه الملفات الصوتية الى مجموعتين: الاولى للتدريب بما مجموعه 4.3 ساعة والثانية للفحص بمقدار 1.1 ساعة صوتية.

لكل البيانات الصوتية تم انتاج النسخة التحريرية المشكلة على شكل ملف نصي بحيث يتم كتابة الكلام (محتوى الملف الصوتي) على شكل نص وبسطر منفصل لكل ملف، حيث استخدمت هذه

النسخة التحريرية المشكلة في انتاج النموذج اللغوي المرجعي. بالاضافة الى انتاج النموذج اللغوي المرجعي فقد تم ايضاً انتاج القاموس الصوتي المرجعي حيث احتوى على ما مجموعه 14,234 كلمة مختلفة. تم فحص الطريقة باستخدام النظام المرجعي والنظام المحسن وكانت النتائج كما هو موضح في الجدول 4:

الجدول 4. قياس أداء النظام المرجعي والنظام المحسن

النظام	نسبة الخطأ (بالمئة)
المرجعي	12.21
المحسن	9.91
مقدار التحسن في الاداء :	
	2.30

وقد تم قياس نسبة الخطأ حسب المعادلة التالية:

$$\text{نسبة الخطأ} = \frac{\text{الكلمات المضافة} + \text{الكلمات المحذوفة} + \text{الكلمات التي تم استبدالها}}{\text{مجموع الكلمات التي تحتويها الجمل التي تم فحصها}} * 100\%$$

وفيما يلي عينة إحصائية (الجدول 5) عن القواعد التي تم استخدامها في النظام، حيث تشير الإحصائية إلى أن قواعد الإدغام قد استخدمت بشكل أكبر من قواعد الانقلاب :

الجدول 5. مقارنة إحصائية لتكرار استخدام كل قاعدة صوتية

القاعدة	الحرف الاخير من الكلمة الاولى	الحرف الاول من الكلمة الثانية	عدد مرات الاستخدام
1	بْ	ب	17
	تْ	ت	38
	ذْ	د	2
	رْ	ر	16
	زْ	ز	1
	سْ	س	7
	غْ	ع	18
	فْ	ف	12
	قْ	ق	3

49	ل	لُ	
42	م	مُ	
=====		الاحرف التي لم تظهر هنا	
205		يعني انها لم تظهر في المدونة	
1531	ي ر م ل و ن	ـُ ، ـُ ، ـُ ، ـُ ، ـُ ، ـُ	2
200	ب	ـُ ، ـُ ، ـُ ، ـُ ، ـُ ، ـُ	3
25	د	تْ	4
4	ط	تْ	5
32	ت	ذْ	6
14	م	بْ	7
0	ظ	ذْ	8
1	ق	كْ	9
6	ر	لْ	10
2018	المجموع		

وبذلك فإن ما مجموعه 2018 حالة اندماج قد وجدت في النظام. ولكن وبسبب تكرار عدد من الكلمات فإن ما مجموعه 1639 كلمة فقط قد اضيفت الى القاموس الصوتي المرجعي. هذا بالإضافة الى 1639 جملة جديدة قد اضيفت الى نسخة البيانات التحريرية التي استخدمت في انتاج النموذج اللغوي المحسن.

كما أن الجدول 6 يوضح الفرق بين النظام المرجعي والنظام المحسن فيما يتعلق بالأداء، حيث انخفضت درجة التعرف قليلاً لبعض البيانات الصوتية الا انها ارتفعت بشكل اكبر لبيانات صوتية اخرى، بحيث يمكن القول ان نسبة التحسن قد طغت على نسبة الانخفاض الذي حصل لبعض البيانات الصوتية. ولا بد من التركيز على أن التحسن في الأداء في أنظمة التعرف الآلي على الصوت يتم من خلال حساب معدل الخطأ للكلمة كما تم في الجدول 4، وان المقارنة التي تم تقديمها في الجدول 6 هي فقط من اجل تعزيز فهمنا للنتائج التي تم الحصول عليها.

الجدول 6. مقارنة أداء النظام المرجعي بالنظام المحسن

عدد الملفات ونسبتها من المجموع العام للملفات	حالة التعرف في النظام المرجعي والمحسن
91.5% = 1047 ملف صوتي	تم التعرف عليها في كلام النظامين بشكل كامل مع تجاهل الاختلافات البسيطة في حركات التشكيل

تم التعرف عليها في النظام المرجعي ولكن لم يتم التعرف عليها في النظام المحسن	2.01% = 23 ملف صوتي
تم التعرف عليها في النظام المحسن و لم يتم التعرف عليها في النظام المرجعي	6.46% = 74 ملف صوتي

7 الخاتمة

تم استخدام عدد من قواعد علم الصوتيات في اللغة العربية - قواعد التجويد - وذلك من اجل نمذجة ظاهرة اندماج الكلمات في اللغة العربية الفصحى، لاستخدامه في أنظمة التعرف الآلي على الكلام العربي. وقد عالج هذا البحث مسألة اندماج كلمتين متجاورتين وأثره على الأداء، كما تم اقتراح طريقة لتقليل الاخطاء الناتجة عن هذا النوع من عوامل ضعف الأداء في أنظمة التعرف الآلي على الكلام العربي المتصل. تم استخدام قاعدة بيانات صوتية مكونة من 5.4 ساعة للتدريب والفحص، وذلك باستخدام نظام سفينكس 3. تم إضافة الكلمات المتوقع اندماجها إلى القاموس الصوتي على شكل كلمة واحدة، بالإضافة إلى تمثيل هذه الكلمات في النموذج اللغوي. وقد أظهرت النتائج حصول تحسن ملحوظ في أداء النظام، كما لوحظ أن هذا التحسن الملحوظ كان مرتبطاً بتمثيل هذه الكلمات المندمجة في النموذج اللغوي. كما أظهرت النتائج أن حدوث حالات الاندماج الناتجة عن الإدغام أكثر من حالات الاقلاب في اللغة العربية المسموعة.

8 شكر وتقدير

يتقدم الباحثون بالشكر والتقدير لمدينة الملك عبد العزيز للعلوم والتقنية على دعمها للمشروع رقم (08-INF100-04) والذي من نتائجه هذه الورقة، والشكر موصول ايضاً لجامعة الملك فهد للبترول والمعادن على توفير بيئة العمل المناسبة لانجاز هذا العمل.

- [1] Kuwabara, H., 1997. "Acoustic and perceptual properties of phonemes in continuous speech as a function of speaking rate". In: Proceedings of Eurospeech, Rhodes, Greece, pp. 1003–1006.
- [2] Martinez, F.; Tapias, D.; Alvarez, J.; "Towards speech rate independence in large vocabulary continuous speech recognition," Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on , vol.2, no., pp.725-728 vol.2, 12-15 May 1998. doi: 10.1109/ICASSP.1998.675367
- [3] Sloboda T., Waibel A., "Dictionary learning for spontaneous speech recognition," Proc. ICSLP-96, Philadelphia (PA), USA, 2328–2331, 1996.
- [4] McAllister D. et al, "Fabricating conversational speech data with acoustic models: A program to examine model-data mismatch," Proc. ICSLP, Sydney, pp.1847-1850, 1998.
- [5] Fosler-Lussier E., Greenberg S. , and Morgan N., "Incorporating Contextual Phonetics into Automatic Speech Recognition", International Congress of Phonetic Sciences (ICPhS '99), San Francisco, California, , pp. 611-614, 1999.
- [6] Saraçlar M., Nock H., and Khudanpur S., "Pronunciation modelling by sharing Gaussian densities across phonetic models," Computer Speech and Language, vol. 14, pp. 137–160, 2000.
- [7] Amdal I., Fosler-Lussier E., "Pronunciation variation modeling in automatic speech recognition", Elektronik, vol. 99, no. 2, 2003.
- [8] أمل خطاب , فن اتقان التجويد , 2004, <http://fanltajweed.jeeran.com/tajweed/index.htm>.
- [9] Elshafei Ahmed M., "Toward an Arabic Text-to-Speech System", The Arabian Journal of Science and Engineering, Vol. 16, No. 4B, pp.565-583, 1991.
- [10] Ali M., Elshafei M., Al-Ghamdi M., Al-Muhtaseb H., and Al Najjar A., "Generation of Arabic phonetic dictionaries for speech recognition," 5th international conference on innovations in information technology, United Arab Emirates, December 2008.
- [11] Tarweed's Rules of the Holy Quran, <http://www.quran-radio.com/tagweed2.htm>

Title: Cross-Word Arabic Pronunciation Variation Modeling for Speech Recognition

Abstract. One of the problems in the speech recognition of Modern Standard Arabic (MSA) is the differences in pronunciations posed by different local Arabic spoken dialects. In addition, cross-word pronunciation variations alter the phonetic spelling of words beyond their listed forms in the phonetic dictionary, leading to a number of Out-Of-Vocabulary (OOV) word forms. The increase in the number of OOV results in degrading the recognition performance. This paper presents a knowledge-based approach to model cross-word pronunciation variation at both phonetic dictionary and language model levels. The proposed approach is based on modeling cross-word pronunciation variation by expanding the phonetic dictionary and corpus transcription. Using Carnegie Mellon University (CMU) Sphinx speech recognition engine, the Enhanced system achieved Word Error Rate (WER) of 9.91 % on a test set of fully diacritized transcription of about 1.1 hours of Arabic broadcast news. The WER is enhanced by 2.3 % compared to the Baseline system.

Keywords: Speech Recognition, Pronunciation Variation, Cross-Word Vocabulary, Phonetic Dictionary, Modern Standard Arabic.

المصطلحات

الخلاصة	Abstract
النموذج الصوتي	Acoustic model
النظام المرجعي	Baseline system
جامعة كارنيجي ميلون	Carnegie Mellon University
مدونة	Corpus
القواعد باستخدام البيانات الصوتية	Data-Driven-based rules
مفك الترميز	Decoder
النظام المحسن	Enhanced System
الكلمات الجوهرية	Keywords

القواعد باستخدام المعرفة اللغوية	Knowledge-based rules
اللغة العربية المعاصرة (الفصحى)	Modern Standard Arabic
حرف صوتي	Phoneme
علم الصوتيات	Phonology
مدونة صوتية	Speech corpus
سفينكس 3	Sphinx 3
معدل الخطأ للكلمة	Word Error Rate

التلخيص الآلي للنصوص العربية: بين نظرية التلقين الرقمي و نظرية البنية البلاغية

محمد المهدي بودبوس¹، إسكندر كسكاس²، محمد الهادي المعلول³، لمياء هدرش بلغيث¹
¹ وحدة البحث أن.أل.ب*، مخبر ميراك، كلية العلوم الاقتصادية و التصرف بصفاقس، جامعة صفاقس، ص ب 3018، صفاقس، تونس.
² مخبر إبيريت، جامعة بول سياتي، تولوز 3، فرنسا.
³ مخبر آل.بي.آل، 5، نهج باستار، ص ب 13604، 80975 أكس آن بروفانس، فرنسا.

mahdiboudabous@gmail.com, iskandarkeskes@gmail.com,
 mohamed.maaloul@lpl-aix.fr, l.belguith@fsegs.rnu.tn

الخلاصة. نهتم في هذه المقالة بالتلخيص الآلي للنصوص العربية. في هذا الإطار نقدم ونقارن طريقتين لتوليد موجز للنص. تعتمد الطريقة الأولى على نظرية البنية البلاغية "RST" التي تستخدم معارف لغوية خالصة لتحليل النص، فيما تعتمد الطريقة الثانية على نظرية التلقين الرقمي القائمة على قواعد حسابية بحتة. ولتقييم مدى فعالية الطريقتين المقترحتين قمنا بإدراجهما ضمن نظامي "ARSTResume" و "R.I.A". وفي تقييمنا لهذين النظامين تمكنا من الوصول إلى جملة من الاستنتاجات والنتائج الهامة التي تميز كلا الطريقتين.

الكلمات الجوهرية: التلخيص الآلي للنصوص العربية، الموجز، نظرية التلقين الرقمي، نظرية البنية البلاغية، خوارزمية "المكان ذات الدعم الإتجاهي"، العلاقات البلاغية، المؤشرات البلاغية.

1 تقديم

ساهم انتشار استعمال الحاسوب و تطور الشبكة العنكبوتية في تضخم حجم النصوص العربية بالصيغة الرقمية، مما أدى إلى طرح العديد من الأسئلة حول كيفية التصرف في هذا الكم الهائل والمتزايد من هذه النصوص. وبالتالي أصبحت الحاجة متأكدة أكثر من أي وقت مضى إلى توفير الوسائل اللازمة للتصفح السريع للنصوص، قصد تمكين المستخدم من تقدير درجة ملائمة الوثيقة للمعلومة المطلوبة. كما أن التصفح السريع للنص يمثل حلا في بعض الأحيان لربح الوقت والجهد، على اعتبار أن المعلومات المطلوبة متوفرة في الموجز المولّد.

لهذا كان التلخيص الآلي للنصوص مجال بحث، شد أنظار العديد من الباحثين على مر السنين الأخيرة. وقد قام الباحثون في هذا المجال بوضع عديد التصورات سواء منها الرمزية أو الرقمية.

تعتمد مناهج التصورات الرمزية على أدوات بلاغية ولغوية لتحليل النص، ويتعلق الأمر هنا برصد العلاقات الدلالية و البلاغية القائمة بين مختلف وحدات النص قصد تحديد الأهمية النسبية لبعضها البعض. ولا يمكن تطبيق هذه المناهج إلا بتحليل مسبق ومعمق لمدونة الدراسة التي تقوم عليها الدراسات اللغوية [1].

بينما تقوم مناهج التصورات الرقمية على النظريات التلقينية و التقنيات الإحصائية [2]. تكمن أهمية هذه المناهج في قدرتها على التأقلم مع أغلب المدونات النصية بمختلف مجالاتها واهتماماتها. إذ يتعلق الأمر هنا بإسناد درجات تبرز أهمية كل جملة في النص، اعتمادا على مجموعة من المقاييس. و لا يحتوي الموجز النهائي إلا على الجمل التي تحصلت على أعلى الدرجات.

تتمحور هذه المقالة في ثلاثة أجزاء. يعرض الجزء الأول طريقة رمزية تعتمد على نظرية البنية البلاغية "RST" [1] و يعرض الجزء الثاني طريقة رقمية تعتمد على نظرية التلقين باعتماد خوارزمية "المكانن ذات الدعم الإتجاهي (م د إ)". فيما يبرز الجزء الأخير تقييم للطريقتين المقترحتين لتلخيص النصوص العربية من خلال نظامي "ARSTResume" و "R.I.A".

2 طريقة التلخيص المعتمدة على نظرية البنية البلاغية

كما أشرنا في التقديم، يتطلب إنجاز التلخيص الآلي باعتماد التصورات الرمزية تحليلا مسبقا لمدونة الدراسة. وفي هذا السياق قمنا بتحليل معمق لمدونة متكونة من مئة مقال صحفي، مكنتنا من رصد مجموعة من العلامات اللغوية والعلاقات البلاغية التي تهدف إلى ربط مختلف وحدات المقال ببعضها البعض قصد إعادة تشكيل النص على نحو هرمي. هذه المؤشرات والعلامات اللغوية تم تنظيمها وفق قواعد بلاغية [3].

2.1 التحليل اللغوي لمدونة الدراسة

تتكون مدونة الدراسة من مئة مقالة عربية مجمعة من الموقع الإلكتروني لجريدة دار الحياة، من صنف "html" وحسب التشفير "UTF-8". وللتذكير فإن تحميل هذه المقالات تم دون التقيد بمضمون أو بحجم المقال. و لعل ما يبرر هذا الاختيار حقيقة أنه كلما كانت المدونة متنوعة كلما كانت أشمل واحتوت على عدد أكبر من العلامات/المؤشرات اللغوية [1].

من خلال دراستنا المعمقة لمجمل المقالات تبين أن بعض الوحدات اللغوية تكون على قدر من الأهمية من حيث إسهامها في توليد الموجز كما في استقلاليتها عن أي موضوع أو مضمون مدقق.

يتمثل دور هذه العلامات اللغوية، التي تكون القواعد أو الأطر البلاغية، في الربط أولا بين أصغر جزأين (قطاعين) متتاليين في النص أحدهما يكون نواة - جزء أساسي من النص من أجل التناسق - والآخر يكون نواة أو فرع - جزء اختياري - [4] وثانيا، في تحديد نوع العلاقة البلاغية التي تربط بينهما.

إن هذه العلامات اللغوية متعددة الأصناف (أسماء، أفعال، أدوات ربط، عبارات نحوية، الخ...) وتقوم دائما بنفس الوظائف الدلالية، وهي تصنف إلى نوعين : مؤشرات محفزة ومؤشرات إضافية [3]. حيث أن المؤشرات المحفزة تعلن عن مفاهيم أساسية ملائمة بالنسبة إلى مهمة التلخيص الآلي. أما المؤشرات الإضافية فيتم البحث عنها في مساحة محددة انطلاقا من المؤشر المحفز وتُستعمل سياقيا لتأكيد العلاقة البلاغية التي أكدها المؤشر المحفز أو نفاها [1].

و فيما يلي قاعدتين بلاغيتين موزعتين بحسب العلاقتين البلاغيتين: "ترجيح" و"شرط"، وهو ما يشير إليه الجدولان التاليين.

اسم العلاقة البلاغية	{ ترجيح }
القيود على الجزء (1)	لاشيء
القيود على الجزء (2)	يحتوى على المؤشر المحفز والمعلن عن العلاقة البلاغية {لعل، ولعل، ولعل، ولعلها، ولعلمهم، الخ...}
مكان المؤشر المحفز والمعلن عن العلاقة البلاغية	أول الجملة
الجزء المعتمد	الجزء (2)

الجدول الأول: القاعدة البلاغية "ترجيح"

اسم العلاقة البلاغية	{ شرط }
القيود على الجزء (1)	يحتوى على واحد أو أكثر من هذه المؤشرات {ما/لا الخ...}
القيود على الجزء (2)	يحتوى على المؤشر المحفز والمعلن عن العلاقة البلاغية {دون/من دون / لو الخ...}
مكان المؤشر المحفز والمعلن عن العلاقة البلاغية	أول الجملة
الجزء المعتمد	الجزء (1) الجزء (2)

الجدول الثاني: القاعدة البلاغية "شرط"

كما نشير إلى أن هذه القواعد التي شكلتها العلامات البلاغية (المؤشرات المحفزة والمعلنة عن العلاقات البلاغية والمؤشرات الإضافية) هي التي تقوم باستخراج العلاقات البلاغية التي يبينها الجدول التالي.

قائمة العلاقات البلاغية	شرط
	استدراك
	تفصيل
	استثناء
	توكيد
	تقليل
	ربط

قاعدة
نفي
تمثيل
تفسير
ترتيب
استنتاج
جزم
تعريف
ترجيح
إمكان
حصر
تخصيص
تعليل

الجدول الثالث: العلاقات البلاغية

علما وأن بعض هذه العلاقات البلاغية مماثلة لما وجد من طرف "مذكور ومن معه" [5].

2.2 الطريقة المقترحة

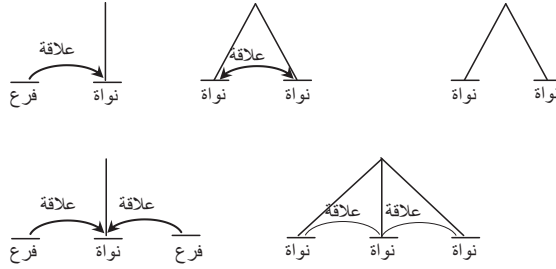
نقوم في هذه الفقرة بتقديم لمحة عامة للطريقة المقترحة المعتمدة على نظرية البنية البلاغية مرتكزين في ذلك على القواعد البلاغية التي وظفناها لبناء الشجرة البلاغية، وذلك لما في رسم الشجرة البلاغية من أهمية في توليد الموجز.

وتكمن أهمية طريقتنا في استعمال المؤشرات المحفزة و الإضافية لتحديد علاقة بلاغية واحدة بصفة قطعية بين قطاعين متتاليين وهو ما يؤسس لبناء شجرة بلاغية "RST" واحدة. وهذا أمر مخالف لنظرية الهيكل البلاغي "RST" لمان وتمسون [6] حيث يُمكن للمؤشر اللغوي أن يشير إلى أكثر من علاقة بلاغية مما يؤدي إلى ضرورة اختيار واحدة من بين هذه العلاقات و مما يؤدي كذلك إلى بناء العديد من الشجرات البلاغية "RST" [7].

هذا ونشير إلى أن بناء رسم البنية الهرمية للنص المصدر - المتكون من قطاعات النص والعلاقات البلاغية التي تربط بينهما - يرتكز على بعض القواعد البلاغية وعلى جملة من المخططات البلاغية.

هذه المخططات أو الأنماط البلاغية تشير إلى التنظيم الهيكلي للنص أيًا كان مستوى التسلسل الهرمي في الشجرة "RST"، إذ يمكن الربط بين نواة وفرع، اثنين أو أكثر من النوى بينها أو بين نواة مع عدة فروع [8]، فبيكلة النص في صيغة شجرة بلاغية يعتمد على تطبيق هذه المخططات بصفة متكررة حتى آخر النص.

تتجسم هذه المخططات البلاغية في خمس نماذج، كما يبين ذلك الرسم التالي، والتي يمكن استخدامها بشكل متكرر في ربط وتنظيم هيكل مختلف قطاعات النص مهما كان حجمه.



الرسم الأول: المخططات البلاغية [6]

إذا كان المؤشر الرئيسي المحفز والمعلن عن العلاقات البلاغية في بداية الجملة فإن العلاقة البلاغية المعلنه تربط هذه الجملة مع الجملة السابقة (وليس القطاع السابق).

الجدول الرابع: مثال لقاعدة بلاغية في بناء شجرة بلاغية

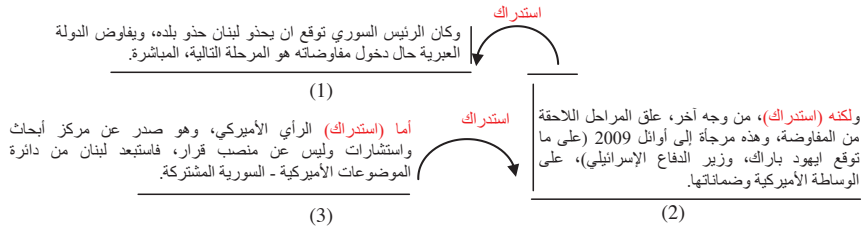
تبين هذه القاعدة أنه إذا كان هناك مؤشر رئيسي في بداية الجملة و يعلن عن علاقة بلاغية، فإن هذه العلاقة تربط الجزء الذي يحتوي على المؤشر الرئيسي و الجملة التي تسبقها.

المثال التالي يوضح فائدة هذه القاعدة :

وكان الرئيس السوري توقع ان يحذو لبنان حذو بلده، ويفاوض الدولة العبرية حال دخول مفاوضاته هو المرحلة التالية، المباشرة. (1) **ولكنه (استدراك)**، من وجه آخر، علق المراحل اللاحقة من المفاوضات، وهذه مرجأة إلى أوائل 2009 (على ما توقع ايهود باراك، وزير الدفاع الإسرائيلي)، على الوساطة الأميركية وضماناتها. (2)

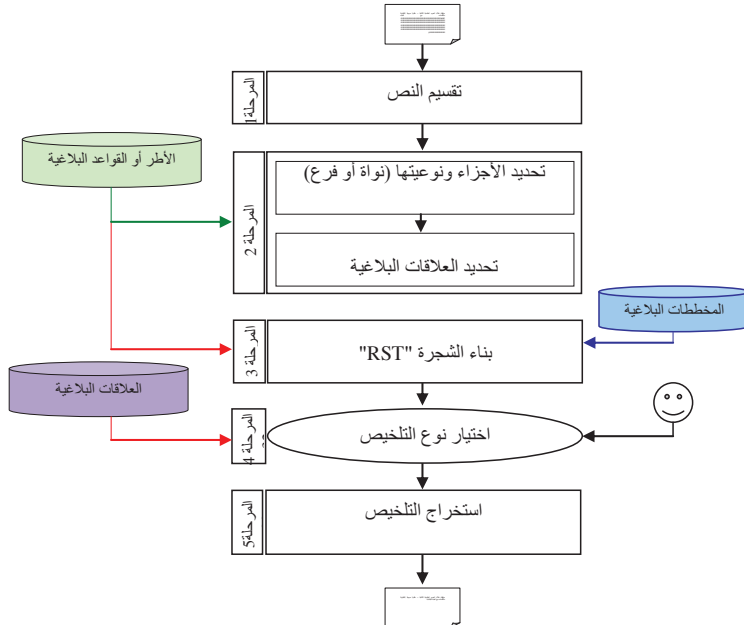
أما (استدراك) الرأي الأميركي، وهو صدر عن مركز أبحاث واستشارات وليس عن منصب قرار، فاستبعد لبنان من دائرة الموضوعات الأميركية - السورية المشتركة. (3)

في هذا المثال العلاقة اللغوية "استدراك" المشار إليها عبر المؤشر "ولكنه" تربط بين الجزء (1) و الجزء (2). أما العلاقة البلاغية الثانية "استدراك" المشار إليها عبر المؤشر "أما" الموجود في بداية الجملة فإنه يشير إلى الربط القائم بين الجزء (3) و الجزء (1) وليس الجزء (2) فقط. وذلك كما يبينه الرسم الثاني للشجرة البلاغية في هذا المثال:



الرسم الثاني: شجرة بلاغية

نحاول الآن الوقوف على مكونات طريقتنا إعتماً على المراحل التالية :



الرسم الثالث: مراحل طريقة التلخيص المعتمدة على نظرية البنية البلاغية

إن هذه الطريقة تهدف إلى تلبية مختلف الإحتياجات الافتراضية للمستخدم. وفي الواقع، يمكن للمعلومة أن تكون غير ذات أهمية في حد ذاتها، إلا أنها قد تتلاءم مع إحتياجات المستخدم. وعلى هذا النحو، فنحن نقدم للمستخدم إمكانية اختيار مسلكه في تلخيص النص من خلال اختيار العلاقات البلاغية التي تهتمه.

وللحد من عدد الجمل المكونة للتلخيص وتأكيد ملاءمتها في ذات الوقت للإحتياجات الافتراضية للمستخدم، فإننا نقوم باختزال الشجرة "RST" عن طريق حذف جميع الفروع التي تشكل علاقات بلاغية لم يقع الاحتفاظ بها في التلخيص النهائي من قبل المستخدم. وهكذا فإن التلخيص النهائي لا

يحتفظ إلا بالجملة المكونة للوحدات الأساسية (نوى) للعلاقات البلاغية الباقية في الشجرة "RST" بعد اختزالها.

يجدر الإشارة أننا قمنا ببرمجة هذه الطريقة في إطار النظام "ARSTRésumé" كما قمنا بتقييم هذا النظام وتحصلنا على نتائج جد مشجعة. وسوف نقوم في الفقرة الرابعة ببسط هاته النتائج وتحليلها.

3 طريقة التلخيص المعتمدة على نظرية التلقين

تندرج الطريقة المقترحة للتلخيص الآلي للمقالات الصحفية باللغة العربية ضمن التصورات الرقمية اعتماداً على نظرية التلقين بشبه الإشراف. تتطلب هذه الطريقة مدونة تلقين متكونة من مجموعة من النصوص وجملة من المقاييس الرقمية وقع الاستناد عليها لتدريج جمل النص. الهدف من هذه الطريقة المقترحة يكمن في إيجاد السبل الكفيلة بتحويل مقاييس التصنيف التي اعتمدها الخبراء لإنجاز الموجز، في شكل درجات تبرز مدى أهمية كل جملة في المقال [9].

تتكون مدونة التلقين من مئة نص، كل نص هو عبارة عن مقال صُحفي مرفق بموجز وقع إنتاجه من قبل ثلاثة خبراء. حيث تتمثل مهمة كل خبير في استخراج موجز لكل نص عن طريق تصنيف الجمل إلى فئتين: فئة "ذات صلة" (+) وفئة "لا صلة" (-). أثر ذلك إعتدنا على مؤشر "كابا" [10] لاستخراج الجمل الأكثر توافقاً بين الخبراء والتي تمثل الموجز، وقد أرفقناه بكل مقال موافق لها.

أما بالنسبة للمقاييس الرقمية فقد قمنا بجمع كل المقاييس المتاحة لنا استعمالها وتطبيقها على مدونة النصوص الكتابية وذلك إستناداً على الأبحاث المنجزة في هذا المجال لكل من [11]، [12]، [13] و [14]، ثم قمنا باختيار المقاييس والمقادير التي يمكن تطبيقها على مدونة التلقين، مع تبين مدى قابليتها للحساب وذلك لتجنب الترميز اليدوي. وتعتبر هذه المقاييس الركيزة الأساسية لعملية التلقين الآلي للطريقة المقترحة.

وفيما يلي يتم تقديم شرح مفصل لبعض المقاييس المعتمدة من خلال الجدول الخامس مع الإشارة إلى أن هذه المقاييس تم تصنيفها إلى صنفين إثنين: مقاييس موضوعية و مقاييس معجمية [15]. وللتذكير فإن المقاييس الموضوعية تقوم بتصنيف الجمل إلى أصناف مختلفة حسب موقعها الترتيبي في المقال، بينما تقوم المقاييس المعجمية بحساب وزن و مقدار كل جملة.

المقاييس	الشرح
موقع الجملة في النص	صنف الجملة يكون حسب موقعها في النص: 1 إذا كانت الجملة موجودة في بداية النص، 2 إذا كانت الجملة في منتصف النص، 3 إذا كانت الجملة في نهاية النص.
	صنف الجملة حسب موقعها في الفقرة: 1 إذا كانت موجودة في بداية الفقرة، 2 إذا كانت الجملة في منتصف الفقرة، 3 إذا كانت الجملة في نهاية الفقرة.
:	:

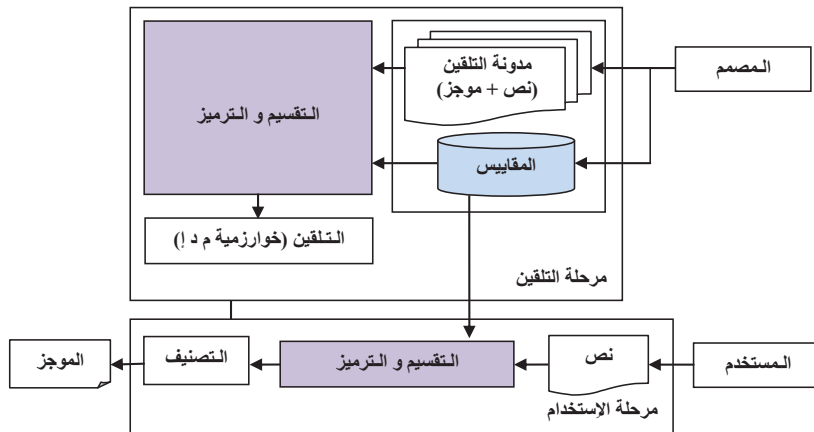
رتبة جزء الجملة	صنف الجملة حسب رتبة الجزء الذي توجد فيه.
كلمات العنوان	يتم حساب وزن الجملة حسب تردد كلمات العنوان في الجملة.
تردد الكلمات	يتم حساب وزن الجملة حسب تردد الكلمات المهمة فيها. (الكلمات المهمة هي الكلمات الأكثر استعمالاً في النص).
:	:
الكلمات الإيجابية	يتم حساب وزن الجملة حسب تردد الكلمات الإيجابية فيها. (الكلمات الإيجابية قمنا باستخراجها استناداً على المدونة).

الجدول الخامس: شرح لبعض المقاييس الموضوعية والمعجمية

3.1 المراحل الأساسية للطريقة المقترحة

نبرز في هذه الفقرة لمحة عامة عن الطريقة المقترحة التي تنقسم إلى مرحلتين: أولاً - مرحلة التلقين و ثانياً - مرحلة الاستعمال. تقوم المرحلة الأولى بعملية تقسيم النصوص إلى وحدات مختلفة ثم ترميزه، من خلال تحليل كل وحداته وتصنيفها أو حساب وزنها حسب المقاييس الموضوعية و المعجمية، للقيام بعملية التلقين، و تضطلع المرحلة الثانية بتوليد الموجز بصفة آلية وفقاً لقاعدة الاستخراج المتأتية عن عملية التلقين.

و في ما يلي تفسير لمرحلتي التلقين و الاستخدام من خلال هذا الرسم:



الرسم الرابع: مراحل طريقة التلخيص المعتمدة على نظرية التلقين

¹ مثال للكلمات الإيجابية المستخرجة من المدونة: {الهدف من هذا المقال، أبرز ما يمكن قوله، التقديرات الرسمية، إلخ...}.

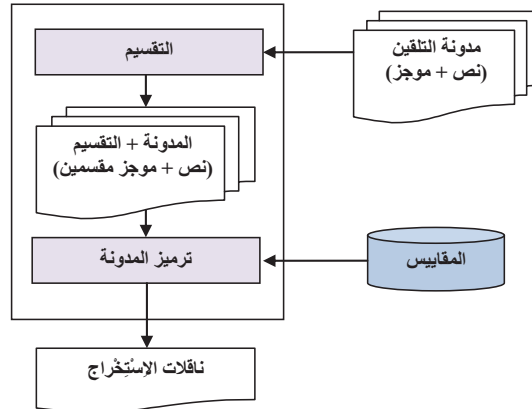
3.1.1 مرحلة التلقين

الهدف من هذه المرحلة هو تقسيم المقالات إلى جمل و ترميزها وفقا لأوزان وأصناف. هدفنا من ذلك هو تحويل مقاييس التصنيف، التي اعتمدها الخبراء لإنجاز الموجز، إلى درجات (أوزان وأصناف) قصد الحصول على مدونة مكوّنة تكون العنصر الأساسي لعملية التلقين. تؤدي هذه المرحلة إلى إيجاد قاعدة الاستخراج، الكفيلة بتحديد مدى أهمية الجملة [9].

3.1.1.1 طور تقسيم و ترميز المدونة

نقوم هنا بتقسيم هرمي للمقالات المكونة لمدونة التلقين و هيكلتها في شكل وحدات (عناوين، أجزاء، فقرات وجمل). نستند في هذه المرحلة على علامات الوقف و مجموعة من علامات "html" (<p>،
، <div>، </div>، <p>، الخ.) لتقسيم النصوص العربية. وتقدم هذه المرحلة، كنتيجة، النص المصدر مقسم و مدعم بعلامات تحيط بالعناوين: <عنوان> ... </عنوان>، الأجزاء: <باب>... </باب>، الفقرات: <فقرة>... </فقرة> والجمل: <جملة>... </جملة> [1].

اثر ذلك تتم عملية ترميز جمل المدونة، حيث يقع تحليل كل جملة و تصنيفها أو حساب وزنها حسب كل المقاييس لنحصل في النهاية على مجموعة من ناقلات الاستخراج، تعود كل واحدة منها بالنظر إلى جملة معينة. ثم يضاف إلى هذه الناقلات قسم الفئة الذي تنتمي إليه الجملة: فئة "ذات صلة" (+) إذا كانت الجملة من بين جمل الموجز و فئة "لا صلة" (-) إذا لم تكن الجملة من بين جمل الموجز. وللإشارة فإن تحديد قسم الفئة لكل ناقلية إستخراج يكون بالاعتماد على الموجز المرافق لكل مقال للمدونة. وتمثل ناقلات الاستخراج نتيجة لهذا الطور و ركيزة أساسية لطور التلقين.



الرسم الخامس: تفصيل مرحلة تقسيم و ترميز المدونة

3.1.1.2 طور التلقين

يمثل هذا الطور القاعدة الأساسية لمرحلة الإستخدام فيما بعد (أنظر فقرة 3.1.2)، حيث تتم خلاله عملية التلقين وذلك باستعمال ناقلات الإستخراج لتوليد قاعدة الاستخراج الكفيلة بتحديد مدى أهمية الجملة في مرحلة الاستخدام.

هذا وللتتويه فإن قاعدة الاستخراج هي في شكل معادلة تجمع بين أفضل المقاييس، وذلك استنادا على خوارزمية "المكائن ذات الدعم الإتجاهي (م د إ)" و ناقلات الاستخراج الخاصة بمدونة التلقين.

إن اختيارنا لهذه الخوارزمية استند أساسا إلى عدة دراسات أثبتت مقدرة خوارزمية "المكائن ذات الدعم الإتجاهي (م د إ)" عندما يكون التصنيف ثنائيا، وهو ركيزة طريقتنا المقترحة، على عدد كبير من الخوارزميات الأخرى [16].

في ما يلي نستعرض مثالا لمعادلة قاعدة الاستخراج:

$$\text{قاعدة الاستخراج} = 1\alpha * \text{مقياس} + 2\alpha * \text{مقياس} + \dots + n\alpha * \text{مقياس} + \beta$$

α : ضارب المقياس i.

β : القيمة المضافة لقاعدة الاستخراج.

3.1.2 مرحلة الاستخدام

تمكننا هذه المرحلة من استخراج موجز للنص، بصفة آلية، اعتمادا على معادلة قاعدة الاستخراج وأبرز أطوار هذه المرحلة التي يمر بها النص هي طور تقسيم ثم ترميز النص و طور تصنيف الجمل.

3.1.2.1 طور تقسيم و ترميز النص

يشبه هذا الطور في شكله طور تجزئة و ترميز المدونة في مرحلة التلقين، حيث تتم في مرحلة أولى عملية تجزئة النص تليها إثر ذلك عملية ترميز النص. لكن هناك نقطتا اختلاف أولا: إن عملية التقسيم تتم إستنادا إلى مقالة واحدة تكون محور الموجز المراد توليده، وليس لمجموعة من المقالات مثلما هو الشأن في مرحلة التلقين. ثانيا: غياب موجز النص و بالتالي غياب صنف الجملة في ناقلات الاستخراج.

3.1.2.2 طور التصنيف

تستخدم معادلة قاعدة الاستخراج لحساب مجموع درجات كل جملة في النص و ذلك بالاعتماد على ناقلة الاستخراج الخاصة بها، واثر ذلك يقع تصنيف الجمل حسب مجموع درجاتها إلى فئتين: فئة "ذات الصلة" (+) و فئة "لا صلة" (-). ويتكون الموجز النهائي من الجمل المصنفة من فئة ذات صلة (+).

قمنا بإدراج الطريقة المقترحة في نظام "R.I.A" (Résumé Intelligent Arabe).

و تكمن أهمية هذا النظام في اعتماد خوارزمية "المكانن ذات الدعم الاتجاهي (م د إ)" للقيام بعملية التلقين، حيث تتميز هذه الخوارزمية بخاصية التصنيف الثنائي للبيانات [9].

4 التقييم

لتقييم مدى فعالية الطريقتين المقترحتين- طريقة التلخيص المعتمدة على نظرية البنية البلاغية و طريقة التلخيص المعتمدة على نظرية التلقين، قمنا باختبار كفاءة كل من النظامين "ARSTResume" و "R.I.A" على مدونة تقييم مكونة من خمسين مقالا مختلفا عن المقالات التي وقع استعمالها في مدونة التلقين. يتمثل هذا الإختبار في مقارنة التلاخيص المولدة من كلا النظامين مع التلاخيص المعدة من طرف الخبراء².

يعرض الجدول السادس شرحا مفصلا لعدد جمل مقالات مدونة التقييم، بالإضافة إلى عدد الجمل المكونة لموجز الخبراء المرافق للمقال. كما يبرز عدد الجمل المهمة (أ) المستخرجة من قبل كل نظام، وللتذكير فان الجمل المهمة هي التي تتوافق مع جمل موجز الخبير. ويبين هذا الجدول أداء كلا النظامين من خلال عرض القيم التي تحصلنا عليها لمعباري التذكير و الدقة.

نظام ARSTResume				نظام R.I.A				عدد جمل الموجز 3	عدد جمل المقال
المقياس- ف	التذكير	الدقة	أ	المقياس- ف	التذكير	الدقة	أ		
0.3	0.25	0.38	1	0.66	0.5	1	6	6	12
0.3	0.23	0.42	2	0.79	0.7	0.9	5	6	13
0.35	0.29	0.44	2	0.48	0.36	0.71	5	7	14
0.32	0.24	0.44	2	0.57	0.47	0.73	4	7	17
0.45	0.41	0.49	2	0.46	0.4	0.53	3	10	20
0.38	0.32	0.47	5	0.74	0.64	0.87	11	13	22
0.39	0.34	0.44	3	0.6	0.52	0.7	8	13	23
0.64	0.61	0.67	11	0.22	0.18	0.27	5	21	33
0.66	0.58	0.76	9	0.8	0.73	0.88	12	15	33
0.54	0.47	0.64	5	0.81	0.73	0.92	12	14	34
:	:	:	:	:	:	:	:	:	:
0.57	0.49	0.69	9	0.22	0.17	0.33	4	18	41
0.63	0.52	0.81	10	0.43	0.34	0.6	5	16	50
0.57	0.46	0.74	12	0.3	0.24	0.39	5	27	59

² هذه الموجزات وقع إعدادها بالاعتماد على ثلاثة خبراء مختصين في اللغة العربية.

³ هذه الجمل وقع استخراجها من طرف ثلاثة خبراء استنادا على مؤشر "كأبا"

0.5	0.39	0.71	16	0.3	0.23	0.4	8	30	64
0.52	0.43	0.65	27	0.32	0.28	0.37	9	43	70
0.5	0.45	0.63	8	0.6	0.42	0.69	10	15	المتوسط

الجدول السادس: مقارنة أداء نظام R.I.A مع نظام ARSTResume

نلاحظ من خلال قراءتنا لنتائج هذا الجدول أن القيمة المتوسطة لمعايير الدقة، التذكير و المقياس-ف هي على التوالي 69%، 42% و 60% بالنسبة لنظام "R.I.A"، و 63%، 45% و 50% بالنسبة لنظام "ARSTResume". نستنتج من خلال هذا التقييم تفوق نسبي لنظام "R.I.A" على نظام "ARSTResume"، إذ أن القيمة المتوسطة لمعاري الدقة و المقياس-ف بالنسبة لنظام "R.I.A" تفوق القيمة التي تحصلنا عليها بالنسبة لنظام "ARSTResume".

كما نلاحظ أن عدد جمل المقال لها دور في تفوق نظام "ARSTResume"، فكلما فاق عدد جمل المقال 35 جملة كلما كانت النتائج أفضل باستخدام نظرية البنية البلاغية، حيث سجلنا نسبة دقة ب 73% و نسبة تذكير ب 45%، فيما سجلنا نسبة دقة ب 49% و نسبة تذكير ب 35% إذا كان عدد الجمل أقل من 35 جملة، تفسيرنا ذلك أنه كلما كان المقال طويلاً كلما احتوى على أكبر عدد من العلاقات البلاغية و المؤشرات اللغوية و بالتالي يستطيع النظام تغطية أكثر قدر ممكن من المقال لتوليد موجز شبيه بموجز الخبر. و لا ينطبق هذا على نظام "R.I.A" فكلما كان عدد الجمل المكونة للمقال محصور بين 10 و 35 جملة كلما كانت النتائج أفضل، حيث سجلنا نسبة دقة ب 80% و نسبة تذكير ب 52%، بينما سجلنا نسبة دقة ب 45% و نسبة تذكير ب 26% إذا كان عدد الجمل أكثر من 35 جملة، ويمكن تفسير ذلك بأنه كلما كان المقال طويلاً من حيث عدد الجمل كلما كانت العملية الحسابية أقل شمولية و تتطلب مدونة تلقين ذو حجم كبير و هو ما يقلل من أداء نظام "R.I.A".

5 الخاتمة

تعرض هذه المقالة طريقتين للتلخيص الآلي للنصوص العربية، تعتمد الطريقة الأولى على نظرية البنية البلاغية "RST" و تعتمد الثانية على نظرية التلقين الرقمي. قمنا في البداية بتقديم طريقة البنية البلاغية و عرض مختلف أطوار نظام "ARSTResume" الذي يقوم بإدراج هذه الطريقة. بعد ذلك قمنا ببسط مختلف مراحل طريقة التلقين الرقمي و استعراض نظام "R.I.A" الذي يقوم بإدراج مختلف مراحل هذه الطريقة.

كما قمنا بتقييم كلا النظامين باستعمال مدونة تقييم موحدة. و قد أثبت هذا التقييم التفوق النسبي لنظام "R.I.A" على نظام "ARSTResume"، بالنسبة للمقالات التي تحتوي على عدد قليل من الجمل - لا يتجاوز 35 - حيث تحصل نظام "R.I.A" على نسبة 64% حسب مقياس-ف مقابل 41% بالنسبة لنظام "ARSTResume".

أما بالنسبة للمقالات الطويلة - التي يتجاوز عدد جملها 35 - فقد لاحظنا تفوق نظام "ARSTResume" حيث تحصل هذا الأخير حسب مقياس-ف على 56% مقابل 33% بالنسبة لنظام "R.I.A".

- [1] محمد الهادي المعلول، إسكندر كسكاس و لمياء هدرش بلغيث. التلخيص الآلي للنصوص العربية اعتمادا على نظرية البنية البلاغية. الدورة السادسة للمؤتمر الدولي لعلوم وهندسة الحاسوب. تونس 2010.
- [2] J-L Minel, Filtrage sémantique : du résumé automatique à la fouille de textes. Paris : Hermès Science Publications, 2002.
- [3] M. H. Maâloul, I. Keskes and L. Hadrich Belguith, Résumé automatique de documents arabes basé sur la technique RST, Conférence international de Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (TALN /RECITAL 2010), 12ème edition, 19-22 juillet, Montréal – Canada, 2010.
- [4] Christophe Luc, Une typologie des énumérations basée sur les structures rhétoriques et architecturales du texte, TALN – Tours, France, 2001.
- [5] H. I. Mathkour, A. Tourir et W. A. Al-Sanea, Parsing Arabic Texts Using Rhetorical Structure Theory, Journal of Computer Science 4 (9): 713-720, 2008.
- [6] William C. Mann and Sandra A. Thompson, Rhetorical structure theory: Toward a functional theory of text organization, Text, 8(3): p. 243 – 281, 1988.
- [7] N. Nicolas, Problems in the Application of Rhetorical Structure Theory to Text Generation, Thesis submitted in partial fulfilment of the M. Eng. Sc. (Cognitive Science) Degree, University of Melbourne, June 1994.
- [8] D. Marcu, Discourse trees are good indicator of importance in text, Advances in Automatic Text Summarization, pp123 – 136, 1999.
- [9] M.M. Boudabous, M.H. Maâloul, L. Belguith Hadrich, Digital Learning for Summarizing ARABIC Documents, IceTAL, Islande, 2010.
- [10] Cohen J., A coefficient of agreement for nominal scales, Educ. Psychol. Meas., 1960, 20, 27-46.
- [11] H. P. Edmunson, New methods in automatic abstracting, Journal of the ACM 16(2): pp264-285, 1969.
- [12] J. Kupiec, J. Pederson, and F. Chen, A trainable document summarizer, In Proceedings of the Annual International ACM SIGIR, pages 68–73, 1995.
- [13] I. Mani and E. Bloedorn, Machine Learning of Generic and User-Focused Summarization, Proceedings of the Fifteenth National Conferences on AI pp. 821—826, 1998.
- [14] J. Goldstein, M. Kantrowitz, V. Mittal and J. Carbonell, Summarizing Text Documents: Sentence Selection and Evaluation Metrics, Proceedings of the 22nd ACM SIGIR conference, 121--127, 1999.
- [15] M.-R. Amini, A. Tombros, N. Usunier and M. Lalmas, Learning Based Summarization of XML Documents, Information Retrieval Journal Springer Netherlands, 10(3):233-255, 2007.
- [16] Rabah DIBOUNE, Mémoire de Projet de Fin d'Etudes d'Ingénieur d'Etat en Electronique : Utilisation des Machines à Vecteurs de Support (SVM) pour la reconnaissance des chiffres manuscrits, Faculté d'Electronique et Informatique Département Instrumentation et Automatique, 2006-2007.

الملخص باللغة الانجليزية

Automatic summarization of Arabic texts: Learning Theory v.s. Rhetorical Structure Theory

Abstract. We are Interested in this paper to automatic summarization of Arabic texts. In this context, we propose two methods to extract a summary from an Arabic text. The

first method relies on the theory of rhetorical structure "RST" that uses linguistic knowledge to analyse the text, while the second method relies on the theory of digital learning which is purely based on mathematical rules. To evaluate the effectiveness of both methods we have implemented them in "ARSTResume" and "RIA" systems. The evaluation results show that RIA system gives better result when the texts are short and that "ARST"Resume" gives better result when the texts are long.

Keywords: Automatic summarization, Arabic text, Support Vector Machines (SVMs) algorithm, Digital learning theory, Rhetorical Structure Theory, Rhetorical relations.

المصطلحات

عربي	انجليزي
ميراكل	MIRACL
الموجز	Abstract
تلخيص	Summary
التلقين الرقمي	Digital learning
البلاغية	Rhetorical
الدرجة	Score
القيم الدلالية	Tagged Values
الإشراف	Supervised
تدوين	Annotation
بشبه الإشراف	semi-supervised
ذات صلة	Relevant
نظرية البنية البلاغي	Rhetorical Structure Theory
قواعد الاستخراج	Rules extraction
مؤشر "كابا"	Indice de kappa
علاقات بلاغية	Rhetorical relations
علامات بلاغية	Rhetorical indexes
قواعد (أطر أو أنماط) بلاغية	Rhetorical frames
مخططات أو أنماط بلاغية	Rhetorical diagrams
قطاعات (وحدات صغيرة)	Minimal units
مقياس - ف	F-measure

النظام العربي للغة التواصل العالمية

سامح الأنصاري¹، مجدي ناجي³، نهى عدلي¹
¹ مكتبة الإسكندرية

ص.ب.: ١٣٨ الشاطبي، الرمز البريدي: 21526 الإسكندرية - جمهورية مصر العربية
² جامعة الإسكندرية، كلية الآداب، قسم الصوتيات واللغات
 ص.ب.: 138 الشاطبي، الرمز البريدي: 21526 الإسكندرية - مصر
³ جامعة الإسكندرية، كلية الهندسة، قسم هندسة الحاسب والنظم
 الشاطبي، الرمز البريدي: 21544 الإسكندرية - مصر

البريد الإلكتروني: sameh.alansary@bibalex.org

Magdy.nagi@bibalex.org

Noha.adly@bibalex.org

الخلاصة :

يتعرض هذا البحث للمكون العربي في طريقة جديدة ومبتكرة للتغلب على حواجز اللغة، وتتمثل هذه الطريقة في نظام لغة التواصل العالمية، وهي لغة اصطناعية تعبر عن المعارف والمعلومات بنفس الطريقة التي تعمل بها اللغات الطبيعية ولكن في شكل تجريدي بحث لا يستخدم أيًا من المفردات أو التراكيب التي تنتمي إلى لغات الإنسان، فيتناول البحث البنية التحتية التي تم بناؤها بالفعل لهذا المكون ويعرض الإنجازات والتطبيقات التي قام بها المركز المتخصص في بنائه، وينتهي البحث بإطلاق دعوة لجميع الدول العربية للمشاركة في بناء وتطوير الأدوات والموارد اللغوية اللازمة لتكوين شبكة لغة تواصل عالمية مختصة باللغة العربية فقط بما تشمله من تنوعات ولهجات مختلفة.

الكلمات الجوهرية: لغة التواصل العالمية، الشبكة الدلالية، التمثيل المعرفي، تحليل اللغة، توليد اللغة، الشبكة العربية للغة التواصل العالمية، اللغويات الحاسوبية.

1 مقدمة

بالرغم من الفرص والإمكانات المذهلة التي أتت بها العولمة، فإن مع اختفاء الحدود القومية ازدادت حدود من نوع آخر سما وعمقا وتأثيرا على المستويات العلمية والثقافية والاجتماعية والاقتصادية وغيرها؛ تتمثل هذه الحدود في حواجز اللغة. فمن ناحية، أصبحت المعارف الإنسانية متاحة للجميع وأصبح من السهل معرفة كل جديد يكتشف أو يستحدث حول العالم بفضل الإنترنت ووسائل الاتصال والإعلام المتطورة، إلا أن- على الجانب الآخر- معظم تلك المعارف والمعلومات متوفرة باللغة الإنجليزية أساسا، تليها اللغات الأكثر انتشارا مثل اللغة الألمانية والفرنسية. فوفقا للدراسات في (UNESCO, 2005) (Lobachev, 2008) (Gerrard and Nakamura 2004)¹ حوالي 68.39% من محتوى الإنترنت- والذي هو مصدر جوهري للمعلومات -

¹ http://portal.unesco.org/ci/fr/files/16771/10917899351/Japanese_Speakers_and_the_Internet_Gerrard-Nakamura.pdf/Japanese+Speakers+and+the+Internet_Gerrard-Nakamura.pdf

مكتوب باللغة الإنجليزية، بينما حوالي 27.3 % فقط من مستخدميها يتحدثون الإنجليزية كلغة أولى² وهذا الرقم في تزايد مستمر مع التوسع في استخدام الإنترنت في دول العالم الثالث، مما يشير إلى عدم تحقق العدالة أو المساواة في الوصول إلى المعلومات. وعلى نحو مماثل، تنشر معظم البحوث العلمية والاكتشافات الجديدة باللغة الإنجليزية، الأمر الذي يحصر منافع تلك المعارف على فئات المتعلمين أو العارفين بتلك اللغات، وبالتالي يؤثر سلباً على العملية التعليمية والبحث العلمي. حتى الآن، فإن الحلول الوحيدة لتجاوز تلك الحواجز اللغوية هي تعلم لغات مختلفة أو الاستعانة بترجمين لترجمة ما يحتاجه المرء من معلومات، ولكن كلا الحلين غير عملي فهما يتطلبان وقتاً ومجهوداً، وتمويلاً، كما أنهما غير قادرين على توفير حل نهائي يزيل الحواجز بين كل اللغات، فلن يستطيع أحد تعلم كل اللغات، ولن يتوافر مترجمون للترجمة من وإلى كل اللغات في الوقت المناسب. لذلك جاءت لغة التواصل العالمية بحل جديد لهذه المشكلة من شأنه أن يهدم تلك الحواجز نهائياً وأن يضمن الحرية والمساواة في الوصول إلى المعلومات والمعارف. يتمثل هذا الحل في تمثيل المعارف للحاسوب في شكل محايد ومستقل عن اللغات الطبيعية يكون بإمكانه نقل المعنى المجرد للنصوص دون أن ينحاز إلى مفردات أو تراكيب لغة بعينها دون أخرى. ينقسم البحث التالي إلى جزئين رئيسيين، يتناول الجزء الأول منهما لغة التواصل العالمية بشكل نظري بحث يناقش ما يميز تلك اللغة من موارد وآليات تمكنها من سد الفجوة بين اللغات الطبيعية وبعضها وبين الحاسوب. أما الجزء الثاني فيتناول لغة التواصل العالمية من منظور لغتنا العربية بما تفرضه من تحديات، مع التعريف بالمركز الذي يعمل على تطوير المكوّن العربي في لغة التواصل العالمية منذ العام 2004 وهو مركز إبراهيم شحاتة في مكتبة الإسكندرية.

2 نحو نظام لحوار الإنسان مع الآلة: لغة التواصل العالمية

إن علاقة اللغة بالحاسوب هي علاقة منفعة متبادلة؛ فعلى جبهة اللغة يُستخدم الحاسوب في العديد من التطبيقات في مجال اللغويات مثل: الصرف الحاسوبي، والنحو الحاسوبي، والدلالة الحاسوبية، والمعجمية الحاسوبية، وعلم النفس اللغوي الحاسوبي. ومن الناحية الأخرى، اقتبس علماء الحاسوب في تطويرهم للغات البرمجة الكثير من أسس اللغات الطبيعية ويسعون بخطى حثيثة إلى التقريب بين هذه اللغات الاصطناعية واللغات الطبيعية بهدف تسهيل التعامل مع الحاسوب دون وسيط برمجي، فالهدف الأسمى لبرمجة الحاسوب هو أن يتعامل معه الفرد مباشرة بلغته الطبيعية بدلاً من اللغات الاصطناعية (نبيل علي، 2001).

ومن هنا ظهرت جهود تحاول إزالة الحواجز بين اللغة التي يستخدمها الإنسان وبين الحاسوب كأداة ذات قدرات عالية في فهرسة وتخزين ومعالجة واسترجاع البيانات والمعلومات بما يجعل الإنسان قادراً على استغلال إمكانيات الحاسوب باستخدام لغته الطبيعية، وبما يجعل الحاسوب قادراً على أن يفهم لغة الإنسان الطبيعية على مستوى الكلمة والجملة والمعنى. وقد شهدت الآونة الماضية هيمنة علماء الحاسب الآلي على معالجة اللغات الطبيعية لإخضاعها لقيود الحاسب الآلي، واتسمت هذه المرحلة بالجمود في التعامل مع اللغات الطبيعية حيث تم وضع اللغات الطبيعية في إطار هندسي بحث لا يعبر اهتماماً لتصميمها اللغوي ولا يراعي المرونة التي تتمتع بها. لذا كان لابد من انتقال الهيمنة إلى اللغويين حيث إنهم هم القادرون على وضع قواعد اللغة في صياغات

² هذه الإحصائية موجودة في <http://www.internetworldstats.com/stats7.htm>

صورية حاسوبية متبنين في ذلك علم اللغويات الحاسوبية والذي يقصد به الدراسة العلمية للغة من وجهة نظر حاسوبية.

2.1 لغة التواصل العالمية: منطق العمل

لغة التواصل العالمية هي تمثيل معرفي مجرد لمحتويات النصوص، ويحمل هذا التمثيل كل ما كان يحويه النص الأصلي من معلومات نحوية وبراجماتية ودلالية وشكلية في شكل شبكة دلالية توضح المعنى الدقيق لكل كلمة في السياق الوارد وماهية العلاقة الدلالية بينها وبين الكلمات الأخرى في النص. ويقوم النظام المبني على لغة التواصل العالمية بتحليل النصوص الواردة واستخراج معناها وتحويله إلى لغة التواصل العالمية في شكل شبكات دلالية، ثم عند الحاجة إلى ذلك، يقوم النظام بتوليد هذه الشبكات الدلالية في شكل جمل طبيعية صحيحة تتماشى مع مفردات وتركيب اللغة الهدف، وقد تكون الجمل المولدة مختلفة مع الجمل الأصلية في اللغة أو الطول أو الأسلوب... إلخ. ومن الجدير بالذكر أن عمليتي التحليل والتوليد مختلفتان كلية ومستقلتان عن بعضهما تمام الاستقلال، لذلك تتولى مؤسسة واحدة مسؤولية تطوير الأدوات اللازمة لتحليل وتوليد لغة واحدة بعينها دون اعتبار للغات الأخرى التي قد تشكل الطرف الآخر في تلك العمليات.

انطلق برنامج تطوير لغة التواصل العالمية في عام 1996 عندما بدأ معهد الدراسات المتقدمة في طوكيو الدعوة لهذا المشروع. وتقوم حالياً مؤسسة لغة التواصل العالمية الرقمية³ (UNDL Foundation) بتطوير هذا المشروع وإدارته والإشراف على تنفيذه. وقد اشترك في هذا المشروع حتى الآن سبع عشرة مؤسسة تعمل كل منها على تطوير الأدوات والوسائل اللازمة لتحليل وتوليد اللغة الخاصة بها، من بين هذه اللغات اللغة الإسبانية، واللغة الفرنسية، واللغة اليابانية، واللغة البرتغالية، واللغة التايوانية، واللغة العربية - التي يتم حالياً بناء المكون الخاص بها في مركز إبراهيم شحاتة في مكتبة الإسكندرية في مصر - وغيرها من اللغات. ومن المتوقع أن يدخل هذا النظام حيز التنفيذ بنهاية العام 2011.

يستخدم نظام لغة التواصل العالمية مكونات تؤهله للقيام بتحليل دلالي للنصوص التي يقوم بمعالجتها. هذه المكونات هي:

- مفردات لغة التواصل العالمية وهي تمثل الوحدات اللغوية المجردة التي تعبر عن المعارف (المفاهيم) الإنسانية.
- العلاقات الدلالية وهي الرموز المستخدمة لتحديد نوع العلاقة بين كلمتين في جملة ما.
- الملامح وهي الرموز الإضافية التي تستخدم في التحليل الدلالي من أجل تعديل الشبكة الدلالية وإضافة المعلومات التي لم يتم التعبير عنها بمفردات لغة التواصل العالمية أو بالعلاقات الدلالية.
- مجموعة الصفات اللغوية وهي قائمة من الصفات اللغوية القادرة على وصف السلوك اللغوي للكلمات والتركيب وصفاً دقيقاً.

بالإضافة إلى تلك المكونات، هناك موارد لا تقل أهميتها وتشمل القواميس والقواعد اللغوية، أما القواميس فتحتوي على جميع المعلومات المطلوبة لإجراء تحليل ناجح من أجل بناء

النصوص المكتوبة بلغة التواصل العالمية، أو توليد جمل صحيحة بأية لغة طبيعية. ويتكون كل مدخل في قاموس من أحد مفردات لغة التواصل العالمية والكلمة الطبيعية التي تقابله، بالإضافة إلى مجموعة من الصفات اللغوية والقواعد الصرفية الخاصة بتلك الكلمة. أما القواعد النحوية الخاصة بالتحليل والتوليد فهي المسؤولة عن تحليل أو توليد الجمل باللغة الطبيعية وضمان سلامتها لغوياً، وهي عادة ما تكون أحادية الاتجاه؛ من اللغة الطبيعية إلى لغة التواصل العالمية (قواعد التحليل) ومن لغة التواصل العالمية إلى اللغة الطبيعية (قواعد التوليد). وتعتمد هذه القواعد على نظرية X-bar كمنطق عمل (Chomsky, 1970).

بالإضافة إلى كل ما سبق من موارد، فإن نظام لغة التواصل العالمية يستخدم العديد من الأدوات والمحركات اللازمة لتحليل وتوليد اللغات الطبيعية وهي كالاتي: أولاً أدوات التحليل وهي IAN, UNL editor أما أدوات التوليد فهي EUGENE وغيرها من الأدوات الثانوية⁴.

2.2 لغة التواصل العالمية: الاستخدامات

كما أوضحنا من قبل، فإن قدرات لغة التواصل العالمية تمكنها من فهم ما يرد إليها من نصوص فهمها تاماً صحيحاً متكاملًا، يمكن أن يستخدم هذا الفهم لاحقاً في كثير من التطبيقات التي يهتم بها مجال معالجة اللغات الطبيعية. فمثلاً يستخدم هذا الفهم لتوليد نص مقابل يختلف مع الأصلي في اللغة فتتحقق الترجمة الآلية بتحويل المفاهيم والتراكيب المشار إليها في التمثيل الدلالي إلى تلك التي تتناسب مع اللغة الهدف، أو يختلف في الطول فيتولد نص مقابل مختصر أو مطول بأن يستخدم التمثيل المعرفي في تحديد المفاهيم الرئيسية والمفاهيم الثانوية وبالإستغناء عن تلك الثانوية يتحقق التلخيص الآلي، وبإضافة بعض المعدلات يتولد نص مطول بنفس المحتوى الأساسي، أو يختلف في التنوع المفرد، فبإبدال بعض المفردات العامة مثلاً بأخرى فصحي يكون النص الخارج نصاً مختلفاً من المنظور الاجتماعي، وبنفس الطريقة يمكن إبدال بعض الكلمات الدارجة بأخرى تقنية فيكون النص الخارج نصاً ذا طابع علمي أو تقني، أو مختلفاً في الأسلوب فيمكن تبسيط بعض التراكيب مثلاً أو إضافة بعض المحسنات البلاغية التي لا تغير في المحتوى الدلالي للنص شيئاً ولكن فقط تغير من أسلوبه الأدبي، وتندرج تلك التغييرات تحت عمليات التنقيح الآلي.

استخدام آخر لا يقل أهمية عن توليد نسخ مختلفة من النصوص الواردة هو استخدام الشبكة الدلالية من أجل البحث داخل محتوى الإنترنت واسترجاع ما يحتاجه المستخدم من معلومات. فبفهمه لما يبحث عنه المستخدم، يستطيع البرنامج المبني على تكنولوجيا لغة التواصل العالمية إظهار نتائج البحث الأقرب صلة بالموضوع المبحوث عنه واستبعاد تلك التي لا تمت له بصلة. فبدلاً من مقارنة الحروف والكلمات والتراكيب في كلمات البحث بتلك الموجودة على الإنترنت لجمع النتائج، سيعتمد برنامج لغة التواصل العالمية على مقارنة المعنى الدلالي.

بالإضافة إلى ذلك، وبما أن لغة التواصل العالمية هي مجرد تمثيل لمعاني الكلمات لا لغتها أو شكلها، فمن الممكن أن يبحث المستخدم عن معلومات مكتوبة بلغة تختلف عن تلك التي يستخدمها في البحث، ويمكن تدعيم ذلك أيضاً باستخدام لغة التواصل العالمية كخلفية لصفحات الإنترنت، وبهذا الشكل تعمل لغة التواصل العالمية في الفضاء الافتراضي فتولد الصفحات التي يحتاجها المستخدم باللغة التي تناسبه، مما يوفر العدالة والمساواة في حرية الوصول للمعلومات.

⁴ مزيد من المعلومات عن مكونات النظام وموارده موجود في www.unlweb.net

3 اللغة العربية ولغة التواصل العالمية

تعد مسألة تطويع الحاسب لمعالجة اللغة العربية أمراً ضرورياً ومستعجلاً لتقليص الفجوة بين التقدم التقني في الغرب وبيننا كأمة لها خصوصيتها وحضارتها. كما أنه الوسيلة الوحيدة التي تستطيع بها الأمة العربية مواكبة الانفجار المعرفي والمعلوماتي في عصر الإنترنت الساحق والفضاء الإلكتروني الواسع. فلا يستطيع أى مجتمع أو أمة - مهما كانت - أن تعيش بمعزل عن هذه الثورة المعرفية أمام اكتساح اللغة الإنجليزية لكل المراجع المعرفية والمعلوماتية الموجودة على الإنترنت وانتشارها بشكل يهدد بسحق كل الأمم واللغات التي لا تستخدم تكنولوجيا المعلومات والحاسوب في اللحاق بعصر المعلومات.

3.1 تحديات اللغة العربية

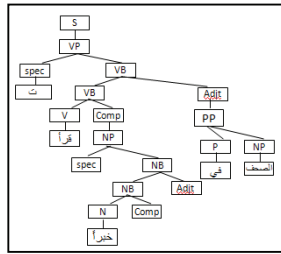
اللغة العربية تتميز بخصائص كثيرة تجعلها لغة قابلة للحوسبة، ومعاصرة للتقنيات الحديثة، وإن كانت معالجة الحاسب الآلي للغة العربية ليست بالأمر المستحيل إلا أنها ليست بالأمر اليسير، فصعوبات المعالجة تكمن في أن اللغة العربية لغة مترابطة ومتشعبة، يكثر فيها الاشتقاق والتصريف ويحتل الإعراب فيها دوراً أساسياً، و تلك تحديات معروفة ومدرسة بشكل عميق لدى المختصين في المعالجة الآلية للغة العربية، لذلك سنتعرض مباشرة إلى دور لغة التواصل العالمية في التعامل مع تلك التحديات على المستوى التركيبي، والدلالي، والصرفي العربي.

3.1.1 أولاً على المستوى التركيبي:

إن تحديد المتعلقات الدلالية للأفعال العربية يعد من الأمور الشاقة حيث إن اللغة العربية تتمتع بمرونة نحوية كبيرة، فيمكن مثلاً أن يقع الفاعل قبل أو بعد الفعل مثل "كتب محمد" أو "محمد كتب" وأيضاً في بعض الأحيان لا يأتي المفعول به مباشرة بعد الفاعل كما هو الحال في جملة مثل "قرأت في الصحف خبراً". لذلك لابد أن يكون نظام المعالجة الآلية قادراً على التعامل مع مثل هذه المرونة، وقادراً على تحديد متعلقات الفعل العربي النحوية والدلالية بدقة. يقوم نظام لغة التواصل العالمية بتحديد الأدوار النحوية بالتعاون مع الصفات الدلالية المخزنة في القاموس أمام كل كلمة ومفهومها؛ مثل: عاقل أو غير عاقل... الخ للأسماء، أو فعل قول أو فعل قراءة... إلخ للأفعال.

فالفعل "قرأ" مثلاً فعل قراءة، والـ"ت" ضمير فاعل عاقل، و"الصحف" شيء غير عاقل، و"في" حرف جر دال على المكان، و"خبر" شيء غير عاقل يمكن أن يُقرأ. لذا لا يمكن وضع هذه الكلمات في أماكنها في الشجرة النحوية إلا بعد النظر في تلك الصفات الدلالية، حيث إن المتمم لفعل القراءة (READV) لابد وأن يكون شيئاً يمكن قراءته (READN)، كما تُملئ القاعدة (أ) في (1)، بينما المخصص لابد وأن يكون اسماً أو ضميراً عاقلاً (HUM، PRON) كما توضح القاعدة (ب). أما الملحق هنا فهو مسبوق بحرف الجر "في"، وفي تلك الحالة يدل الملحق على المكان ويحمل الرمز (LOCPP) كما هو موضح في القاعدة (ج). وبناءً على تلك القواعد يتم تحديد العلاقات التركيبية كما هي موضحة في الشجرة النحوية في شكل 1.

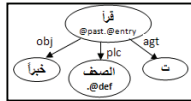
- (1) أ - (V, READV, %x)(READN, %w) := Comp(V, %x; READN, %w);
 ب - (V, READV, %x)(PRON, HUM, %y) := spec(V, %x; PRON, %y);
 ج - (V, READV, %x)(LOCP, %z) := Adj(V, %x; LOCP, %z);



شكل 1

بعد ذلك يمكن بناء الشبكة الدلالية انطلاقاً من الشجرة النحوية في شكل 1 عن طريق مجموعة من القواعد الصورية. فمثلاً يكون إدراك علاقة المخصص كفاعل دلالي كما في (أ) من (2)، والمتمم على أنه مفعول دلالي كما في (ب)، والمكان على أنه مكان قراءة الخبر كما في (ج) لتتكون بذلك الشبكة الدلالية في الشكل 2:

- (2) أ - $\text{spec}(V, \%x; \text{PRON}, \%y) := \text{agt}(V, \%x; \text{PRON}, \%y);$
 ب - $\text{comp}(V, \%x; \text{READN}, \%w) := \text{obj}(V, \%x; \text{READN}, \%w);$
 ج - $\text{adjt}(V, \%x; \text{LOCPP}, \%z) := \text{plc}(V, \%x; \text{LOCPP}, \%z);$



شكل 2

وبالرغم من أن تلك المرونة النحوية التي تتمتع بها اللغة العربية تُعد تحدياً في المعالجة الآلية كما رأينا في عملية التحليل، إلا أن الصرامة في ترتيب الكلمات داخل بعض التراكيب النحوية تمثل – في المقابل – تحدياً من نوع آخر في عملية توليد اللغة العربية، فمثلاً عندما يأتينا تعبير بلغة التواصل العالمية مثل الموجود في شكل 3.

<pre>{uni} agt(202316868 32.@past.@entry,110287213 51.@def) obj(202316868 32.@past.@entry,107679356 85) gol(202316868 32.@past.@entry,109847010 72.@def) {/uni}</pre>	أعطى : 202316868 الرجل : 110287213 خبر : 107679356 السائل : 109847010
---	---

شكل 3

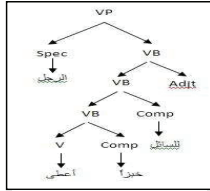
يكون توليده محتملاً على وجهين:

- أعطى الرجل للسائل خبراً.
- أعطى الرجل خبراً للسائل.

وكلاهما صحيح ومقبول في تركيب اللغة العربية نظراً للمرونة النحوية التي تتمتع بها اللغة. فالفاعل "أعطى" ينصب مفعولين بغض النظر عن الترتيب فيما بينهما داخل الجملة الواحدة. والنظام اللغوي الذي تعتمد عليه لغة الشبكات العالمية يمنح عالم اللغة حرية ترتيب الكلمات أثناء التوليد حيث يتم التوليد وفقاً لمرحلتين أساسيتين، المرحلة الأولى يتم فيها تحويل العلاقات الدلالية بين الكلمات إلى علاقات نحوية باستخدام قواعد مثل تلك الموجودة (3) حيث تقوم القاعدة (أ) بتحويل العلاقة الدلالية "agt" والتي تعبر عن العلاقة بين الفعل وفاعله إلى علاقة نحوية تسمى "VS" أي مخصص الفعل، وتقوم القاعدة (ب) بتحويل العلاقة الدلالية "obj" والتي تعبر عن العلاقة بين الفعل ومفعوله إلى علاقة نحوية تسمى "VC" أي متمم الفعل، وتقوم القاعدة (ج) بتحويل العلاقة الدلالية "gol" والتي تعبر عن العلاقة بين الفعل والمفعول الثاني إلى علاقة نحوية تسمى "VC" أي المتمم الثاني للفعل، لتصبح الجملة مترابطة بعلاقات نحوية بدلاً من العلاقات الدلالية تمهيداً لتوليدها في سياق عربي.

- (3) أ - $\text{agt}(\text{VER}, \%01; \%02, \text{NOU}) := \text{VS}(\%01; \%02);$
 ب - $\text{obj}(\text{VER}, \%01; \text{NOU}, \%02) := \text{VC}(\%01; \%02)$
 ج - $\text{gol}(\text{VER}, \%01; \text{NOU}, \%02) := \text{VC}(\%01; \%02);$

بعد ذلك يتم تمثيل العلاقات النحوية في شكل شجري اعتمدا على نظرية X-Bar النحوية كما هو واضح في شكل 4 باستخدام مجموعة القواعد الموضحة في (4). حيث تقوم القاعدة (أ) بالانتقال بمتعم الفعل الأول بين الفعل ومفعوله (VC) إلى مستوى أعلى في البناء الشجري لتصبح (VB)، وتقوم القاعدة (ب) بالانتقال بكل من متعم الفعل الأول بين الفعل ومفعوله والذي أصبح في مستوى VB، ومتعم الفعل الثاني بين الفعل ومفعوله الثاني VC إلى مستوى أعلى في البناء الشجري وهو (VB)، وأخيرا تقوم القاعدة (ج) بالانتقال بكل من مخصص الفعل VS ومتعمي الفعل VB إلى أعلى مستوى في البناء الشجري وهو VP، ليتكون لدينا بذلك البنية التركيبية العميقة للجملة كما تظهر واضحة في شكل رقم 4.



شكل 4

- (4) أ - $\text{VC}(\%x; \%y) := \text{VB}(\%x; \%y);$
 ب - $\text{VB}(\%x; \%y) \text{VC}(\%x; \%z) := \text{VB}(\text{VB}(\%x; \%y); \%z);$
 ج - $\text{B}(\text{VB}(\%x; \%y); \%z) \text{VS}(\%x; \%w) := \text{VP}(\text{VB}(\text{VB}(\%x; \%y); \%z); \%w);$

تتعامل المرحلة الأولى مع البنية التركيبية العميقة للجملة وسنجد فيها أن كلا الاحتمالين يشتركان في نفس البنية العميقة، أما المرحلة الثانية فيتم فيها الترتيب الحر بين الكلمات وإتمام العمليات الصرفية من تطابق وجمع وتأنيث ... إلخ. أي أنه بعد تحديد البنية التركيبية العميقة يكون للقائمين على التصميم حرية الترتيب بين الكلمات في المرحلة الثانية والتي تتعامل مع البنية التركيبية السطحية للجملة وفقا لما يضعوه من قواعد وخصائص للكلمات. ولأن الفعل "أعطى" يمنح مفعوله حرية التبادل فيما بينهما فإن القائم على التصميم لن يوسمه بخصائص تميزه عن غيره من الأفعال. لكن بالمقابل عندما يأتي التعبير بلغة التواصل العالمية الموجود في شكل 5.

<pre>{/uni} agt(200873682:29.@past.@entry,109983572:82.@def) obj(200873682:29.@past.@entry,109923673:85.@pl.@def) obj(200459776:36,100295701:14.@def) gol(200873682:29.@past.@entry,200459776:36) {/uni}</pre>	نصح : 200873682 الوزير : 109983572 المواطن : 109923673 تأجيل : 200459776 السفر : 100295701
--	---

شكل 5

لن يكون لها سوى احتمال توليد واحد فقط وهو:

- نصح الوزير المواطنين بتأجيل السفر.

وذلك لأن الفعل "نصح" ينصب مفعولا واحدا فقط وهو "المواطنين" وتأخيره سيُخل بالسياق النحوي للجملة وقد يضيفي بعضا من اللبس في المعنى إذا تم توليد الجملة على هذا النحو:

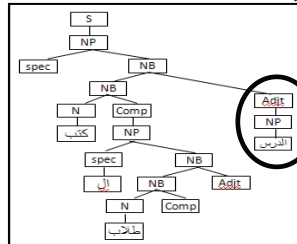
- نصح الوزير بتأجيل السفر المواطنين.

لذلك يوسم الفعل "نصح" بخاصية تحدد موضع مفعوله حتى لا يتأخر.

3.1.2 ثانياً: على المستوى الدلالي:

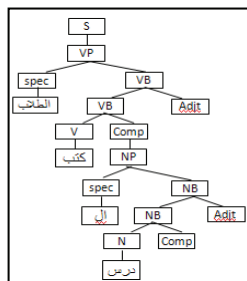
تعد اللغة العربية من أوضح اللغات وأدقها فمواطن اللبس فيها محدودة وذلك لثراء اللغة بالمفردات وإمكانية استغلال الاشتقاق للتعبير عن المعاني، لذلك فاللبس الدلالي الذي يحدث غالبا ما يرجع إلى غياب التشكيل الذي ينتج عنه ظهور حالات معقدة من اللبس نظرا لتعدد احتمالات قراءة الكلمة.

ومشكلة الليس الناتج عن غياب التشكيل ليست مشكلة سهلة أو هينة بالنسبة للتحليل اللغوي. فعلى سبيل المثال، كلمة "كتب" يمكن قراءتها لتعني الفعل "كَتَبَ" أو الاسم الجمع "كُتُبٌ"، وبالتالي فإن معناها في جملة مثل "كتب الطلاب درس" يختلف عن معناها في الجملة "كتب الطلاب مفيدة". ويفيد السياق اللغوي الذي يحيط بالكلمة في تحديد معناها المقصود، ففي الأولى يمكن أن تكون "كتب" مبتدأ حيث يتوافر فيها شروط المبتدأ في كونها اسما مخصوصا (مضافا إلى كلمة "الطلاب"). ولكن لكل مبتدأ خبر فماذا عن الخبر في الجملة؟ فلنفترض أنه هنا كلمة "الدرس"، إذن فهو خبر مفرد أي ليس بجملة ولا شبه جملة وهذا غير جائز في السياق الوارد حيث أن الخبر في هذا السياق لابد له أن يكون صفة (نعت مفرد أو جملة). وهذا يمكن صياغته عن طريق المعالجة الآلية نحويا في الشجرة النحوية في الشكل 6.



شكل 6

إذن فإن كلمة "الدرس" لا تصلح لأن تكون خبرا هنا، حيث إن الخبر يُدرك نحويا في إطار نظرية X-bar على أنه ملحق يحوي صفة إذا كان مفردا، وهو هنا يحوي اسما، وهذه الشروط يمكن صياغتها صوريا بواسطة القواعد في 5. فالقاعدة (أ) ترفض احتمال أن يكون الخبر جملة اسمية إذا كان اسمها غير مضاف إلى ضمير، أما القاعدة (ب) فتؤيد بقوة احتمالية وجود الخبر كجملة نعت. وهذه القواعد ترغم النظام على أن يختار كلمة بديلة من القاموس المخزن به كلمة "كتب" ليختار الفعل بدلا من الاسم ويحلل الجملة من جديد على أنها جملة فعلية بها فعل وفاعل ومفعول. عندئذ يتمكن النظام من تنفيذ القواعد الموضوعية فيوسم كل كلمة ويضعها تحت ما يناسبها في الشجرة النحوية كما هو مبين في شكل 7.



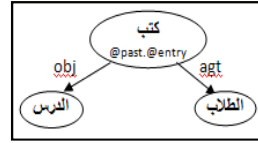
شكل 7

$$(5) \text{AN}(\text{NB}; \text{NP}^{\wedge}(\text{N}; \text{PRON}))=0;$$

$$\text{AN}(\text{NB}; \text{ADJP})=225;$$

يلي هذه المرحلة مرحلة التحليل الدلالي وتتم بواسطة القواعد الدلالية في (6) عن طريق تحديد العلاقات الدلالية بين الكلمات حسب الوسم النحوي في الشجرة. فالقاعدة (أ) في (6) تنص على أن علاقة المخصص بين الفعل والاسم الذي يليه تُدرك دلاليا على أنها فاعل دلالي، وفي القاعدة (ب) علاقة المتمم بين الفعل والاسم الآخر تُدرك دلاليا على أنها مفعول دلالي. وبعد تنفيذ هذه القواعد يكون شكل الشبكة الدلالية كما هو مبين في شكل 8:

- (6) أ-Spec(NP,%x;VB,%y):=agt(VB,%y;NP,%x);
 ب-Comp(VB,%y;NB,%z):=obj(VB,%y;NB,%z);



شكل 8

هذا بالنسبة للدلالة على مستوى تحليل اللغة العربية. أما على مستوى التوليد فثراء اللغة العربية بالمفردات يعد سلاحا ذا حدين؛ فهذا الثراء يعطي اللغة قوة وثراء معجميا واضحا دون غيرها من اللغات، ولكن لأن هذا الثراء لا تتمتع به غالبية اللغات التي ننقل عنها إلى العربية فنجد أن وحدة دلالية واحدة من الممكن أن تقابلها في العربية أكثر من مفردة، فعلى سبيل المثال، الوحدة الدلالية الممثلة بالتمثيل الرقمي "107942152" في لغة التواصل العالمية يقابلها في اللغة العربية مترادفين: "الناس"، و"نسمة" كما يخزنهما قاموس لغة التواصل العالمية في شكل 9.

{س}	{107942152}"107942152"(LST=WRD,POS=NOU)<ara,2,1>;
{ن}	{107942152}"107942152"(LST=WRD,POS=NOU)<ara,2,1>;

شكل 9

والاختيار فيما بين هذين المترادفين يعتمد على السياق المصاحب لكل منهما. فكلمة "نسمة" لا تأتي إلا إذا كان يسبقها رقم في سياقات التعداد. لذلك توسم كلمة "نسمة" بخاصية (AFTNUM) أي أنها تأتي بعد رقم. وبالتالي تخرج "نسمة" في سياقها الصحيح، وكذلك "الناس". فتتمكن قواعد لغة التواصل العالمية للغة العربية من توليد الجملتين التاليتين:

- يبلغ عدد سكان ليبيا حوالي سبعة ملايين نسمة.
- يجلب الناس الماء من مصدر بعيد.

3.1.3 ثالثاً: على المستوى الصرفي:

تتسم اللغة العربية بالاشتقاق الصرفي الغزير، وبالرغم من غزارته فهو اشتقاق شبه منتظم. وهذا الانتظام يجعل المستوى الصرفي أكثر المستويات اللسانية قابلية للحوسبة دوناً عن المستويات الأخرى الدلالية والتركيبية خاصة في مرحلة التوليد الآلي للغة العربية. فعلى سبيل المثال، الفعل الأجوف "قال" يتغير شكله طبقاً لزمان الفعل نفسه وطبقاً للفاعل الذي يأتي معه من حيث النوع والعدد ولذلك يتم تخزين الفعل بالشكل "قال" في القاموس ويُدرج معه مجموعة من الصفات التي يتم من خلالها توليد جميع الصيغ الصرفية لهذا الفعل: قال- يقول- قالاً- يقولان- قالوا- يقولون- قالت.

فمثلاً لتصريف الفعل "قال" في الزمن المضارع المعلوم في حالة ما إذا كان الفاعل مفرداً مذكراً أو ضميراً غائب (أي: هو يقول) تستخدم القواعد الصرفية في (7) لتوليد الشكل "يقول" عن طريق تفعيل القاعدة (8) المخزنة في القاموس:

- (7) أ-(%01,@present):=(%01,+@present,+PRS);
 ب-(%x,M103,PRS):=(%x,-M103,+FLX(PRS:="و":"ا","0>"));
 ج-(%x,^inflected,PRS):=(%x,!FLX,+inflected);

(8) PRS&3PS&SNG&MCL&ACT:="و":"ا","0>";

القاعدة في (7) (أ) مسئولة عن إعطاء الفعل الوسم PRS والمستخدم في القاموس في القاعدة (8)، وتستخدم هذه الصفة في القاعدة (ب) لكي يتم تحميل الفعل أيضا بمهمة أخرى وهي إضافة ياء في بداية الفعل وتحويل الحرف الثاني إلى "واو"، أما القاعدة في ج فهي المسئولة عن إعطاء محرك التوليد أمرا بتنفيذ مهمة التصريف.

يتضح مما سبق أن اللغة العربية تتميز بمجموعة من الخصائص على المستوى التركيبي والدلالي والصرفي والتي إما أن تسهم في تسهيل أو تعقيد معالجتها آليا بشكل يتفاوت من خاصية إلى أخرى. ولغة التواصل العالمية لديها من الأدوات والخصائص ما يمكنها من التعامل مع تلك التعقيدات بشكل أو بآخر، إلا أن كل هذا يعد في مرحلة المهد وبحاجة لمزيد من الجهد والمتابعة لنصل إلى أفضل النتائج.

3.2 المركز العربي للغة التواصل العالمية

في يوليو عام 2004 وبالاتفاق بين مؤسسة لغة التواصل العالمية ومكتبة الإسكندرية، تم إنشاء المركز العربي للغة التواصل العالمية ليكون مسئولا عن تصميم وتنفيذ المكون العربي في مشروع لغة التواصل العالمية، وقد أطلق عليه مركز إبراهيم شحاتة، ويضم المركز فريق عمل مؤهلا تأهلا عاليا من مهندسي البرمجة واللغويين الحاسوبيين للعمل على بناء أدوات اللغة العربية.

3.2.1 البنية التحتية للمكون العربي في لغة التواصل العالمية:

نجح المركز العربي في بناء البنية التحتية لنظام لغة التواصل العالمية العربية، فقد قام ببناء عدد من الأدوات والمصادر التي تمكنه من تحليل وتوليد اللغة العربية بكفاءة ودقة، تتمثل هذه البنية في القواميس، وقواعد التحليل، وقواعد التوليد.

(أ) القاموس العربي: صُمم القاموس العربي ليكون المصدر الرئيسي للمعلومات ، ونقطة الانطلاق لعمليتي التحليل والتوليد. فالقاموس العربي هو وسيلة الربط بين مفردات اللغة العربية ومفردات لغة التواصل العالمية حيث يتكون مدخل القاموس من كلمة عربية وما يقابلها من تمثيل رقمي، وهذا الربط يمنع حدوث أي لبس أو غموض في معنى الكلمة العربية، فكل مدخل في القاموس يعبر عن مفهوم محدد لكلمة عربية محددة، وقد تتكرر الكلمة في أكثر من مدخل ولكن بمعاني مختلفة تمثلها تمثيلات رقمية مختلفة لمفردات من لغة التواصل العالمية.

وتصميم القاموس يسمح له بالتعامل مع الكلمة العربية أيا كانت طبيعتها، فهو يسمح بتخزين إما كلمة أو كلمات مركبة وإما تراكيب معقدة ويستطيع التعامل مع كل تلك الأشكال بنجاح. يضم القاموس العربي حاليا حوالي 130 ألف مدخل يغطي مختلف الفئات المعجمية (الأسماء، الأفعال، الصفات، الظروف) كما هو موضح في الجدول 1.

الأسماء	الأفعال	الصفات	الأحوال	العدد الكلي
60700	40500	23500	5300	130000

جدول 1

ويحتوي القاموس أيضا على كل المعلومات اللغوية الخاصة بالكلمة العربية والتي يمكن بواسطتها التعرف على نوع الكلمة وطبيعتها وسلوكها اللغوي مما يساعد على نجاح عمليتي التوليد والتحليل.

هناك ثلاثة أنواع من المعلومات تخزن مع مداخل القاموس:

- 1- معلومات صرفية وهي المعلومات الخاصة بالكلمة والتي تحدد سلوكها الصرفي في السياقات المختلفة مثل التذكير والتأنيث والعدد وذلك بالنسبة إلى الأسماء، وأيضا معلومات عن نوع الفعل (سالم، مهموز، معتل، مضعف) ووزنه وذلك بالنسبة للأفعال.
 - 2- معلومات دلالية وهي معلومات عن التصنيف الدلالي للكلمات.
 - 3- معلومات تركيبية وهي مسئولة عن وصف سلوك الكلمة داخل الجملة مثل التعدي واللزوم في الأفعال.
- بالإضافة إلى تلك المعلومات اللغوية، يحتوي القاموس أيضا على مجموعة من القواعد المسؤولة عن توصيف السلوك اللغوي للكلمة ويصل عددها إلى 290 قاعدة، وتنقسم تلك القواعد إلى نوعين:
- 1- القواعد الصرفية وهي المسؤولة عن الحصول على التصريفات المختلفة للكلمة العربية وفقا للسياق الذي وردت فيه
 - 2- القواعد النحوية وهي المسؤولة عن وصف السلوك النحوي للكلمة وعن تحديد عدد ونوع المتعلقات النحوية اللازمة لتلك الكلمة.

ب) قواعد التحليل العربية: وهي القواعد المسؤولة عن تحليل اللغة العربية ونقلها إلى لغة التواصل العالمية لتكون جاهزة للتعامل معها وفكها إلى أي شكل أو لغة طبيعية أخرى (Alansary et al., 2006a). تتم عملية التحليل عبر ثلاث مراحل: الأولى هي مرحلة التحليل الصرفي ويتم فيها تحليل كلمات الجملة إلى سوابق وجذوع ولواحق أي إلى وحدات صرفية، وتحديد أيها يمثل مفهوما مجردا وأيها لا يمثل مفهوما، فمثلاً "الولدان" بعد تحليلها إلى وحدات صرفية ستكون "ال" + "ولد" + "ان". فتمثل الودحتان "ولد" و "ان" مفاهيم في حين أن "ال" لا تمثل مفهوما، وأحيانا لا يستطيع النظام التعرف على المفاهيم ذات السوابق واللواحق إلا بواسطة قواعد فك اللبس الصرفي التي تجيز بعض التتابعات الصرفية ولا تجيز البعض الآخر. على سبيل المثال، عند اختيار المفهوم المقصود بالكلمة "ولد"، تحل قواعد فك اللبس الصرفي هذا اللبس بين الفعل "وَلَدَ" والاسم "وَلَدٌ" بأن لا تجيز أن تسبق أداة التعريف "ال" فعلاً وعلى ذلك يتم اختيار الاسم. ولكن وبهذا المنهج يقوم النظام بالتحليل الصرفي بدقة تبلغ 95% ويساهم أيضاً في فك اللبس الدلالي للكلمات ذات السوابق واللواحق بنسبة 40% تقريباً. فنظراً لأن كل مرحلة تعتمد على الأخرى فبدون التحليل الصرفي لا نصل للتحليل النحوي وبدون التحليل النحوي لا يمكن إقامة العلاقات الدلالية لبناء الشبكة الدلالية. وعلى هذا النحو فقد تم بناء نموذج عام لتحليل اللغة العربية يقوم ببناء العلاقات الدلالية في نصوص من الصحف، ويتضمن هذا النموذج حوالي 1000 قاعدة نحوية دلالية، كما تم بناء نموذج قواعد متخصص للتعامل مع عناوين الكتب بما تحويه من محذوفات، وقد ضم هذا النموذج حوالي 500 قاعدة تقريباً، وكذلك فقد تم بناء الشبكات الدلالية لتقرير اقتصادي كانت جملة مثالا على طول جمل اللغة العربية، كما كان النص الأول الذي تم التعامل معه في مجال الاقتصاد.

ج) قواعد التوليد العربية: وهي القواعد المسؤولة عن توليد الجمل العربية من أية شبكة دلالية بنظام لغة التواصل العالمية (Alansary et al., 2006b). ولقواعد توليد اللغة العربية

نسختان أساسيتان: النسخة الأولى تم الانتهاء منها أثناء المرحلة الأولى من المشروع وتتضمن 1427 قاعدة تنقسم إلى مرحلتين رئيسيتين: المرحلة الأولى تختص بتركيب مكونات الجملة وتتابع الكلمات داخل الجملة الواحدة وتوليد الأنواع المختلفة من الجمل سواء كانت اسمية أو فعلية ... الخ، أما المرحلة الثانية فتختص بالقواعد الاشتقاقية التي تتعامل مع بنية الكلمات (سواء كانت أفعالا أم أسماء أم صفاتا أم أحوالا) وتحقيق المطابقة في العدد والنوع والتعريف وإدراج المسافات وعلامات الترقيم.

ولأن النسخة الأولى من القواعد كانت تعاني بعض المشكلات في التعامل مع بعض ظواهر اللغة العربية وتعقيدها، فقد بدأنا في محاولات لإطلاق النسخة الثانية من قواعد التوليد العربية والتي تُعد بمثابة نسخة معدلة ومطورة من النسخة الأولى تعمل على توليد الجملة العربية عبر أربعة مراحل رئيسية، حيث تهتم المرحلة الأولى بتحويل العلاقات الدلالية إلى علاقات نحوية ، وتحديد البنية التركيبية العميقة للجملة في شكل شجري ، وتحتوي النسخة على 80 قاعدة حتى الآن. والمرحلة الثانية تهتم بتحديد البنية التركيبية السطحية للجملة في شكل شجري مع مراعاة ترتيب الكلمات داخل الجملة الواحدة، وتحتوي النسخة على 50 قاعدة قابلة للزيادة، أما المرحلة الثالثة فتهتم بالانتقال بالتركيب من شكله الشجري إلى قائمة متتابعة من الكلمات ، مع وضع المسافات بين الكلمات ، وتحتوي النسخة على 50 قاعدة قابلة للزيادة، أما المرحلة الرابعة والأخيرة فتهتم ببنية الكلمة وتحقيق التطابق في العدد والنوع والتعريف بين الكلمات داخل الجملة الواحدة وتحتوي على 132 قاعدة قابلة للزيادة. وما زال العمل قائما على بناء وتطوير النسخة الثانية من قواعد التوليد العربية. ومن الجدير بالذكر أنه قد تم استخدام النسخة الأولى في عدد من المشروعات الخاصة بالترجمة الآلية، وجاري اختبار النسخة الثانية على مشروعات أخرى.

(د) المدونة اللغوية الدولية للغة العربية: بدأت مكتبة الإسكندرية في مشروع كبير لبناء المدونة اللغوية الدولية للغة العربية، وهو محاولة حقيقية لبناء مدونة لغوية عربية حديثة ومتطورة تحتوي على 100 مليون كلمة، وقد تم تجميع العينات المكتوبة للغة العربية المعاصرة من مصادر متعددة روعي فيها أن تكون ممثلة لقطاع إقليمي كبير من الدول الناطقة باللغة العربية وعاكسة بشكل حقيقي وواقعي لأنماط استخدام اللغة العربية في أنحاء العالم العربي.

والمصادر التي اعتمدت عليها المدونة اللغوية الدولية للغة العربية تشمل الصحف، والمجلات العامة والمتخصصة، والصحافة الإلكترونية، ومقالات الإنترنت، والكتب، والمصادر الأكاديمية. كما تضمنت المدونة عددا كبيرا من الموضوعات مثل العلوم الإستراتيجية والاجتماعية، والرياضيات، والدين، والأدب، والإنسانيات، والعلوم الطبيعية والتطبيقية، وذلك بهدف تحقيق شروط التمثيل والانتشار والتوازن في الحجم والمجالات والمصادر (Alansary et al., 2007).

وقد قامت المدونة اللغوية الدولية للغة العربية بجمع حوالي 75 مليون كلمة من العينة المراد تجميعها ومن ثم ، تم تحليلها صرفيا وتركيبيا، وقد روعي في ذلك أن تكون العينة متوازنة بما يتلاءم مع الواقع العربي المعاصر ، وبحيث تكون أيضا ممثلة لهذا الواقع. وقد قامت المدونة اللغوية الدولية للغة العربية ببناء برنامج إلكتروني لعرض جميع النصوص المجمعة بطريقة تسهل على الباحث البحث عن أي كلمة في أي مجال يختاره، وتحتوي المدونة أيضا على التحليل الصرفي والتركيبي لكل كلمة موجودة داخلها، الأمر الذي يساعد الباحث اللغوي على وجه الخصوص.

وقد تم الاعتماد في مرحلة التحليل الصرفي على أحد المحللات الصرفية الشهيرة وهو تيم باك والتر (Buckwalter) ، ويوضح التحليل الصرفي عدد من المعلومات كالمسابق واللاحق والجذع الخاص بكل كلمة بالإضافة إلى نوع الكلمة من حيث التأنيث والجنس والعدد والتعريف، كما يقوم المحللون أيضا بتحديد المعاني المختلفة لكل كلمة تبعا للسياق الذي وردت فيه. وبانتهاء التحليل الصرفي لكل الكلمات نجد أن كل كلمة سيكون أمامها حوالي 17 معلومة لغوية، وقد تم الانتهاء حتى الآن من تحليل حوالي نصف مليون كلمة، وتعتبر هذه الكلمات المحللة حجر الأساس الذي سيتم بناءً عليه تحليل باقي العينة اللغوية تباعا وبطريقة آلية تعتمد على بعض الطرق الإحصائية وبعض القواعد اللغوية التي تضمن تحليل المتبقي من العينة بطريقة بالغة الدقة. وقد تم كل ذلك خلال الأربع سنوات الماضية بداية من عام 2007 (Alansary et al., 2008).

3.2.2 إنجازات المركز العربي

في هذه الفقرة سنعرض باختصار بعض الإنجازات التي حققها المركز العربي للغة التواصل العالمية وتشمل:

(أ) نظام مكنتات غير معتمد على لغة: قام المركز العربي للغة التواصل العالمية بتصميم وتنفيذ نموذج لنظام مكنتات غير معتمد على لغة يسمح بترجمة المعلومات الخاصة بالكاتب إلى اللغات الست الرسمية للأمم المتحدة (العربية- الصينية- الإنجليزية- الفرنسية- الروسية- الأسبانية) بالإضافة إلى اللغة البرتغالية. وهذا النظام مصمم لكي يسمح للمستخدم باستدعاء وتصفح المعلومات الخاصة بالكاتب الموجودة في فهارس المكنتات الإلكترونية باللغة التي يطلبها بصرف النظر عن اللغة المخزن بها هذه الكتب، كما يسمح للمستخدم المتخصص (المكتبي) بفهرسة الكتب وإضافة أو تعديل المعلومات المختلفة الخاصة بكل كتاب، كما يوفر هذا النظام أيضا معلومات إحصائية عن عدد الكتب التي تم تخزينها. ويتم العمل داخل النظام على النحو التالي: يتم تحليل اسم الكتاب تحليلًا دلاليًا لبناء العلاقات الدلالية بين كلماته باستخدام قواعد التحليل، ثم يتم تخزين مخرجات التحليل وهي في صورة نصوص مكتوبة بلغة التواصل العالمية في الفهارس كي تكون جاهزة لعملية التوليد في أية لغة طبيعية مرة أخرى (Alansary et al., 2009a).

(ب) التقرير السنوي للصندوق العربي للإنماء الاقتصادي والاجتماعي: وأيضًا من الإنجازات التي قام بها المركز العربي للغة التواصل العالمية باستخدام قواعد التحليل القيام بنقل التقرير السنوي للصندوق العربي للإنماء الاقتصادي والاجتماعي لعام 2008 إلى لغة التواصل العالمية حتى يكون جاهزًا للترجمة إلى أية لغة طبيعية أخرى، ويتكون التقرير من حوالي 800 جملة تقريبًا.

(ج) الأمير الصغير: تمكنت كذلك البنية التحتية لقواعد التحليل العربية من التعامل مع نوعية مختلفة من النصوص العربية وهي النصوص الأدبية، فقد تم تحليل النسخة المترجمة إلى العربية من قصة الأمير الصغير الفرنسية *Le Petit Prince* لإثبات قدرة قواعد التحليل على التعامل مع هذه النوعية من النصوص، وبالفعل تم تكويد الفصول الخمسة الأولى من القصة إلى لغة التواصل العالمية لكي تكون متاحة في الشكل المطلوب لترجمتها فورًا إلى أية لغة أخرى، وجاري إعداد قواعد التوليد لترجمتها مرة أخرى إلى العربية كاختبار لجودة قواعد التحليل العربية.

(د) موسوعة دعم نظم الحياة (EOLSS): هي موسوعة متعددة التخصصات وتعد أضخم موسوعة إلكترونية مكتوبة باللغة الإنجليزية حيث تتكون من 123000 صفحة انترنت أي ما يعادل حوالي 250000 صفحة مطبوعة. وقد قام المركز العربي للغة التواصل العالمية بالاشتراك في المشروع الذي طُرح من قبل منظمة اليونسكو بهدف ترجمة هذه الموسوعة إلى اللغات الست الرسمية للأمم المتحدة، وكان المركز العربي مسئولاً عن إنتاج النسخة العربية⁵. وتضمنت المرحلة الأولى من المشروع ترجمة 25 نصاً (حوالي 1300 جملة) من الموسوعة مكوداً من قبل مؤسسة لغة الشبكات العالمية ومعداً للترجمة إلى اللغات المطلوبة، وتمكنت قواعد التوليد العربية باستخدام قاموس متخصص من ترجمة النصوص المطلوبة، وتم تقييم الترجمات المولدة مقارنة بأنظمة الترجمة الآلية الأخرى وكانت جودة النسخة العربية المولدة مرضية، وبالفعل تم إنشاء موقع على الانترنت⁶ يضم النصوص المتفق على ترجمتها وتتصل فيه اللغات المشتركة في المشروع ببعضها، ويتم من خلال الموقع ترجمة النصوص على الانترنت مباشرة⁷، وبعد نشر نتائج هذا المشروع تلقى المركز العربي للغة التواصل العالمية دعوة من القائمين على بناء وتطوير موسوعة الحياة (EOL) لإصدار النسخة العربية من الموسوعة. موسوعة دعم الحياة هي موسوعة إلكترونية متاحة مجاناً على الإنترنت باللغة الإنجليزية وتضم معلومات عن 1.8 مليون كائن حي. وتهدف ترجمة الموسوعة إلى نشر المعرفة عن هذه الكائنات للمحافظة عليها، وقد تم اختبار مدى قدرة قواعد التوليد العربية على التعامل مع نصوص هذه الموسوعة، وكانت النتائج جيدة مقارنة بأنظمة الترجمة الآلية الأخرى (Adly and Alansary, 2009, (Alansary et al., 2007, (2009).

(هـ) إنجازات أخرى: تم استخدام قواعد التحليل العربية في تحليل عدد كبير من النصوص العربية الأخرى مأخوذة من مصادر مختلفة في عدة مجالات مثل مقالات متنوعة من الجرائد، وكذلك عدد من الكتب مثل كتاب موسوعة المشاهير الذي يتناول السيرة الذاتية لمشاهير الشخصيات العربية والعالمية، وكتاب مكتبة الإسكندرية، وأيضاً كتب عن حضارة مصر مثل كتاب أبو سميل: معبد الشمس المشرقة، وكذلك تم تحليل السيرة الذاتية للرئيس الراحل جمال عبد الناصر مأخوذة من المكتبة الرقمية لمكتبة الإسكندرية.

3.3 الشبكة العربية للغة التواصل العالمية: دعوة للمشاركة

وجاري استكمال العمل داخل المركز العربي للغة التواصل العالمية في جميع الاتجاهات لبناء ما هو مطلوب للغة العربية من مصادر وأدوات للتحليل والتوليد، ولكننا بدأنا نواجه سؤالين مهمين؛ الأول هو: أية لغة عربية نقصد ونهدف إلى العمل عليها؟ وليس أية لهجة فنحن جميعاً نعرف أن اللهجات العربية تختلف من دولة إلى أخرى داخل الوطن العربي، ولكن مع ظهور اللغة الفصحى المعاصرة والتي تختلف عن فصحي التراث، تأثرت هذه اللغة باختلاف اللهجات الإقليمية حيث أدخلت بعض المفردات من تلك اللهجات إلى الفصحى المعاصرة وبالتالي أصبحت لغة الكتابات الرسمية تختلف أيضاً من دولة عربية إلى أخرى على مستوى المفردات والتراكيب، ومن هنا أصبحنا نواجه مشكلة تعدد اللغات العربية المستخدمة في الكتابات الرسمية، فوجب علينا التعامل مع هذه التعددية وإلا لما استطعنا القول أن المركز العربي للغة التواصل العالمية مسئول عن اللغة

⁵ نتائج هذا المشروع موجودة في <http://www.eolss.net/Eolss-Definition-Context.aspx>

⁶ هذا الموقع هو <http://www.undl.org/unleolss/unleolss.htm>

العربية وممثل لها حيث سيكون هناك جزء مفقود لم يتم التعامل معه، لذلك نتحدث عن أية لغة عربية يجب تحليلها وتكويدها باستخدام نظام لغة التواصل العالمية، وأيضاً إلى أية لغة عربية يمكن فك النصوص المكتوبة بلغة التواصل العالمية، فليس من المنطق التركيز على العربية المصرية مثلاً وتجاهل باقي اللغات العربية في سوريا والكويت والسعودية وغيرها من البلاد. خصوصاً بعد أن اتضح لنا هذا الاختلاف جلياً على مستوى المفردات والتراكيب من خلال أمثلة مأخوذة من بلاد عربية مختلفة كما يلي:

(أ) الاختلاف على مستوى المفردات: تستخدم بعض الدول العربية مفردات مختلفة للتعبير عن نفس المعنى مثل كلمتي "خَبَر" و"صنع" فهما كلمتان مختلفتان ولكن لهما نفس المعنى كما في المثال (9):

(9) إن أباه خَبَر أشكالاً وأنواعاً مختلفة من المراكب الفضائية حتى أن إحداها تشبه شيئاً غريباً قطعة شوكولا محشوة بالبندق.
إن أباه صنع أشكالاً وأنواعاً مختلفة من المراكب الفضائية حتى أن إحداها تشبه شيئاً غريباً قطعة شوكولا محشوة بالبندق

وهناك نوع آخر من الاختلاف على مستوى المفردات وهو استخدام نفس المفردة مع اختلاف طريقة صياغتها كما في المثال (10)؛

(10) مسيرات طالبة ضخمة في الجامعات الأردنية. (الشام)
أكد أهمية وتدعيم الأنشطة الطلابية. (شمال أفريقيا)

(ب) الاختلاف على مستوى التركيب: كذلك نجد بعض التراكيب تختلف من بلد إلى آخر حيث تستخدم الأزمنة بطريقة مختلفة، فمثلاً بعض البلاد تستخدم كلمة "منذ" التي تدل على الماضي للتعبير عن المستقبل كما في مثال (11):

(11) منذ الموسم القادم سأبدأ تدريبيه، لكن ليس هذا ما يشغل بالي. (الشام)
بحلول الموسم القادم سأبدأ تدريبيه، لكن ليس هذا ما يشغل بالي. (شمال أفريقيا)
ويظهر الاختلاف على مستوى التركيب أيضاً في استخدام الأفعال مع المتعلقات النحوية، فمثلاً الفعل "أعطى" يحتاج بعده إلى حرف جر "لـ" للإشارة إلى من قام الفاعل بإعطائه، ولكن بعض الدول تستخدم هذا الفعل بدون حرف جر كما في مثال (12)

(12) جَلَبَتِ الأمُّ برتقالةً، أعطتها ابنها.
جَلَبَتِ الأمُّ برتقالةً، أعطتها لـ ابنها.

ومن هنا ظهرت الحاجة إلى بناء مراكز خاصة باللهجات المختلفة للبلاد العربية تكون مسؤولة عن بناء الأدوات والمصادر المطلوبة للغة التي يستخدمها شعب هذه البلد حتى يتمكن من توصيف اللغة العربية كاملة مكتمة.

أما السؤال الثاني ففي أي مجال أو تخصص من التخصصات المختلفة يتم العمل؟ فهل نقوم مثلاً ببناء المصادر اللغوية من قواميس وقواعد نحوية تختص بالتعامل مع النصوص الأدبية أو النصوص الطبية أو غيرها من المجالات المختلفة؟ أم أننا يجب أن نعمل لكي نغطي كل المجالات والتخصصات؟ في واقع الأمر إن خطة عمل المركز العربي للغة التواصل العالمية مدرج فيها بناء قواميس ومصادر لغوية تتعامل مع كافة المجالات قدر الإمكان، وفعلاً بدأ تنفيذ ذلك من خلال بناء قواميس متنوعة متخصصة، ولكن إذا نظرنا للوضع بشكل موضوعي فسوف نجد الآتي؛ فلكي يتمكن المركز العربي من بناء مصادر لغوية لكل مجال فإنه سيستغرق دهوراً.

بالإضافة إلى ذلك، فإن بعض الدول العربية تشتهر بتخصصات معينة، فمثلاً دول الخليج تشتهر بالنفط وسوريا بالطب... إلخ. وبالتالي سوف يصبح من الأفضل لو تقاسمت الدول العربية العمل كل حسب تخصصه.

لذلك فإننا نطرح دعوة للمشاركة في بناء الشبكة العربية للغة التواصل العالمية تهدف إلى بناء المصادر اللغوية الخاصة باللغة العربية على اختلاف مجالاتها واختلاف لهجاتها في البلاد العربية المختلفة. هذه الشبكة سوف يكون لها دور كبير في المحافظة على اللغة العربية ونشر ثقافة الدول العربية بين دول العالم، وسيكون لهذه الشبكة فور اكتمالها صدى كبير على بناء وتطوير الجزء الخاص باللغة العربية في المشروع وكفاءته، حيث سيصبح توصيف اللغة العربية كاملاً مكتملاً، فعندما يعهد لكل دولة ببناء الجزء الذي تبرع فيه سيكون من الممكن تطوير كل العناصر السابقة في نفس الوقت، فعلى سبيل المثال يمكن لسوريا بناء القواميس الخاصة بالمصطلحات الطبية حيث إنها تشتهر بتدريس الطب باللغة العربية، وعلى نفس المنوال تتولى المملكة العربية السعودية ترجمة المصطلحات المتعلقة بالدين الإسلامي، بينما تتولى دولة الإمارات بصفتها رائدة في مجالات التجارة والأعمال مهمة بناء القواميس الخاصة بتلك المجالات. بالإضافة إلى ذلك، تتولى كل دولة إدخال المتغيرات التي تحويها اللهجة المستخدمة من قبل شعبها، وبذلك تتم الوحدة العربية في مشروع لغة التواصل العالمية، وسوف يقدم المركز العربي للغة الشبكات العالمية بمكتبة الإسكندرية الدعم المطلوب من تدريب للمراكز الجديدة المختلفة حيث إن لديه فريقاً من اللغويين الحاسوبيين والمهندسين مدرباً على أعلى مستوى يمكنه تقديم أفضل برامج التدريب ونقل الخبرات التي تم اكتسابها خلال السنوات الماضية إلى كل من يرغب في الانضمام إلى الشبكة العربية، وكذلك سوف يقوم المركز العربي بتقديم الأدوات المطلوبة التي تم تطويرها وبنائها من خلال الخبرات المكتسبة لتسهيل العمل وتوفير الوقت والجهد ولمواجهة أية صعوبات.

4 الخاتمة

في هذا البحث تناولنا موضوع حواجز اللغة وكيف تؤدي إلى عدم العدالة في الوصول إلى المعارف، وعرضنا باقتضاب المحاولات التي ظهرت لهذه الحواجز باستخدام الحاسوب. كانت هذه المحاولات غير كافية وغير قادرة على معالجة التعقيدات التي تتجلى في اللغات الطبيعية، إلى أن ظهر منهج جديد يتمثل في لغة التواصل العالمية. تستطيع لغة التواصل العالمية القيام بدور الوسيط بين كل اللغات لقدرتها على التعامل مع اللغات المختلفة وتحليلها إلى شكل دلالي مجرد وقدرتها على توليد جمل طبيعية من مثل هذا الشكل المجرد. ولتوضيح إمكانات هذا النظام عرضنا قدرته على معالجة اللغة العربية والتي هي من أصعب اللغات الطبيعية نظراً لظواهرها التركيبية والصرفية والاشتقاقية المعقدة. وهذا النظام ليس خيالاً بل هو نظام ملموس حقيقي وعامل، فقد قام المركز العربي للغة التواصل العالمية ببناء البنية التحتية من القواميس والقواعد اللغوية والمدونات التي تمكنه من تحليل وتوليد الجمل العربية بدقة وبشكل صحيح. ويقترح المركز العربي تعزيز هذه البنية بإنشاء شبكة عربية للغة التواصل العالمية بمقدورها أن تغطي جميع اللهجات العربية المختلفة وجميع المجالات والتخصصات بما يحقق اكتمال المكوّن العربي في مشروع لغة التواصل العالمية في أسرع وقت وبأقصى كفاءة.

1. UNESCO (United Nations Educational, Scientific and Cultural Organization). Report: Measuring Linguistic Diversity on the Internet. 2005.
2. S. Lobachev. Top languages in global information production. *The Canadian Journal of Library and Information Practice and Research*. vol. 3, no. 2. 2008.
3. نبيل علي، كتاب الثقافة العربية وعصر المعلومات، عالم المعرفة، المجلس الوطني للثقافة والفنون والآداب، "سلسلة عالم المعرفة" العدد 265، الكويت، 2001
4. Noam Chomsky. Remarks on nominalization. In: R. Jacobs and P. Rosenbaum (eds.) *Reading in English Transformational Grammar*, 184-221. Waltham: Ginn. 1970.
5. Sameh Alansary, Magdy Nagi, and Noha Adly, Processing Arabic Text Content: The Encoding Component in an Interlingual System for Man-Machine Communication in Natural Language, 6th International Conference on Language Engineering, December 6-7 2006a, Cairo.
6. Sameh Alansary, Magdy Nagi, and Noha Adly, Generating Arabic Text: the Decoding Component in an Interlingual System for Man-Machine Communication in Natural Language, 6th International Conference on Language Engineering, December 6-7 2006b, Cairo.
7. Sameh Alansary, Magdy Nagi, and Noha Adly, Building an International Corpus of Arabic (ICA): Progress of Compilation Stage, 7th International Conference on Language Engineering, December 5-6 2007, Cairo.
8. Sameh Alansary, Magdy Nagi, and Noha Adly, Toward Analyzing the International Corpus of Arabic (ICA): Progress of Morphological Stage, 8th International Conference on Language Engineering, December 18-19 2008, Cairo.
9. Sameh Alansary, Magdy Nagi and Noha Adly, A Library Information System (LIS) based on UNL knowledge infrastructure, Seventh International Conference on Computer Science and Information Technologies, 28 September - 2 October, 2009a, Yerevan, Armenia.
10. Sameh Alansary, Magdy Nagi, and Noha Adly, A Semantic-Based Approach for Multilingual Translation of Massive Documents , 7th Symposium of Natural Language Processing, Thailand, December 13 - 15 2007.
11. Sameh Alansary, Magdy Nagi, and Noha Adly, The Universal Networking Language in Action in English-Arabic Machine Translation, 9th Conference on Language Engineering- December 23-24 2009b, Cairo.
12. Noha Adly, Sameh Alansary , Evaluation of Arabic Machine Translation System based on the Universal Networking Language , 14th International Conference on Applications of Natural Language to Information Systems (NLDB 2009), Saarland University, Saarbrücken-Germany, June 24 - 26 2009.
13. Gerrard Helen and Sachiko Nakamura Japanese Speakers and the Internet. *IJMS: International Journal on Multicultural Societies*. 2004, vol. 6, no.1, pp. 173-183. UNESCO.

المخلص باللغة الانجليزية

Title: The Arabic Universal Networking Language System

Abstract. This research puts forth a new and creative method to overcome language barriers; the Universal Networking Language (UNL) System. The Universal Networking Language is an artificial language that expresses human knowledge and information in the same way natural language expresses them, but in an abstract language-independent form that does not use any of the vocabulary or structures of human languages. The paper gives special attention to the Arabic component in this system; it discusses the components and resources already built in its infrastructure and briefly overviews the achievements and applications the center devoted to building this

component has accomplished so far. The paper ends in an invitation for all Arabic countries to participate in building the linguistic tools and resources required to analyze to start a Universal Networking Language Network devoted to the Arabic language only with all its varieties and dialects.

Keywords: Universal Networking Language (UNL), Semantic Network, Knowledge Representation, Language Analysis, Language Generation, Arabic UNL network, computational linguistics.

المصطلحات

الأدوار الدلالية	Semantic roles
أنتولوجيا	Ontology
البنية التركيبية السطحية	Surface syntactic structure
البنية النحوية العميقة	Deep syntactic structure
التلخيص الآلي	Automatic summarization
التنقيح الآلي	Automatic proofreading
التوصيف الدلالي	Semantic annotation
التوليد الآلي للغات	Natural Language Generation (NLG)
الدلالة الحاسوبية	Computational Semantics
شبكة دلالية	Semantic network
الشجرة النحوية	Syntactic tree
الصرف الحاسوبي	Computational Morphology
علم اللغويات الحاسوبية	Computational Linguistics
علم النفس اللغوي الحاسوبي	Computational Psycholinguistics
العلوم الإدراكية	Cognitive Sciences
عملية تحويل النصوص المكتوبة باللغات الطبيعية إلى نصوص مكتوبة بلغة التواصل العالمية	UNLization process
عملية توليد نصوص مكتوبة باللغات الطبيعية من النصوص المكتوبة بلغة التواصل العالمية	NLization process
الفاعل الدلالي	Agent
الفهم الآلي للغات	Natural Language Understanding (NLU)
قاعدة معارف	Knowledge base
قائمة بمجموعة الصفات اللغوية المستخدمة في نظام لغة التواصل العالمية	Tagset
قواعد التحليل	Analysis rules
قواعد التوليد	Generation rules
اللبس (الغموض)	Ambiguity
المتعلقات الدلالية	Semantic arguments
متعلقات المسند	Argument predicates

متمم	Complement
مخصص	Specifier
المعجمية الحاسوبية	Computational Lexicography
مفردات لغة التواصل العالمية	Universal Words (UWs)
الملامح	Attributes
الملحق	Adjunct
الموارد اللغوية	Linguistic resources
النحو الحاسوبي	Computational grammar
نظام مكتبات	Library Information System (LIS)



الجلسة الثانية عشرة

خوارزم تطوري تفاضلي كمي وخصب لحل مشاكل التحسين التوافقي

عامر ذراع¹، سهام مسهول²، سليم شيخي¹
¹ مخبر الأنظمة المعقدة، معهد الإعلام الآلي، جامعة منتوري، الجزائر
² قسم تقنيات المعلومات، جامعة الملك سعود، المملكة العربية السعودية

البريد الإلكتروني: ¹draa_amer@yahoo.fr، ²smeshoul@ksu.edu.sa، ³slchikhi@yahoo.com

الخلاصة. يقدم هذا المقال خوارزما تفاضليا كميًا وخصبا لمعالجة مشاكل التحسين التوافقي. ويرتكز الخوارزم المقترح على استعمال تمثيل كمي للأفراد واستعمال عمليتين كميتين مستنبطتين من الحوسبة الكمية: القياس والتداخل، بالإضافة إلى العملية الأساسية للخوارزمات التطورية التفاضلية. كما يقوم الخوارزم المقترح بإعادة توليد مجموعة من الأفراد دوريا لتوفير تنوع أكبر للمجموعة. ولقد تم استعمال الخوارزم المقترح لحل مشكلة تحسين بعض الدوال المرجعية في مجال التحسين التوافقي وقد أعطى ذلك نتائج جد واعدة.

الكلمات الجوهرية. الخوارزمات التطورية التفاضلية، الحوسبة الكمية، التحسين التوافقي، تحسين الدوال.

1 مقدمة

تمثل مشاكل التحسين التوافقي مجموعة من المشكلات التي تحتاج جهدا ووقتا معتبرين للتمكن من حلها. وتكمن الصعوبة فيها في عدم وجود صيغة واضحة ودقيقة لحلها، إذ انه يتوجب معاينة كل الاحتمالات ومقارنتها للحصول على أحسن حل للمشكلة. ويتميز عدد الحلول الممكنة بأنه يتضاعف أسيا كلما زاد حجم المعطيات. ولهذا نجد أن الخوارزمات الكلاسيكية، التي تدرس وتقرن كل الحلول الممكنة، محدودة الفاعلية لحل هذا النوع من المشكلات [1].

نتيجة لذلك ظهرت مؤخرا مجموعة من الخوارزمات التي تعرف بالخوارزمات التقريبية والتي تكفي بالحصول على حلول مقبولة، ليست بالضرورة الأمثل، لأجل توفير الوقت والجهد؛ حيث أن الخوارزمات الكلاسيكية قد تحتاج إلى عشرات ومئات السنين لحل بعض مشكلات التحسين التوافقي حين يتعلق الأمر بحجم كبير من المعطيات. ومن بين هذه الخوارزمات نجد الأكثر رواجاً، نظراً لما حققته من نتائج مشجعة، الخوارزمات الجينية، الشبكات العصبية الصناعية، محاكاة التلدين، البحث المحظور، والخوارزمات التطورية التفاضلية.

والخوارزميات التطورية التفاضلية خوارزميات سكانية مشابهة للخوارزميات الجينية، إذ تعتمد على الطفرة والعبور إضافة إلى الانتقاء. ويكمن الاختلاف الرئيس بين الخوارزمين في أن الجيني يعتمد بالأكثر على عملية العبور، فيما يولي التفاضلي أهمية أكثر للطفرة، والتي تتمثل في القيام بتشويش فرد من الأفراد بحاصل الفرق بين فردين آخرين يختاران عشوائيا من بين أفراد المجموعة [2، 3].

من جهة أخرى، نجد أن التمثيل الكمي المستمد من الحوسبة الكمية يوفر إمكانيات هائلة للتغلب على تعقيد المشكلة. والحوسبة الكمية، كمجال جديد بصدد التطوير، تتمثل في استعمال عمليات مستمدة من الميكانيكا الكمية كالأكوان الاحتمالية، التداخل والتراكب؛ وذلك لغرض معالجة المعلومات. وتتمثل المفاهيم الكمية الأساسية في البيت الكمي (أو الكيوبت (*qubit*))، متعدد الكيوبتات (*quregister*)، القياس، إضافة إلى العمليات الكمية [4، 5، 6].

يعتمد الخوارزم التطوري التفاضلي المقترح في هذا المقال على الطفرة كأساس للتطور ويستعمل تمثيلا كميا للأفراد، إذ يسمح هذا التمثيل، والمستمد من الحوسبة الكمية، بتنوع أكثر للحلول. إضافة إلى ذلك، يستعمل هذا الخوارزم عمليتي تداخل كمي، واحدة موضعية والأخرى كلية، واللذان تسمحان لأحسن الحلول بقيادة باقي الحلول نحو وضع أفضل في فضاء البحث. كذلك، ننبئ في هذا الخوارزم عملية توليد تقوم بتعويض بعض الأفراد بأخرى تولد عشوائيا لزيادة التنوع في المجموعة.

ويجدر بنا أن نذكر أن هذا النوع من التهجين بين الخوارزميات السكانية والحوسبة الكمية قد تم استعماله لحل عدة مشكلات في مجال الحوسبة، نذكر منها، على سبيل المثال لا الحصر، الخوارزميات الجينية الكمية لحل مشكلة الوزراء [1] ولحل مشاكل التحسين التوافقي [4]، الخوارزميات التفاضلية الكمية لحل مشكلة الوزراء [8]، والتلقائيات الخلوية الكمية لحل مشكلة التاجر المتجول [9].

يهدف تجسيد هذا النموذج وتقييمه، قمنا باختيار مشكلة تحسين الدوال، والتي تمثل إحدى أشهر مشكلات التحسين التوافقي. وبما أن هذا النوع من المشكلات يتطلب عادة مقاديرا هائلا من موارد الحساب، فاستغلال التمثيل الكمي يسهم بصورة كبيرة في تخطي مشكلة التعقيد لأنه يمنحنا إمكانية التمثيل التراكمي للحلول، ومنه تقليص حجم المجموعة مع المحافظة على تنوع الحلول. ولقد تم اختيار ثلاث دوال مرجعية شهيرة في المجال تتميز بصعوبة إيجاد الحل الأمثل فيها نظرا لكثرة الحلول الموضعية فيها، وهي: دالة روزنبروك (*Rosenbrock*)، دالة غريوانك (*Griewank*) ودالة الكرة (*Sphere*)، مع أخذ حالتين في كل دالة: من أجل 10 متغيرات و 30 متغير على التوالي.

ولقد تم مقارنة النتائج المحصل عليها مع تلك المقدمة في عمل "هوانغ ومن معه" [7] أين نجد مقارنة بين الخوارزم الكمي السربي والخوارزم التطوري الكلاسيكي، الشيء الذي أبدى فاعلية النموذج المقترح في هذا المقال. وتم اختيار عمل هوانغ ومن معه كأساس لمقارنة عملنا مع الأعمال السابقة في المجال لسببين رئيسيين. يتمثل الأول في أن هوانغ وزملاؤه يقترحون نظاما سربيا سكانية مما يجعله قريبا من النموذج المقترح في هذا العمل، و أما الثاني فيتمثل في أنهم قاموا بمقارنة فعلية مع الخوارزميات التطورية الكلاسيكية مثبتين تفوق خوارزمهم على هذه الأخيرة فتمكنا بذلك من المقارنة مع خوارزمين في آن واحد.

استنادا لما سبق، يتلخص محتوى باقي المقال فيما يلي؛ نستعرض في الفقرة الثانية المفاهيم الأساسية للخوارزمات التطورية التفاضلية والحوسبة الكمية. ونورد في الفقرة الثالثة تفاصيل النموذج المقترح في هذا العمل: الخوارزم التطوري التفاضلي الكمي. أما في الفقرة الرابعة، فنقدم ونحل بعض النتائج المحصل عليها عند اختبار الخوارزم المقترح. لنختم المقال في الفقرة الخامسة.

2 مفاهيم أساسية

1.2 الخوارزمات التطورية التفاضلية

تمثل الخوارزمات التطورية التفاضلية أحد أنواع الخوارزمات السكانية التطورية. وهي عبارة عن خوارزمات عشوائية تطورية تهدف إلى تحسين الوسائط المكونة للأفراد. ويتمثل المبدأ الأساسي لعمل هذه الخوارزمات، كما يوضحه الشكل 1، فيما يلي. يتم ابتداء تشكيل مجموعة سكانية من الأفراد تمثل نقطة انطلاق الخوارزم. بعد ذلك يتم بصورة دورية تطبيق العمليتين الأساسيتين للخوارزم وهما العبور والطفرة التفاضلية. وذلك حتى بلوغ شرط توقف الخوارزم و الذي عادة ما يكون عددا محددا من التكرارات [2، 10].

ويتمثل العبور في المزج بين فردين من مجموعة الأفراد للحصول على فردين جديدين. وهو العبور ذاته المعروف في الخوارزمات الجينية. وأما الطفرة التفاضلية، فتتمثل في تشويش فرد من الأفراد يختار عشوائيا بحاصل الفرق المثلل بين فردين آخرين يتم اختيارهما عشوائيا أيضا. ويعبر عن هذه العملية بالمعادلة (1). حيث يمثل $\vec{a}$ الفرد المراد تشويشه، ويمثل كل من $\vec{a}_{r1}$ و $\vec{a}_{r2}$ الفردين المختارين عشوائيا من المجموعة، والمعامل F عددا حقيقيا ينتمي إلى المجال $[0,2]$ ، وأما $\vec{v}$ فهو الفرد الجديد الذي يضاف إلى المجموعة معوضا الفرد القديم (حسب كفاءته طبعا).

$$\vec{V} = \vec{a} + F * (\vec{a}_{r1} - \vec{a}_{r2}) \quad (1)$$

مؤخرا، صارت معظم الخوارزمات التطورية التفاضلية تعتمد على الطفرة التفاضلية فقط، فقد تم الاستغناء عن العبور، إذ توفر هذه الطفرة الاستكشاف والاستغلال في آن واحد، فاستعمل ثلاث أفراد من المجموعة عند كل عملية يسمح باستغلال ما توصلت إليه المجموعة من حلول، وتسمح عملية الطفرة ككل باستكشاف مناطق جديدة في فضاء البحث.

2.2 الحوسبة الكمية (Quantum Computing)

ترتكز الحوسبة الكمية أساسا على تطبيق نوع جديد من العمليات، والتي تعرف بالعمليات الكمية، والتي تعمل على تمثيل جديد للمعلومات، حيث تتمثل المعلومة الأساسية في البت الكمي [10، 11].

• البت الكمي (Quantum Bit)

تستعمل الحواسيب الكلاسيكية البت كوحدة أساسية لتمثيل المعلومات، والذي يأخذ إما القيمة 1 أو 0. أما الحواسيب الكمية فتستعمل البت الكمي (*Quantum Bit*) qubit أو الكيوبت. وهو عبارة عن شعاع وحدة في فضاء هيلبرت (*Hilbert Space*). إذن فالكيوبت لا يمثل فقط 0 أو 1 ولكن

تراكبا للقيمتين معا. ويمكن إعطاء قيمته بالمعادلة (2). حيث يمثل $|0\rangle$ و $|1\rangle$ البتتين الكلاسيكيين 0 و 1 على التوالي، أما α و β فهما عدنان مركبان يحققان العلاقة الممثلة بالمعادلة (3) [12].

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (2)$$

$$|\alpha|^2 + |\beta|^2 = 1 \quad (3)$$

```

Generate random solutions that cover the given space.
Evaluate each solution.
g=1
While (convergence is not reached)
  For i=1 to Population size
    Apply differential mutation.
    Execute differential crossover
    Clip the solutions if necessary
    Evaluate the new solution.
    Apply differential selection.
  End
  g=g+1;
End

```

الشكل 1. الخوارزم التطوري التفاضلي

إذا تم قياس هذا التراكب بالنسبة للأساس $\{|0\rangle, |1\rangle\}$ ، فإن احتمال الحصول علو 0 هو $|\alpha|^2$ واحتمال الحصول على 1 هو $|\beta|^2$.

• السجل الكمي (Quantum Register)

يمكن لسجل ذي n بيت في النظام الكلاسيكي أن يأخذ قيمة واحدة من بين 2^n قيمة ممكنة. أما في النظام الكمي، فسجل كمي ذو n كيوبيت يحتوي على 2^n قيمة في آن واحد، وذلك نظرا لأن كل بيت كمي يحتوي على القيمتين 0 و 1 في نفس الوقت. هذا التراكب لـ 2^n قيمة هو ما يفسر إمكانات السرعة التي يوفرها هذا النوع من الحوسبة، فبعملية واحدة يمكن معالجة 2^n معلومة في آن واحد. على نحو مشابه للكيوبيت، فإن أساس القياس لسجل كمي هو $\{|00\dots0\rangle, |00\dots1\rangle, \dots, |11\dots1\rangle\}$ [5,4].

• القياس (Measurement)

عند قياس بيت كمي (أي ملاحظته)، يتم إسقاط الحالة الكمية على أحد حالتي الأساس المرفق بجهاز القياس. ونتيجة القياس هي نتيجة احتمالية. وتغير عملية القياس الحالة الكمية للبيت الكمي المقاس. وبطريقة مماثلة، يتم قياس سجل كمي كسلسلة متتابعة من قياسات الكيوبيت [4].

• العمليات الكمية (Quantum Gates)

تعمل العمليات الكمية على بيتات وسجلات كمية. وتأتي هذه العمليات على شكل مصفوفات وحدوية تحقق العلاقة: $M.M' = I$ حيث M هي العملية الكمية و I هي مصفوفة الوحدة. كون العمليات وحدوية يجعلها عكسية. ولقد تم إثبات أن كل عملية كلاسيكية يمكن انجازها بعملة كمية ما [11]، [12].

3 الخوارزم التطوري الكمي المقترح

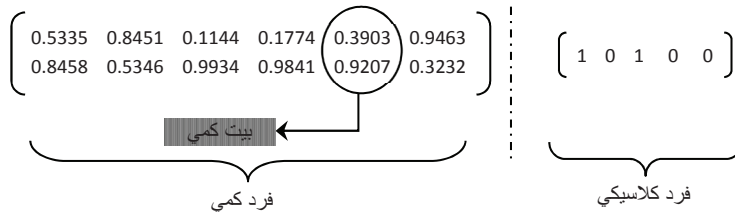
تعمل الخوارزمات التطورية التفاضلية الكلاسيكية على مجموعة من الأفراد تمثل عادة على شكل أشعة ذات عناصر ثنائية أو حقيقية. ولتمثيل أحسن لفضاء البحث، لا بد من استعمال عدد مناسب من الأفراد والذي عادة ما يكون كبيرا مما يجعل هذه الخوارزمات، كغيرها من الخوارزمات السكانية، مكلفة من حيث وقت المعالجة وحجم الذاكرة.

في هذا الإطار، و لتقليص عدد أفراد المجموعة، وبالتالي جعل الخوارزم أسرع، نعتمد في هذا العمل تمثيلا كميا للأفراد. وهو ما يقلص بصورة كبيرة عدد الأفراد ويوفر تنوعا أكبر للخوارزم. كما يستعمل الخوارزم المقترح عمليات مستمدة من الميكانيكا الكمية. ونشرح فيما يلي التمثيل الكمي للأفراد والعمليات الكمية المطبقة عليها.

1.3 التمثيل الكمي للأفراد

بطريقة مماثلة لمثيله الكلاسيكي، يستعمل الخوارزم التطوري التفاضلي الكمي تمثيلا شعاعيا للأفراد، حيث يحتوي كل شعاع على مجموع الوسائط المراد تحسينها. ويمكن الاختلاف الرئيس بينهما في أن الأول يستعمل ببتات كلاسيكية، أما الثاني فيستعمل ببتات كمية. وبذلك يمكن لفرد كمي واحد أن يمثل كل أفراد المجموعة في آن واحد. ويقدم الشكل 2 مثالا عن فرد كلاسيكي (يمينا) وآخر كمي (يسارا).

كما أسلفنا، يمكن لكل بيت كمي أن يمثل 0 و 1 في نفس الوقت كلا باحتمال. وعليه، يمثل فرد كمي في المجموعة كل الأفراد في آن واحد، كلا باحتمال. مثلا، البيت الكمي المشار إليه بالدائرة في الشكل 1. يمثل 0 باحتمال $(0.3903)^2$ و 1 باحتمال $(0.9207)^2$ ، إذن، يسمح هذا التمثيل لأفراد قليلة بتغطية كل فضاء البحث، ومنه، لن يكون ضروريا استعمال عدد أكبر من الأفراد.



الشكل 2. التمثيل الكلاسيكي (يمينا) والتمثيل الكمي (شمالا) للأفراد

2.3 الخوارزم المقترح

اعتمادا على التمثيل الكمي الذي تم شرحه في الفقرة السابقة، نقوم فيما يلي بشرح كيفية عمل الخوارزم التطوري التفاضلي الكمي لحل مشاكل التحسين التوافقي.

بداية، يتم اختيار المجموعة السكانية الابتدائية بصورة عشوائية، والمكونة من عدد من الأفراد الكمية، ويتم تقسيم هذه الأفراد إلى مجموعات جزئية نصلح عليها هنا بالقبائل. ثم يقوم الخوارزم بتنفيذ العمليات التالية بصورة دورية.

- القياس
- التقييم
- المقارنة والتحديث
- الطفرة التفاضلية الكمية
- التداخل الكمي الموضعي والتداخل الكمي الكلي
- إعادة التوليد

طيلة عملية المعالجة، يتم الاحتفاظ بأحسن نتيجة كلية (لكامل المجموعة) وبأحسن نتيجة موضعية (لكل قبيلة) لاستعمالهما في عمليتي التداخل.

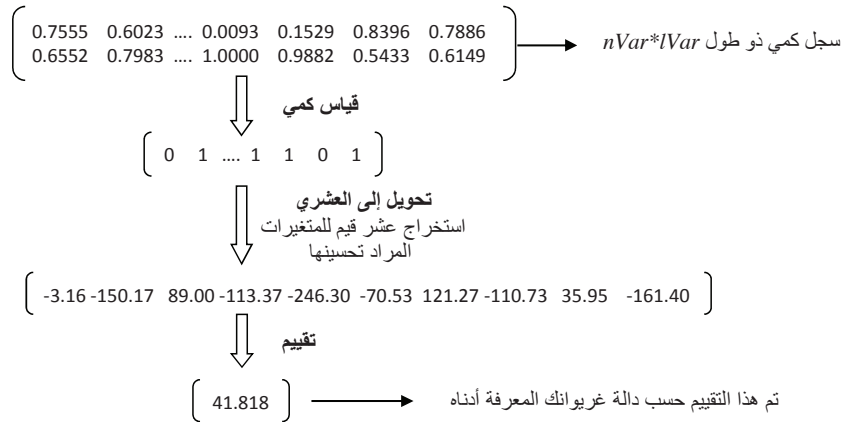
تسمح عملية **القياس** بالحصول على فرد ثنائي (كلاسيكي) انطلاقاً من فرد كمي وذلك اعتماداً على التمثيل الكمي للبيئات. علماً أن قياس كل بيت كمي للحصول على بيت ثنائي يستند إلى قانون احتمالي. ليكن A الفرد الكمي المراد قياسه، يتم الحصول على فرد ثنائي كلاسيكي بتطبيق قياسات متوالية على البيئات الكمية المشكلة لهذا الفرد. ويتم قياس كل بيت كمي بمقارنة مربع المركبة الثانية β مع عدد عشوائي r فإذا كان $r \geq |\beta|^2$ فالنتيجة هي 1 وإلا فالنتيجة تكون 0. وعلى عكس القياس الكمي الخالص الذي نجده في الآلات الكمية، والذي يحطم حالة التراكب، فإن القياس في هذا الخوارزم يسمح لنا بالحفاظ عليها، لأنه يتم في آلة كلاسيكية. وهذا يخدم الخوارزم لأنه يحتاج لهذا التراكب في تكرارات قادمة.

أما عملية **التقييم**، فنقوم فيها بحساب جودة الحل المقدم من المصفوفة الثنائية الناتجة عن القياس، وذلك استناداً لنوع المشكلة التوافقية المعالجة. ففي حالة تحسين الدوال المدروسة في هذا المقال مثلاً، يتم حساب قيمة الدالة بعد تحويل الوسائط من الشكل الثنائي إلى الشكل العشري. ويعطي الشكل 3 مثلاً عن عمليتي القياس والتقييم. حيث يمثل $nVar$ و $IVar$ عدد المتغيرات (أي بعد الدالة) و طول التمثيل الثنائي لكل متغير على التوالي.

بعد ذلك يقوم الخوارزم **بمقارنة** نتيجة التقييم المحصل عليها بكفاءة أحسن حل كلي $globalBest$ و أحسن حل في قبيلة الفرد الحالي $localBest(k)$ ، أين يمثل $flokNumber$ و $k=1$. عدد القبائل في المجموعة. في حالة وجود تفوق في الكفاءة، يتم **تحديث** أحسن حل كلي أو أحسن حل موضعي. وتجدر الإشارة إلى أن أحسن حل كلي هو عبارة عن فرد ثنائي، و أما أحسن حل موضعي فهو فرد كمي. وذلك لغاية إمكانية تطبيق التداخلين الكمييين الكلي والموضعي والذين نشرحهما لاحقاً.

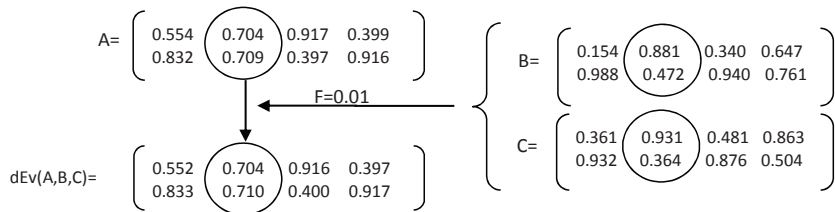
بعد هذا يتم تطبيق **الطفرة التفاضلية الكمية**. وتقوم هذه العملية بتشويش فرد من الأفراد يتم اختياره عشوائياً بحاصل الطرح المثل لفردين كمييين يتم اختيارهما عشوائياً أيضاً. وذلك مع مراعاة الشرط المعبر عنه سابقاً في المعادلة (3). و لهذا تقوم هذه العملية بالعمل على مركبة واحدة في كل بيت كمي، ويتم استنتاج المركبة الثانية انطلاقاً من المعادلة (3). و يعطي الشكل 4 مثلاً عن تشويش الفرد A بالفرق المثل (بالمعامل $F=0.01$) للفردين B و C .

و تسمح عملية التداخل الكمي الموضوعي بتقريب كل فرد كمي A من المجموعة نحو أحسن حل موضوعي في قبيلته. أي أن كل كيوبت في الفرد A يقرب للكيوبت ذي نفس الموقع في الفرد الكمي B . وذلك باستعمال معامل حقيقي يتم اختياره حسب نوع المشكلة المعالجة. ويتم هذا حسب الاحتمال $locIntProb$ الذي يمثل احتمال تطبيق التداخل الكمي الموضوعي.



الشكل 3. القياس والتقييم في حالة تحسين الدوال

وتقوم عملية التداخل الكمي الكلي بتقريب أحسن حل موضوعي لكل قبيلة نحو أحسن حل كلي للمجموعة. وتختلف هذه العملية عن مثيلتها الموضوعية في كون التقريب يحدث نحو بيئات كلاسيكية. فإذا كان A الفرد المراد تقريبه نحو أحسن فرد كلي $globalBest$ ، يتم دراسة حالة كل بيت في هذا الأخير، واستنادا لذلك ترفع قيمة إحدى المركبتين في البيت الكمي وتخضع الأخرى. فيرفع احتمال الحصول على الصفر (أي المركبة الأولى فيه) إذا كان البيت ذو نفس الرتبة في أحسن حل كلي مساويا للصفر، وإلا يحدث العكس. مع العلم أن رفع قيمة مركبة أو خفضها لا بد أن يكون متبوعا بتعديل لقيمة المركبة الأخرى للحفاظ على الشرط الوارد في المعادلة (3). ويعطي الشكل 5 مثالين لتداخل كمي موضوعي (أ) وتداخل كمي كلي (ب).



أخيراً، يقوم الخوارزم بتطبيق عملية إعادة التوليد، مما يجعل منه خوارزماً خصباً. في هذه العملية، يتم تبديل بعض الأفراد من المجموعة بأفراد جدد وذلك لزيادة التنوع.

مع مرور الوقت، يؤدي كلا من التداخلين الموضوعي والكلي إلى الحصول على أفراد متشابهة كثيراً، ومنه تصبح الطفرة التفاضلية بلا فائدة لأن الفرق بين فردين يعطي دائماً الصفر أو الصفر تقريباً. إذن، يقع الخوارزم في مشكلة الحلول الموضوعية (*local minima*). إذن، التبديل الكلي لبعض الأفراد بأخرى يتم توليدها عشوائياً يسمح بتفادي النضوج المبكر للنظام. وكغيرها، تطبق هذه العملية باحتمال معين والذي عادة ما يكون ضئيلاً من 2 إلى ثلاث مرات في عمر الخوارزم.

4 النتائج التجريبية

1.4 المشكلة المعالجة

لدراسة مدى نجاعته، تم تجسيد الخوارزم لحل مشكلة تحسين الدوال وتحديد الدوال التالية [7].

أ. دالة روزنبروك (Rosenbrock Function)

$$f(x) = \sum_{i=1}^{n-1} \left[(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2 \right]$$

حيث x_i عدد حقيقي في المجال $[-30, +30]$ ، و n عدد المتغيرات. ويتمثل الحل الأمثل في $(0, 0, \dots, 0)$ ذي الكفاءة $f(\vec{x}^*) = 0$

ب. دالة غريوانك (Griewank Function)

$$f(x) = \frac{1}{4000} \left(\sum_{i=1}^n (x_i - 100)^2 \right) - \prod_{i=1}^n \cos\left(\frac{x_i - 100}{\sqrt{i}}\right) + 1$$

حيث x_i عدد حقيقي في المجال $[-600, +600]$ ، و n عدد المتغيرات. ويتمثل الحل الأمثل في $(100, 100, \dots, 100)$ ذي الكفاءة $f(\vec{x}^*) = 0$.

ج. دالة الكرة (Sphere Function)

$$f(x) = \sum_{i=1}^n (x_i)^2$$

حيث x_i عدد حقيقي في المجال $[-100, +100]$ ، و n عدد المتغيرات. ويتمثل الحل الأمثل في $(0, 0, \dots, 0)$ ذي الكفاءة $f(\vec{x}^*) = 0$.

$$A = \begin{pmatrix} 0.4484 & \mathbf{0.8753} & 0.3950 & 0.2323 \\ 0.8938 & \mathbf{0.4836} & 0.9187 & 0.9726 \end{pmatrix} \quad B = \begin{pmatrix} 0.6470 & \mathbf{0.2212} & 0.2062 & 0.8955 \\ 0.7625 & \mathbf{0.9752} & 0.9785 & 0.4452 \end{pmatrix}$$

$$QuInterfere(A,B) = \begin{pmatrix} 0.4815 & \mathbf{0.7663} & 0.3636 & 0.3428 \\ 0.8764 & \mathbf{0.6425} & 0.9316 & 0.9394 \end{pmatrix}$$

أ. التداخل الكمي الموضعي

$$A = \begin{pmatrix} 0.4484 & \mathbf{0.8753} & 0.3950 & 0.2323 \\ 0.8938 & \mathbf{0.4836} & 0.9187 & 0.9726 \end{pmatrix} \quad S = \begin{pmatrix} 1 & \mathbf{0} & 1 & 0 & 0 \end{pmatrix}$$

$$Interfere(A,S) = \begin{pmatrix} 0.4113 & \mathbf{0.9152} & 0.3619 & 0.5857 \\ 0.9115 & \mathbf{0.4030} & 0.9322 & 0.8105 \end{pmatrix}$$

ب. التداخل الكمي الكلي

الشكل 5. التداخل الكمي الموضعي (أ)، و التداخل الكمي الكلي (ب)

2.4 اختيار المعلومات

من أجل كل دالة تم أخذ عدد المتغيرات 10 ثم 30 على التوالي، وتم تشكيل مجموعة السكان ب 5 قبائل ذات 8 أفراد في كل قبيلة، ثم ب 8 قبائل ذات 10 أفراد في كل قبيلة. أما عدد التكرارات فقد أختير 1000 تكرار من أجل 10 متغيرات و 2000 من أجل 30 متغير. مع تنفيذ البرنامج 50 مرة لكل حالة. ولقد تم اختيار هذه القيم للمعلومات للتمكن من مقارنة موثوقة مع نتائج عمل [7] حيث يتم استعمال خوارزم سربي كمي لتحسين نفس الدوال. ويلخص الجدول الأول المعلومات التي تم اختيارها في كل حالة من حالات التجارب.

الجدول الأول: المعلومات المختارة لكل تجربة

الدالة المحسنة	معامل التداخل	معامل الطفرة	احتمال الطفرة	احتمال إعادة التوليد	احتمال التداخل الموضعي	احتمال التداخل الكلي
روزنبروك	5	0.001	0.01	0.01	0.1	0.1
غريوانك	5	0.001	0.01	0.01	0.1	0.1
الكرة	5	0.001	0.01	0.1	0.5	0.1

3.4 النتائج التجريبية

يلخص الجدول الثاني نتائج تنفيذ الخوارزم التطوري التفاضلي الكمي ومقارنتها مع نتائج [7] ونتائج الخوارزمات التطورية الكلاسيكية المقررة في نفس المقال. في كل حالة يتم حساب المعدل والانحراف المعياري بعد 50 تنفيذ للخوارزم، مع الملاحظة أنه قد تم اختيار طول كل متغير 18 كيوبيت للدالتين الأوليين و330 للدالة الثالثة للحصول على نفس الدقة لإجراء مقارنة عادلة مع [7].

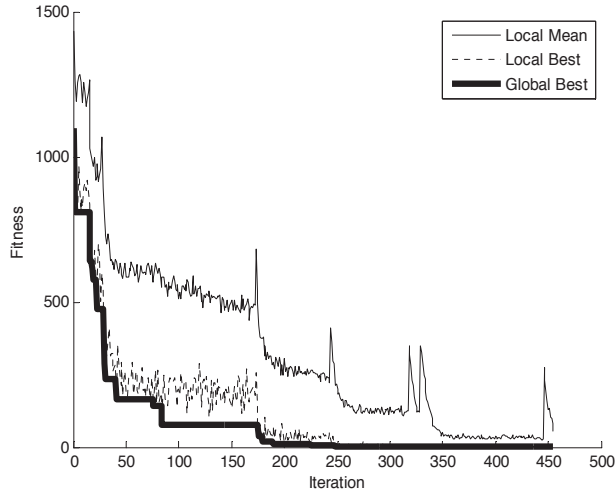
كما هو واضح في الجدول الثاني، فإن الخوارزم المقترح في هذا المقال (التفاضلي الكمي) تفوق على نظيريه، الجني والسريبي الكمي في كل الدوال؛ فقد تحصلنا دائماً على معدل أحسن للـ 50 تنفيذ. كما تم الحصول على الحل الأمثل في الدالة الثالثة (Sphere). كذلك، يبين الانحراف المعياري، أن الحلول دائماً جد مقبولة إذ تدور بالقرب من المعدل.

الجدول الثاني: النتائج التجريبية

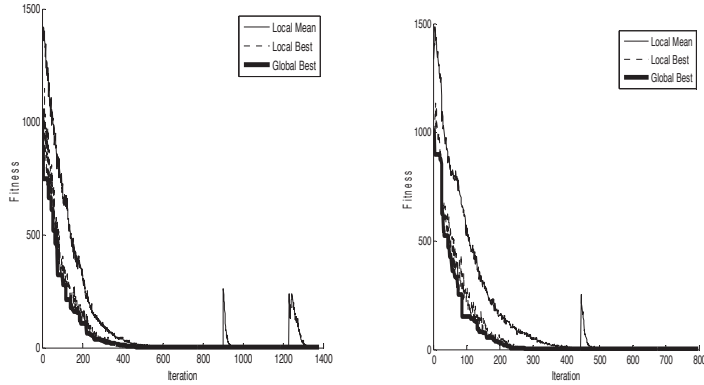
التطوري الكلاسيكي		السريبي الكمي		التفاضلي الكمي		التكرارات	القبائل * الأفراد	عدد المتغيرات	الدالة المحسنة
الانحراف المعياري	المعدل	الانحراف المعياري	المعدل	الانحراف المعياري	المعدل				
174.110	71.023	14.479	10.423	1.667	8.609	1000	8*5	10	دالة روزنبروك
478.625	289.583	63.485	59.028	35.582	34.771	2000	8*5	30	
57.431	37.365	16.589	8.634	1.559	8.448	1000	10*8	10	
289.968	202.456	40.843	51.565	23.348	31.741	2000	10*8	30	
0.073	0.085	0.051	0.069	0.0442	0.006	1000	8*5	10	دالة غريوانك
0.015	0.013	0.017	0.012	0.1331	0.0397	2000	8*5	30	
0.071	0.075	0.033	0.034	0.0777	0.0173	1000	10*8	10	
0.014	0.013	0.009	0.011	0.0346	0.0073	2000	10*8	30	
8.01E-23	3.11E-23	5.82E-71	8.25E-72	0	0	1000	8*5	10	دالة الكرة
5.08E-10	2.26E-10	6.71E-28	1.85E-28	0	0	2000	8*5	30	
2.64E-27	6.13E-28	2.09E-99	3.09E-100	0	0	1000	10*8	10	
7.16E-12	2.47E-12	2.66E-48	1.09E-48	0	0	2000	10*8	30	

ويمثل الشكل 6 كيفية تطور أحسن نتيجة موضعية، معدل النتائج الموضعية وأحسن نتيجة كلية تبعاً للزمن (عدد التكرارات). حيث يمثل الخط الغليظ أحسن نتيجة كلية ويمثل الخط المنقط أحسن نتيجة موضعية، فيما يمثل الخط المستمر الرفيع معدل النتائج الموضعية.

كما هو واضح في الشكل 7، يسمح تطبيق الطفرة التفاضلية الكمية (يمينا) بايجاد الحل الأمثل في عدد أقل من التكرارات مقارنة مع حالة عدم تطبيقها (شمالاً). كذلك، عند مقارنة تطور أحسن حل كلي بين حالة تطبيق الطفرة الكمية التفاضلية وحالة عدم تطبيقها، يتضح جلياً مدى الريح الذي نجنيه فيما يتعلق بتقارب الخوارزم إجمالاً، كما هو ظاهر في الشكل 8. وعلى نحو مماثل توضح الأشكال من 9 إلى 14 الفائدة من تطبيق كل عملية من العمليات الأخرى للخوارزم. ففي كل من الأشكال 9، 11 و13، نجد يمينا كيفية تطور معدل الحلول، أحسن حل موضعي (لكل تكرار) وأحسن حل كلي؛ ونجد شمالاً هذه التطورات عند عدم تطبيق العملية المعنية بالدراسة. كما تمثل الأشكال 10، 12 و14 مقارنة لتطور أحسن حل كلي للخوارزم عند تطبيق العملية المعنية و عند عدم تطبيقها.

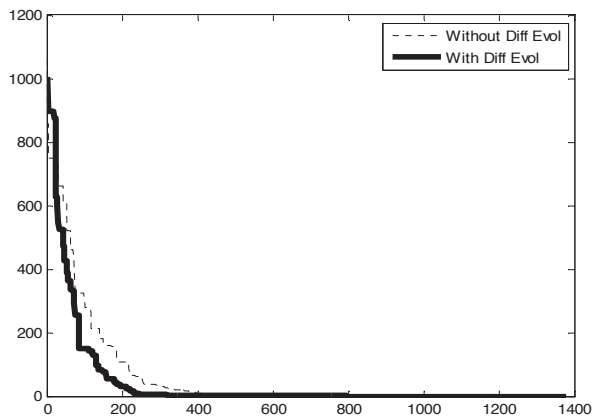


الشكل 6. تطور أحسن حل كلي (الخط الغليظ)، أحسن حل موضعي (الخط المنقط) ومعدل الحلول الموضعية (الخط المستمر الرفيع)

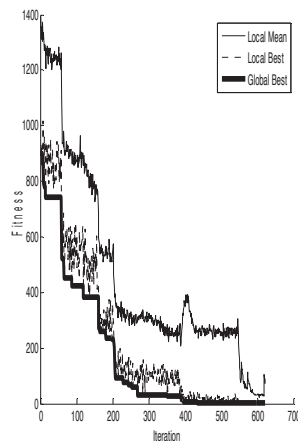
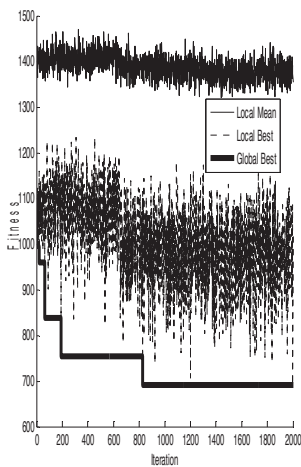


الشكل 7. مقارنة تطور معدل الحلول، أحسن حل موضعي وأحسن حل كلي عند تطبيق الطفرة التفاضلية الكمية (على اليمين) وعند عدم تطبيقها (على الشمال)

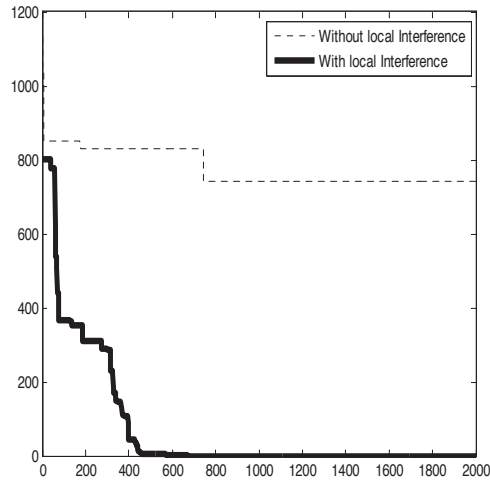
نرى في الشكل 9 أهمية عملية التداخل الكمي في قيادة البحث، يمثل الشكل شمالا بحثا عشوائيا لا يسمح بتقارب الخوارزم إلى بعد وقت طويل. ونرى في الشكل 11 أن التداخل الكلي يسمح بالانجراف نحو حلول أفضل (يمين)، فيما يطول الأمر بالخوارزم من دونه في نقاط موضعية (شمالا).



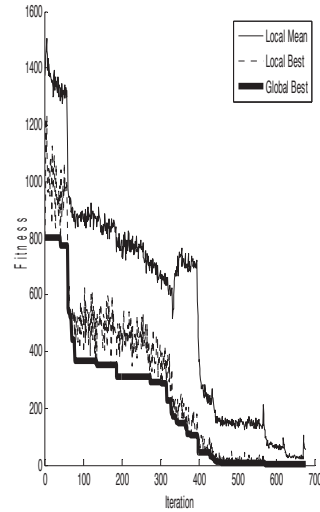
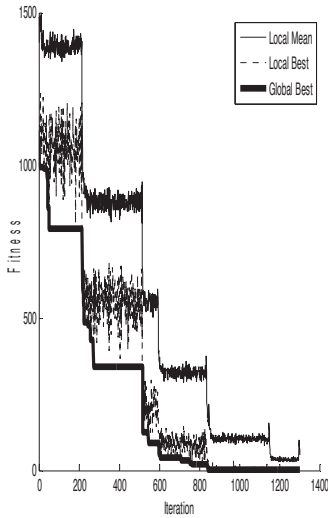
الشكل 8. تطور أحسن حل كلي بتطبيق الطفرة التفاضلية الكمية (الخط الغيظ) وعدم تطبيقها (الخط المتقطع)



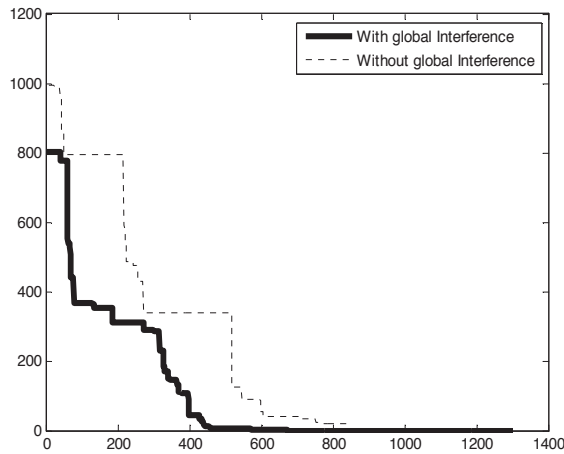
الشكل 9. مقارنة تطور معدل الحلول، أحسن حل موضعي وأحسن حل كلي عند تطبيق التداخل الموضعي (على اليمين) وعند عدم تطبيقه (على الشمال)



الشكل 10. تطور أحسن حل كلي بتطبيق التداخل الكلي (الخط الغيظ) وعدم تطبيقه (الخط المتقطع)

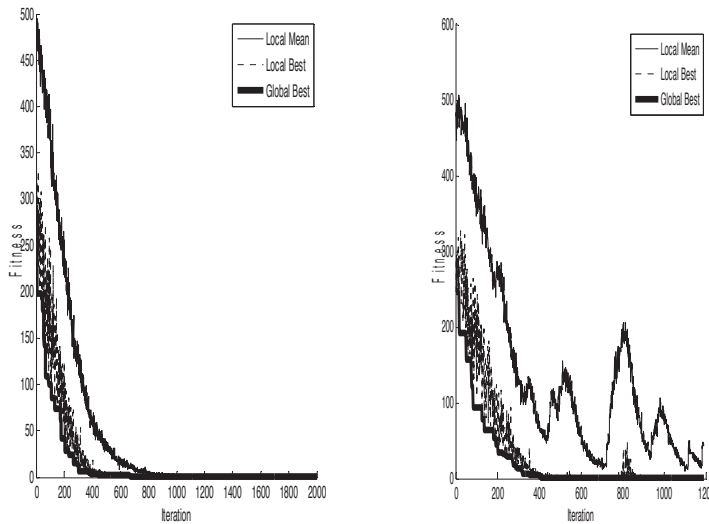


الشكل 11. مقارنة تطور معدل الحلول، أحسن حل موضعي وأحسن حل كلي عند تطبيق التداخل الكلي (على اليمين) وعند عدم تطبيقه (على الشمال)

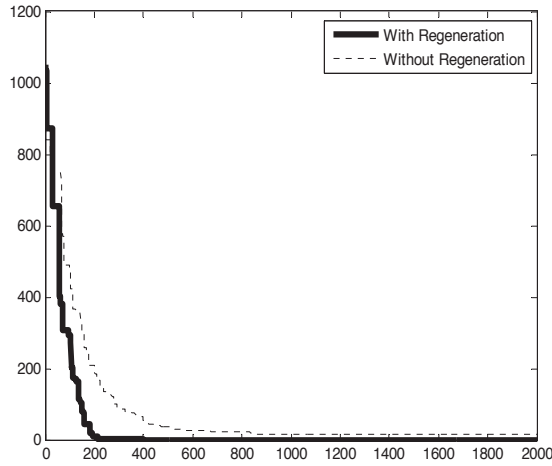


الشكل 12. تطور أحسن حل كلي بتطبيق التداخل الكلي (الخط الغيظ) وعدم تطبيقه (الخط المتقطع)

كذلك الحال فيمل يخص إعادة التوليد، والتي تسمح للخوارزم بتنوع أكبر كما هو مبين بالشكل 13.



الشكل 13. مقارنة تطور معدل الحلول، أحسن حل موضعي وأحسن حل كلي عند تطبيق إعادة التوليد (على اليمين) وعند عدم تطبيقها (على الشمال)



الشكل 14. تطور أحسن حل كلي بتطبيق إعادة التوليد (الخط الغيظ) وعدم تطبيقها (الخط المتقطع).

5 الخاتمة

تم في هذا المقال تقديم خوارزم جديد لحل مشاكل التحسين التوافقي، وهو عبارة عن خوارزم تطوري تفاضلي كمي وخصب. يركز هذا الخوارزم أساساً على استعمال تمثيل كمي للأفراد مستمد من الحوسبة الكمية، والذي يوفر تنوعاً أكبر للخوارزم. كما تم تعديل العملية الأساسية للخوارزمات التفاضلية بتعريف عملية تفاضلية كمية حل محلها. كذلك، تم استغلال عملية التداخل المستمدة من الحوسبة الكمية بشكلها الموضعي والكلّي. إضافة لهذه العمليات، تم تبني عملية جديدة تتمثل في عملية إعادة التوليد، وتسمح هذه الأخيرة بتفادي الوقوع في النضج المبكر للخوارزم، وهذا ما يجعل الخوارزم خصباً.

ولتجريب مدى نجاعة هذا النموذج، قمنا بتجسيده لحل مشكلة تحسين الدوال كمسألة كلاسيكية مرجعية في مجال التحسين التوافقي. ولقد أثبتت النتائج المحصل عليها والتي قدمت على شكل مقارنة مع خوارزم كمي آخر ومع خوارزم جيني سابقين لحل نفس المشكلة، مدى قدرات هذا الخوارزم لحل مشاكل التحسين التوافقي. إضافة إلى ذلك، بيّنا من خلال الأشكال الممثلة لتطور كفاءة الأفراد، الفائدة المجنية من تنفيذ كل عملية.

كأعمال مستقبلية، نطمح لاستغلال النموذج المصمم لحل مشاكل أخرى في مجال التحسين التوافقي أو المعالجة الآلية للمعلومة بصفة أعم.

- [1] A. Draa, H. Talbi, and M. Batouche, **A Quantum-Inspired Genetic Algorithm for Solving the N-Queens Problem**, in the *Proceedings of the 7th International Symposium on Programming and Systems, (ISPS'2005)*, Algiers, May 2005.
- [2] M. Salomon, **Etude de parallelisation de méthodes heuristiques d'optimisation combinatoire, Application au recalage d'images médicales**, doctorate thesis, Luis pasteur university, Strasbourg I, 2001.
- [3] D. Karaboga, S. Ökdem, **A Simple and Global Optimization Algorithm for Engineering Problems: Differential Evolution Algorithm**, *Turk J Elec Engin*, vol.12, no.1, 2004.
- [4] K. Han, J. Kim, **Genetic Quantum Algorithm and its Application to Combinatorial Optimization Problem**, In: *Proceedings of The 2000 Congress on Evolutionary Computation*, pp. 1354-1360, IEEE Press, 2000.
- [5] K. Han., and J. Kim, **Quantum-inspired evolutionary algorithm for a class of combinatorial optimization**, *IEEE transactions on evolutionary computation*, vol. 6, no. 6, 2000.
- [6] C. Chunlin, D. Daoyi, **Quantum Intelligent Mobile System**, *Quantum Inspired Intelligent Systems, Studies in Computational Intelligence*, Vol. 121, ISSN 1860-949X, 2008, Springer-Verlag Berlin Heidelberg.
- [7] Y.R. Huang, C.L. Tang, and S. Wang, **Quantum-Inspired Swarm Evolution Algorithm**, *Journal of Communication and Computer*, Vol.5, no.5, ISSN 1548-7709, 2008, USA.
- [8] A. Draa, S. Meshoul, H. Talbi, M. Batouche, **"A Quantum-Inspired Differential Evolution Algorithm for Solving the N-Queens Problem"**, *The International Arab Journal of Information Technology (IAJIT)*. Vol. 7 N°.1, pp. 21-27, 2010
- [9] A. Draa, S. Meshoul, **"A Quantum-Inspired Learning Cellular Automaton for Solving the Travelling Salesman Problem"**, In the *Proceedings of the 12th United Kingdom Simulation, Implementation and Modelling (UKSIM 2010)*, March 2010, Cambridge, UK.
- [10] A. Draa, H. Talbi, and M. Batouche, **A Quantum-Inspired Differential Evolution Algorithm for Rigid Image Registration**, In: *Proceedings of The International Conference on Computational Intelligence (ICCI'2004)*, Istanbul, ISBN 975-98458, 2004.
- [11] P. Shor, **Algorithms for quantum computation: discrete logarithms and factoring**, In: *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, IEEE Press, 1994.
- [12] E. Rieffel, W. Polak, **An introduction to quantum computing for non-physicists**, arxiv.org, quant-ph/9809016 v2, 2000.

Title: A Fertile Quantum-Inspired Differential Evolution Algorithm for Combinatorial Optimisation

Abstract. In this paper, a Fertile Quantum-inspired Differential Evolution approach (FQDE) for solving combinatorial optimisation problems is presented. The proposed approach is a hybridisation between a differential evolution algorithm and quantum computing. In addition, to the operations inspired from quantum computing, an operation of regenerating the population or some of its individuals is applied periodically in order to escape local minima, which makes the approach a fertile one. The algorithm has been used to solve a well-known combinatorial optimisation problem, that of functions optimisation. Three benchmark functions with two instances in each have been used to confirm the abilities of the algorithm. Very promising results have been obtained. Comparisons of the obtained results with those obtained by a quantum swarm algorithm and a conventional evolutionary algorithm proved that the fertile quantum inspired evolutionary algorithm is better.

Keywords: Differential Evolution, Quantum Computing, Combinatorial Optimization, Functions Optimisation

اللفظ العربي	اللفظ الانجليزي
إعادة توليد	Regeneration
البيت الكمي (الكيوبيت)	Quantum bit
التداخل	Interference
التراكب	Superposition
التشويش	Perturbation
التقييم	Evaluation
التنوع	Diversity
الحوسبة الكمية	Quantum computing
خوارزم تطوري تفاضلي كمي وخصب	A Fertile Quantum Differential Evolution Algorithm
خوارزم سكاني	Population-based algorithm
الخوارزمات الجينية	Genetic Algorithms
السجل الكمي	Quantum register
طفرة تفاضلية	Differential mutation
العبور	Crossover
الفرق المثقل	Weighted difference
القياس	Measurement
كلي	Global
مجموعة سكانية	Population
مشاكل التحسين التوافقي	Combinatorial optimisation problems
مقارنة وتحديث	Comparison and update
موضعي	Local
وسائط	Parameters



الدورة السابعة لل مؤتمر الدولي لعلوم وهندسة الحاسوب باللغة العربية

تحسين مطابقة الصور باستخدام معيار القرب

إبراهيم كلزيم¹، أحمد حمّوش²، إدريس أبو تاج الدين¹

¹ جامعة محمد الخامس أكادال، كلية العلوم. مختبر الأبحاث في الإعلاميات و الإتصالات. وحدة 29 مرتبطة بالمركز الوطني للبحث العلمي. الرمز البريدي 1014 الرباط، المغرب.

² جامعة محمد الخامس السوسي، المدرسة العليا للتعليم التقني. الرباط، المغرب.

البريد الإلكتروني: ibr_guelzim@yahoo.fr – guelzim@cnrst.ma

الخلاصة. الهدف من هذا العمل هو تقديم خوارزمية جديدة لمطابقة الصور المجسمة لإعادة بناء صورة ثلاثية الأبعاد. نقترح استخدام معيار القرب لتحسين أداء المطابقة مع استعمال مبدأ الفرصة الثانية. مقياس التشابه المستعملين هما مقياس المعلومة المتبادلة و معامل الارتباط. تتم المطابقة بين جوارات النقط المميزة المستخرجة من الصور. نعتبر أن الصور المستعملة التقطت بألة تصوير واحدة في أوقات متقاربة. أجريت الإختبارات على صور رمادية التلوين.

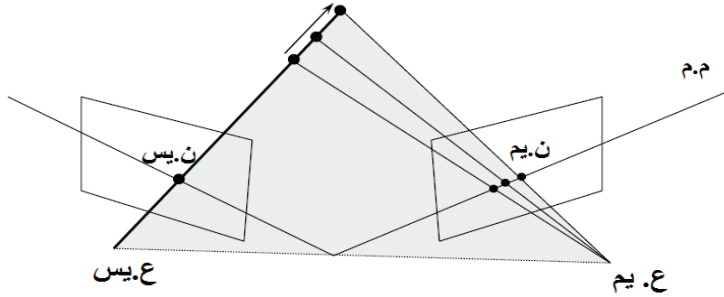
الكلمات الجوهرية: الرؤية بالحاسوب، بناء الصور ثلاثية الأبعاد، الصور المجسمة، علم الروبوت

1 مقدمة

إن محاكاة النظام البصري البشري هو أحد الأهداف الرئيسية للرؤية بالحاسوب. في هذا السياق، كانت معالجة الصور المجسمة موضوع كثير من البحوث خلال العقود الأخيرة. إن النظام البصري البشري يستطيع تكوين صورة ثلاثية الأبعاد بضم صورتين الملتقطتين بالعينين ثم المطابقة بينهما. بهذه الطريقة تقوم الرؤية بالحاسوب بتكوين الصور ثلاثية الأبعاد بالإعتماد على صورتين على الأقل لنفس المشهد. يمكن الحصول على هتين الصورتين إما باستعمال التنين للتصوير ثلثتقطان المشهد من زاويتين مختلفتين و إما باستعمال آلة متحركة للتصوير ثلثتقط المشهد في لحظتين متقاربتين. هذه الوسيلة الأخيرة هي المستعملة غالبا في مجال الروبوتات المتحركة. تتم المطابقة بين الصورتين بالمطابقة بين النقط المميزة أو بين الخطوط المميزة المستخرجة من كلتا الصورتين. المقاربة المعتمدة على الخطوط المميزة تستهلك وقت أكبر من تلك المعتمدة على النقط المميزة نظرا لعدد النقط الهائل الذي يجب مطابقتها. لهذا السبب، تركز معظم البحوث على مطابقة النقط المميزة والتي بدأ فكرتها مورافيك[10].

في هذه الحالة و بفضل خاصية إستواء النقط المتطابقة تُحسب الإحداثيات ثلاثية الأبعاد بالمطابقة بين جوار النقط المعنية. إذا كان حجم الجوار المستعمل صغير تتقلص كمية المعلومة المتوفرة للمطابقة أما إذا كان الحجم كبيرا تكون مصداقية المعلومة أكبر من الناحية الإحصائية، لكن قد يتسبب في احتجاب معلومات مهمة[11]. خلال [6] ، اقترح الكُتاب ضبط حجم الجوار حسب طبيعة الصورة المعالجة.

يُستمد قيد إستواء النقط المتطابقة مباشرة من هندسة جهاز الإستقبال. بفضل هذا القيد، نقصر في البحث عن مطابق نقطة في الصورة اليسرى على مستقيم المطابقة في الصورة اليمنى عوض البحث في الصورة بأكملها. (الشكل 1).



الشكل 1. قيد إستواء النقط المتطابقة: مطابق النقطة اليسرى ن.يس يوجد على مستقيم المطابقة ع.يم

ويبقى هذا القيد صالحا سواء خضع جهاز الإستقبال للمعايرة أو لا [8][20]. و هو أيضا صالح لمطابقة الخطوط المميزة [2] [15] أو حتى مطابقة أجزاء مميزة من الصورة [5] [17] [19]. لكن حساب مستقيم المطابقة لكل نقطة يستهلك وقتا كبيرا، بالإضافة إلى ضرورة معرفة مسبقة لمعطيات جهاز الإستقبال.

من بين المجالات التي تستعمل المطابقة بين الصور نجد الروبوت المتحرك [10]، التعرف على الوجه [13] و تكوين الصور ثلاثية الأبعاد.

في هذا المقال نقترح طريقة جديدة لتحسين أداء المطابقة معتمدة على إدراج معيار القرب، ثم نقارن بين المطابقة التي استعمل فيها المعيار والتي لم تستعمل. إعتدنا أيضا في الطريقة المقترحة على مبدأ الفرصة الثانية. مقياس التشابه المستعملين هما مقياس المعلومة المتبادلة و معامل الارتباط.

في ما يلي نبدؤ بتقديم مقياس التشابه المستعملين، ثم نقدم الطريقة المقترحة للمطابقة بين الصور، لنخلص إلى خاتمة ورؤى مستقبلية.

2 مقياس التشابه المستعملين

2.1 المعلومة المتبادلة

المعلومة المتبادلة (MI) بين متغيرين عشوائيين تقيس كمية المعلومة المتوفرة لمعرفة متغير بالإطلاع على المتغير الآخر. المعلومة المتبادلة بين المتغيرين العشوائيين $\{x_1, x_2, \dots, x_k\} = X$ و $\{y_1, y_2, \dots, y_n\} = Y$ هي:

$$MI(X, Y) = H(X) - H(X/Y) \quad (1)$$

$$= H(Y) - H(Y/X) \quad (2)$$

$$= H(X) + H(Y) - H(X, Y) \quad (3)$$

علما أن H تمثل دالة الأنتربي و هي تساوي:

$$H(X) = E[h(x_i)] = -\sum_{i=1}^k p_i \log_2(p_i(x_i)) \quad (4)$$

مع $h(x) = -\log(p(x))$ و $p_i = P(X = x_i) / i \in \{1, 2, \dots, k\}$

المعلومة المتبادلة هي كمية موجبة، متماثلة تنعدم إذا كان المتغيرين مستقلين. وهي تخضع لمبدأ عدم خلق المعلومات (أو نظرية معالجة البيانات)

إذا كانت g_1 و g_2 دالتين قابلتين للقياس، فإن :

$$MI(g_1(X), g_2(Y)) \leq MI(X, Y) \quad (5)$$

المتفاوتة (5) تعني أن أي معالجة للمعلومة الخام لا يمكن أن تعطي كمية معلومات أكبر.

المعلومة المتبادلة هي مقياس تشابه متفق عليه [13] [3] [23] وهي مستخدمة في المطابقة [21]، وتسجيل الصور [1]، واختيار المعلمة [22] وغيرها من التطبيقات.

2.2 معامل الارتباط

معامل الارتباط ρ بين متغيرين عشوائيين بحسب درجة الاعتماد الخطية بينهما. معامل الارتباط بين اثنين من المتغيرات العشوائية X و Y يقوم بحساب درجة الاعتماد الخطية بينهما. وهو يساوي قسمة التغاير بين المرتبطين σ_{xy} على انحراف معياري لكل متغير $\sigma_x \sigma_y$.

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (6)$$

معامل الارتباط هو قياس متماثل يأخذ قيمته بين -1 و 1 حيث تكون الصلة الخطية بين المتغيرين كاملة. إذا كان المتغيرين مستقلين فإن معامل ارتباطهما يكون منعدما لكن العكس ليس صحيح. الفرق بين معامل الارتباط والمعلومة المتبادلة هو أن هذه الأخيرة تقيس درجة الارتباط الخطي و غير الخطي بين المتغيرين في حين أن معامل الارتباط لا يقيس إلا درجة الارتباط الخطي.

3 الطريقة المقترحة

اخترنا في هذا المقال المطابقة بين النقط المميزة المستخرجة من الصور.

3.1 كشف النقط المميزة

لاختيار طريقة استخراج النقط المميزة، وجب مراعاة معيارين اثنين: جودة الكشف و الوقت اللازم للكشف. حسب نوعية التطبيق، يتم اختيار معيار على حساب الآخر. هناك فصيلتين مهمتين لكشف النقط المميزة:

- فصيطة تعتمد على المعاملات الرياضية : هاريس[4]، شي و تومازي[14]، ليندبارك[7] و هاريس-لابلاس[9]...
- فصيطة تعتمد على تغير المظهر: مورافيك[10]، سوزان[16] و فاست[12]...

لقد اخترنا استخدام كاشف هاريس لأنه مستقر، وثابت بالنسبة الدوران و له نسبة تكرر جيدة. ويمكننا أيضا التحكم في عدد نقاط الكشف عن كل صورة عن طريق تغيير معالم كشف هاريس لغرض التخفيف من العمليات الحسابية.

3.2 الخواريزمية المتبعة

ابتداء، نقيم معيار القرب باعتبار أن الصورتين تم التقاطهما في وقت متقارب. نعتبر المسافة بين النقطة P على الصورة اليسرى و النقطة Q على الصورة اليمنى على النحو التالي:

$$d(P, Q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2}$$

حيث إن (x_p, y_p) و (x_q, y_q) هي إحداثيات P و Q على التوالي.

يصبح مقياس التشابه على الشكل التالي:

$$MID(P, Q) = MI(P, Q) / \sqrt{d(P, Q)}$$

و

$$CCD(P, Q) = \rho(P, Q) / \sqrt{d(P, Q)}$$

ثم لنقطة مميزة P في الصورة اليسرى، نبحث عن النقطة المطابقة في الصورة اليمنى. لهذا الغرض، نقوم بحساب قياس التشابه بين جوار النقطة P و جوارات النقط المميزة في الصورة اليمنى (الشكل 2). النقطة Q الخاصة بأكبر قياس للتشابه هي مرشحة لتكون النقطة المطابقة للنقطة P إذا كانت قدر قياس التشابه يتعدى عتبة نعيثها بالتجربة. نقوم بإعادة الشيء نفسه بالنسبة للنقطة Q، بالبحث عن مطابق لها في الصورة اليسرى (الشكل 3). إذا وجدنا أن النقطة P هي المرشحة للمطابقة نقرر أن النقطتين P و Q متطابقتين و إلا نعد إلى مبدأ الفرصة الثانية باستعمال درجة الثقة و ذلك بقياس الفرق النسبي بين القيمة القصوى لقياس التشابه و القيمة الأصغر منها مباشرة. نلمس فائدة مبدأ الفرصة الثانية في الحالة التي يتم فيها تحقيق أقصى قدر التشابه بالنسبة لنقطتين معا أو إذا كانت أقصى قيمتين متقاربتين جدا. نعتبر النقطة R، النقطة الثانية المختارة في الصورة اليمنى باعتبار مبدأ الفرصة الثانية. إذا كانت النقطة المرشحة في الصورة اليسرى هي النقطة P نقرر أن النقطتين P و R متطابقتين و إلا نقرر أن النقطة P لا تملك أي مطابق في الصورة اليمنى.

4 النتائج التجريبية و المناقشة

لقد قمنا بالاختبارات على ثمانية أزواج من الصور المجسمة (الشكل رقم 5). لقد اخترنا صوراً مختلفة، والتي تحتوي على هياكل مختلفة ليكون التقييم أفضل.

عدد النقاط المميزة المستخرجة في الصور اليسرى هو 113. حجم الجوار المستخدم هو 9×9 . استخدمنا هذا الحجم للجوار لنحصل على عينة غنية (81 نقطة) لحساب الاحتمالات من أجل أن يكون قياس التشابه صحيحاً. العتبة المستخدمة تختلف حسب نوع من الصور.

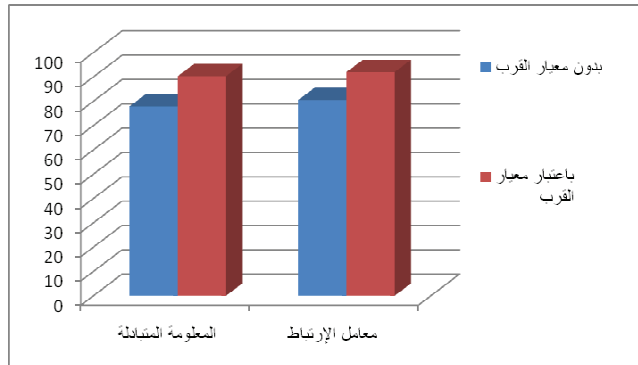
يكون الحكم على مطابقة نقطة في الصورة اليسرى صحيحاً إذا عثرنا على النقطة المطابقة الحقيقية أو إذا حكمنا على النقطة أنها لا تملك مطابقاً وهي فعلاً لا تملكه.

في حالة الشكل 4، الخوارزمية دون إدراج معيار القرب تختار نقطة بعيدة جداً (R) من النقطة المطابقة الحقيقية (Q)، ولكن عن طريق إدخال المعيار تحدد الخوارزمية المقابل الصحيح.

المعلومة المتبادلة	معامل الارتباط	
بدون معيار القرب	77,9%	80,5%
باعتبار معيار القرب	90,3%	92,0%

الجدول 1. نتائج المطابقة الصحيحة بين الصور اليسرى واليمنى

بمراعاة شرط أن النقاط الصور تم اتخاذها في أوقات متقاربة جداً، يُحسّن إدخال معيار القرب نتائج المطابقة بنسبة 15.9 % في حالة المعلومة المتبادلة و 14.3 % في حالة معامل الارتباط. ويرجع ذلك إلى أن أقرب نقطة لديها نسبة أكبر ليتم اختبارها للمطابقة (الجدول 1 و الرسم البياني 1).



الرسم البياني 1. نتائج المطابقة الصحيحة بين الصور اليسرى واليمنى

نضيف إلى الصور إشارات مشوشة و نقوم بنفس العمل أعلام. نتيجة التحسن في حالة المعلومة المتبادلة هي 32,1 % و في حالة معامل الارتباط هي 20,0 % (الجدول 2)

معامل الارتباط	المعلومة المتبادلة	
14,3%	15,9%	صور غير مشوشة
20,0%	32,1%	صور مشوشة

الجدول 2. نسبة تحسين المطابقة في حالة الصور العادية و المشوشة

5 خاتمة

في هذا المقال، قدمنا طريقة جديدة للمطابقة بين الصور المجسمة تعتمد على معيار القرب. قياسا التشابه المستعملين هما المعلومة المتبادلة و معامل الارتباط. باعتبار قيد أن النقاط الصورتَم اتخاذاها في أوقات متقاربة، فإن معيار القرب حسن بشكل ملموس نتائج المطابقة بين الصور.

النتائج المحصل عليها مشجعة مما يدفعنا لتطبيق طريقتنا في مجال الروبوت المتحرك بعد مقارنتها مع طرق موجودة.

المراجع

- [1] J. Atif. "Recalage non-rigide multimodal des images radiologiques par information mutuelle quadratique normalisée". Université de Paris XI – Orsay. 29 Octobre 2004
- [2] N. Ayache. Construction et fusion de représentations visuelles 3D. Applications à la robotique mobile. PhD thesis, Université Paris-Sud, 1988.
- [3] Guoyan Zheng and Xuan Zhang. A Unifying MAP-MRF Framework for Deriving New Point Similarity Measures for Intensity-based 2D-3D Registration. IEEE conference. ICPR 2006.
- [4] C. Harris & M. Stephens (1988). A combined corner and edge detector Proceedings of the 4th Alvey Vision Conference: pages 147--151.
- [5] A. Gagalowicz and L. Vinet. Regions matching for stereo pairs. In Sixth scandinavian conference on image analysis, pages 63–70, Juin 1989.
- [6] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptative window : Theory and experiment. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(9) :920–932, 1994.
- [7] T. Lindeberg, "Feature detection with automatic scale selection". International Journal of Computer Vision 30 (2). 1998
- [8] Q.T. Luong and O. Faugeras. Self-calibration of a stereo rig from unknown camera motions and points correspondences. Technical Report 2014, INRIA, Juillet 1993.

- [9] K. Mikolajczyk, C. Schmid(2004). "Scale and affine invariant interest point detectors" IJCV, Volume 60, Number 1 - 2004
- [10] Moravec H. "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover", Ph.D. thesis, Stanford University, Stanford, California, May 1980. Available as Stanford AIM-340, CS-80-813 and CMU-RI-TR-3
- [11] C. Rabaud, "Une nouvelle approche de mise en correspondance stéréoscopique dense par méthodes possibilistes". Université Montpellier II. 15 Juillet 2005.
- [12] Edward Rosten (2006), Tom Drummond."Machine Learning for High-Speed Corner Detection". ECCV 2006. p 430-443
- [13] D.B. Russakoff, C.Tomasi, T.Rohlfing, Calvin and R.Maurer, Jr. Image Similarity Using Mutual Information of Regions. ECCV 2004.
- [14] J. Shi and C. Tomasi (June 1994). "Good Features to Track,". 9th IEEE Conference on Computer Vision and Pattern Recognition, Springer.
- [15] T. Skordas and R. Horaud. Mise en correspondance structurelle pour la vision stéréoscopique.TSI, 7(6):591–608, 1988.
- [16] S. M. Smith and J. M. Brady (May 1997). "SUSAN - a new approach to low level image processing.". International Journal of Computer Vision 23: 45-78.
- [17] L. Sommellier, "Mise en correspondance d'images stéréoscopiques utilisant un modèle topologique". Université CLAUDE BERNARD - LYON 1. 23 septembre 1997
- [18] R. VAILLANT, I. SURIN, "Reconstruction de visages par stéréovision active". Traitement du Signal [Trait. Signal], Vol. 12, N° 2-NS, p. 201-211. 1995
- [19] L. Vinet. Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques. PhD thesis, Université Paris IX Dauphine, juillet 1991.
- [20] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report 2273, INRIA, Mai 1994.
- [21]Yong Seok Heo, Kyoung Mu Lee, Sang Uk Lee: Mutual information-based stereo matching combined with SIFT descriptor in log-chromaticity color space. CVPR 2009: 445-452.
- [22] M.A. Kerroum, and A. Hammouch, and D. Aboutajdine. Textural feature selection by joint mutual information based on Gaussian mixture model for multispectral image classification. Pattern Recognition Letters. Volume 31. N°10. pp : 1168-1174. July 2010.
- [23] Skerl D, Tomazevic D, Likar B, Pernus F. Evaluation of similarity measures for reconstruction-based registration in image-guided radiotherapy and surgery. Int J Radiat Oncol Biol Phys. Volume 65, Issue 3 , Pages 943-953, 1 July 2006.

Title: Improvement of image matching by using the proximity criterion

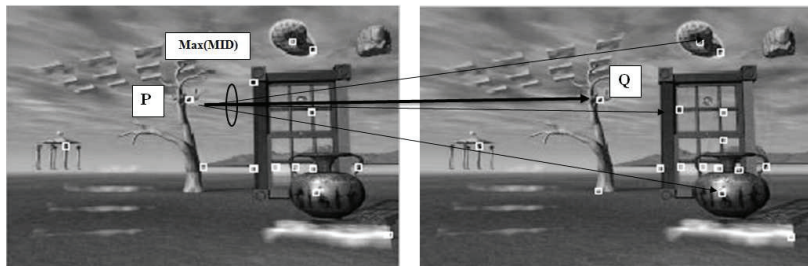
Abstract. The aim of this work is to develop a new algorithm for matching stereo images to the 3D reconstruction. We propose the use of a proximity criterion, to improve performance, and applying of the second chance algorithm. The similarity measures used are mutual information and correlation coefficient. The matching is done between neighborhoods of points of interest extracted from the images. We work in case which the sensor has a slight displacement between two images. The tests are performed on grayscale images.

Keywords: computer vision; 3D reconstruction; stereoscopic images; robotics;

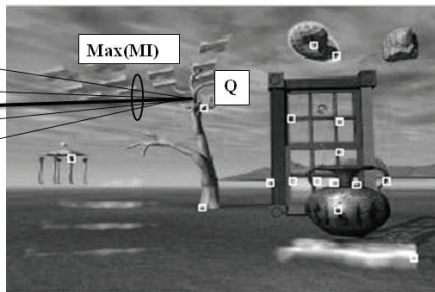
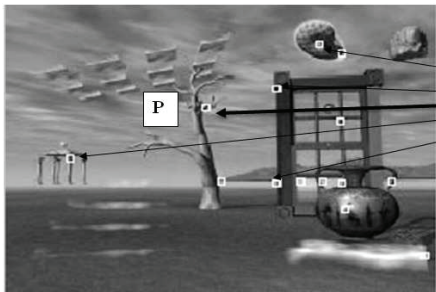
المصطلحات

توضع هنا ترجمة أكبر عدد ممكن من المصطلحات إلى اللغة العربية.

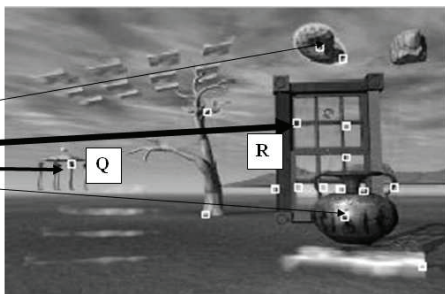
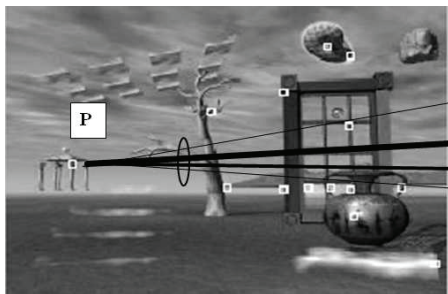
المعلومة المتبادلة	Mutual Information
النقط المميزة	Points of interest
كشف الخطوط المميزة	Edge detection
صور رمادية التلوين	grayscale images



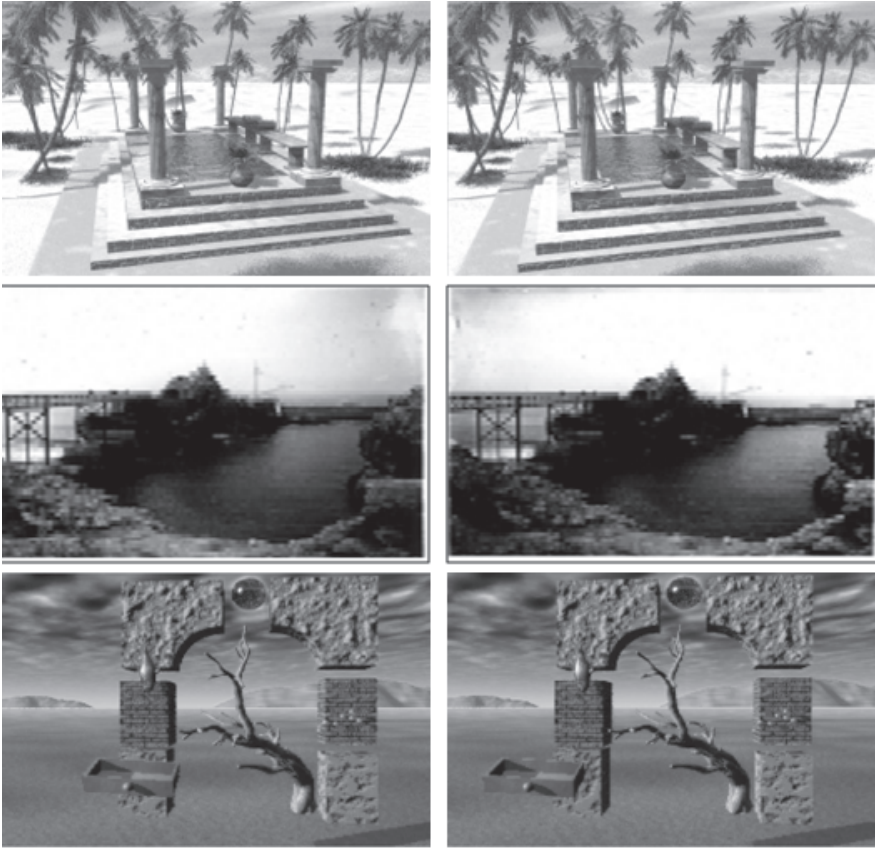
الشكل 2. البحث عن مطابق النقطة P في الصورة اليمنى باستعمال المعلومة المتبادلة (MID)



الشكل 3. تأكيد المطابقة بين النقطتين P و Q.



الشكل 4. الحاجة إلى إدراج معيار القرب



الشكل 5. مثال للصور المستعملة

- مقر العمل الحالي: المركز الوطني للبحث العلمي و التقني. الرباط - المغرب
- الإقامة: قطاع 11 بلوك أ رقم 6 حي الرياض- الرباط - المغرب

3 صورة صاحب الجواز Photographie du titulaire  امضاء صاحب الجواز signature du titulaire 	2 الاسم الشخمي: ابراهيم Prénom(s) Ibrahim الاسم العائلي: كزلزم Nom GUELZIM الجنسية مغربية Nationalité Marocaine تاريخ الميلاد: 11.01.1979 Né le 11.01.1979 مكان الميلاد: الرباط Lieu de naissance Rabat المهنة: طالب Profession Etudiant عنوان السكن: الرباط، قطاع 11 بلوك أ رقم 6 Domicile Rabat, Sect. 11, N°6 Bloc A, Hay Ryad. رقم بطاقة: A340732 Numéro C.I.N. A340732
--	---

خوارزمي التصنيف باستعمال المعلومات السياقية تطبيق في مجال الصور الملونة النسيجية

موسى أحمد¹، الغشام العمراني عبداللطيف¹، الاطا أوليفي²
¹ جامعة عبدالمالك السعدي، المدرسة الوطنية للعلوم التطبيقية، مختبر التكنولوجيا المبتكرة
 ص.ب. 1818، الرمز البريدي 90000، طنجة، المملكة المغربية
² جامعة بواتي، مختبر الإشارة و الصورة، الرمز البريدي 86000، بواتي، فرنسا

البريد الالكتروني: amoussa@uae.ac.ma aghacham@uae.ac.ma

الخلاصة. نريد عبر هذا المقال اقتراح خوارزمي جديد للتصنيف باستعمال المعلومات السياقية للمعطيات اعتمادا على القوة التفاعلية فيما بينها. هذا المنهج يستخدم الاحتمال الشرطي لجيبس للتصنيف التدريجي للمعطيات المطروحة باستعمال عينات تمثيلية للمجموعات. هذه التقنية الغير المعلمة تعتمد على مفاهيم النظرية المركوفية المرتكزة على نظام الجوار، وذلك من أجل تطوير نموذج فعال لتصنيف المعطيات المعقدة. من بين مجالات تطبيق هذا الخوارزمي، نقترح في هذه الورقة مجال تجزيء الصورة الملونة النسيجية. في هذا السياق، نستعمل جزيئات من نسيج الصورة كعينات ممثلة بالمعطيات المستخرجة من تخمينات النموذج التقديري الخطي ثنائي الأبعاد متعدد القنوات و المعقد، وهذا في مجال ألوان الصورة التالية: أحمر-أخضر-أزرق و المجال صبغة-إضاءة-إشباع مطور و المجال إضاءة-تلون. الأمر يقف على تعليم كل خانة من الصورة حسب ألوانها و نسيجها. النتائج التجريبية المعتمدة على نسبة أخطاء تصنيف الصور تبين فعالية أكثر للنموذج المقترح مقارنة بنماذج أخرى في هذا الباب.

الكلمات الجوهرية: خوارزمي التصنيف- الاحتمال الشرطي لجيبس- الصورة الملونة النسيجية- النموذج التقديري الخطي

1 مقدمة

تعتبر دراسة التجميع بمثابة تقنية مهمة في التحليل التقديري للمعطيات المتوفرة خاصة عند جهل كيفية توزيعها. يمكن طرح مشكل التجميع على النحو التالي : نعتبر Q معطيات على شكل نقط في المجال المتعدد الأبعاد R^N ، مسألة التجميع تعود إلى تقسيم هذه النقط على عدد K من المجموعات بحيث أن نقط نفس المجموعة متقاربة أكثر من نقط تنتمي لمجموعات مختلفة.

عدة أبحاث استعملت التصنيف في مجال تجزيء الصور وذلك عن طريق تصنيف الخانة نظرا للتجانس بين تصنيف الصورة و تجميع المعلومات [1]. في هذا السياق اقترحت مجموعة من المقاربات تتعلق بتوزيع المعطيات منها : استخراج النواة [2]، تحديد المحيط [3]، الشكل الرياضي [4] الشبكة العصبية [5] المنهج المركوفي [6]، لاستخراج نواة المجموعات المكونة للمعطيات.

لكن تصنيف جميع المعطيات المطروحة يستلزم مرحلة ثانية نستعمل خلالها النوات المستخرجة سابقاً كعينات، وذلك باعتبار النقط الممثلة للمعطيات و الموجودة داخل النواة كنقط مصنفة أما النقط الخارجة عن النوات يمكن تصنيفها باعتماد نظرية الجوار الأقرب [7]. هذه النظرية الأخيرة تبقى محدودة الفعالية في حالات التجميع المعقدة كالمعطيات المتداخلة أو المعطيات ذات التوزيع المعقد. في هذا الإطار نقترح من خلال هذا المقال خوارزمي جديد لتصنيف النقط الخارجة عن النوات باستعمال مفاهيم النظرية المركوفية. الجزء الثاني من هذا المقال يبين المراحل المختلفة لهذه الطريقة و يجسد أهميتها في تحليل المعطيات عن طريق أمثلة محاكاة. الجزء الثالث يستعرض تطبيق هذا الخوارزمي على المشكل المعقد المتمثل في تصنيف الصور الملونة النسيجية، بحيث أن المعطيات المتوفرة تستخرج من تخمينات النموذج التقديري الخطي المرتبط بأبعاد الألوان أحمر- أخضر- أزرق و صبغة-إضاءة-إشباع مطور و إضاءة-تلون [8,9].

2 وصف منهجية التصنيف المقترحة

نعتبر Γ مجموعة من الملاحظات D_q عددها Q بحيث أن: $D_q = (d_{q,1}, d_{q,2}, \dots, d_{q,n}, \dots, d_{q,N})^T$ ، $d_{q,n}$ تمثل إحداثيات الملاحظة D_q في المجال R^N . بعد استخراج وتحديد نواة الأقسام المكونة للمجموعة Γ ، يمكن تقسيم هذه الأخيرة إلى مجموعتين: المجموعة Γ_p المكونة من معطيات النواة المستخرجة (أو الملاحظات المصنفة) و المجموعة Γ_{np} المكونة من المعطيات الباقية (أو الملاحظات الغير المصنفة). الرسم 1-ب يبين المجموعة Γ_p المستخرجة باستعمال الخوارزمي [6] المطبق على المجموعة Γ (الرسم 1-أ)، بينما المجموعة Γ_{np} مبنية في الرسم 1-ج. لنعتبر X مجال مركوفي عشوائي مؤشر على المجموعة Γ ومعرف بالطريقة التالية :

$$(1) \quad X = (X_q)_{q=1, \dots, Q}, \quad \begin{cases} X_q = \omega_k / k \neq 0, & D_q \in \Gamma_p \\ X_q = \omega_0 & D_q \in \Gamma_{np} \end{cases}$$

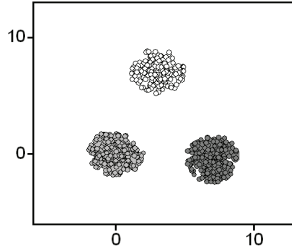
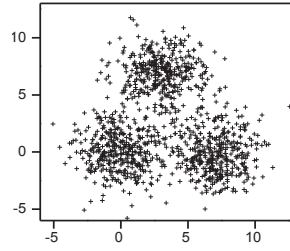
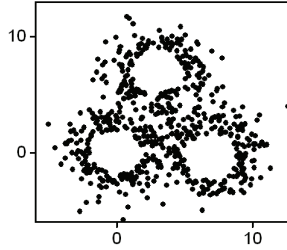
حيث أن $\Omega = \{\omega_0, \omega_1, \dots, \omega_k, \dots, \omega_K\}$ علامة متعلقة بالمجموعة k .

الهدف الأساسي لمنهجية التصنيف المقترحة هنا هو تصنيف كل ملاحظة D_q ، منتمية إلى Γ_{np} ، آخذين بعين الاعتبار المعلومات داخل جوار D_q بصفة حصرية. من أجل هذا الهدف و في إطار الفرضية المركوفية للمجال X ، ندخل مفهوم الشلة المرتبط بالجوار V_q لحساب الدالة الطاقية [10] (انظر المعادلة 2). الاحتمال الشرطي لجيبس المحصل من هذه الطاقة، يعكس عضوية الملاحظة الغير المصنفة في الأقسام المختلفة الموجودة في الجوار V_q . الرسم 2 يبين مثال لجوار كروي وكل شلة c من الدرجة الثانية منتمية إلى C_2 متعلقة به.

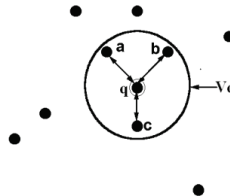
$$(2) \quad U(X_q = \omega_k / X_r = \omega_{k'}, \forall r \in V_q) = \sum_{c \in C_2} \varphi_c(X_q = \omega_k, X_r = \omega_{k'}, (q, r) \in c)$$

$$\varphi(X_q = \omega_k, X_r = \omega_{k'}) = \frac{\gamma_{q,r}}{1 + \text{dist}(q, r)} \quad \gamma_{q,r} = \begin{cases} \beta & \text{إذا } \omega_k \neq \omega_{k'} \\ -\beta & \text{إذا } \omega_k = \omega_{k'} \end{cases} \quad \text{حيث أن}$$

$$(3) \quad P(X_q = \omega_k / X_r, \forall r \in V_q) = \frac{\exp[-U(X_q = \omega_k / X_r, \forall r \in V_q)]}{\sum_{j=0}^K \exp[-U(X_q = \omega_j / X_r, \forall r \in V_q)]}$$

1- ب: المجموعة Γ_p 1- أ: المجموعة Γ 1- ج: المجموعة Γ_{np}

الرسم 1. مثال لوضع معطيات مكونة من معطيات مصنفة و غير مصنفة.



الرسم 2. الشلة من الدرجة الثانية ل V_q $C_2 = \{(q,a), (q,b), (q,c)\}$

نشير إلى أن دالة الطاقة الكامنة المستعملة في [10]، تأخذ بعين الاعتبار القرب من الملاحظات داخل الجوار وعلاماتها(المعادلة 2). الخوارزمي المقترح في هذه الورقة يحقق الحد الأقصى للإحتمال الشرطي لجيبس لكل ملاحظة غير مصنفة (المعادلة 3). المنهجية تتبدى بتصنيف الملاحظات المجاورة للنواة، ثم تكرر هذه العملية إلى حين تصنيف كل الملاحظات.

الخوارزمي:

تهيئة

$\leq$ معطى متوفر: $(\Gamma, \Gamma_p, X^0, \Omega, V)$, $\Gamma_{np} = \Gamma / \Gamma_p$ و $NP = \text{Card}(\Gamma_{np})$

بداية

$\leq$ ما دام ; $(NP > 0)$

عداد = 0،

لكل $q \in \Gamma_{np}$ و لكل قسم $\omega_k \in \Omega$ احسب :

$$P(X_q = \omega_k / X_r, \forall r \in V_q) \quad \text{و} \quad U(X_q = \omega_k / X_r, \forall r \in V_q) -$$

$$\hat{\omega}_j = \text{Arg} [\text{Max}_{\omega_k \in \Omega} P(X_q = \omega_k / X_r, \forall r \in V_q)] -$$

$$- \text{ إذا } \omega_0 \neq \hat{\omega}_j \text{ ف } \{ \omega_0 \neq \hat{\omega}_j, NP-1 = NP, \text{ احسب } + \}$$

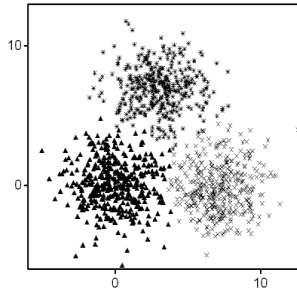
$\leq$ نهاية

- إذا عداد = 0 و $0 < NP$ وسع حجم V_q ،

$\leq$ نهاية ما دام

نهاية الخوارزمي

تطبيق هذا الخوارزمي على مجموعة ملاحظات الرسم 1-أ يسمح بتصنيف كل معطيات المجموعة Γ_{np} . نسبة الخطأ، المقدرة بقسمة عدد المعطيات المصنفة خطأ على عدد جميع المعطيات، تقدر ب: 2.25% في حين أن خوارزمي الجوار الأقرب [7]، يصنف هذه المعطيات بنسبة خطأ تقدر ب: 2.50%. هذا المثال النموذجي يبين مدى أهلية الخوارزمي المقترح لتصنيف جيد للملاحظات انطلاقاً من عينات.



الرسم 3. نتيجة تصنيف المجموعة Γ

3- تطبيق المنهجية لتجزئ الصورة الملونة النسيجية.

لتطبيق طريقة التصنيف المقترحة لتجزئ الصور الملونة النسيجية، يتعين علينا احتساب المعطيات المستخرجة من مجال ألوان الصورة ومن نموذج نسيجها. في الفرعيتين التاليتين نذكر بمجالات الألوان وكذا المنهج التقديري الخطي المستعمل في هذا الإطار.

التحليل المقدم في هذه الورقة يستعمل ثلاث مجالات مختلفة من الألوان: المجال أحمر-أخضر-أزرق و المجال صبغة-إضاءة-إشباع مطور و المجال إضاءة-تلون. المجال أحمر-أخضر-أزرق هو الأكثر استعمالا في علوم الحاسوب و معالجة الصورة (الرسم 4-أ)، نظرا لسهولة إنتاجه للألوان يمزج بعضها ببعض [11] ، إلا أن هذا المجال لا يتناسب مع الإدراك البصري للإنسان. الورقة [11] تعرض أيضا كيفية الانتقال من مجال الألوان أحمر-أخضر-أزرق إلى نظيره صبغة-إضاءة-إشباع مطور (الرسم 4-ب). مؤخرا، استعمل المجال صبغة-إضاءة-إشباع مطور في عدة مجالات نظرا للميزات التالية:

- إعطائه قيمة صغيرة للإحداثيات اللونية،
- مستقل عن قيمة الإضاءة.

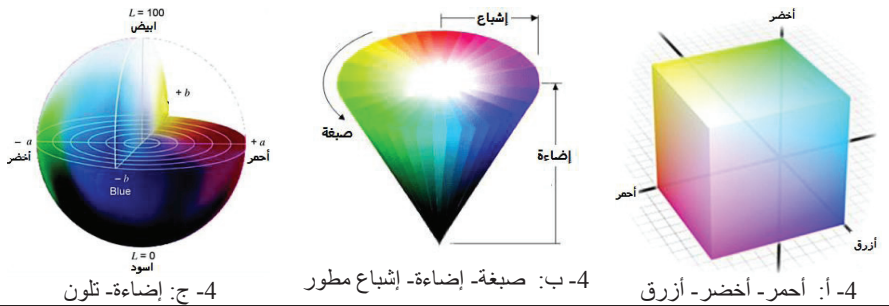
ندرس كذلك بشكل تجزيء الصورة الملونة النسيجية في المجال إضاءة-تلون (الرسم 4-ج) الذي يعتبر مجالا موحدا مرتكزا على الإدراك البصري للإنسان [12].

نستعمل المعلومة اللونية المستخرجة من مجال أحمر-أخضر-أزرق لبناء قناتين للصورة تحتوي الأولى على معلومة الإضاءة والثانية على المعلومة اللونية. عند الانتقال إلى مجال صبغة-إضاءة-إشباع مطور، القناة اللونية تحسب بالطريقة التالية: $C = S \times \exp(jb \cdot H)$ مع H قيمة الصبغة و S قيمة الإشباع محتسبة بالانتقال من مجال أحمر-أخضر-أزرق إلى مجال صبغة-إضاءة-إشباع مطور. بينما في الانتقال إلى مجال إضاءة-تلون تحسب القناة اللونية بالطريقة الآتية $C = a^* + jb^*$ مع a^* و b^* متغيرتين لونيتين محتسبتين بالانتقال من مجال أحمر-أخضر-أزرق إلى مجال إضاءة-تلون [13].

على إثر هذا، تصبح الصورة المطروحة للتحليل تحتوي على قناتين: الأولى تمثل المعلومة الضوئية على شكل عدد حقيقي و الثانية تمثل المعلومة اللونية على شكل عدد معقد، وذلك تمثل أي

خانة في الصورة ب: $x_n = \begin{bmatrix} I_n \\ c_n \end{bmatrix}$

بحيث أن $I_n \in \mathbb{R}$ و $n = (n_1, n_2) \in \Lambda \subset \mathbb{Z}^2$ مع $c_n \in \mathbb{C}$ منطقة ثنائية الأبعاد بحجم $|\Lambda| = M$



الرسم 4. مجالات الألوان المستعملة

3-2 النموذج التقديري الخطي

نعتبر $Y = (Y_n)_{n \in \mathbb{Z}^2}$ حدث تلقائي ثنائي الأبعاد متعدد القنوات عددها P . النموذج التنبؤي الخطي $\hat{Y}$ يحدد بـ:

$$\hat{Y}_n = - \sum_{m \in D} A_m (Y_{n-m} - \mu) \quad \text{والخطأ التقديري للسلسلة هو: } E_n = (Y_n - \mu) - \hat{Y}_n$$

بحيث أن $m = (m_1, m_2) \in D \subset \mathbb{Z}^{*2}$. عامل المصفوفات $P \times P$. يفترض في سلسلة الخطأ التقديري $E = \{E_n\}_{n \in \mathbb{Z}^2}$ أن يكون حدثاً ثابتاً متعدد القنوات. تفاصيل النماذج التقديرية الخطية متعددة القنوات المستعملة في هذه الورقة موضحة في الورقة [8] وهي: نصف مسطح الغير التماثلي ذاتي التراجع - ربع مسطح ذاتي التراجع و المجال العشوائي الكوس-المركوفي. سنمتحن النماذج الثلاثة السالف ذكرها في كل مجالات الألوان المذكورة في الفرعية 3-1.

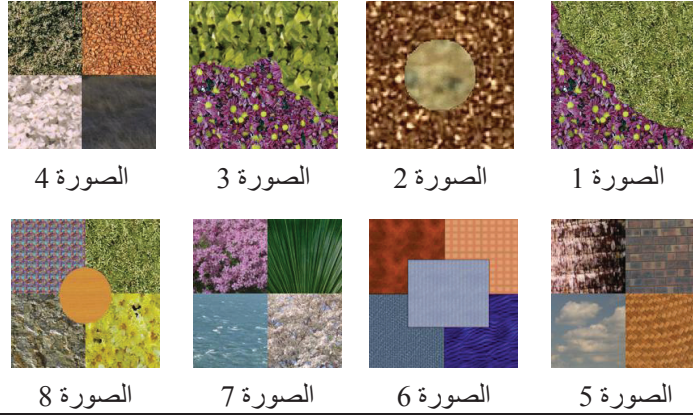
3-3 تجزيء الصورة الملونة النسيجية

يصعب تقدير و استخراج المعطيات الحقيقية المتعلقة بالصور الطبيعية المعقدة لارتباطها بذاتية المحلل، لذلك ومن أجل تقييم خوارزمي تصنيف الصورة الملونة النسيجية المقترح، سنستعمل صوراً ملونة نسيجية اصطناعية وطبيعية ذات معطيات جد معقدة. كل الصور المعالجة لها حجم 184×184 خانة مأخوذة من قاعدة البيانات المستعملة في الورقة [14].

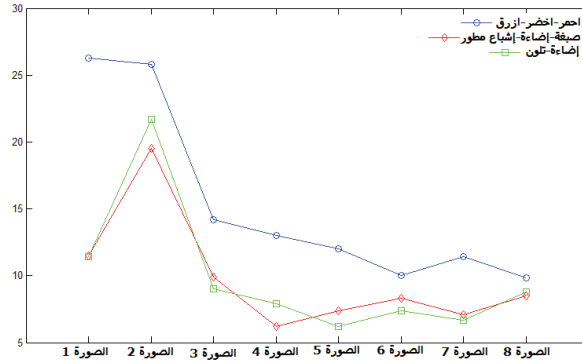
في المرحلة الأولى، لخوارزمي تجزيء الصورة الملونة النسيجية، نستعمل صويرات من حجم 32×32 خانة كعينة لكل قسم من أقسام الصورة. نستعمل الخطأ التقديري متعدد القنوات لكل هذه الصويرات لاستخراج كل ملاحظات المجموعة Γ_p (فرعية 2).

في المرحلة الثانية، نستخرج بنفس الطريقة المعطيات المتعلقة بالصورة المطروحة للتحليل (إحدى صور الرسم 5) و التي تمثل المجموعة Γ_{np} . خوارزمي التصنيف المقترح يطبق على هذه المعطيات باستعمال نظام الجوار المبني على مسافة جيب التمام بين الملاحظات. المبتغى من هذه المحاكاة هو من جهة تأكيد النتائج المتعلقة باختيار المجال اللوني الأنسب [9]، ومن جهة أخرى تبيان قدرة خوارزمي التصنيف المقترح في هذه الورقة على تصنيف المعطيات المعقدة المستخرجة من الصور الملونة النسيجية.

النتائج المحصل عليها مبينة في الرسم 6، هذا الأخير يبين نسبة معدل الخطأ باستعمال مجالات الألوان الثلاث و المتعلقة بنماذج التقدير المعرفة سابقاً (نصف مسطح الغير التماثلي ذاتي التراجع - ربع مسطح ذاتي التراجع و المجال العشوائي الكوس-المركوفي). الرسم 7 يبين نفس النتائج المجمعة بطريقة التصنيف التالية: الجوار الأقرب [7]، طريقة باييس [15] و الخوارزمي المقترح في هذا الملف.



الرسم 5. مجموعة الصور الملونة النسيجية المستعملة

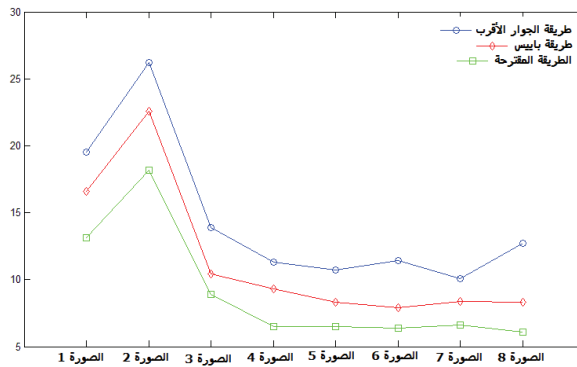


الرسم 6. نسبة خطأ التصنيف عند التجميع بمجالات الألوان.

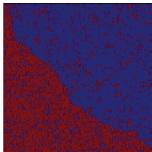
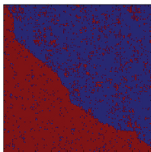
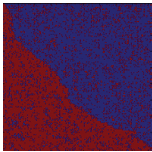
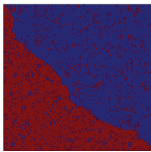
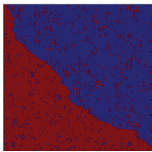
4- مناقشة و خاتمة

النتائج الأولية تؤكد أن تركيبة المجال اللوني تساهم بصفة واضحة في أداء خوارزمي تحليل الصورة الملونة النسيجية [9,8]. إذا حللنا نتائج نسبة الخطأ اعتمادا على تركيبة المجال اللوني، نلاحظ أن المجالين، إضاءة-تلون و صبغة-إضاءة-إشباع مطور، يعطيان نتائج متقاربة و يستطيعان تمييزا جيدا لتركيبية المعلومة بصفة شاملة. لتوضيح أكثر، نبين في الرسم 7 نتائج تصنيف الصورة الأولى باستعمال ثلاثة طرق للتصنيف في المجالين، إضاءة-تلون و صبغة-إضاءة-إشباع مطور. النتائج الممثلة في الرسم 7 تبين الأداء المتميز للخوارزمي الجديد المقترح للتصنيف بحيث أن نسبة خطأ التصنيف لهذه التقنية أصغر من خطأ التصنيف المتعلق بطريقة الجوار الأقرب و طريقة باييس. نبين في الرسم 9 تجزئ الصورة 8 باستعمال النموذج الخطي التقديري في المجال اللوني

صبغة-إضاءة-إشباع مطور. نسبة خطأ الطريقة المقترحة قدرت ب 1,15 % بينما نسبة خطأ طريقة الجوار الأقرب و طريقة باييس هي على التوالي 7,8% و 7,5%.

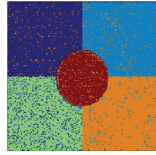


الرسم 7. نسبة خطأ التصنيف عند التجميع بالطريقة المقترحة.

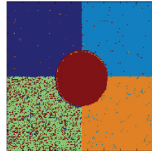
مجال اللون	طريقة الجوار الأقرب	طريقة باييس	الطريقة المقترحة
صبغة-إضاءة-إشباع مطور			
	نسبة الخطأ = 13.8%	نسبة الخطأ = 9.3%	نسبة الخطأ = 8.1%
إضاءة-تلون			
	نسبة الخطأ = 14.5%	نسبة الخطأ = 8.9%	نسبة الخطأ = 7.4%

الرسم 8. تجزئ الصورة 1 (الرسم 5) باستعمال ثلاث طرق للتصنيف في مجال الألوان صبغة-إضاءة-إشباع مطور و إضاءة-تلون

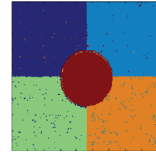
هذه النتائج تؤكد قدرة خوارزمي التصنيف المقترح في هذه الورقة على تصنيف معطيات معقدة كالتي تستخرج من الصور الملونة النسيجية، و كمنظور لهذا العمل، نقترح إدماج معلومات متعلقة بتفاعل الخانات فيما بينها كمعلومة إضافية لطريقة التصنيف المقترحة.



طريقة الجوار الأقرب



طريقة باييس



الطريقة المقترحة

الرسم 9. تجزيء الصورة 8 (الرسم 5) باستعمال ثلاثة طرق للتصنيف في مجال الألوان صبغة-إضاءة-إشباع مطور.

المراجع

- [1] Webb, A. : Statistical Pattern Recognition. Second Edition, John Wiley & Sons, Ltd, 2002.
- [2] Touzani, A., Postaire, J.-G. : Mode Detection by Relaxation. IEEE Trans. Pattern Anal. and Machine Intell. 10 (6) (1988) 970-978.
- [3] Touzani, A., Postaire, J.-G. : Clustering by Mode Boundary Detection. Pattern Recognition Letters, 9 (1) (1989) 1-12.
- [4] Postaire, J.-G., Zhang, R.-D., Botte-Lecocq, C. : Cluster Analysis by Binary Morphology. IEEE. Trans. Pattern Anal. and Machine Intell., 15 (2) (1993) 170-180.
- [5] Zhang, G.-P. : Neural Networks for Classification : a survey. IEEE Transaction On Syst. Man. And Cyber.- Part C: Application and Review 30 (4), (2000) 451-462.
- [6] Moussa, A., Sbihi, A., Postaire, J.-G. : A Markov Random Field Model for Mode Detection in Cluster Analysis, Pattern Recognition Letters, 29 (9), (2008) 1197-1207.
- [7] Cover, T. -M., Hart, P. E.: Nearest Neighbor Pattern Classification. IEEE Trans. on Information Theory, 13 (1), (1967), 21-27.
- [8] Qazi, I.-U.-H, Alata, O., Burie, J.-C., Fernandez-Maloigne, C.: Color Spectral Analysis for Spatial Structure Characterization of Textures in IHLS Color Space. Pattern Recognition, 43(3), (2010), 663 – 675.
- [9] Qazi, I.-U.-H, Alata, O., Burie, J.-C., Moussa, A., Fernandez-Maloigne, C.: Choice of a Pertinent Color Space for Color Texture Characterization Using Parametric Spectral Analysis. Pattern Recognition, 44, (2011), 16 – 31.
- [10] Derin, H., Elliot, H.: Modelling and Segmentation of Noisy and Textured Images using Gibbs Random Field. IEEE Trans. Pattern Anal. and Machine Intell., 9(1), (1982), 39-55.
- [11] Hanbury, A.: A 3D-Polar Coordinate Colour Representation well Adapted to Image Analysis. In Proceedings of the Scandinavian Conference on Image Analysis, (2003), 804-811.
- [12] Commission Internationale de l'Eclairage, Colorimetry, CIE 15.2, (1986).
- [13] Paschos, G.: Perceptually Uniform Color Spaces for Color Texture Analysis: An empirical evaluation, IEEE Transactions on Image Processing, 10 (6), (2001), 932-937.
- [14] Ilea, D., Whelan E. P. F.: An Adaptive Unsupervised Segmentation Algorithm Based on Color-Texture Coherence. IEEE Transactions on Image Processing, 17(10), (2008), 1926-1939.
- [15] Postaire, J. -G : De l'Image à la Décision – Analyse des Images Numériques et Théorie de la Décision, Dunod Informatique, Paris, France, 1987, p. 43-53

المخلص باللغة الانجليزية

Title : Classification Algorithm Using Contextual Information Application to Textured Color Images.

Abstract. The present paper proposes a new classification method which takes into account the contextual information of the observations. Based on the energy interaction model, this method maximizes the Gibbs conditional probability to classify progressively the observations available from samples to clusters represented by their prototypes. This non parametric method is based on the potential function, clique concept, and adapted neighborhood system that allow a suitable model for classifying complex data. Application of this new method concerns segmentation of color textured images. In fact, we use a small training sub image as prototypes, the available sample is represented by data belonging from estimated 2D multichannel complex linear prediction model, used in RGB, IHLS and $L^*a^*b^*$ color spaces. The problem stands in assigning, to each pixel of the available image, the label corresponding to its texture. Experimental results show a better performance by the proposed method in terms of percentage segmentation error for the used color textures, as compared to classical approach, in all studied color spaces.

Keywords: classification – Textured Color Image – Gibbs Conditional Probability-Linear Prediction Error

المصطلحات

Gibbs conditional probability	الاحتمال الشرطي لجيبس
Markov's theory	النظرية الماركوفية
Potential Function	دالة الطاقة الكامنة
Clique	شلة
Interaction energy	طاقة التفاعل
Core Extraction/Mode detection	استخراج النواة
Edge Detection	تحديد المحيط
Nearest Neighbour	الجوار الأقرب
Neural Network	الشبكة العصبية
Neighbourhood System	نظام الجوار
Neighbour Support Region	مرتکز جوار المنطقة
Clustering	تجميع
Contextual Information	المعلومة السياقية

Classification	التصنيف
2D multichannel complex linear prediction model	النموذج التقدير الخطي ثنائي الأبعاد متعدد القنوات والمعد
RGB: Red Green Blue	أحمر-أخضر-أزرق
IHLS: Improved Hue Luminance Saturation	صبغة-إضاءة-إشباع مطور
$L^*a^*b^*$	إضاءة-تلون
NSHP: Non Symetric Half Plane	نصف مسطح الغير التماثلي
AR : AutoRegressive	ذاتي التراجع
QP: Quart Plane	ربع مسطح
GMRF: Gauss Markov Random Field	المجال العشوائي الكوس-المركوبي

نظرية فحص برامج بالمقيدات

١٠٣٧

لزعر نجيب¹، لباح يحي²، عربي نورالدين²، بلعيد محمد السعيد³، حمزة دريد¹
¹ جامعة رين، معهد بحوث المعلوماتية
الرمز البريدي 35000، رين، فرنسا
² جامعة السانتيه، كلية العلوم، فرع الإعلام الالي
ص.ب. 1524، الرمز البريدي 31000، وهران، الجزائر
² جامعة نيس، معهد بحوث المعلوماتية
الرمز البريدي 06000، نيس، فرنسا

البريد الالكتروني :

¹nadjib.lazaar@inria.fr, ²{ylebbah, aribi_noureddine}@yahoo.fr, ³msbelaid@i3s.unice.fr

الخلاصة. إن أي عملية تطوير للبرمجيات في إطار صناعي بما في ذلك البرمجة بالمقيدات، تركز على مرجعية رسمية و تشمل مرحلتي الفحص و الاختبار. و استعمل هذا الأسلوب بكثرة لمعالجة الإشكاليات الموجودة في ميدان الذكاء الاصطناعي. ولقد اتسع نطاق استعمل هذا النوع من البرمجة في السنوات الأخيرة ليشمل ميادين حساسة كال دفاع، الملاحة الجوية و التجارة الإلكترونية. نستعرض من خلال هذه الورقة العلمية نظرية أولى في فحص و اختبار البرامج التي تعتمد على المقيدات حيث تعرف أول أداة فحص، باسم *CPTEST*، التي تتخذ من هذه النظرية إطاراً و قاعدة أساسية.

الكلمات الجوهرية: مقيد، برمجة بالمقيدات، فحص البرامج، كشف عن الأخطاء، علاقة توافق.

1 مقدمة

عرفت البرمجة بالمقيدات (ب.م) (بالإنجليزية : *Constraint Programming*) نجاحاً حقيقياً في حلّ الإشكاليات تحت إطار صناعي وذلك في شتى المجالات كترتيب المهام (*planning*)، تنظيم الجداول (*scheduling*)... [11]، كما أكدت ال (ب.م) على حضورها في مجالات حساسة كالتجارة الإلكترونية [4]، مراقبة و إدارة الملاحة الجوية [2,5].

و نجد العديد من اللغات و أدوات البرمجة (ب.م) المُسوَّقة، و لعل من أشهرها *COMET* [13]، *OPL* [12]، *SICStus Prolog*. إضافة إلى الأكاديمية منها مثل *Choco*، *Gecode*، *Essence*. وبالرغم من أن كشف و تصحيح الأخطاء في برامج (ب.م) تلقى إهتماماً كبيراً من طرف الباحثين، على سبيل المثال المشروع الأوروبي (*DiCSiPL*) [1]، فإننا لا نجد إلى يومنا هذا أي أداة فحص فعالة آليّة (*Automatic Testing Tool*) خاصة ببرامج (ب.م).

إذ لا يمكن فحص البرامج (ب.م) بشكل آلي اعتماداً على النظريات القائمة [3] (المستعملة لفحص برامج التقليدية ذات التعليمات المتسلسلة مثل (C++/Java)) وذلك لأسباب مختلفة من بينها :

- عملية البرمجة (ب.م) تختلف بشكل كامل عن البرمجة التقليدية و من ثمّ، فإن الأخطاء المرتكبة في البرامج الأولى تكون بطبيعة الحال مختلفة عنها في الثانية،
- برنامج (ب.م) يمثل مجموعة الحلول التي ترضي المقيدات بينما البرنامج التقليدي يمثل طريقة الوصول إلى حل معين...

إذن في إطار هذه الدراسة سنقدم نظرية و إطار أولي في فحص البرامج (ب.م) الذي يعتمد على علاقة التوافق (Conformity Relation) بين المرجع و البرنامج.

لفحص برنامج (ب.م) يعد هذا المرجع بمثابة النموذج الأول الذي يمثل ترجمة مباشرة ذات ولاء لمواصفات المشكلة الأساسية والذي سيكون في أغلب الأحيان نقطة إنطلاق للمبرمج من أجل الحصول على برنامج فعال و ذلك بتطبيق عدد من التحسينات كاستعمال مقيدات خاصة (Global Constraint)، تكرارية (Redundant)، إستلزامية (Implied) و إضافية (Surrogate)، كما يتم استعمال متغيرات جديدة. كل هذه التحسينات تزيد في سرعة الوصول إلى الحل ، كما أنها قد تكون السبب في تبلور أخطاء في البرنامج.

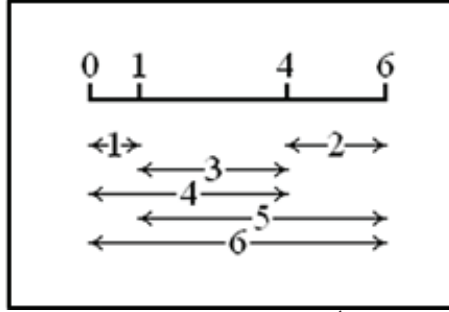
في الفقرة الموالية سنتطرق إلى مثال توضيحي لنشرح بشكل عام نظرتنا في فحص (ب.م). بينما نتناول في الفقرة الثالثة شرح بعض المفاهيم و التعاريف الأساسية المتعلقة بفحص البرمجيات، كذلك نقدم الخوارزمية الأساسية التي استخدمناها في عملية فحصنا للبرامج (ب.م). بينما تقدم الفقرة الرابعة عرضاً متواضعاً لأداة الفحص الآلي (CPTEST) مع عرض بعض النتائج الأولية على مشكل أكاديمي (Golomb Rulers) معروف بين الباحثين في (ب.م). وأخيراً ننهي هذا المقال بخاتمة تشمل ما تطرقنا إليه، بالإضافة إلى الأعمال التي هي قيد البحث.

2 مثال توضيحي

سننترق في هذه الفقرة إلى مثال حول التحسينات الممكنة للحصول على برنامج فعال ومثالي إنطلاقاً من نموذج-مرجعي. اخترنا مشكلاً أكاديمياً (Golomb Rulers) [10] الذي يمثل مشكلة الحصول على أداة قياس خاصة (Ruler). الخاصية في هذه الأداة أنها تمثل مجموعة من النقاط $(0 = x_1 < x_2 < \dots < x_m)$ حيث أن كل مسافة بين نقطتين مختلفتين تختلف عن الأخرى

$(x_j - x_i \neq x_l - x_k)$. المشكل يعود إلى الحصول على أصغر أداة قياس ممكنة ($\text{minimize } x_m$). يستعرض الشكل 1 مثالا عن أصغر أداة قياس ذات 4 نقاط.

أما في الشكل 2 الجزء (أ) فنقدم نموذجاً لـ (Golomb Rulers) الذي يعد ترجمة أولى ومباشرة للمشكل المطروح سابقاً. في حين يمثل الجزء (ب) من الشكل 2 برنامجاً مكرراً و مطوّراً بعد تطبيق سلسلة من التحسينات. و هنا نشير إلى أن كل من النموذج و البرنامج مكتوبين بلغة البرمجة : (OPL: Optimization Programming Language).



الشكل 1: أداة قياس Golomb ذات 4 نقاط.

إذ من السهل إقناع شخص بأن النموذج (أ) يمثل مشكل (Golomb Rulers)، لكن من الصعب تحديد ذلك مع البرنامج (ب). حيث نجد في البرنامج المطور متغيرات جديدة عن الموجودة في النموذج كالمصفوفة ($d[indexes]$) ، مقيد (cc6) الذي يكسر التناظر بشكل ثابت (static symmetry breaking). كما نجد مقيدات تكرارية ، إستراتيجية وإضافية (cc7, cc8, cc9). مقيد (cc5) يعتبر مقيد خاص "الاختلاف الكلي" ($allDifferent$). السؤال الذي يطرح نفسه هنا ، هل البرنامج (ب) يوافق النموذج (أ) ؟ الطريقة الأولى التي تخطر على بال فاحص البرامج هي أخذ حل للبرنامج (ب) والتحقق إذ كان حلاً للنموذج (أ). لو قدمنا البرنامج (ب) ، بعد حذف مقيد (cc5) ، لمكتبة برمجية (solver) (بعدد نقاط : $m=8$) ، بإمكان هذه المكتبة إعطاء الحل التالي :

$$x=[0 \ 1 \ 3 \ 6 \ 10 \ 26 \ 27 \ 28]$$

هذا الحل يعد خاطئاً و ليس بأداة قياس (Golomb) وذلك لوجود مسافتين متساويتين بين نقاط مختلفة ($28-27 = 27-26$).

إذن هذا وإن دلّ على شيء فإنما يدل على وجود خطأ في البرنامج (ب). سنتطرق الفقرة التالية إلى طريقة أوتوماتيكية للكشف عن وجود أخطاء في برامج (ب.م).

3 إطار فحص البرمجة بالمقيدات (ب.م)[7]

3.1 النموذج-المرجعي و البرامج بالمقيدات

النموذج-المرجعي (M) و ما يندرج عنه من برامج مطوّرة (P) ، يعدان تزامناً (Conjunction) من مجموعة من المقيدات C_i التي تربط عدد من المتغيرات x و التي بدورها ترتبط بعوامل تغير k . على سبيل المثال ، مشكل Golomb Rulers ، يتعلق عدد المتغيرات بعدد نقاط الأداة :

$$m=3 \rightarrow x=[0 \ 1 \ 3]$$

$$m=4 \rightarrow x=[0 \ 1 \ 4 \ 6]$$

تمثل المجموعة $sol(M)$ (على التوالي $sol(P)$) مجموعة الحلول للنموذج M (على التوالي للبرنامج P).

<pre> 1 using CP; 2 3 int m=...; 4 5 dvar int x[1..m] in 0..m*m; 6 7 minimize x[m]; 8 9 subject to{ 10 11 c1: forall(i in 1..m-1) 12 x[i] < x[i+1]; 13 14 c2: forall(ordered i,j,k,l in 1..m: 15 (i!=k j!=l)) 16 (x[j]-x[i])!=(x[l]-x[k]); 17 18 }</pre>	<pre> 1 using CP; 2 3 int m=...; 4 5 tuple indexerTuple { 6 int i; 7 int j; 8 } 9 10 {indexerTuple} indexes = {<i,j> i,j in 1..m : 11 i<j}; 12 13 dvar int x[1..m] in 0..m*m; 14 dvar int d[indexes]; 15 16 minimize x[m]; 17 18 subject to{ 19 20 cc1: forall (i in 1..m-1) 21 x[i] < x[i+1]; 22 23 cc2: forall(ind in indexes) 24 d[ind] == x[ind.j]-x[ind.i]; 25 26 cc3: x[1]==0; 27 28 cc4: x[m] >= (m * (m - 1)) / 2; 29 30 //cc5: allDifferent(all(ind in indexes)d[ind]); 31 32 cc6: x[2] <= x[m]-x[m-1]; 33 34 cc7: forall(ind1,ind2,ind3 in indexes: 35 (ind1.i==ind2.i) && (ind2.j==ind3.i) && 36 (ind1.j==ind3.j) && (ind1.i<ind2.j<ind1.j)) 37 d[ind1]==d[ind2]+d[ind3]; 38 39 cc8: forall(ind1,ind2,ind3,ind4 in indexes: 40 (ind1.i==ind2.i) && (ind1.j==ind3.j) && 41 (ind2.j==ind4.j) && (ind3.i==ind4.i) && 42 (ind1.i<m-1) && (3<ind1.j<m+1) && 43 (2<ind2.j<m) && (1<ind3.i<m-1) && 44 (ind1.i<ind3.i<ind2.j<ind1.j)) 45 d[ind1]==d[ind2]+d[ind3]-d[ind4]; 46 47 48 cc9: forall(i,j in 2..m, k in 1..m : i < j) 49 x[i]==x[i-1]+k => x[j] != x[j-1]+k; 50 }</pre>
--	--

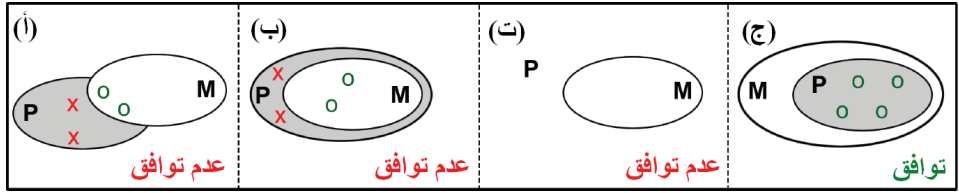
(أ)

(ب)

الشكل 2 : نموذج و برنامج Golomb Rulers.

3.2 علاقة التوافق

إن فحص برنامج (ب.م) و محاولة كشف الأخطاء فيه يعتمد على مدى توافقه مع النموذج-المرجعي المحدد. إذ كلما كانت حلول البرنامج تنتمي إلى حلول النموذج، كلما كان التوافق كبيراً.



الشكل 3 : حالات توافق و عدم توافق البرنامج P مع النموذج-المرجعي M.

يمثل الشكل 3 الحالات الممكنة لتوافق أو عدم توافق البرنامج P مع النموذج-المرجعي M. الحالة (أ) و (ب) هي حالات عدم توافق بسبب وجود حلول للبرنامج P التي لا يقبلها النموذج-المرجعي M. كما أنّ الحالة (ت) تعد أيضاً حالة عدم توافق بما أنّ الخطأ المرتكب في البرنامج P أدى لإنعدام الحلول. الحالة (ج) تعد حالة توافق حيث أنّ كل الحلول P هي أيضاً حلول للنموذج-المرجعي M.

تعريف 1 : (علاقة توافق $conf$)

$$P \text{ conf } M \Leftrightarrow \emptyset \subsetneq sol(P) \subseteq sol(M)$$

لكي يتحقق التوافق بين البرنامج P والنموذج-المرجعي M يجب أن تكون مجموعة الحلول $sol(P)$ غير فارغة و أن تكون جزء من المجموعة $sol(M)$.

3.3 عملية الفحص

الهدف من عملية الفحص هو الحصول على حل ل P الذي لا يكون حلاً ل M. و بذلك يمكننا الكشف عن أخطاء في P. إذن للحصول على حل مخالف يلزمنا حل المشكل التالي :

$$P \wedge \neg M$$

$$M \equiv C_1 \wedge C_2 \dots \wedge C_n \quad \text{بما أن :}$$

فيمكننا اختيار مقيد C_i من M و محاولة حل المشكل : $P \wedge \neg C_i$

في حالة انعدام الحلول نقوم بالرجوع إلى الخلف (*Backtracking*) و اختيار مقيد آخر C_j و حل المشكل :

$$P \wedge \neg C_j$$

الخوارزمية 1: $one_negated(P, M)$

لإجراء عملية الفحص ، تقوم الخوارزمية 1 باختيار مقيد C_i من M لتقوم بنفيه ومن ثم البحث عن حل للبرنامج P بعد أن يضاف له نفي C_i . في حالة عدم الحصول على حل ، تقوم الخوارزمية بالرجوع إلى الخلف و اختيار مقيدة أخرى للنفي و هكذا حتى نحصل على الحل الذي يؤكد وجود خطأ في البرنامج.

الخوارزمية 1 : $one_negated(P, M)$

Input : P : البرنامج $M = \{C_1, \dots, C_n\}$: النموذج المرجعي
Output: $conf$ when P conform M , $\neg conf$ (+ non-conformity point)
 otherwise

$nc \leftarrow \emptyset$
foreach $C_i \in M$ **do**
 $nc \leftarrow Solve(B \wedge \neg C_i)$
 if nc **then return** $\neg conf(nc)$
end
return $conf$

4 النتائج والتحقق التجريبي**4.1 أداة فحص آلية CPTEST [6]**

قمنا خلال دراستنا السابقة ببرمجة الخوارزمية 1 وإضافتها إلى الأداة التي سميت $CPTEST$ ، و التي تعتمد على المكتبة البرمجية $IBM Ilog CP Optimizer 2.1$. بحيث تقوم بفحص برامج (ب.م) المكتوبة بلغة $OPL : Optimization programming language$ [12] التي تعتبر من أحسن برامج المقيدات و أكثرها انتشاراً في الأوساط الصناعية. يعتمد $CPTEST$ على المكتبة البرمجية $IBM Ilog CP Optimizer 2.1$. يمكنكم الإطلاع على مجموعة من النتائج التي تخص هذه الأداة كما يمكنكم تحميل الأداة $CPTEST$ من خلال موقع الانترنت التالي:

(www.irisa.fr/celtique/lazaar/CPTEST)

إذن الخوارزمية 1 تعتمد على نفي المقيدات للكشف عن الأخطاء و في هذا الإطار يمكننا القول أن $CPTEST$ بإمكانه نفي أغلبية المقيدات حتى الخاصة منها مثل : $allDifferent$, $bin\ packing$, $allMinDistance...$

4.2 التطبيق على مشكل Golomb Rulers

لقد جربنا *CPTEST* على أربعة برامج (*CPUT: Constraint Program Under Test*) تتضمن أخطاء متعددة و توصلنا إلى مجموعة من النتائج نعرضها من خلال الجدول التالي الذي يوضح النتائج التي تم الحصول عليها مع قياس للوقت اللازم للفحص.

m = 8		conf
Non-conf points		[0 7 8 18 24 26 35 44]
CPUT1	T(s)	4.29s
Non-conf points		[0 4 5 26 28 31 47 63]
CPUT2	T(s)	5.62s
Non-conf points		[0 4 5 26 28 31 47 63]
CPUT3	T(s)	9.53s
Non-conf points		[0 12 18 20 29 33 34 39]
CPUT4	T(s)	12.60s

الجدول 1: فحص و كشف عن أخطاء في برامج *Golomb Rulers*

لنأخذ على سبيل المثال البرنامج *CPUT2* ، نرى ان *CPTEST* إستطاع الحصول على حل لأداة قياس *Golomb* ذات 8 نقاط لكنه لا يلبي مطالب النموذج-المرجعي في ظرف وقت أقل من ستة ثواني. إذن هذا الحل ([0 4 5 26 28 31 47 63]) لا يعد أداة قياس صحيحة و ذلك لوجود مسافتين متساويتين بين أربعة نقاط مختلفة (31-47 = 47-63). كما يدل على وجود خطأ في البرنامج *CPUT2* الذي يجب إعادة النظر فيه و تصليحه.

5 الخاتمة

لقد قدمنا من خلال هذه الورقة العلمية دراسة حول فحص برامج بالمقيدات (ب.م) كما استعرضنا قوة و فعالية الخوارزمية في الكشف عن وجود الأخطاء المحتملة. تمّت برمجة هذه الخوارزمية في أداة فحص آلية تدعى *CPTEST* و التي تمّ اختبارها على المثال الأكاديمي *Golomb Rulers*.

عملية الفحص المقترحة في هذه الورقة تعتمد على نفي المقيدات، التي ليس من السهل الحصول عليها. و لهذا فإن مشكل نفي المقيدات يبقى قيد البحث. كما أنّ عملية كشف الأخطاء في البرامج تستدعي بعد ذلك عملية كشف موقع الخطأ [8] و تصحيحه [9].

6 شكر وتقدير

أولاً وقبل كل شيء، الحمد و الشكر لله على هذا التوفيق بتقديم دراسة علمية بلغة القرآن الكريم. نقدم أيضاً كل الشكر و التقدير لمحمد الخطيب على أرائه و تصحيحاته اللغوية.

المراجع

- [1] P. Deransart, M. V. Hermenegildo, and J. Maluszynski, editors. Analysis and Visualization Tools for Constraint Programming, Constrain Debugging (DiSCiPl project), volume 1870 of Lecture Notes in Computer Science. Springer, 2000.
- [2] P. Flener, J. Pearson, M. Agren, Garcia-Avello C., M. Celiktin, and S. Dissing. Air-traffic complexity resolution in multi-sector planning. Journal of Air Transport Management, 13(6):323-328, 2007.
- [3] J. B. Goodenough and Susan L. Gerhart. Toward a theory of test data selection. In Proceedings of the international conference on Reliable software, pages 493–510, New York, NY, USA, 1975. ACM.
- [4] A. Holland and B. O'Sullivan. Robust solutions for combinatorial auctions. In ACM Conference on Electronic Commerce (EC-2005), pages 183-192, 2005.
- [5] U. Junker and D. Vidal. Air traffic ow management with ilog cp optimizer. In International Workshop on Constraint Programming for Air Traffic Control and Management, 2008. 7th EuroControl Innovative Research Workshop and Exhibition (INO'08).
- [6] N. Lazaar. CPTEST: A framework for the automatic fault detection, localization and correction of constraint. In Proc. of the 3rd Workshop on Constraints in Software Testing, Veri_cation, and Analysis programs (CSTVA'11), held within ICST'11, Mar. 2011, Berlin (Germany).
- [7] N. Lazaar, A. Gotlieb, and Y. Lebbah. On testing constraint programs. In Proc. of Principles of Constraint Programming, CP'2010, Sept. 2010.
- [8] N. Lazaar, A. Gotlieb, and Y. Lebbah. Fault localization in constraint programs. In Proc. of the 2010 IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2010, Oct. 2010.
- [9] N. Lazaar, A. Gotlieb, and Y. Lebbah. A framework for the automatic correction of constraint programs. In Proc. of the 2011 IEEE International Conference on Software Testing, Veri_cation and Validation, ICST'11, Mar. 2011, Berlin (Germany).
- [10] W. T. Rankin. Optimal golomb rulers : An exhaustive parallel search implementation. Master's thesis, Duke University, Durham, 1993.
- [11] F. Rossi, P. Beek and T. Walsh. Handbook of Constraint Programming (Foundations of Artificial Intelligence). Elsevier Science Inc. 2006.
- [12] P. Van Hentenryck. The OPL optimization programming language. MIT Press, Cambridge, MA, USA, 1999.
- [13] P. Van Hentenryck and L. Michel. Constraint-Based Local Search. The MIT Press, 2005.

الملخص باللغة الانجليزية

Title: Testing Theory on Constraints Programs

Abstract. The success of several constraint-based modeling languages such as OPL, ZINC, or COMET, appeals for better software engineering practices, particularly in the testing phase. This paper introduces a testing framework enabling automated test case generation for constraint programming. We propose a general framework of constraint program development which supposes that a first declarative and simple constraint model is available from the problem specifications analysis. Then, this model is refined using classical techniques such as constraint reformulation, surrogate, redundant, implied and global constraint addition, or symmetry-breaking to form an improved constraint model that must be thoroughly tested before being used to address real-sized problems. We think that most of the faults are introduced in this refinement step and propose a process which takes the first declarative model as an oracle for detecting non-conformities and derive practical test purposes from this process. We implemented this approach in a new tool called CPTEST that was used to automatically detect non-conformities on classical benchmark program : the Golomb rulers.

Keywords: Constraint programming, software testing, conformity relation, constraint negation.

المصطلحات

الذكاء الاصطناعي	Artificial Intelligence
الرجوع إلى الخلف	Backtracking
علاقة توافق	Conformity Relation
مقيد	Constraint
نفي مقيد	Constraint Negation
مقيد خاص	Global Constraint
مقيد إستلزامي	Implied Constraint
نموذج-المرجعي	Model-Oracle
عامل متغير	Parameter
البرمجة بالمقيدات	Constraint Programming
مقيد تكراري	Redundant Constraint
فحص البرمجيات	Software Testing
المكتبة البرمجية	Solver
كسر التناظر بشكل ثابت	Static Symmetry Breaking
مقيد إضافي	Surrogate Constraint
متغير	Variable

بناء معجم لكلمات الآراء قائم على تعليم الآلة بهدف تحليل الآراء

علاء الدين حموده ، محمد رحيم
جامعة الأزهر ، كلية الهندسة، قسم هندسة النظم والحاسبات

البريد الإلكتروني: mrohaim@azhar.edu.eg ، alaa_ham@gega.net

الخلاصة. ينطوي تحليل الآراء على تصنيف الآراء الواردة في التعليقات إلى آراء إيجابية " و أخرى "سلبية". أحد الأساليب المتبعة لإجراء تصنيف الآراء يتم باستخدام معاجم كلمات الآراء. تهدف هذه الورقة إلى بناء معجم كلمات آراء لا يعتمد على مجال معين . حيث أننا نقترح بناء معجم كلمات آراء باستخدام طريقة تعليم الآلة القائم على تغذية الآلة بمجموعة تعليقات في مجالات مختلفة ومصنفة سلفاً إلى إيجابي و سلبي. ولقد قدم هذا المعجم المقترح نتائج أفضل من المعاجم السابقة المنشورة سواء المعاجم المنشأة يدوياً أو المعاجم المنشأة بطريقة آلية مثل قاموس كلمات المشاعر (SentiWordNet). كما قدمنا تحسناً في طريقة الحساب المستخدمة في تحليل الآراء الخاصة بالتعليقات.

الكلمات الجوهرية: تحليل الآراء، معجم كلمات الآراء، تعليقات العملاء، أنظمة التوصيات.

١ مقدمة

يحتوى الإنترنت في هذه الأيام على عدد كبير من التعليقات الخاصة بالمنتجات سواء كانت موجودة في المدونات أو المنتديات المختلفة. لذلك يهتم محللو الأعمال بالإنترنت للحصول على التعليقات و الآراء الخاصة بأداء الشركات و كفاءة المنتجات. يمكن لعملية التنقيب عن الآراء أن تستخدم في تطبيقات مختلفة مثل محركات البحث ، وأنظمة التوصيات والإقتراحات و بحوث التسويق.

تعتبر عملية تصنيف تعليقات المنتجات مشكلة مهمة في مجال التنقيب في الآراء ولذلك إستخدمت تقنيات مختلفة لمعالجة هذه المشكلة. يمكن تقسيم هذه التقنيات إلى نهجين رئيسيين. نهج قائم على استخدام المعاجم الخاصة بالآراء مع معالجة اللغات الطبيعية، والنهج الآخر قائم على استخدام الخوارزميات الخاصة بتعليم الآلة.

تعتمد أساليب تعليم الآلة- سواء الموجه وغير الموجه- على خصائص مختلفة للنص المراد تصنيفه. في البداية هناك عمل موضح في [1] يستخدم عدة خوارزميات خاصة بأسلوب تعليم الآلة الموجه، والتي تعتمد على إستخراج حقبة كلمات مميزة للنصوص. وقد أوضح هذا العمل أن أفضل أداء لعملية التصنيف كان باستخدام خوارزم التعليم "المكائن ذات الدعم الإتجاهي (م د إ)" مع إستخدام خاصية أن تكون الكلمات مفردة. إن تصنيف التعليقات حسب الخصائص النحوية لكلماتها تم

إستخدامه في عدة مقالات سابقة، أحد هذه الأعمال قام بإستخدام الخصائص النحوية للكلمات لتصنيف المقالات الإخبارية [٢] . كذلك في [٣] إستخدم أسلوب مماثل على مجموعة مقالات أخرى. في [٤] إستخدم عدة خصائص للكلمة وهي لفظها ومصدرها وإعرابها بتركيبات مختلفة لتصنيف مجموعات مختلفة لتعليقات عملاء. هناك دراسات أخرى إستخدمت معاجم المشاعر مثل قاموس كلمات المشاعر لإنشاء مجموعة خصائص مختلفة مستقاة من القياسات الموجودة في المعجم، ثم إستخدام هذه الخصائص في تنفيذ خوارزم التعليم "المكانن ذات الدعم الإتجاهي" ليستخدم في عملية التصنيف كما هو موضح في [٥] و [٦].

معاجم الآراء هي مصادر تحدد الدلالات الشعورية للكلمات. وإستخدام هذه المعاجم في مجال بحوث التنقيب في الآراء يبني على افتراض مؤداه أن الكلمات يمكن اعتبارها كوحداث معلوماتية تقدم أفكار لتوثيق الرأي الذاتية [٥]. إنشاء معاجم الآراء يدويا قد تم تطبيقه كما هو موضح في [١]، بحيث تم توقع قطبية الوثيقة عن طريق عد الكلمات الموجبة والكلمات السالبة. كذلك تم تقديم منهج مشابه في [٧] ، حيث تم إستخدام معجم آراء يستند على الجمع بين عدة مصادر مجمعية موجودة مسبقاً.

إن البناء اليدوي لمعاجم الآراء يستهلك الكثير من الوقت. وللتغلب على هذه المعضلة تم إقتراح منهج آخر لبناء معجم المشاعر، وذلك عن طريق زيادة حجم المعجم الذي يحتوى على الكلمات الشعورية المشهورة بإضافة كلمات أخرى ذات دلالات شعورية مشابه لهذه الكلمات . أحد معاجم الآراء الشائعة الإستخدام في هذا المنهج هو قاموس كلمات المشاعر [٨] والذي أعطى دلالات شعورية للكلمات الموجودة في قاموس الكلمات (WordNet) بإستخدام طريقة التعليم شبه الموجه، وهو متاح مجاناً للأغراض البحثية. ويقدم قاموس كلمات المشاعر قاعدة بيانات للكلمات الشعورية الخاصة باللغة الإنجليزية تصلح لأن تكون بديلاً للمعاجم المنشأة يدوياً. والطريقة المستخدمة لتطوير هذا المعجم

نقدم في هذه البحث معجم للكلمات الخاصة بإبداء الرأي المكون بإستخدام أسلوب تعليم الآلة والمبني على تعليقات من مجالات مختلفة. ولقد قمنا بتقسيم البحث كالتالي : يعرض القسم الثاني الخطوات اللازمة لبناء معجم كلمات آراء قائم على تعليم الآلة (م ك أ ق ت أ). بينما القسم الثالث يعرض نتائج إستخدام هذا المعجم. ثم في القسم الرابع نقدم عدة إقتراحات لتحسين أداء المعجم، وفي النهاية نضع بين أيديكم الاستنتاج النهائي.

٢ معجم كلمات الآراء المقترح

لقد قمنا بإقتراح معجم كلمات آراء قائم على تعليم الآلة والذي يعتمد في تكوينه على حقيقة الكلمات الناتجة من تطبيق خوارزم التعليم "المكانن ذات الدعم الإتجاهي" على التعليقات. التقنية المقترحة لبناء هذا المعجم تتكون من مرحلتين رئيسيتين كما هو مبين في شكل ١.

٢.١ مرحلة تجهيز البيانات

لقد استخدمنا فى عملنا تعليقات المنتجات الخاصة بحزمة نصوص الأمازون والمتاحة فى الموقع الإلكتروني [٩]. تحتوى هذه الحزمة على ما يقرب من ٥,٠٠٠,٠٠٠ تعليق للعديد من المنتجات المختلفة (كتب، كاميرات، إلخ) مما يعنى أنها تحتوى على تعليقات فى مجالات مختلفة. يتكون كل تعليق من جزء نصى و تقييم رقمى من ١ إلى ٥. تم تنفيذ خطوات التجهيز التالية على هذه المجموعة قبل تطبيق خوارزم التعليم.

١. تحويل تعليقات الحزمة إلى تصنيفين (موجبة و سالبة) بإعتبار التعليقات ذات التقييم (١ و ٢) سالبة والتعليقات ذات التقييم (٤ و ٥) موجبة وإهمال التعليقات ذات التقييم ٣ ، "تعليقات عادية".
 ٢. إختيار ٢٥٠٠٠ تعليق عشوائياً من 756,958 تعليق، بحيث يكون عدد التعليقات الموجبة مساوى لعدد التعليقات السالبة وأن لا يزيد عدد أحرف التعليق عن ٥٠٠ حرف.
 ٣. تطبيق عملية تقسيم الكلمات على التعليقات المختارة لتقسيم النص الخاص بالتعليق إلى مجموعة من الكلمات والأرقام وعلامات الترقيم.
 ٤. تطبيق التحليل الصرفى للحصول على مصادر الكلمات.
- لقد قمنا باستخدام أداة الهيكل العام لهندسة النص [١٠] فى تنفيذ الخطوات ٣ و ٤. و تستخدم هذه الأداة فى تطوير وتطبيق تكنولوجيا معالجة النصوص.

٢.٢ مرحلة تطوير المعجم

لقد قمنا باستخدام حزمة البيانات المجهزة من المرحلة السابقة فى تعليم المكانن ذات الدعم الإتجاهي. ويقوم هذا التعليم على أساس تقسيم النص إلى وحدات لغويه (كل وحدة تحتوى على لفظه واحده) مع استخدام الخصائص اللغويه لهذه الوحدات. تستخدم تقنية تعليم الآله فى حساب قيم للكلمات حسب معدل تواجدها فى البيانات المستخدمة للتعليم. ولقد تم تطبيق المكانن ذات الدعم الإتجاهي باستخدام العوامل الآتية:

- السمة اللغوية المستخدمة: هى لفظة الكلمات كما هى.
- خاصية (ن-وحدة): هى خاصية تستخدم لإنشاء سلسلة من اللفظات التى سيتم إستخدامها فى التعليم. فى تقنيتنا قمنا بمزج أحادى-الوحدة (الذى يحتوى على لفظه واحدة) و ثنائى-الوحدة (الذى يحتوى على لفظتين) باستخدام أوزان بنسبة ١:٥ على الترتيب. وتم إستخدام خاصية ثنائى-الوحدة مع إعطائها وزن أكبر أثناء عملية التعليم لكى نأخذ فى الإعتبار التعبيرات المركبة مثل "جيد جداً" بدلاً من أخذ كل كلمة منها على حده "جيد" و "جداً".
- طريقة حساب أوزان الألفاظ المستخدمة: هى تردد اللفظة-معكوس تردد الوثيقة (ت ل - م ت و). و فيها وزن اللفظة يساوى حاصل ضرب تردد اللفظة فى معكوس تردد الوثيقة. ويعرف تردد اللفظة بأنه عدد مرات تكرار اللفظة فى الوثيقة. ولتفادى حدوث نوع من التحيز للفظه معينة بسبب تكرارها كثيراً فى الوثائق كبيرة الحجم يتم حساب تردد اللفظة بالمعادلة الآتية.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

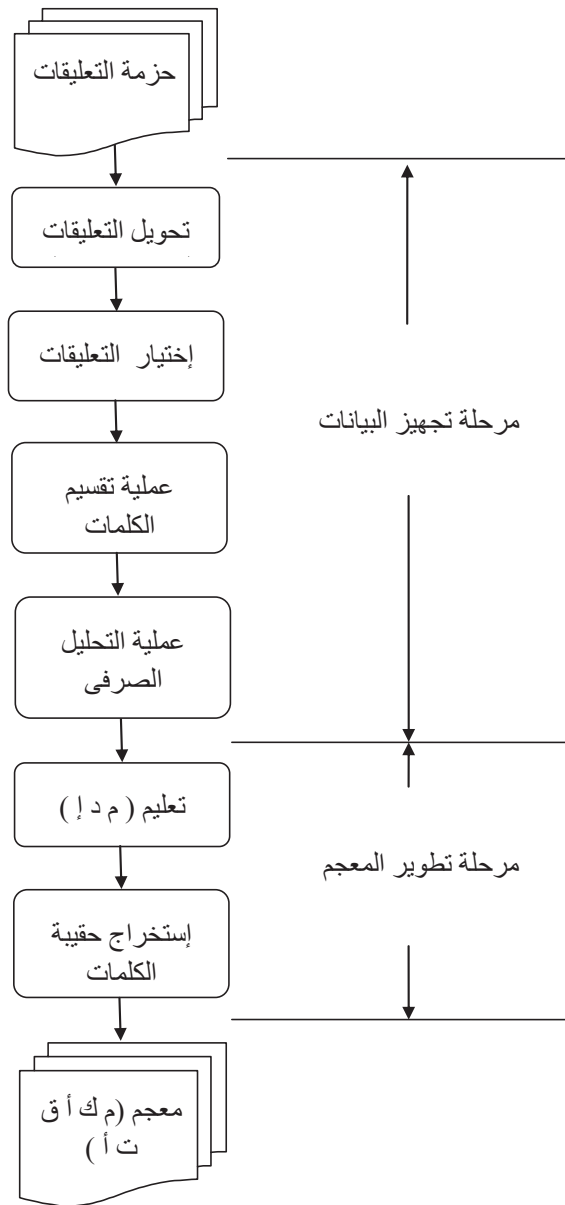
حيث أن: $n_{i,j}$ هي عدد مرات ظهور اللفظة t_i في الوثيقة d_j والمقام هو عدد الألفاظ الموجودة في الوثيقة

أما معكوس تردد الوثيقة فهو مقياس للتعبير عن أهمية اللفظة ويتم حسابه بالمعادلة الآتية.

$$idf_i = \log \frac{|D|}{|\{d_j: t_i \in d_j\}|}$$

حيث أن: $|D|$ هو العدد الإجمالي للوثائق المستخدمة، و $|\{d_j: t_i \in d_j\}|$ عدد الوثائق التي ظهرت فيها اللفظة t_i .

إن النموذج المولد من هذا التعليم سوف يحتوى على حقيقة-كلمات تحتوى على العديد من الألفاظ ذات أوزان موجبة وسالبة. وقد اعتبرنا أن هذه الأوزان تمثل درجات شعوريه لهذه الألفاظ ، وتصنف اللفظة على أنها موجبة او سالبة حسب الإشارة.



الشكل ١. إنشاء معجم كلمات آراء بتطبيق تقنية تعليم الآلة

٣ استخدام معجم كلمات الآراء القائم على تعليم الآلة لتصنيف تعليقات العملاء

من خلال حقبة الكلمات المولدة في القسم السابق قمنا باختيار الكلمات الأعلى حسب القيم المطلقة الخاصة بها وتم إعتبارها كمحتوى معجم كلمات الآراء القائم على تعليم الآلة المقترح، وقد قمنا بتطبيق هذا المعجم على اثنتين من الحزم النصية، الحزمة الأولى هي حزمة نصوص الأمازون، حيث اخترنا أربعة آلاف تعليق (٢٠٠٠ موجب، ٢٠٠٠ سالب) هذه التعليقات مختارة بحيث تكون مختلفة عن التعليقات المستخدمة للتعليم وذلك كي نضمن واقعية النتائج. الحزمة الثانية هي حزمة نصوص الأفلام المستخدمة في [١]، [٤]، [٥]. هذه المجموعة تتكون من ١٠٠٠ تعليق موجب و ١٠٠٠ تعليق سالب، وهي متاحة على الموقع [١١].

ولقد استخدمنا طريقة "عد الكلمات" في تطبيق معجم (م ك أ ق ت أ) لتصنيف التعليقات. في هذه الطريقة يتم تطبيق المعجم عن طريق عد الكلمات الموجبة والسالبة الموجودة في التعليق ومن ثم تحديد قطبية الآراء استناداً إلى أي النوعين قد حقق النتيجة الأعلى. تم تطبيق هذه الطريقة باستخدام أحجام مختلفة للمعجم (حجم المعجم يعني عدد الكلمات التي يحتويها) والنتائج مبينة في الجدول الآتي. من الجدول نرى أن (م ك أ ق ت أ) يمتلك أداء (دقة) جيد عندما يكون حجمه ما بين ١٥٠٠٠ و ٢٠٠٠٠ كلمة، بينما الأداء الأعلى عند ١٥٠٠٠ كلمة.

حجم (م ك أ ق ت أ) (كلمة)	الدقة (%)		
	الأمازون	الأفلام	المتوسط
٢٠٠٠٠	٧٢,١٣	٧٠,١٥	٧١,١٤
١٥٠٠٠	٧١,٣٢	٧١,٧٥	٧١,٥٥
١٠٠٠٠	٧٠,٤٢	٦٧,٦	٦٩,٠١
٥٠٠٠	٦٨,٤٢	٦٥	٦٦,٧١

جدول ١: نتائج تصنيف التعليقات باستخدام (م ك أ ق ت أ) المبني على التعليقات العادية مع إعتبار لفظة الكلمات كسمات لغوية واستخدام طريقة العد.

وللمقارنة بين المعجم الخاص بنا والمعجم الأخرى المنشورة، استخدمنا نفس حزمة النصوص (حزمة نصوص الأفلام) المستخدمة سابقاً. وتظهر النتائج أن استخدام المعجم الخاص بنا يثمر عن تحسن هام عن المعجم الأخرى، كما هو مبين في الجدول الآتي:

الطريقة	الدقة
إستخدام (م ك أ ق ت أ) مع طريقة عد الكلمات (هذا البحث).	٧١,٧٥ %
إستخدام قاموس كلمات المشاعر مع طريقة عد الكلمات [١].	٦٥,٥٨ %
إستخدام قائمة من الكلمات الموجبة والسالبة المكونة يدوياً مع طريقة عد الكلمات [١].	٦٩,٣٥ %
إستخدام قائمة كلمات مكونة من عدة معاجم مع طريقة عد الكلمات [٤].	٦٧,٨٠ %

جدول ٢ : مقارنة النتائج

٤ تحسين معجم كلمات الآراء القائم على تعليم الآلة

في القسم السابق إستخدمنا المكانن ذات الدعم الإتجاهى لبناء معجم مشاعر قائم على تعليم الآلة بإستخدام حزمة البيانات المستخرجة من حزمة نصوص الأمازون (التعليقات بنسبة تصويت ١ و ٢ تكون سالبة ، والتعليقات بنسبة تصويت ٤ و ٥ تكون موجبة) مع إعتبار لفظة الكلمات كسمة مميزة للتعليم. وفي هذا القسم نقترح عدة تحسينات لتحسين أداء المعجم. وفي كل من هذه التحسينات يتم تسجيل النتائج للتحقق من تحسن الأداء.

فى التحسين الأول: إستخدمنا حزمة نصوص بديلة مع معنى قوى للإيجابية والسلبية. تم تكوين هذه الحزمة من البيانات باستخراجها من حزمة نصوص الأمازون بحيث تكون التعليقات بنسبة تصويت ١ سالبة وبنسبة تصويت ٥ موجبة "تعليقات قوية". ثم إستخدام هذه التعليقات لبناء معجم محسن بالإعتماد على الآتي:

- حزمة البيانات عبارة عن ٢٥٠٠٠ تعليق قوى تم استخراجها عشوائيا من ١,٢٦٠,٤٩٩ تعليق ذات أحجام ليست أكبر من ٥٠٠ حرف.
- السمة اللغوية المستخدمة: هى لفظة الكلمات كما هى. مع باقى الخصائص المستخدمة لتعليم الآلة كما هو موضح فى القسم الثالث.

كما هو موضح فى القسم السابق ، أن الأداء الأفضل للمعجم كان عند حقبة كلمات ذات حجم ١٥٠٠٠. ولذلك قمنا باستخدام معجم مولد يحتوي على ١٥٠٠٠ كلمة لإختبار التحسينات الخاصة بالمعجم (م ك أ ق ت أ). وطبقاً للنتائج الموضحة فى الجدول ٣ فإن إستخدام حزمة البيانات القوية فى التدريب قد أحدث تحسن ضئيل قدره ٣,٣٣ % .

نوع (م ك أ ق ت أ) مع طريقة الحساب			الدقة (%)
الأمازون	الأفلام	المتوسط	
٧٢,٢٥	٧١,٥	٧١,٨٨	المعجم القائم على التعليقات القوية مع لفظة الكلمات مع إستخدام طريقة العد.

جدول ٣ : نتائج تصنيف التعليقات باستخدام (م ك أ ق ت أ) المبنى على التعليقات القوية مع إعتبار لفظة الكلمات كسمة لغوية وإستخدام طريقة العد.

أما فى التحسين الثاني: تم استخدام طريقة "جمع قيم الكلمات" والتي يتم فيها حساب مجموع قيم الكلمات الموجبة ومجموع قيم الكلمات السالبة لتعليق ما، ويتم تحديد الرأي الخاص بالتعليق (موجب أم سالب) حسب النوع صاحب المجموع الأعلى. وتتميز هذه الطريقة عن طريقة عد الكلمات في أنها تأخذ في الإعتبار القيم المقدرة للكلمات والتي تعبر قوة تعبير الكلمة الموجبة عن الإيجاب أو تعبير الكلمة السالبة عن السلب [١٢].

النتائج في الجدول ٤ تبين أن إستخدام طريقة التجميع ينتج تحسن هام في أداء تصنيف التعليقات يصل الى ٥.٨٧ % . في الواقع هذ التحسن ليس مكتسباً من المعجم نفسه ولكن من كيفية إستخدامه في تحديد قطبية الآراء الخاصة بالتعليقات.

نوع (م ك أ ق ت أ) مع طريقة الحساب			الدقة (%)
الأمازون	الأفلام	المتوسط	
٧٨,٩٩	٧٦,٥	٧٧,٧٥	المعجم القائم على التعليقات القوية مع لفظة الكلمات مع إستخدام طريقة التجميع.

جدول ٤ : نتائج تصنيف التعليقات باستخدام (م ك أ ق ت أ) المبنى على التعليقات القوية مع إعتبار لفظة الكلمات كسمة لغوية وإستخدام طريقة الجمع.

أما فى التحسين الثالث: قمنا بإستخدام مصادر الكلمات كسمة لغوية عوضاً عن لفظة الكلمة. ثم نطبق نفس التقنية الموصوفة في القسم الثالث لبناء معجم محسن ، مبني على الآتي:

- حزمة البيانات عبارة عن ٢٥٠٠٠ تعليق قوى تم إستخراجها عشوائياً من ١,٢٦٠,٤٩٩ تعليق ذات أحجام ليست أكبر من ٥٠٠ حرف.
- السمة اللغوية المستخدمة: هي مصادر الكلمات. مع باقى الخصائص المستخدمة لتعليم الآلة كما هو موضح فى القسم الثالث.

جدول ٥ يوضح أن إستبدال سمة لفظة الكلمة بسمة مصدرها ينتج زيادة فى الدقة بنسبة ٠.٨٣ % .

الدقة (%)			نوع (م ك أ ق ت أ) مع طريقة الحساب
المتوسط	الأفلام	الأمازون	
٧٨,٥٨	٧٧,٨	٧٩,٣٦	المعجم القائم على التعليقات القوية مع مصادر الكلمات مع إستخدام طريقة التجميع.

جدول ه : نتائج تصنيف التعليقات بإستخدام (م ك أ ق ت أ) المبني على التعليقات القوية مع إعتبار مصادر الكلمات كسمة لغوية وإستخدام طريقة الجمع.

٥ الخاتمة

معاجم الآراء تعتبر واحدة من المصادر القيمة المستخدمة لتحليل آراء الأشخاص. لذلك إقترحنا إستخدام تعليم الآلة القائم على السمات اللغوية للتعليقات لبناء معجم كلمات آراء. حيث يقدم هذا المعجم المقترح تحسينات تفوق المعاجم المنشورة سابقاً. كما نقدم أيضاً تحسينات أخرى لبناء المعجم بإستخدام التعليقات القوية مع إستخدام مصادر الكلمات كسمة أساسية تستعملها مكائن الدعم الإتجاهي. كذلك تم تقديم تحسين آخر على دقة تصنيف التعليقات بإستخدام طريقة (تجميع قيم الألفاظ) كطريقة حسابية. حيث أدت كل هذه التحسينات إلى دقة قدرها ٧٩.٣٦% و ٧٧.٨% لحزم نصوص الأمازون والأفلام على الترتيب.

المراجع

- [1] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. Proceedings of EMNLP, 2002.
- [2] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Proceedings of HLT/EMNLP, Vancouver, Canada, 2005.
- [3] Franco Salvetti, Stephen Lewis and Christoph Reichenbach. Automatic Opinion Polarity Classification of Movie Reviews. Colorado Research in Linguistics. Volume 17, Issue 1 June 2004.
- [4] Alistair Kennedy and Diana Inkpen. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. Computational Intelligence, Vol. 22, 110–125, 2006.
- [5] B. Ohana and B. Tierney. Sentiment Classification of Reviews Using SentiWordNet. 9th. IT&T Conference, Dublin Institute of Technology, Dublin, Ireland, 22nd.-23rd. October, 2009.
- [6] Horacio Saggion and Adam Funk. Interpreting SentiWordNet for Opinion Classification. Proceedings of the Seventh conference on International Language Resources and Evaluation LREC10, 2010.
- [7] Adam Funk, Yaoyong Li, Horacio Saggi, Kalina Bontchevaon and Christian Leibold. Opinion Analysis for Business Intelligence Applications. Proceedings of the first international workshop on Ontology-supported business intelligence, 2008.
- [8] Andrea Esuli and Fabrizio Sebastiani, SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. Proceedings from International Conference on Language Resources and Evaluation (LREC), Genoa, 2006.

- [9] <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.htm>
- [10] Hamish Cunningham, Diana Maynard, Kalina Bontcheva, and Valentin Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), 2002.
- [11] <http://www.cs.cornell.edu/~people/pabo/movie-review-data>
- [12] Alaa Hamouda and Mohamed Rohaim. Reviews Classification Using SentiWordNet Lexicon. World Congress on Computer Science and Information Technology, January, 2011.

الملخص باللغة الانجليزية

Title: Building Machine Learning Based Senti-word Lexicon for Sentiment Analysis

Abstract. Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative". One of approaches used to make sentiment classification is using sentiment lexicon. This paper aims to build a sentiment lexicon which is domain independent. We propose a Machine Learning Based Senti-word Lexicon (MLBSL) based on reviews from different domains. Our proposed MLBSL yields an improvement over previous published manual and automatic-built lexicons like SentiWordNet. We also provide an improvement in calculation method used in reviews sentiment analysis.

Keywords: Sentiment Analysis, Sentiment Lexicon, Customers Reviews, Recommender Systems

المصطلحات

Bag-of-Words	حقيبة كلمات
Bigram	ثنائي الوحدة
Blog	مدونة
Corpora	حزم نصية
Corpus	حزمة نصية
Customers Reviews	تعليقات العملاء
Document polarity	ميل الوثيقة
General Architecture for Text Engineering tool	أداة الهيكل العام لهندسة النص
Machine Learning	تعليم الآلة

Morphological Analysis	التحليل الصرفي
N-grams	ن-وحدة
Products Reviews	التعليقات الخاصة بالمنتجات
Recommender Systems	أنظمة التوصيات
Reviews	تعليقات
Sentiment Analysis	تحليل الآراء
Sentiment Classification	تصنيف الآراء
Sentiment Lexicon	معجم كلمات الآراء
SentiWordNet	قاموس كلمات المشاعر
Term	لفظة
Term weight	وزن اللفظة
Unigram	أحادى الوحدة
WordNet	قاموس الكلمات



الجلسة الثالثة عشرة

تتبع سريع للأشياء في سلاسل الصور الطبية باستخدام الحدود النشطة

عبدالقادر فقير¹، نصيرة بن عمران²

¹ جامعة معسكر، قسم الرياضيات والإعلام الآلي

ص.ب. 763، الرمز البريدي 29000، طريق المامونية، معسكر، الجزائر

² جامعة العلوم والتكنولوجيا وهران، قسم الإعلام الآلي، ص.ب. 1505 المناور، الرمز البريدي 31000، وهران، الجزائر

aekefir@univ-mascara.dz aekefir@gmail.com nabenamarne@yahoo.com

الخلاصة: نقدم في هذه الورقة العلمية طريقة لتجزئة سلاسل الصور الرقمية الطبية وذلك باستخدام نموذج الحدود النشطة. فكرتنا تقوم على استخدام دالة متغيرة لقوى الحد النشط بدل الدالة الثابتة المعروفة في النموذج. في المرحلة الأولى، نعرف مجموعة قوى الحد النشط من أجل استخراج محيط الشيء المراد تتبعه، ثم بعد ذلك نقوم بإقحام قوة جديدة لتغيير دالة القوى الأولى. الدالة الثانية تمكننا من تتبع سريع للشيء محل الاهتمام. لزيادة سرعة الطريقة قمنا باقتراح تنفيذ متعدد المستويات يميز مختلف لنموذج الحدود النشطة. فعالية الفكرة المقدمة أثبتت بمجموعة من النتائج التجريبية وذلك باستخدام مجموعة من سلاسل صور الفيديو: سلسلة صور لخلية حية متحركة، سلسلة صور جراثومة وسلسلة صور صدى القلب .

الكلمات الجوهرية: الحدود النشطة، تجزئة الصور الطبية، سلاسل صور فيديو طبية، التتبع السريع للأشياء، الثعابين، ميز متعدد

1 مقدمة

مما مكن منه الإنسان في العقود الأخيرة هو هذه الخطوة الهائلة في التحول الآلي لمختلف النشاطات اليومية. من ذلك ما أصبح يسمى الإبصار الاصطناعي و الإبصار بالحاسوب.

فقد أصبح استخدام أنظمة الإبصار واسع شمل العديد من الميادين مثل الميدان الصناعي (مراقبة النوعية مثلا)، الميدان الأمني (التعرف على هويات الأشخاص ومراقبة الأماكن العامة) الميدان الطبي (إنتاج الصور ثلاثية الأبعاد في بعض تقنيات التصوير الطبي مثلا).

تعتبر تجزئة الصور عملية أساسية في أنظمة الإبصار وفي ميدان تحليل الصور الرقمية بصفة عامة إذ أنها تمثل حجر الأساس لكل العمليات اللاحقة العالية المستوى التي تتعامل مع المعطيات الرمزية الممثلة لأجزاء لصورة. تشمل تجزئة الصور تقنيات عديدة يمكن جمعها في قسمين [1]:

الاول: اكتشاف حدود الأشياء وحواها باستغلال عدم استمرارية مستويات الألوان أو مستويات الرمادية أو غيرها من الخصائص.

الثاني: استخراج الأجزاء المتجانسة المتماثلة التي لها خصائص متشابهة

نجد أن نموذج الحدود النشطة المقترح من طرف كاص وفريقه [2] إحدى طرق تجزئة الصور التي عرفت استخداما واسعا في العقدين الأخيرين من أجل استخراج حواف الأشياء بشكل عام وفي تجزئة الصور الطبية بشكل خاص [3][4].

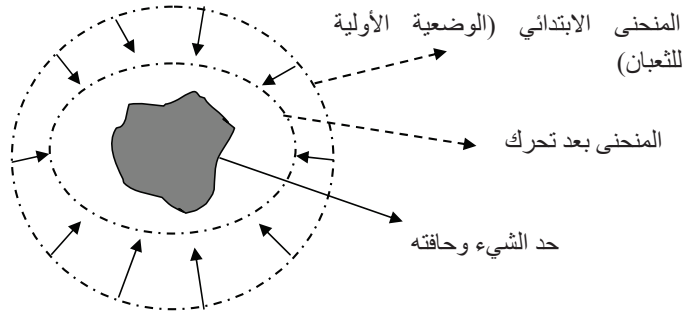
الطبيعة الحيوية والنشطة لحدود النموذج ابتداء من الوضعية الأولية للحد إلى غاية الوضعية النهائية جعلت المختصين يطلقون عليها إسم الثعابين إذ أن تحركها يشبه تحرك الثعابين [5].

في هذه المقالة نقترح طريقة لاستخراج وتتبع الحدود في سلسلة صور رقمية وذلك باستعمال نموذج الحدود النشطة أو الثعابين. مقالتنا منظمة كما يلي: نجد في الفقرة الثانية توضيحا لنموذج الحدود النشطة. في الفقرة الثالثة نعطي تفصيلا للطريقة المقترحة. مجموعة من النتائج التجريبية ومناقشتها ستكون في الفقرة الرابعة والتي تكون متبوعة بخاتمة للمقالة.

2 نموذج الحدود النشطة

2.1 المبدأ العام

نموذج الحدود النشطة أو الثعابين هو بكل بساطة خط أو منحنى معلّم يكون مغلقا ويكون مفتوحا ذا طرفين. فكرة الثعابين تقوم على وضع الخط بجوار الشيء المراد استخراج حدوده. يتغير شكل هذا الخط وحجمه تحت تأثير قوى تدفعه أو تجذبه إلى حواف الشيء المراد كما هو مبين في الشكل الأول.



الشكل الأول. تحرك وتطور الحد النشط

هذه القوى مقسمة إلى ثلاث أنواع مرتبطة بالثعابين [5]:

أ. طاقة ذاتية تتعلق بشكل الحد النشط وتسمى طاقة داخلية

ب. طاقة افتراضية مصدرها الصورة تسمى طاقة خارجية. هذه الطاقة هي التي تقوم بدفع منحني الثعبان إلى حدود الأشياء الموجودة في الصورة

ج. طاقة السياق والتي من خلالها يمكن للمستعمل إضافة بعض الشروط الخاصة للنموذج حسب الصور والتطبيقات وغيرها

بالنسبة للصورة الرقمية، الحد النشط هو مجموعة من النقاط الممثلة لمنحناء، الموضوع ابتداء بجوار الشيء المراد تعيين حوافه، تُحرَّك هذه النقاط حتى تستقر عند حدود وحواف الشيء المقصود ، وتحركها هذا يتم باستخدام طريقة لتخفيض قيمة الطاقة المرفقة بالحد النشط.

2. 2 طاقات الحدود النشطة

دالة طاقات الحد النشط هي تركيبة لمجموعة من الطاقات التي تم الإشارة إلى أنواعها ، وهنا نضع مزيد توضيح لها :

1. 2. 2 الطاقة الداخلية

الهدف من هذه الطاقة هو تسيير الحالة الذاتية للثعبان، بحيث تكون متماسكة. لذلك فهذه الطاقة معرفة بمجموع قوتين:

أ. قوة الاستمرارية والتي تسمح بالمحافظة على البعد المتساوي بين نقاط الثعبان

ب. قوة الإنكسارية والتي تجنب الحصول على نقاط معزولة والتي لن تكون منسجمة مع الشكل

2. 2. 2 الطاقة الخارجية

هذه الطاقة خارجية بالنسبة للحدود النشطة لأنها لا تتعلق بالخصائص الذاتية لها وإنما تأخذ بعين الاعتبار خصائص الصورة المعالجة كاللون والرمادية وغيرها. لكن أهم خصائص الصورة المستعملة والتي وظفناها في بحثنا هذا هي درجة الميل والتي تمثل المشتقة الأولى للصورة. طاقة الميل لها أهمية بالغة في إستقرار الحد النشط على حواف الشيء المقصود لأن هذه الحواف تتميز بقيمة قصوى للفروق بين رمادية النقاط الصورية المتجاورة (أي الميل).

3. 2. 2 طاقة السياق

كما أشرنا هذا النوع الاختياري من الطاقة يسمح بإدخال خصائص ومعارف مسبقة للشيء المراد في الصورة. سنستعمل في بحثنا طاقتين اثنتين

أ. طاقة النفخ المقترحة من طرف كوهين[6]. تسمح هذه الطاقة بتحديد إتجاه انتشار نقاط الحد النشط إلى الداخل او إلى خارجه مثل كرة الأطفال التي تنفخ فتتحرك أجزاؤها إلى الخارج وبالعكس.

ب. طاقة الإتجاه والمقترحة من طرف [7] . بعكس الأولى هذه الطاقة تمكننا بتعريف إتجاه واحد لحركة نقاط الشعبان يمينا يسارا وغيرها من زوايا الإتجاهات.

باستعمال تركيبة طاقات الحد النشط، يمكن أن نتناول مسألة تخفيض لهذه الطاقة الكلية. حل هذه المسألة (تخفيض الطاقة) هو حدود وحواف الشيء المراد. توجد عدة تنفيذات لحل هذه المسألة وفي عملنا هذا قمنا باستخدام خوارزمية الشره [8]

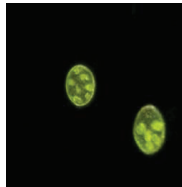
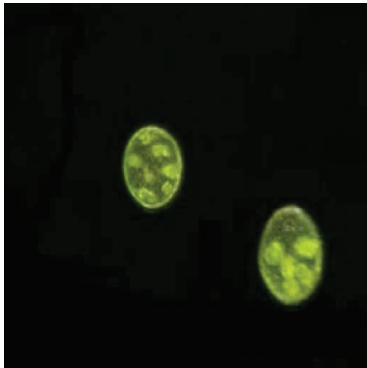
3 شرح الطريقة المقترحة

يمكن أن نقسم طريقتنا المقترحة في هذه الورقة إلى المراحل التالية:

- أ. تحضير سلاسل الصور ذات ميز متعدد
- ب. استخراج حواف الشيء المراد باستخدام الحد النشط المختار
- ج. تغيير دالة الطاقة للحد النشط وتتبع الشيء باستخدام الحد النشط بدالة الطاقة الجديدة

3. 1 الميز المتعدد لسلاسل الصور

في هذه العملية، نقوم بتصغير حجم الصور إلى مستويات متعددة: أول، ثاني، حتى نصل إلى المستوى المطلوب. الصور الأصلية تمثل المستوى رقم صفر. في عملنا هذا، كل نقطة صورية من صورة المستوى الأول تمثل أربع نقاط من الصورة الأصلية. وعليه مستوى رماديتها يساوي معدل مستوى رماديات النقاط الأربع. إذن الصورة الأصلية تنقسم إلى النصف طولاً وعرضاً للحصول على صورة المستوى الأول التي سيكون حجمها يساوي ربع الأصلية. وهكذا حتى الوصول إلى المستوى الأعلى المطلوب. (انظر الشكل 2).



أ. المستوى الثاني

صورة 128×128 نقطة

ب. المستوى الأول

صورة 256×256 نقطة

ج. المستوى الأصلي:

صورة 512×512 نقطة

الشكل 2. ثلاث مستويات لصورة بميز مختلف

كل عمليات معالجة الحد النشط ستتم في المستوى الأعلى، وبعد استقرار الثعبان نقوم بنشر النتائج (عرض الحواف المستخرجة) إلى المستوى الأدنى.

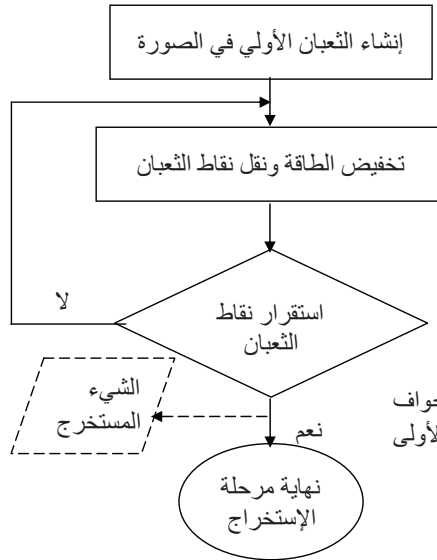
هذه العملية لها فائدتان :

أ. سرعة التنفيذ: فبذل التحرك في مساحة 512×512 مع ما يصاحب ذلك من حسابات تصغير دالة الطاقة (نفصلها فيما بعد) نكتفي بالتحرك في مساحة 64×64 أو 128×128 للقيام بنفس المهمة.

ب. تسمح بإزالة بعض آثار التشويش و الغيش الموجود في الصورة والذي يمكن من أن يعرقل حركة الثعبان.

3. 2 استخراج حواف الشيء المراد

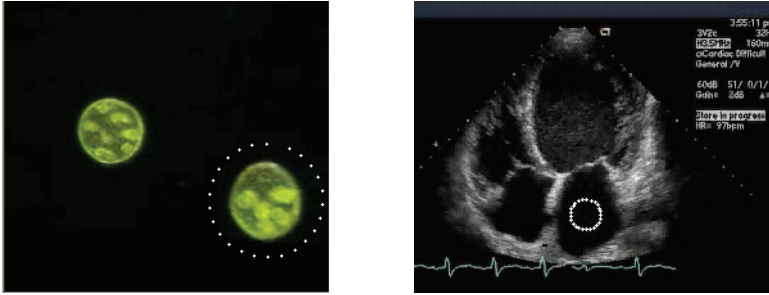
يمكننا أن نلخص هذه المرحلة في الشكل الثالث



الشكل 3. استخراج حواف الشيء في الصورة الأولى من السلسلة

3. 2. 1 إنشاء الثعبان الأولي

يتم إنشاء الثعبان بداخل الشيء المراد أو حوله محيطا به. لذلك قمنا بتعيين مجموعة نقاط ثعبانية على شكل دائرة (انظر الشكل الرابع).



الشكل الرابع. حالتى إنشاء الثعبان الأولي (في داخل الشيء المراد أو حوله)

يكفينا لتحقيق ذلك إدخال مركز الدائرة ونصف قطرها وعدد النقاط الثعبانية. إحداثيات النقاط يمكن حسابها بالصيغة رقم 1

$$\begin{aligned} x_i &= x_c - r * \sin(2\pi i/N) \\ y_i &= y_c - r * \cos(2\pi i/N) \end{aligned} \quad i=0..n-1 \quad (1)$$

بحيث : x_i, y_i إحداثيات النقطة الثعبانية رقم i

x_c, y_c إحداثيات مركز الدائرة

r هو نصف قطرها

N هو عدد النقاط الثعبانية المستعملة

3. 2. 2 الحد النشط المستعمل لاستخراج الحواف

قمنا باستخدام الحد النشط المَعْلَمي، ودالة الطاقة المرفقة يمكننا تعريفها بالصيغة رقم 2

$$E_{snake} = \sum_{i=1, N} (a E_{continuity}(P_i) + b E_{curvature}(P_i) + c E_{gradient}(P_i) + d E_{Balloon}(P_i)) \quad (2)$$

$P_i = (x_i, y_i)$ هي النقاط الثعبانية $i=1..N$

a, b, c, d هي المعاملات المرفقة بكل طاقة

لمزيد من تفاصيل صيغة كل طاقة وعبارتها يمكن الرجوع إلى عملنا السابق [9]

لكل نقطة ثعبانية، نأخذ مجموعة من النقاط الصورية المجاورة لها (في ورقتنا هذه استعملنا جوار 3×3) ونقوم بحساب الطاقة لكل هذه النقاط الجارة. النقطة الجارة التي لها أضعف قيمة للطاقة تمثل الموضع الجديد للنقطة الثعبانية.

استعملنا خوارزمية الشَّره خوارزمية لتخفيض الطاقة (انظر الخوارزمية رقم 1).

Algorithm 1: Greedy Algorithm

BEGIN

Repeat

for all points of snake do

Begin

for all points of neighborhood do

Calculate the energies

for all points of neighborhood do

Normalization

Minimize to get the new position of snake (point of snake)

End

Until Stop Criterion

عامل التوقف المذكور في الخوارزمية رقم 1 هو استقرار النقاط الثعبانية الموجود كشرط في الشكل 3. إنه يمثل نسبة النقاط الثعبانية المستقرة خلال ثلاث تكرارات تعاقبية.

3.2.3 عرض النتيجة في المستويات الدنيا

عند نقطة الاستقرار السالفة الذكر في الخوارزمية رقم 1، يكون الحد النشط متطابقا مع حواف الشيء المراد في الصورة الأولى من السلسلة، وبما أن العمليات تتم في المستوى الأعلى (الصورة الأصغر حجما) نقوم بنشر النتيجة إلى المستويات الدنيا.

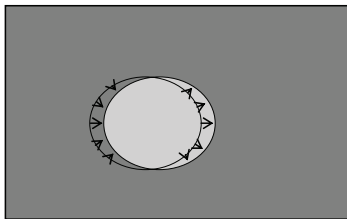
مما سبق نعلم أن كل نقطة ثعبانية هي في الحقيقة نقطة صورية في المستوى الأعلى إذن فهي تمثل أربع نقاط صورية من المستوى الأدنى المباشر. فلا بد من اختيار نقطة واحدة من المستوى الأدنى لتمثل النقطة الثعبانية هناك. تحديد ذلك يتم بانتقاء النقطة من النقاط الأربعة التي تملك أضعف قيمة للطاقة المعرفة في الصيغة 2 كما مر معنا في الخوارزمية 1. عملية نشر النتائج هذه تتكرر حتى نصل إلى المستوى الأدنى أي الصورة الأصلية.

3.3 تغيير دالة الطاقة للحد النشط وتتبع الشيء المراد

بعد التعرف وتحديد حواف وحدود الشيء المراد تتبعه في الصورة الأولى من السلسلة، قمنا بتأهيل ألي للحد النشط في الصورة اللاحقة. اعتبرنا في عملنا هذا أن الحد النشط النهائي (أي الممثل لحواف الشيء المراد) في الصورة المعالجة هو الثعبان الأولي في الصورة اللاحقة.

لكن هذه الطريقة الآلية للتأهيل مع الطاقة المستخدمة في المرحلة السابقة (الصيغة 2) تطرح لنا إشكالا: الشيء محل الإهتمام سيتحرك يمينا أو شمالا أو غيرها من الإتجاهات، مع إمكانية تغيير حجمه وشكله (انظر الشكل الخامس). وطاقة النفخ المستعملة في دالة الطاقة تسمح باتجاه واحد لكل النقاط الشعبانية: إلى الداخل أو إلى الخارج، بينما في الحالة المذكورة سيصبح دور هذه الطاقة غير مجديا هنا من أجل تتبع مباشر للشكل محل الإهتمام. لذلك اقترحنا هنا اقحام بدل طاقة النفخ طاقة جديدة تسمى طاقة الإتجاه. تفاصيل صيغتها نجدها في [7]

مع الاحتفاظ بنفس الدالة وباستعمال طاقة النفخ، هناك حل آخر لهذه المعضلة اقترحناه في ورقتنا السابقة [9]



الشكل الخامس. الاتجاه المرجو
لنقاط الشعبانية يؤكد محدودية طاقة
النفخ في مرحلة التتبع

4 النتائج التجريبية ومناقشتها

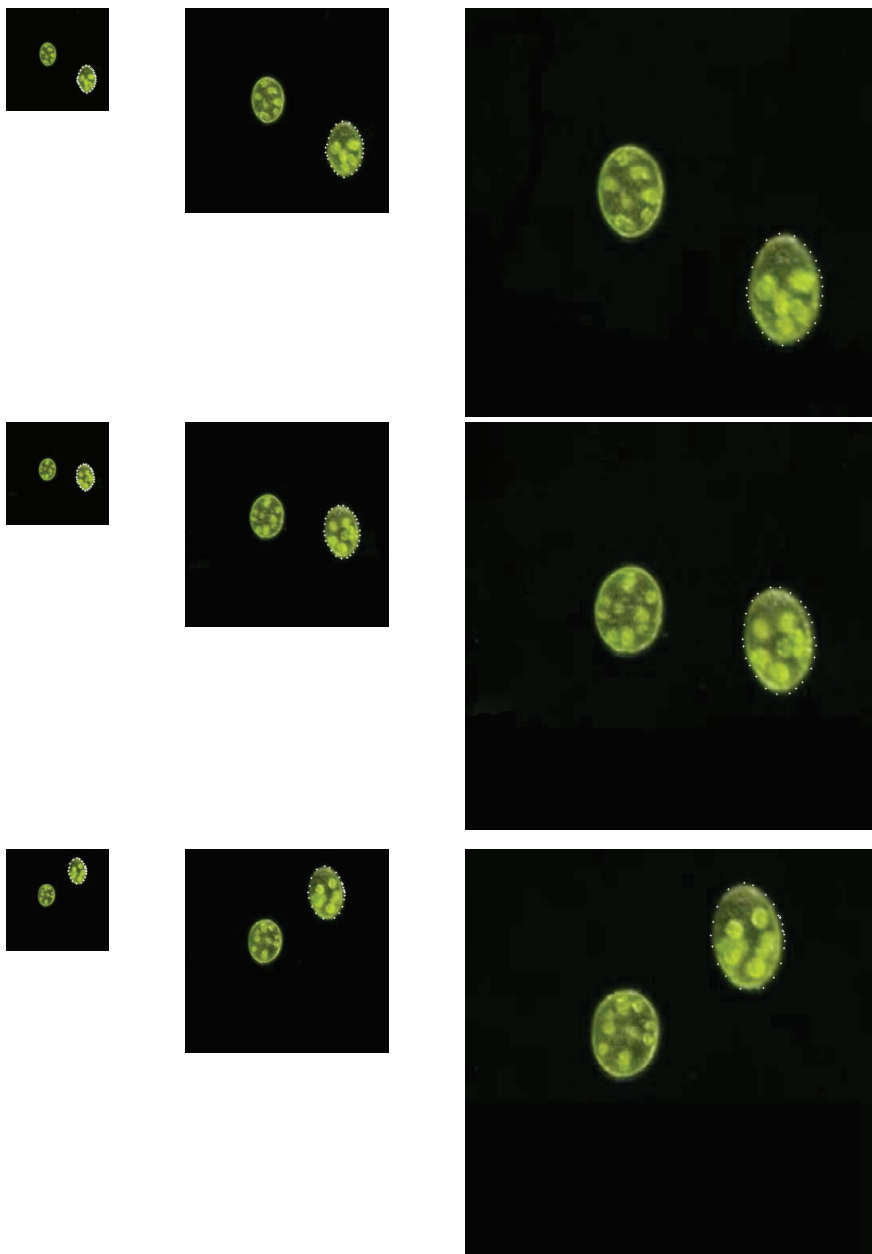
قمنا بتجربة الطريقة المقترحة باستخدام مجموعة من سلاسل الصور، ونكتفي في هذا المقال بعرض ثلاثة منها: سلسلة الخلية الحية المتحركة، سلسلة صور الجرثومة وسلسلة صور صدى القلب.

بعد عدة تجارب، قمنا بضبط مَعلَومات النموذج لكل سلسلة ويمكن اختصارها في الجدول الأول. ونتائج التجزئة ممثلة في الأشكال 6 ، 7 و 8 التي أعطينا فيها نماذج لصور النتائج من أول، وسط وآخر السلسلة. الحد النشط النهائي في كل صورة ممثل فيها بمجموعة من النقاط كما هو ملاحظ. من أجل تناسب صور الخلية المتحركة مع حجم المقالة تم تقليص أبعادها إلى النصف في إظهارها وعرضها هنا لا في التجارب. كما أنه تعمدنا إضافة بعض التشويش اليدوي لسلسلة صور الجرثومة كما هو واضح. الجرثومة حركتها جد بطيئة في السلسلة لكن كان لها تغيير شكل أسرع وبعض أشكالها المتقاربة متكررة في السلسلة فكان هدفنا إذن تتبع جسمها المتغير الشكل. في هاتين السلسلتين كان إنشاء الشعبان الأولي حول الجسمين و خارجهما، بينما في سلسلة صور صدى القلب كان الشعبان الأولي داخل التجويف (البطين) والذي كانت حركته دورية تقريبا كما هو معلوم.

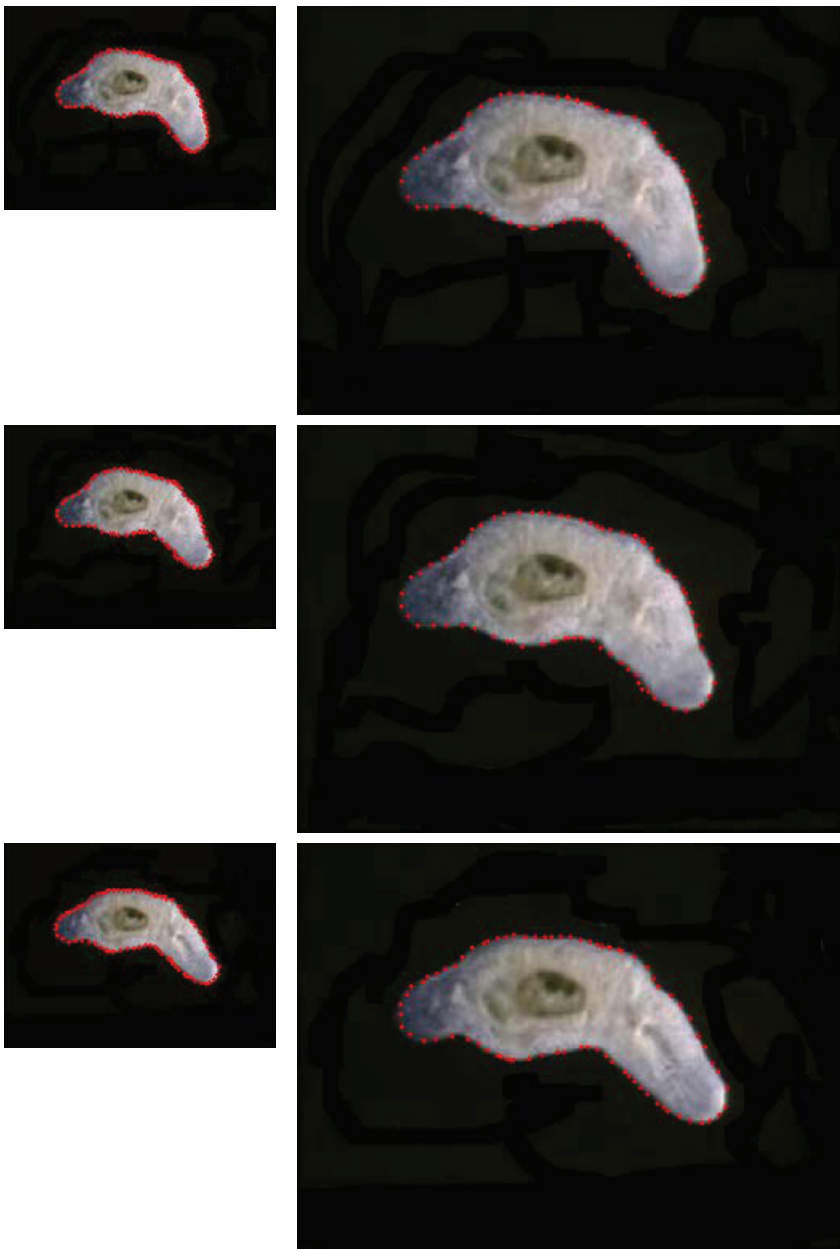
الجدول الأول. مَعلَومات التجارب

المعلومات	سلسلة الخلية	سلسلة الجرثومة	سلسلة القلب
المعلومات a, b, c, d	0.1, 0.5, 1, 0, 1	0.1, 1, 1, 0, 1	0.1, 0.5, 1, 0, 1
عدد النقاط الشعبانية	25	80	36

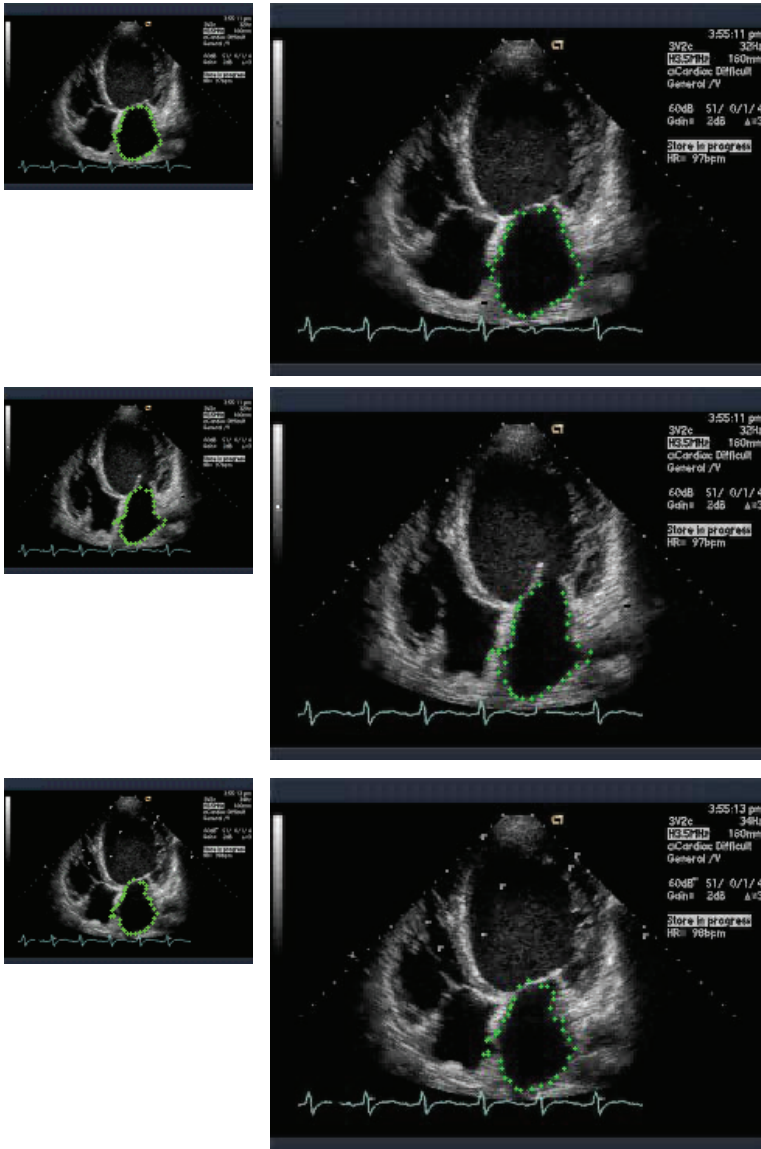
240×320	255×340	512×512	أبعاد الصور الأصلية
---------	---------	---------	---------------------



الشكل السادس. نتائج تتبع الخلية المتحركة باستعمال ثلاثة مستويات بميز متعدد



الشكل السادس. نتائج تتبع الجرثومة باستخدام مستويين



الشكل السادس. نتائج تتبع حركة البطين في سلسلة صور صدى القلب باستعمال مستويين

لاحظنا في الجملة أنه تم تحديد حواف الشيء المراد في الصورة الأولى ومن ثم تتبعه في بقية صور السلاسل الثلاثة. بالنسبة للسلسلتين الأوليتين (الخلية المتحركة والجراثومة) السلسلة كانت بسيطة لذلك تمت عملية التجزئة (تحديد حواف الشيء المراد) في كامل السلسلة بدقة جيدة. التشويش الموجود في سلسلة الجراثومة تم تجاوزه برفع معامل طاقة الإستمرارية مقارنة بالسلسلتين إضافة إلى معامل الطاقة الخارجية الذي كان في أعلى قيمه في كل السلاسل.

في السلسلة الثالثة (صور صدى القلب) تم إجراء عدة تجارب لضبط مَعْلَمَات النموذج. رغم ذلك نلاحظ أن بعض النقاط الثعبانية لم تكن في الموضع المطلوب. يرجع ذلك إلى عاملين أساسيين. الأول منهما هو نوعية الصور فهي أكثر تشويشا ومهما استعملنا المعالجات الأولية لتحسينها ستبقى نسبة التشويش فيها معتبرة. الأمر الثاني هو نسبة النقاط الثابتة المختارة كعامل للتوقف (انظر الشكل الثالث و الخوارزمية الأولى). غالبا ما نختار نسبة 90% إلى 95%. فهناك مجموعة من النقاط إذن لن تؤخذ بعين الإعتبار في استقرار الثعبان (نقطتان إلى ثلاث نقاط في النتائج المعروضة هنا). إذن حتى ولو لم تكن في الموضع المطلوب سنتجاوزها ونعتبر الثعبان مستقرا. في الجهة المقابلة قمنا في بعض التجارب برفع النسبة إلى 100% لكن لاحظنا أن ذلك قد رفع من زمن التنفيذ بنسبة مهمة مع أن النتيجة لم تتحسن بالقدر المطلوب مقارنة بالزيادة في الوقت وذلك يرجع إلى العامل الأول المذكور.

بالنسبة للمقارنة بين تجزئة سلاسل الصور الأصلية مباشرة وطريقتنا هذه التي استخدمنا فيها المستويات المتعددة لسلاسل الصور، لاحظنا أن نوعية التجزئة (تحديد الحواف والتتبع) كانت متماثلة في كل السلاسل بينما تم تقليص زمن التنفيذ إلى نحو 80% باستعمال المستويات المتعددة (أدخلنا في زمن التنفيذ بالنسبة للطريقة المعروضة في هذه المقالة تجزئة الصور ونشر النتائج أيضا).

5 الخاتمة

اقترحنا في هذه المقالة طريقة لتجزئة سلاسل الصور الطبية وذلك باستخدام نموذج الحدود النشطة. قمنا أولا بتحديد حواف الشيء المراد ومن ثم تتبعه في باقي صور السلسلة. اقترحنا يعتمد على ثلاث أسس: الأول طريقة الإنشاء الآلي للثعبان الأولي في كافة صور السلسلة ماعدا الصورة الأولى. الثاني تغيير دالة طاقات الحد النشط التي كانت ثابتة في كل ما اطلعنا عليه من أبحاث متقدمة، بينما هنا اعتمدنا على تغيير ذلك من مرحلة تحديد الحواف إلى مرحلة التتبع. لذلك قمنا بإدراج طاقة الإتجاه المعتمدة على الفارق بين اللقطتين المتتابعتين. من إيجابيات هذه الطاقة إزالة تأثير التشويش الثابت في الصورة نتيجة للفارق المستعمل. الثالث: استعمال تعدد المستويات من أجل تسريع العملية والتقليل من تأثير التشويش الذي يضمحل نسبيا في المستوى الأعلى ذو الميز الأدنى. تم تجربة الطريقة باستخدام عدة سلاسل صور. قمنا هنا بعرض نتائج سلسلتين بسيطتين (الخلية الحية والجراثومة) والسلسلة الطبية المقصودة (صدى القلب). كانت النتائج مرضية وثبت الهدف المنشود

بعد إعداد مقارنة زمنية بين هذه الطريقة المقترحة هنا وطريقتنا السابقة [9]، إثبتنا أن التتبع كان سريعاً هنا إلى حدود الضعفين وأكثر ولم تسمح حجم المقالة بزيادة تفاصيل المقارنة.

كمشاريع مستقبلية، إننا بصدد اقتراح طاقة جديدة، بدل طاقة الإتجاه المستعملة هنا، تساعد على تتبع أسرع.

يمكن أيضاً تسريع الطريقة باقتراح المعالجة المتوازية للعمليات الثلاثة: إنشاء سلاسل صور المستويات المتعددة، التجزئة في المستوى الأعلى ونشر النتائج إلى المستوى الأدنى. كما أننا نقترح إدخال نظام متعدد المتعاملين كل ذلك من أجل تجزئة وتتبع أي الأشياء.

6 شكر وتقدير

نتقدم بالشكر لكل من جوتسي و بن والي للمساهمة في تنفيذ التجارب.

المراجع

- [1] Coquerez JP, Philipp S. Analyse d'images: filtrage et segmentation. Edition Masson, Paris, 1995.
- [2] M. Kass, A. Witkin and D. Terzopoulos, « Snakes : Active contour models » International journal of computer vision, vol. 55, pp 321-331 , 1988
- [3] T. McInerney and D. Terzopoulos, « Deformable Models in Medical Image Analysis: A Survey » Medical Image Analysis, 1(2), pp 91-108, 1996
- [4] L. He et al. « A comparative study of deformable contour methods on medical image segmentation. » Image and Vision Computing 26, pp 141-163, 2008
- [5] J. Rousselle : Les contours actifs, une méthode de segmentation. Application à l'imagerie médicale. Thèse de doctorat, Université de François Rabelais de Tours, Soutenue le 9 juillet 2003, France
- [6] L. Cohen : On active contour models and balloons. Computer vision, graphic, and image processing : Image Understanding, Vol. 53, N° 2, pp 211-218, Mars 1991.
- [7] B. Lee, I. Choi et G. Jeon : Motion-based moving object tracking using an active contour. IEEE International Conference on Acoustics, Speech and Signal Processing, pp 649-651, May 2006
- [8] D. Williams et M. Sham : A fast algorithm for active contour and curvature estimation. Computer vision graphic image process : Image understanding, vol. 55, n°1, pp 14-26, 1992.
- [9] A. Fekir, N. Benamrane, A. Taleb-Ahmed, Détection et suivi d'objet dans une séquence d'images échocardiographique par les contours actifs, CIIA, May 2009, Saïda, Algeria.

المخلص باللغة الانجليزية

Title: Fast object tracking in medical image sequence using active contours

Abstract. This paper presents approach of medical image sequence segmentation based on active contour models or snakes. Active contours have usually permanent energy function. In our approach, we propose snake with changeable energy function. Hence,

we define combination of forces in first step to extract object in first frame. Then, we introduce a new force in energy function to track quickly our object in other sequence frames. Also, we propose a multi-resolution implementation of this snake to accelerate our approach. The efficiency of our approach is shown through some experimental results using three image sequences: moving biological cell sequence, microbe sequence and echocardiograph image sequence.

Keywords: Active contours, medical image segmentation, multi-resolution, medical image sequence, fast tracking, snakes

المصطلحات

حرصنا على أن يكون ظهور المصطلحات في الجدول حسب أول ظهور لها في النص.

تتبع	Tracking, track
تتبع سريع	Fast tracking
تتبع الأشياء	Object tracking
سلسلة صور	Image sequence
الحد النشط	Active contour
نموذج الحدود النشطة	Active contour model
الحد النشط المعلمي	Parametric active contour
تجزئة الصورة	Image segmentation
دالة القوى	Force function
استخراج المحيط	Contour extraction
الشيء المراد ، المقصود، محل الإهتمام	Object of interest
تنفيذ	implementation
تعدد المستويات	Multi-level, multiscale
ميز ، ميز متعدد	Multi-resolution
صورة صدى القلب	Echocardiography, Echocardiograph image
الثعابين	snakes
الإبصار الإصطناعي	Artificial vision
الإبصار بالحاسوب	Computer vision
نظام الإبصار	Vision system
تحليل الصور الرقمية	Image analysis
الحد، الحاف، المحيط	Edge, contour

عدم استمرارية	Discontinuity
مستويات الرمادية	Gray-levels, grayscale
الأجزاء المتجانسة	Homogeneous region
منحنى مَعْلَمِي	Parameter Curve
المنحنى الابتدائي، الأولي	Initial curve
طاقة داخلية، خارجية، السياق	Internal, external, context energy
تخفيض الطاقة	Energy minimization
تركيبية	Combination
طاقة الإستمرارية	Continuity energy
طاقة الإنكسارية	Curvature energy
النقطة الصُّورِيَّة	Pixel (P icture e lement)
نقطة الثَّعبان، نقطة ثَّعبانية	Point of snake (snaxel : snake element)
الميل	gradient
طاقة النفخ	Balloon energy
طاقة الإتجاه	Direction energy
خوارزمية الشَّره	Greedy algorithm
معالجة	Treatment
التشويش، الغيش	Noise
استقرار	Stability
صيغة	Formula
تكرارات تعاقبية	Iterations
أبعاد الصورة	Image dimensions
نشر النتيجة	Result propagation
تأهيل آلي	Automatic initialization
الثَّعبان الأولي	Initial snake
مَعْلَمَات	Parameters
زمن التنفيذ	Runtime
الفارق بين اللقطتين	Frame difference
مقارنة زمنية	Temporal comparison
المعالجة الموازية	Parallel treatment
نظام متعدد المتعاملين	Multi agent system
تتبع أي	Tracking in real time

تطبيقات معالجة الإشارات الرقمية في مجال التحليل الطيفي للمتسلسلات الجينية

د. أحمد علي رشدي

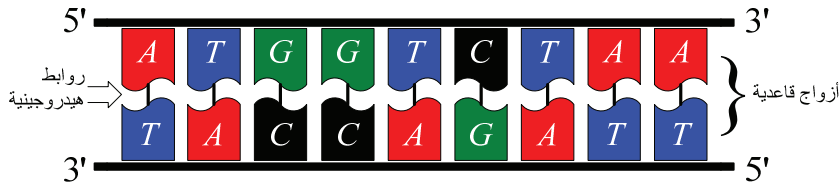
P.O.Box 5542
Santa Clara CA 95056 USA
aarushdi@ieee.org

الخلاصة. برز علم المعلوماتية الحيوية مؤخراً كعلم هجين يسعى لاستشراف أفاق جديدة من علوم الأحياء والطبيعة بمشتقاتها المختلفة. يستخدم هذا العلم برمجيات الحاسوب وطرائق الهندسة الرصينة بغية زيادة سرعة عمليات البحث وتعظيم كفاءتها. فلي بحثنا هذا نتناول جانباً من جوانب هذا العلم. تحديداً ندرس ما يتعلق بتطبيقات علم معالجة الإشارات الرقمية لاستكشاف خصائص طيف الدنا والربط بينها وبين صفات جينية. نقدم أولاً عرضاً لأبرز طرائق مقابلة أبجدية الدنا الرمزية بأبجديات عديدة يمكن التعامل معها بطرائق معالجة الإشارات الرقمية. ثانياً نعرف طيف الدنا ونستنتج نموذجاً رياضياً لحسابه يعتمد على مبادئ معالجة الإشارات متعددة المعدلات. ثالثاً نقدم بعض النتائج التجريبية لتعضيد الحسابات النظرية. ونخلص أخيراً إلى مقترحات مستقبلية للبحث في هذا المجال الواعد.

كلمات مفتاحية – معالجة الإشارات الرقمية – التحليل الطيفي – معالجة إشارات الجينوم – تحليل المتسلسلات الجينية – مقابلة الأبجديات الرمزية بالعديدة – طيف الدنا.

1. مقدمة

الدنا أو الحامض الخلوي النووي كما قدمه الأمريكي جيمس واطسون والبريطاني فرانسس كريك [1] هو مخزن معلومات طويل المدى يحتوي على كافة المعلومات الحيوية والجينية الكفيلة بالتوصيف الدقيق لكيفية تشكل الكائن الحي في أطواره المختلفة وكيفية قيامه بوظائفه الحيوية. يُعرف الحامض الخلوي النووي بشريطيه المتوازيين الملتقيين حول بعضهما ككولب مزدوج في شكل بدعي الانتظام. ويمتد هذا الالتفاف بطول الشريطين حتى النهايتين المرموز لهما بـ (5') و (3'). كل نهاية لأحدى الشريطين تقابلها النهاية الأخرى على الشريط الآخر. الشريط الذي يبدأ بـ 5' يُعرف باسم الشريط السابق أو المتسلسلة السابقة أما الشريط المتلازم معه فيُعرف باسم الشريط اللاحق أو المتسلسلة اللاحقة. يمكن توصيف شريطي الدنا كمتسلسلات حرفية تنتمي وحداتها البنائية إلى أبجدية القواعد النووية $F = \{A, C, G, T\}$ س. ج. ث أو $\mathbb{F} = \{A, C, G, T\}$ ، التي تشمل أربع قواعد تنقسم إلى نوعين هما البيورينات وتشمل الأدينين: A والجوانين: G ، والبايريميدينات وتشمل السيتوزين: C والثيمين: T . يرتبط كل نوع من القواعد على جانب من السلسلتين بنظير له على الأخرى عن طريق رابطة هيدروجينية. فترتبط كل قاعدة من البيورينات بأخرى من البايريميدينات وهو ما يُعرف باسم تزاوج القواعد. تحديداً ترتبط القاعدة A بالقاعدة T وترتبط القاعدة C بالقاعدة G . يُعرف كل زوجين منهما بزوجين قاعديين كما في شكل 1.



شكل 1. مخطط تمثيلي للولب المزدوج وعناصره من القواعد النووية.

جرت محاولات متواترة لفك غموض وسحر المدونة الجينية أسهمت كثيراً في التعرف على الخصائص الجينية المتلازمة مع بعض الصفات الوراثية [2] كما اهتم العديد من الباحثين بالتعرف على ما تحتويه هذه المدونة من معلومات ومدى تباينها عبر أجزاء الدنا المختلفة [3]. تطور مجال الدراسات الجينية مؤخراً وانبثق عنه عدد من العلوم المستحدثة ذات الطابع المزجي بين العلوم الطبيعية والهندسية بغية استكشاف وتحليل المعلومات الكامنة في جينات البشر والكائنات الحية الأخرى. من هذه العلوم علم المعلوماتية الحيوية وعلم معالجة إشارات الجينوم وعلم دراسات المدونة الجينية وعلم نظرية المعلومات الجينية [4-8].

من ناحية هندسية أثبت العديد من الباحثين وجود مكونات ترددية ذات قيم مختلفة في متسلسلات الدنا [9]. يمكن الاستدلال على تلك الترددات وتحديد أماكنها بقياس طيف الدنا والذي يعرف بأنه المقياس المربع لمتسلسلة الدنا الرمزية عقب تحويلها إلى متسلسلة عددية كما نشرح لاحقاً. بتحديد المواقع المحلية لهذه المكونات الترددية توصل العديد من الباحثين إلى خصائص بيولوجية هامة أمكن استخدامها في التعرف على الطابع الجينية لعدد من الأمراض [10] وفي تطبيقات الطب الشرعي المعتمدة على تحليل الدنا [11] وفي التعرف على حوامل الأمراض كالفيروسات والبكتيريا [12].

تنوعت المكونات الترددية في الدنا ما بين مقاطع قصيرة متتابعة ومقاطع مشتتة وثلاثيات تكرارية في حين تتشكل أحياناً أنماط أكثر وضوحاً في تكراريتها مثل الأنساق المتناظرة [13] ونسق التردد المقابل للدورة 3- [14-16] التي توجد بشكل أساسي في الجينات وأيضاً أشباه الجينات [17]. تعتمد الكثير من طرائق اكتشاف مواقع تلك المكونات الترددية على تحويل فورييه الرقمي [18-19] ونسخته قصيرة المدى المبينة على تحريك نوافذ جزئية عبر متسلسلة الدنا وهو ما ندرسه في بحثنا هذا.

ما بقي من هذا البحث مقسم كما يلي: في الفصل الثاني نقدم استعراضاً لأبرز أنظمة مقابلة أبجدية القواعد النووية الرمزية بأبجديات رقمية. نتعرض في الفصل الثالث لكيفية حساب طيف الدنا باستخدام تحويل فورييه الرقمي ونقدم عرضاً مختصراً لكيفية أمثلة طريقة حسابها باستخدام نموذج متعدد المعدلات. في الفصل الرابع نقدم بعض النتائج التجريبية المعضدة للنتائج الرياضية ونخلص باستنتاجات وأفكار مستقبلية في الفصل الخامس.

نستخدم فيما يلي الرموز الغليظة للإشارة إلى المتجهات والمصفوفات. أما المتغيرات أحادية القيمة فنستعمل الرموز المائلة لتمثيلها. كما نصطلح على الرموز الآتية: المصفوفة A^{-1} هي المعكوس الضربي للمصفوفة A والمصفوفة A^T هي المصفوفة المبدلة للمصفوفة A والمصفوفة A^H ترمز للمرافق المركب للمصفوفة المبدلة للمصفوفة A وهو ما يعرف اختصاراً باسم المقابل الهرميتي للمصفوفة A .

2. مقابلة أبجدية الدنا الرمزية بأبجديات عديدة

يمكن التعامل مع متسلسلات الدنا حسابيا بعد مقابلتها بمتسلسلات عديدة أو رقمية قابلة للمعالجة باستخدام طرائق حسابية بسيطة أو طرائق هندسية أكثر تعقيدا مثل طرائق معالجة الإشارات الرقمية. بديهي أن هناك عدداً لا نهائياً من خطط المقابلة الممكن استخدامها لتحويل متسلسلة دنا إلى متسلسلة رقمية أو أكثر تعبر عنها. بعض هذه الخطط تتمتع بما يميزها عن غيرها إما من ناحية الحساب أو من ناحية تمثيلها لخصائص حيوية بما يجعلها حاملة لمعان غير المعاني الحسابية المجردة [20-21].

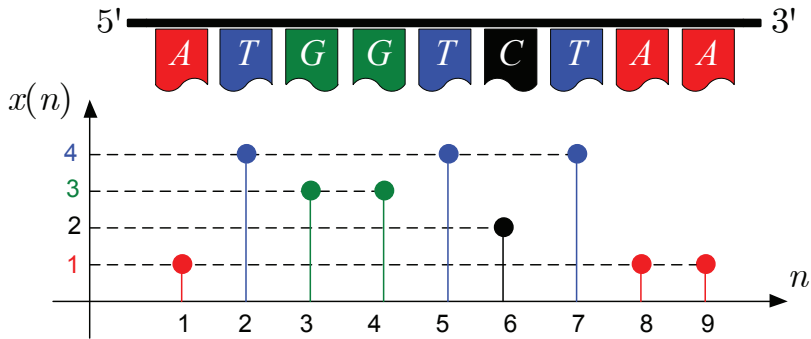
يمكن تعريف عملية مقابلة أبجدية الدنا $\mathbb{F}$ بأبجدية عديدة $\mathbb{Q}$ بأنها الدالة الرياضية أو الاقتران الرياضي $\mathbb{F} \rightarrow \mathbb{Q}$. بافتراض متسلسلة دنا $d(n)$ حيث n هي مؤشر الترتيب فإن العملية $\mathbb{F} \rightarrow \mathbb{Q}$ ينتج عنها المتسلسلة $x(n)$ أو مجموعة المتسلسلات $x_i(n)$ حيث $i = 1, \dots, \gamma$ والقيمة γ هي عدد المتسلسلات الناتجة عن التحويل $\mathbb{F} \rightarrow \mathbb{Q}$. يلزمنا بالتالي لتعريف أي عملية مقابلة تعريفا كاملا تحديد أبعاد وعناصر المجموعة $\mathbb{Q}$ وكذلك عدد المتسلسلات الناتجة عن تلك المقابلة γ . في إطار هذا التعريف نمر في هذا الفصل على بعض خطط المقابلة الأكثر شيوعا.

2.1. مقابلة أبجدية الدنا بأبجدية رباعية من خلال متسلسلة واحدة: $\gamma = 1$

الطريقة الأبسط لتحويل متسلسلة دنا إلى متسلسلات عديدة تأتي من خلال إقران كل قاعدة من قواعد أبجدية الدنا برقم صحيح وتكوين متسلسلة من تلك الأرقام الصحيحة بذات الترتيب الذي ظهر فيه أصولها في متسلسلة الدنا الأصلية. بافتراض الأعداد الصحيحة a, c, g, t يمكن تعريف هذه المقابلة كما في (1):

$$\mathbb{F} = \{A, C, G, T\} \rightarrow \mathbb{Q} = \{a, c, g, t\} \quad \dots (1)$$

على سبيل المثال يشيع استخدام المقابلة: $\mathbb{F} = \{A, C, G, T\} \rightarrow \mathbb{Q} = \{1, 2, 3, 4\}$. شكل 2 يوضح كيفية استخدام هذه المقابلة لاستنباط متسلسلة عديدة $x(n)$ مقابلة للشريط العلوي من متسلسلة الدنا في شكل 1.



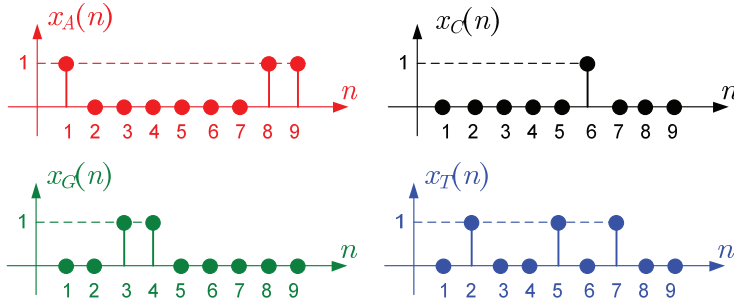
شكل 2. مقابلة الشريط العلوي من متسلسلة الدنا في شكل 1 بمتسلسلة عديدة $x(n)$.

من الواضح أن هذه الطريقة بسيطة حسابيا ولا تحمل الكثير من التعقيدات غير أنها تعطي أوزانا مختلفة للقواعد المختلفة بما قد يميز إحداها على الأخرى حال معالجة المتسلسلة العددية الناتجة $x(n)$ حسابيا.

2.2. مقابلة فوس ثنائية المدى رباعية المتسلسلات $\gamma = 4$

طريقة فوس ثنائية المدى [22] من أكثر الطرق شيوعا لمقابلة أبجدية الدنا بأبجديات عددية. وفيها يتم مقابلة متسلسلة الدنا بأربعة متسلسلات عددية يميز كل منها واحدة من القواعد النووية الأربعة. عناصر هذه المتسلسلات العددية هي عناصر ثنائية المدى حسب التعريف الآتي

$$\mathbb{F} = \{A, C, G, T\} \rightarrow \mathbb{Q} = \{0, 1\} \quad \dots (2)$$



شكل 3. مقابلة الشريط العلوي من متسلسلة الدنا في شكل 1 بأربعة متسلسلات عددية باستخدام مقابلة فوس.

في المتسلسلة العددية الخاصة بأي قاعدة تكون قيمة المتسلسلة العددية المقابلة: الوحدة إذا وجدت هذه القاعدة تحديدا عند نفس الترتيب في متسلسلة الدنا الأصلية $d(n)$ وصفرًا إن لم توجد. ويوضح شكل 3 مقابلة الشريط العلوي من متسلسلة الدنا في شكل 1 بأربعة متسلسلات عددية ثنائية المدى باستخدام مقابلة فوس. من السهل ملاحظة وفرة أو فائض المعلومات الناتج عن تشكيل أربعة متجهات لا ثلاثة فقط بملاحظة أن

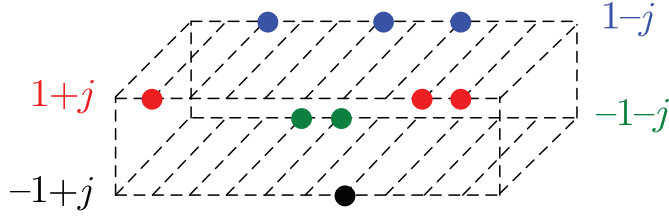
$$x_A(n) + x_C(n) + x_G(n) + x_T(n) = 1 \quad \forall n$$

2.3. المقابلة رباعية الأسطح رباعية المتسلسلات $\gamma = 4$

في طريقة المقابلة رباعية الأسطح يتم إقران القواعد النووية الأربعة بمتجهات أربعة متساوية الطول تتجه نحو الأركان الأربعة لشكل رباعي منتظم. يعني ذلك إقران كل قاعدة بعدد مركب كما يلي:

$$\mathbb{F} = \{A, C, G, T\} \rightarrow \mathbb{Q} = \{1 + j, -1 + j, -1 - j, 1 - j\} \quad \dots (3)$$

لكل قاعدة من القواعد الأربعة تكون قيمة المتسلسلة العددية المقابلة هي القيمة المركبة المناظرة لها إذا وجدت هذه القاعدة عند نفس الترتيب في متسلسلة الدنا الأصلية $d(n)$ وصفرًا إن لم توجد كما في شكل 4.



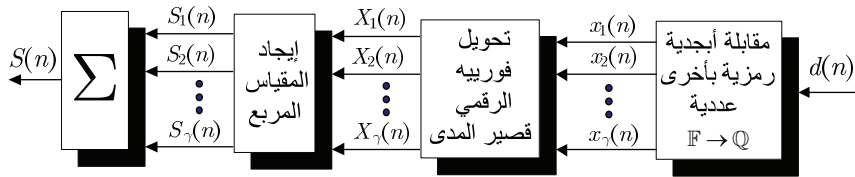
شكل 4. مقابلة الشريط العلوي من متسلسلة الدنا في شكل 1 بأربعة متسلسلات عددية مركبة باستخدام المقابلة رباعية الأسطح.

بالإضافة إلى المقابلات السالف ذكرها يوجد عدد لا يستهان به من المقابلات الأخرى مثل مقابلة منحنى-ز [23] ومقابلة رباعي الأسطح الأكثر تقدماً بالإضافة إلى المقابلات الثنائية ($\gamma = 2$) التي تسعى للتمييز بين معدلات البيورينات والبايريميدينات بربط عناصر كل منهما في متجه واحد [24].

3. حساب طيف الدنا

في هذا الفصل نباشر استعراض كيفية حساب ما يُعرف بطيف الدنا $S(n)$ وهو عبارة عن مجموع الأطياف الخاصة بالمتسلسلات العددية $x_i(n)$ المقابلة لمتسلسلة الدنا $d(n)$ الناتجة من إنجاز المقابلة $\mathbb{F} \rightarrow \mathbb{Q}$.

يمكن حساب طيف الدنا بعد حساب تحويل فورييه الرقمي قصير المدى $X_i(n) \equiv Xi(Rn, k)$ لكل متسلسلة عددية $x_i(n)$. ليتمكن حساب الطيف الخاص به $S_i(n)$ وجمع النواتج للحصول على طيف الدنا $S(n)$ كما نوضح إجمالاً في شكل 5.



شكل 5. مخطط تمثيلي لكيفية حساب طيف الدنا الخاص بتسلسلة دنا.

نقدم فيما يلي من هذا الفصل موجزاً تمهيدياً عن تحويل فورييه الرقمي ونسخته قصيرة المدى قبل إيضاح كيفية حساب طيف الدنا بناءً على تلك الكميات.

3.1. تحويل فورييه الرقمي

يقوم تحويل فورييه الرقمي بتحليل مجموعة متتابعة من الأرقام إلى مكوناتها الترددية المختلفة وذلك بواسطة عناصر تحويل تقوم بنقلها من مجال الزمن إلى مجال التردد. على سبيل المثال نفترض متسلسلة عددية $x(n)$ تحتوي عدد N من العناصر $(n = 0, 1, \dots, N-1)$ ، وتنتمي هذه المتسلسلة إلى مجال الزمن. نعرّف تحويل فورييه الرقمي لهذه المتسلسلة $X(k)$ كما يلي:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}, \quad k = 0, \dots, N-1 \quad \dots (4)$$

وتُسمى المتسلسلة (4) متسلسلة مجال التردد ولها نفس الطول N الذي لمتسلسلة مجال الزمن. كلتا المتسلسلتين تمثل حزمة معلومات تنتمي عناصرها لأبجدية رقمية محددة. أما عناصر التحويل $e^{-j2\pi kn/N}$ فما هي إلا الجذور النونية للوحدة ويمكن تمثيلها كمجموعة من الأعداد المركبة أولها الوحدة الحقيقية وجميعها واقعة على دائرة الوحدة متباعدة عن بعضها بزوايا متساوية قدر إحداها: $2\pi/N$. لتيسير المعالجة الرياضية نجمع عناصر متسلسلة مجال الزمن $x(n)$ في المتجه $\mathbf{x}$ وعناصر متسلسلة مجال التردد $X(k)$ في المتجه $\mathbf{X}$. المتجهان لهما الأبعاد ذاتها $N \times 1$ ويمكن تمثيلهما كما يلي:

$$\mathbf{x} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix}$$

وهو ما يسمح بصياغة تحويل فورييه الرقمي باستعمال المتجهات والمصفوفات بعد تشكيل عناصر التحويل على هيئة مصفوفة فائدرموند كما في المعادلة (5):

$$\begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix} = \begin{bmatrix} 1 & e^{-j\frac{2\pi}{N}} & \dots & e^{-j\frac{2\pi}{N}(N-1)} \\ e^{-j\frac{2\pi}{N}} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & e^{-j\frac{2\pi}{N}} \\ e^{-j\frac{2\pi}{N}(N-1)} & \dots & e^{-j\frac{2\pi}{N}} & 1 \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} \quad \dots (5)$$

نرمز لمصفوفة عناصر تحويل فورييه بالرمز $\mathbf{F}$ وأبعادها $N \times N$ وهو ما يبسط المعادلة (5) إلى

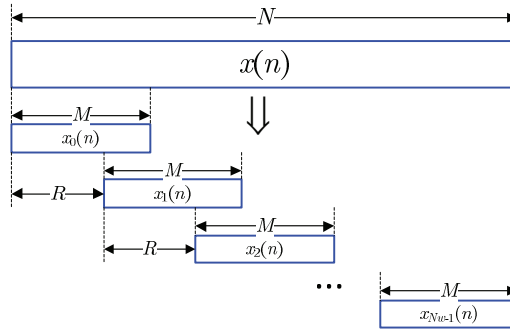
$$\mathbf{X} = \mathbf{F} \mathbf{x} \quad \dots (6)$$

تتشكل المعادلة (6) بتجميع عناصر المتسلسلة الرقمية في مجال الزمن في متجه $\mathbf{x}$ وعناصر ما بعد التحويل إلى مجال التردد في متجه $\mathbf{X}$. بينما تتشكل المصفوفة $\mathbf{F}$ من عناصر التحويل $e^{-j2\pi kn/N}$ عند نقاط الزمن والتردد المختلفة، وهي مصفوفة دورانية.

تحويل فورييه الرقمي قصير المدى هو نسخة من تحويل فورييه الرقمي السالف ذكره تختص بتحليل أجزاء محلية من متسلسلة عددية إلى مكوناتها الترددية بدلاً من تحليل المتسلسلة ككل. تفيد هذه الطريقة في عدد من العمليات منها التعرف على قوة مقياس مكونة ترددية محددة وتغيراتها عبر متسلسلة عددية وذلك بتمرير نافذة متحركة ذات طول ثابت M عبر المتسلسلة العددية ومقياس هذه المكونة الترددية لكل وضع من أوضاع النافذة. يمكن تعريف تحويل فورييه الرقمي قصير المدى لمكونة ترددية قيمتها $1/R$ Hz (ما يقابل دورة زمنية طولها R وهو ما يُعرف باسم دورة- R) على النحو التالي

$$X(Rn, k) \triangleq \sum_{m=0}^{M-1} x(Rn-m)h(m)e^{-j2\pi k \frac{m}{M}}, \quad k = 0, \dots, N-1 \quad \dots (7)$$

في المعادلة (7): القيمة n هي نقطة بداية الدورة وتتحرك بطول المتسلسلة بخطوات إزاحة متساوية طول كل منها R من القواعد النووية، والقيمة N هي طول المتسلسلة العددية الأصلية، والمتجه $h(m)$ هو متجه كل عناصره الوحدة في حالة التعريف الأساسي لتحويل فورييه قصير المدى ويمكن تعديله كما سنرى لاحقاً. للتوضيح الشكل 6 يمثل عملية تمرير نافذة ذات طول M عبر متسلسلة عددية $x(n)$ ذات طول N . يتم تقسيم المتسلسلة $x(n)$ إلى نوافذ متداخلة بمقدار R لقياس المحتوى الترددي لكل منها عند التردد $1/R$ أو قيمة الدورة المناظرة لها: دورة- R .



شكل 6. مخطط تمثيلي لعملية تمرير النوافذ المتداخلة عبر متسلسلة عددية $x(n)$.

يُلاحظ أن طول المتجه الناتج يعتمد على قيمة إزاحة النافذة R وقيمة طول النافذة M بالإضافة إلى طول المتسلسلة الأصلية N وذلك من خلال العلاقة (8).

$$N_w = \left\lceil \frac{N-M+1}{R} \right\rceil \quad \dots (8)$$

حيث $\lceil Y \rceil$ هو سقف العدد الحقيقي Y أي هو العدد الصحيح الذي يفوق أو يساوي Y .

3.3. حساب طيف الدنا

كما أوضحنا في شكل 5 يمكن حساب طيف الدنا بعد حساب تحويل فورييه الرقمي قصير المدى $X_i(n) \equiv$ لكل متسلسلة عددية $x_i(n)$ ليكن حساب الطيف الخاص به $S_i(n)$ والمعرف على النحو التالي

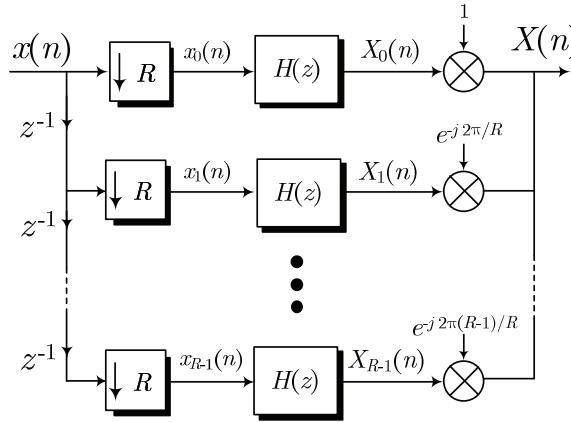
$$S_i(n) = |X_i(Rn, k)|^2 \quad \dots (9)$$

وحينها يمكن تعريف طيف الدنا بأنه

$$S(n) \triangleq \sum_{i=1}^{\gamma} S_i(n) = \sum_{i=1}^{\gamma} \sum_{m=0}^{M-1} x(Rn-m)h(m)e^{-j2\pi k \frac{m}{M}} \quad \dots (10)$$

3.4. نموذج متعدد المعدلات لحساب تحويل فورييه الرقمي قصير المدى وطيف الدنا

لتبسيط الحسابات سألغة الذكر إلى صيغة بسيطة تعتمد على المصفوفات والمتجهات، اقترح المؤلف نموذجا متعدد المعدلات لحساب تحويل فورييه الرقمي قصير المدى موضحاً في الشكل (7). يمكن الرجوع لتفاصيل استنتاج هذا المخطط رياضياً إلى [23, 25]. المكون $\downarrow R$ يرمز لاستخلاص عينات مُقْتَر بنسبة R والمرشح الرقمي $H(z)$ هو تحويل-زرد للمتجه $h(m)$ المذكور في (7).



شكل 7. مخطط متعدد المعدلات لحساب تحويل فورييه الرقمي قصير المدى $X(n)$ لمتسلسلة عددية $x(n)$.

تُعرف المكونات $\{x_0(n), x_1(n), \dots, x_{R-1}(n)\}$ باسم المكونات متعددة الطور للمتسلسلة $x(n)$ ، في حين تُعرف المكونات $\{X_0(n), X_1(n), \dots, X_{R-1}(n)\}$ باسم المكونات المرشحة متعددة الطور للمتسلسلة $x(n)$ نظراً لكونها مخرجات المرشحات الرقمية $H(z)$ عقب تمرير المكونات متعددة الطور خلالها.

بترصيص المكونات المرشحة متعددة الطور للمتسلسلة $x(n)$ في متجه:

$$\Gamma(n) = [X_0(n) \ X_1(n) \ \dots \ X_{R-1}(n)]^T \quad \dots (10)$$

باستخدام (10) يمكن إثبات أن طيف المتسلسلة $x(n)$ هو [20, 21]

$$S(n) = \Gamma^H(n) \mathbf{D} \Gamma(n) \quad \dots (11)$$

حيث المصفوفة $\mathbf{D}$ لها الأبعاد $R \times R$ وهي مصفوفة تتساوى فيه عناصر المحاور وتعرف كما يلي

$$\mathbf{D} = \begin{bmatrix} 1 & e^{-j\frac{2\pi}{R}} & \dots & e^{-j2\pi\frac{R-1}{R}} \\ e^{j\frac{2\pi}{R}} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & e^{-j\frac{2\pi}{R}} \\ e^{j2\pi\frac{R-1}{R}} & \dots & e^{j\frac{2\pi}{R}} & 1 \end{bmatrix} \quad \dots (12)$$

تبرز أهمية طريقة الحساب المعتمدة على النموذج متعدد المعدلات في سهولة تنفيذها حسابيا وفي إمكانية أمثلته بتعديل المرشح الرقمي $H(z)$ المعروف باسم مرشح النافذة المستطيلة إلى مرشحات رقمية أخرى ذات استجابة محدودة كمرشحات نوافذ هامنج وبلاكان للحصول على طيف دنا خال من التشويش. يلاحظ سهولة تعديل مرشح ذي استجابة دفعية محدودة مقارنة بتعديل مرشح ذي استجابة دفعية غير محدودة عبر تغيير تحويل فورييه الرقمي إلى تحويلات رقمية أخرى [26].

4. نتائج تجريبية

لإيضاح الجانب التطبيقي للفصل السابق وبدون إخلال بالعمومية نركز في هذا الفصل على استخلاص المركبة الترددية $f = 1/3$ ويمرر لها عادة بمركبة دورة-3 حيث تكون قيمة $R = 3$. أثبت الباحثون أن لدى حساب طيف الدنا باستخدام تحويل فورييه الرقمي قصير المدى عند هذه المركبة فإن الطيف الناتج له قيم عالية في مناطق خاصة من متسلسلات الدنا هي المناطق المسؤولة عن تشكيل البروتينات في حين تنخفض قيمه في المناطق الأخرى التي لا تتدخل في تشكيل البروتينات [27-31].

من ثم، أصبح تغير طيف الدنا المقيس عند هذه المركبة مقياسا للتفرقة بين أجزاء الدنا من حيث مساهمتها في عملية تكوين البروتينات. لقياس هذه المركبة نفترض متسلسلة دنا $d(n)$ تمت مقابلتها باستخدام مقابلة فوس إلى المتسلسلات العددية التالية: $\{x_A(n), x_C(n), x_G(n), x_T(n)\}$. باستخدام (10) يمكن حساب متجهات المكونات المرشحة متعددة الطور لكل من تلك المتسلسلات على النحو التالي

$$\Gamma_i(n) = [X_{i0}(n) \ X_{i1}(n) \ X_{i2}(n)]^T, \ i \in \{A, C, G, T\} \quad \dots (13)$$

وباستخدام (11) و(13) يكون طيف الدنا الخاصة بمركبة دورة-3 هو

$$S(n) = \sum_{i \in \{A, C, G, T\}} \Gamma_i^H(n) D\Gamma_i(n) \quad \dots (14)$$

وهي صيغة تعتمد على المصفوفات ويسهل حسابها بسرعة كما يسهل إعادة حساب مكوناتها $\Gamma_i(n)$ بعد تعديل النافذة المستطيلة $H(z)$ كما أسلفنا إلى نوافذ أخرى تجريبيا للتخلص من التأثيرات التثويشية.

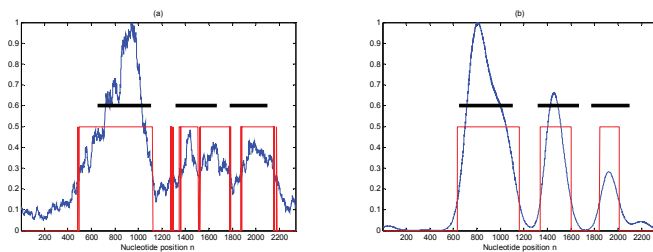
مثال. للتوضيح نتناول على سبيل المثال ثلاثة متسلسلات لها المعرفات التالية (HROMA4A, ECZGL2, ALOEGLOBIM) ويمكن الحصول عليها من قاعدة بيانات المركز القومي لمعلومات التكنولوجيا الحيوية (NCBI) [32] ومعرفة أماكن تشكيل البروتينات بها بشكل المركز القومي لمعلومات (14) لرسم طيف الدنا عند مركبة دورة-3 لهذه المتسلسلات في الأشكال 8 إلى 10 على الترتيب باستخدام النافذة المستطيلة ونافذة بلاكمان، حيث تشير الشرائح السوداء إلى أماكن تشكيل البروتينات الصحيحة ويمكن أن نرى توافقها غالبا مع مساحات ارتفاع طيف الدنا عن حد يسبق تحديده تجريبيا مبينا بالأحمر. يُلاحظ كذلك جودة طيف الدنا المرشحة مكوناته بنافذة بلاكمان مقارنة بتلك التي احتفظت بالنافذة المستطيلة لترشيحها. ■

5. استنتاجات وأفكار مستقبلية

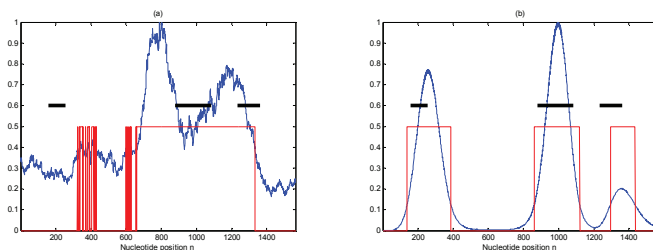
في هذا البحث قدمنا للقارئ العربي نبذة عن مجال بحثي واعد وهو تطبيق أنظمة معالجة الإشارات الرقمية على الإشارات الحيوية ممثلة في متسلسلات الدنا. قدمنا أولا نبذة عن طرائق مقابلة أبجدية الدنا الرمزية بأبجديات عديدة يمكن التعامل معها حسابيا أو هندسيا.

تطرقنا بعدها لتعريف طيف الدنا وكيفية حسابة باستخدام تحويل فورييه الرقمي ونسخته قصيرة المدى. وأوجزنا بعدها عرضا لنموذج حسابي متعدد المعدلات لحساب تحويل فورييه الرقمي قصير المدى وبالتالي طيف الدنا بشكل كفاء. أخيرا أوردنا مجموعة من أمثلة المحاكاة الحاسوبية لمتسلسلات دنا حقيقية.

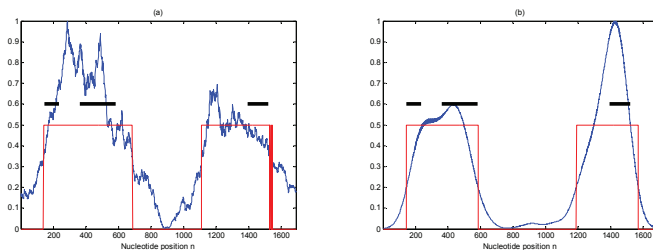
كما هو جلي فإن هذا المجال ما زال مفتوحا لضبط وأمثلة اختيارات نوافذ الترشيح وأطوالها وكذلك للبحث عن ترددات ذات معان أوضح أو للسعي خلف خصائص جينية أدق. يمكن كذلك تطبيق بعض أفكار معالجة الإشارات الرقمية في عمليات البحث عبر قواعد بيانات الدنا الضخمة وذلك لما لها من كفاءة في سرعة المعالجة واستخلاص النتائج.



شكل 8. طيف الدنا الخاص بالمتسلسلة ذات المعرّف HROMA4A باستخدام مرشح نافذة (a) مستطيلة (b) بلاكمان.



شكل 9. طيف الدنا الخاص بالمتسلسلة ذات المعرّف ECZGL2 باستخدام مرشح نافذة (a) مستطيلة (b) بلاكمان.



شكل 10. طيف الدنا الخاص بالمتسلسلة ذات المعرّف ALOEGLOBIM باستخدام نافذة (a) مستطيلة (b) بلاكمان.

- [1] Watson, J. D., Crick, F. H. C., Molecular Structure of Nucleic Acids, *Nature* 171, (1953), 737–738.
- [2] Hayes, B., The Invention of the Genetic Code, *American Scientist* 86, (1998).
- [3] Gatenby, R. A., Frieden, B. R., Information Theory in Living Systems, Methods, Applications, and Challenges, *Bulletin of Mathematical Biology*, 69 (2007), 635–657.
- [4] Moore, J. H., Bases, bits and disease: a mathematical theory of human genetics, *European Journal of Human Genetics*, 16 (2008), 143–144.
- [5] Figureau, A. Information theory and the genetic code, *Origins of Life and Evolution of the Biosphere*, 17 (2006), 439–449.
- [6] Sanchez, R. and Grau, R., A genetic code Boolean structure. II. The genetic information system as a Boolean information system, *Bulletin of Mathematical Biology*, 67 (2005), 1017–1029.
- [7] Alvager, T., Graham, G., Hilleke R., Hutchison, D., and Westgard J., A generalized information function applied to the genetic code, *Biosystems*, 24 (1990), 239–244.
- [8] Dawy, Z., Goebel B., Hagenauer, J., Andreoli, C., Meitinger, T. Mueller, J. C., Gene Mapping and Marker Clustering Using Shannon's Mutual Information, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3 (2006), 47–56.
- [9] J. A. Berger, S. K. Mitra, and J. Astola, "Power spectrum analysis for DNA sequences," in *Proc. of the Int. Sym. on Signal Processing and its App.*, July 2003, pp. 29–32.
- [10] G. Benson, "Tandem repeat finder: A program to analyze DNA sequences," *Nuc. Acids Res.*, vol. 27, no. 2, pp. 573–580, 1999.
- [11] J. Butler, *Forensic DNA Typing: Biology and Technology behind STR Markers*, Academic Press, 2003.
- [12] C. Cummings and D. Relman, "Microbial forensics: crossexamining pathogens," *Science*, vol. 296, pp. 1976–1979, 2002.
- [13] R. Gupta, A. Mittal, and S. Gupta, "An efficient algorithm to detect palindromes in DNA sequences using periodicity transform," *IEEE trans. Signal Processing*, vol. 18, no. 4, pp. 8–20, July 2001.
- [14] B. D. Silverman and R. Linsker, "A measure of DNA periodicity," *Jnl. of Theo. Biology*, vol. 118, no. 3, pp. 295–300, Feb 1986.
- [15] D. Holste et al., "Repeats and correlations in human DNA sequences," *Physical Review*, vol. E 67, no. 06913, 2003.
- [16] D. Anastassiou, "Genomic signal processing," *IEEE Signal Processing Magazine*, vol. 18, no. 4, pp. 8–20, July 2001.
- [17] V. R. Chechetkin and V. V. Lobzin, "Anticodons, frameshifts, and hidden periodicities in tRNA sequences," *Journal of Biomolecular Structure and Dynamics*, vol. 24, no. 2, pp. 189–202, 2006.

- [18] B. Issac et al., "Locating probable genes using Fourier Transform approach," *Bioinformatics*, vol. 18, no. 1, 2002.
- [19] S. Tiwari et al., "Prediction of probable genes by Fourier analysis of genomic sequences," *CABIOS*, vol. 13, pp. 263–270, 1997.
- [20] A. Rushdi and J. Tuqan, "The Role of the Symbolic-to-Numerical Mapping in the Detection of DNA Periodicities," in *the proc. of the IEEE Int. Workshop on Genomic Sig. Proc. and Stat.*, pp 1-4, June 08.
- [21] A. Rushdi, J. Tuqan, and T. Strohmmer, "Identification of Hidden DNA Periodicities by Map Invariant Spectral Analysis," *to appear in the IEEE Transactions on Signal Processing*.
- [22] R. F. Voss, "Evolution of long-range fractal correlations and 1/f noise in DNA base sequences," *Phy. Rev. Lett.*, vol. 68, no. 25, pp. 3805–3808, June 1992.
- [23] A. Rushdi and J. Tuqan, "Gene Identification Using the Z-Curve Representation," in *the proc. of the IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, vol. II, pp 1024-1027, May 2006.
- [24] P. D. Cristea, "Conversion of nucleotides sequences into genomic signals," *Journal of Cellular Molecular Medicine*, vol. 6, no. 2, pp. 279–303, April 2002.
- [25] A. Rushdi and J. Tuqan, "An Efficient Algorithm for DNA Discrete Fourier Analysis," in *the IEEE Cairo International Biomedical Engineering Conference*, vol. BI, pp 1-4, December 2006.
- [26] A. Rushdi and J. Tuqan, "Trigonometric Transforms for Finding Repeats in DNA sequences," in *the proc. of the IEEE Int. Workshop on Genomic Signal Processing and Statistics*, pp 1-4, June 2008.
- [27] P. P. Vaidyanathan and B.-J. Yoon, "Gene and exon prediction using all pass-based filters," in *Gensips Proc.*, May 2003.
- [28] P. P. Vaidyanathan and B. Yoon, "Digital filter for gene prediction applications," in *Proc. Asilomar conference*, November 2003.
- [29] M. Akhtar and E. Ambikairajah, "Time and frequency domain methods for gene and exon prediction in eukaryotes," in *Proceedings of ICASSP*, 2007, pp. 573–576.
- [30] J. Tuqan and A. Rushdi, "A DSP Perspective to the Period-3 Detection Problem," in *the proceedings of the IEEE International Workshop on Genomic Signal Processing and Statistics*, pp 53-54, May 2006.
- [31] J. Tuqan and A. Rushdi "A DSP Approach for Finding the Codon Bias in DNA Sequences," in *the proceedings of the IEEE Transactions on Selected Areas in Signal Proc.*, vol. 2, no. 3, pp 343-356, June 2008.
- [32] The National Center for Biotechnology Information: <http://www.ncbi.nlm.nih.gov/>.

ترجمة المصطلحات

DNA Sequence	متسلسلة دنا
double helix	اللولب المزدوج
leading strand	الشريط السابق
lagging strand	الشريط اللاحق
Purines	البورينات
Adenine	الأدينين
Guanine	الجوانين
Pyrimidines	البايريميدينات
Cytosine	السييتوزين
Thymine	الثايمين
Hydrogen bond	رابطة هيدروجينية
Base pair	زوجين قاعديين
Genetics	علم الجينات
Bioinformatics	المعلوماتية الحيوية
Genomic signal processing	معالجة الإشارات الجينية
Genetic code	المدونة الجينية
Molecular information theory	نظرية المعلومات الجزيئية
DNA forensics	طب الشرعي باستخدام الدنا
Pathogen exposure	التعرض لحوامل الأمراض
Palindromes	نسق متناظر
The period-3 component	مركبة دورة-3
Pseudogenes	أشباه الجينات
Symbolic to Numeric mapping	مقابلة أبجدية رمزية بأبجدية عددية
Binary Voss Mapping	مقابلة فوس ثنائية المدى
Tetrahedral Mapping	مقابلة رباعية الأسطح
DNA Spectrum	طيف الدنا
Discrete Fourier Transform (DFT)	تحويل فورييه الرقمي
Short-time Discrete Fourier Transform (ST-DFT)	تحويل فورييه الرقمي قصير المدى
Multirate Model	نموذج متعدد المعدلات
Downsampling	استخلاص عينات مقتر
Digital Filter	مرشح رقمي
Polyphase Components	المكونات متعددة الطور
Filtered Polyphase Components	المكونات المرشحة متعددة الطور
Rectangular Window Filter	مرشح النافذة المستطيلة
Blackman Window Filter	مرشح نافذة بلاكمان
Finite Impulse Response (FIR) Filter	مرشح ذو استجابة دفعية محدودة
Infinite Impulse Response (IIR) Filter	مرشح ذو استجابة دفعية غير محدودة
Period-3 Component	مركبة دورة-3
Protein Coding Regions	أماكن تشكيل البروتينات

DSP Applications in the Spectral Analysis of Genomic Sequences

Ahmad Ali Rushdi

P.O.Box 5542

Santa Clara CA 95056 USA

aarushdi@ieee.org

Abstract. Bioinformatics has been an attractive research area recently thanks to its interdisciplinary nature that explores new horizons of biological and physical sciences and their various derivatives. This novel field employs computer programming and engineering algorithms in order to tackle biological problems and solve them with computationally efficient approaches. In this work, we address the application of digital signal processing methods to the spectral analysis of genetic data. First, we survey several symbolic to numeric mappings that have been proposed in order to bring DNA sequences into the computer world. Second, we derive expressions for the DNA spectrum and discuss a multirate model to compute it. Third, we introduce simulation data to support the mathematical results. We finally conclude with some future research ideas in this promising area.

Keywords. Digital Signal Processing – Spectral analysis – Genomic Signal Processing – Analysis of Genomic Sequences – Symbolic to Numeric Mapping – DNA Spectrum.

الجغرافية الرقمية، مجالاتها في علوم الأرض و الكون وعلاقتها بالتطوير المعلوماتي " تطبيقات نظم المعلومات الجغرافية في انجاز خرائط الاخطار الطبيعية أنموذجا"

حاجي رحاب¹ ، بومزبر عبد الرحمان² ، مدموم عبد السلام³ ، ضيف الله بسمه⁴.
¹ كلية الهندسة المعمارية و علوم الأرض، جامعة فرحات عباس، سطيف، 19000، الجزائر.
² كلية العلوم الدقيقة، ع. الأرض الطبيعة و الحياة، جامعة الشيخ العربي التبسي، 12002، تبسة.
^{3,4} كلية علوم الأرض و التهيئة العمرانية، جامعة الامير عبد القادر، قسنطينة، 25000، الجزائر.

البريد الالكتروني: boumezbeura@yahoo.fr, slimdem@yahoo.fr, hadjirihab@yahoo.fr, difallahbasma@yahoo.fr.

الخلاصة. تطمح مختلف مخابر البحث الوطنية الى التحكم في استخدام تقنيات الحاسب الآلي والبرمجيات وقواعد المعلومات في تطوير أبحاثها. من بين هذه البرمجيات نذكر تقنية نظم المعلومات الجغرافية، والتي أستعملت في هذه الدراسة من أجل وضع خارطة مخاطر الإنزلاقات و الإنهيارات الأرضية لشمال شرق الجزائر. تحصر هاته الخارطة الأولى من نوعها في الجزائر توزع المناطق الأكثر عرضة لهاته الفئة من المخاطر، حيث تشير الى أن هذه الحوادث تتكرر كل سنة مع حلول الفصل الماطر في الأقاليم الجبلية المتواجدة في شمال ولاية سوق أهراس.

أثبتت هذه الدراسة مدى أهمية تضم المعلومات الجغرافية كتقنية لا غنى عنها في جمع، إدارة و مطابقة عدد هائل من البيانات الجيولوجية المنتشرة في بيئة رقمية من أجل التوصل إلى وثائق تقدم المعلومة في شكل بسيط و حديث.

إن استخدام هاته البرامج لتسيير الدراسات على أرض الواقع يؤدي للرفع من نوعية القرارات وفاعلية الخطط المستقبلية الناتجة عنها و إنجاح مخططات التنمية المستدامة في مجالات تهيئة الإقليم.

الكلمات الجوهرية: نظم المعلومات الجغرافية، خرائط الأخطار الطبيعية، أرك جيس 9.3، النماذج المتعددة المتغيرات، برامج التنمية المستدامة، التطبيقات و البرمجيات.

1 مقدمة

يشهد القرن الحادي والعشرين قفزة نوعية في مجالات استخدام وتطوير تقنية المعلوماتية والحاسب الآلي والبرمجيات وقواعد المعلومات والإنترنت. في سياق ذلك تسعى الجهات البحثية والأكاديمية في الجامعة الجزائرية لمواكبة كبرى الجامعات و المدارس الدولية في تعزيز وإستخدام هذه التقنيات في مشاريع بحثها. الشيء الذي ينعكس إيجابا على مرافقة المستخدمين والمسؤولين في جميع المستويات من محلية إلى وطنية في إتخاذ القرارات المناسبة حيال مشاريع التطوير في مجالات تهيئة الإقليم. من بين هذه التقنيات الحديثة المتركرة على علوم الحاسب الآلي والمعلوماتية نذكر تقنية إستعمال نظم المعلومات الجغرافية GIS لدراسة مواضيع الأخطار الجيولوجية.

إن النمو السريع الذي تشهده المجالات الشمالية للجزائر و الكم الهائل من المعلومات المطروحة للتحليل و المقاربة، إستلزم إستعمال هذه النظم لتخزين ومعالجة وتمثيل المعلومات ذات الطبيعة الفضائية لإنجاز الدراسات وتنفيذ الأعمال التخطيطية والإدارية والهندسية... تهدف هذه الدراسة أساسا إلى الإبتعاد عن المقاربات النظرية و الوصفية في البحث والإستعانة بنظم وتطبيقات المعلومات الجغرافية في رسم خرائط الأخطار الطبيعية، و بشكل خاص خريطة إنزلاقات الأرضيات في الجهة الشمالية الشرقية للجزائر، لذلك ركزت الدراسة على إنجاز دراسات وأبحاث تطبيقية مستفيضة في تحديد نوعيات الأجهزة والبرامج والمعلومات والخرائط من حيث الدقة والمحتوى. هذه الخرائط المشتقة ذات النوعية العالية من المعلومات تهدف إلى مواجهة الأخطار الجيولوجية المتكررة سنة تلو الأخرى و التي تعيق بشكل حقيقي أي تطور أو توسع لهاته المنطقة.

2 تكنولوجيا و نظم المعلومات الجغرافية، إستخداماتها وآفاقها المستقبلية

أتاحت أنضمة المعلومات الجغرافية إمكانية دراسة و إدارة العديد من الظواهر الطبيعية المعقدة، مما يسمح بتوقعها، فهمها وتوجيهها لخدمة المجتمع أو لتجنيبه المصاعب و الكوارث. هذا الأمر جعل من هذه التقنيات مطلبا لا غنى عنه للحياة الحديثة. في هذا السياق سعى المهتمون بالبحث في مجالات الجيولوجيا بعد التيقن من ضرورة لإستخدام نظم المعلومات الجغرافية و تطبيقات الإستشعار عن بعد إلى إتباع إستراتيجيات وأساليب بحثية تتناسب و التقدم العلمي المتزايد في ميدان المعلومات الجغرافية و في كيفية توظيفها كتقنية، مفهوم ، ومكونات في دراسة الإشكاليات الطبيعية من ناحية والإستفادة منها كأدوات تكنولوجيا حديثة من ناحية أخرى. تُعد هذه النظم الجغرافية والمعروفة بـ GIS إحدى أهم المكنان، التي يمكن توظيفها في فهم الظواهر بما يتناسب مع فلسفتها وأهدافها ، ذلك لأنها تمتلك القدرة على تخزين و إستكشاف وتحليل البيانات المكانية بسرعة فائقة ودقة متناهية و إصدار قرارات سليمة بشأن ترشيد إستخدام الموارد المتاحة في إنجاز مخططات التنمية المستدامة. يمثل التحكم في هذه النظم أهم متطلبات التناول في مجال معالجة البيانات ذات البعد الفضائي. حيث تتكون مما يلي : الآلات .Hardware. الوسائل Procedure. البيانات Graphical and Attribute Data. البرامج Software. الأشخاص People.

- إدخال المعلومات وفرزها

تم إدخال جميع المعلومات المتوفرة في برنامج أرك جيس 9.3 ARCGIS لمجموعة اسري الامريكية ESRI لقدرته على تخزين، معالجة، تحليل ورسم المعلومات والخرائط ببراعة ودقة متناهية حيث خصص طبقة من المعلومات لكل موضوع (DigitalMaps, Spatial Data, Tabular Data, Arial Photos, Satellite Images)، ومن ثم تم فرزها في ملفات مخصصة لإستغلال نتائجها من أجل تتابع نظم المعلومات الجغرافية.

تمت الاستعانة بعدة وثائق وخرائط أساسية والتي تم تصنيفها إلى نوعين: الأولى خرائط ورقية أساسية مثل الخرائط الجيولوجية رقم 55, 77, و 78، و الطوبوغرافية بمقياس 1/50000، والثانية مشتقة مثل خرائط توزيع المغياثية ، جرد توزيع الانزلاقات السابقة وأقسام المنحدرات...

إن أهم المراحل لإنجاز هذا النوع من المشاريع تتسلسل في: نسخ و رقمنة الخرائط الأساسية كالجيوولوجية و الطبوغرافية... إشتقاق النموذج الرقمي للارتفاعات ، النموذج الرقمي للأرضية ، و خريطة المنحدرات من صور الأقمار الصناعية. إنجاز خارطة التوزيع المتوسط للأمطار. إنجاز خارطة توزيع و اتجاهات الإنزلاقات. إنجاز خارطة استخدام الأراضي من توزيع المناطق الحضرية و الطرقات إلخ، ومن ثم إدخال كل هاته المواضيع في نفس النظام الجغرافي Arcview. وضع نموذج جبري منطقي من أجل تعيين مصادفة حصول إنزلاق و الذي يجمع عدة عوامل، من ثم مطابقتها مع فرضية التعرض لهذه الظواهر من أجل تعيين المناطق ذات الخطورة. إن الإطلاع على الأبحاث و الدراسات المنجزة للمقارنة و الاستفادة منها في عرض المعلومات سمح بإنجاز خارطة التوزيع الجغرافي للمناطق من حيث خطورتها، إتجاه الإنزلاقات و عدم الثبات...

3 أهمية البحث:

يهدف البحث الحالي إلى تجسيد و توظيف أهم الطرائق و التكنولوجيات الجديدة لمعالجة القضايا و الظواهر الجغرافية و إتخاذ قرارات واقعية بشأنها، الأمر الذي ينعكس بصورة أو بأخرى على إعداد جيل جديد من الباحثين لا يكتفي بمجرد الاستخدام الميكانيكي لهذه التقنية بل يطورها و يثريها و ذلك بالمشاركة في تصميم و تنفيذ و توجيه بعض برمجياتها وفقا لمتطلباته الميدانية. تتمثل الأهمية التطبيقية لجغرافية الكوارث و المخاطر البيئية في التطلع للسيطرة و التصدي لهذه الأخطار بوضع حلول تسهم في تجنبها أو التخفيف من آثارها. تم إختيار هذه الدراسة التطبيقية في ضوء الأثر المباشر و غير المباشر للإنزلاقات الأرضية كأحد أكبر المشاكل التي كثيرا ما تتكرر في شمال شرق الجزائر. إقتصرت الدراسة على محافظة سوق أهراس و المستوطنات الريفية المحيطة بإعتبارها الأشد تضررا من هذه الظاهرة.

1-3 حدود البحث:

إلتزم البحث الحالي بالحدود التالية :

صياغة و تنفيذ الأنشطة المتضمنة برمجية، بإستخدام أرك فيو جي أي اس 9.3 Arc View 9.3 GIS. إعداد و عرض الأشكال البيانية و الجداول الإحصائية و النصوص الجغرافية بإستخدام حزم مايكروسوفت أوفيس 2007. و إعداد و تجهيز الخرائط بإستخدام أوتوكاد ماب 2010 .

2-3 أسئلة البحث:

تحددت أهداف و متطلبات البحث في الاسئلة التالية :

ما فعالية إستخدام تكنولوجيا نظم المعلومات الجغرافية و توليفة من معطيات الإستشعار عن بعد في معالجة هذا الموضوع الجيولوجي المتشعب الأسباب على نطاق جغرافي واسع ؟ كيف يمكن إستخدام تكنولوجيا نظم المعلومات الجغرافية في بناء برمجية كفيلة بتحليل و معالجة هذا الكم الهائل من المعلومات النقطية. ما فعالية إستخدام تكنولوجيا نظم المعلومات الجغرافية في تبسيط المعلومة الجغرافية كنتاج نهائي للاداء عند المستغلين من غير أصحاب المهارات ؟ ، و ما هي أكثر الأماكن تضررا في السنوات الأخيرة بمنطقة الدراسة؟ ما مميزات مناطق الخطر ؟ ما أوقاتها؟ ما الآثار البيئية الناجمة عنها؟ و هل يمكننا التنبؤ بها؟

3-3 إجراءات وخطة البحث :

تحدد خطة وإجراءات الدراسة باتباع التسلسل التالي :

- الجانب النظري :

التعرف و التعريف بنظم المعلومات الجغرافية من حيث تاريخ نشأتها وماهيتها وطريقة عملها ، أدواتها ومكوناتها وعلاقتها بالمجالات العلمية الأخرى.

إستخدام نظم المعلومات الجغرافية و تطبيقاتها في حل المسائل الجيولوجية الرئيسية من حيث: التطبيقات ، المتطلبات ، الأنماط ، أساليب إستخدامها في التحليل وبناء البرمجيات.

حصر المهارات الجيولوجية، الجغرافية، و المعلوماتية اللازمة في معالجة البيانات وإنجاز الدراسات الميدانية من حيث إستخدام المصطلحات الجغرافية كالخرائط، الجداول الإحصائية والرسوم البيانية، صور الأقمار الصناعية...

التحكم في مصادر البيانات و تكنولوجيات المعلوماتية بإستعمال التقنيات المتقدمة للإستشعار عن بعد و كفاءة النقل المتزامن للمعلومات والتكنولوجيا عن طريق الأنشطة الميدانية التعاونية و نشر النتائج على نطاق واسع بإستغلال الشبكات المتاحة.

إدارة قواعد البيانات (Data Management) بالنسبة للمشروعات الصغيرة أين تكون الملفات العادية كافية للتخزين. الشيء الذي لا ينطبق على البيانات ذات الحجم الكبير حيث تستخدم برامج (DBMS) لتساعد على تخزين و تنظيم البيانات.

-الجانب التجريبي: في ضوء ما تم إستخلاصه من الجانب النظري للبحث تم القيام بما يلي :

إعداد قائمة مبدئية بالمهارات الجغرافية.

إستخدام نظم المعلومات الجغرافية في بناء خطة بحث وفقاً للخطوات التالية :

ماهية وأهداف محتوى الدراسة لتحديد المفاهيم والمهارات الجغرافية اللازمة.

تحديد أهداف البرمجية وإعادة تنظيم وصياغة محتواها.

تحديد الخطوات التقنية لإعداد برمجية إستخدام نظم المعلومات الجغرافية : من تصميم البرمجية ، إعداد خريطة تدفق البرمجية، تصميم المراحل الأساسية للتابع المنطقي، تصميم واجهات المستخدم، تحديد الوظائف الأساسية لمكونات البرمجية، إجراء تجربة استطلاعية للبرمجية، تنفيذ البرمجية.

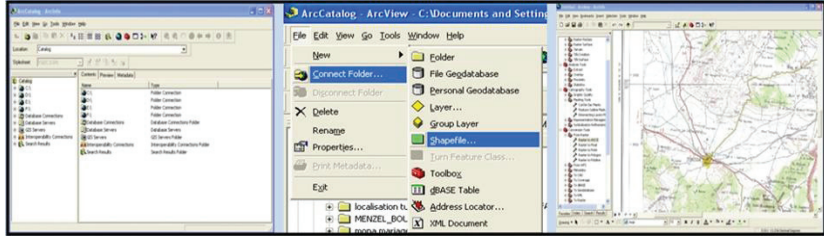
التجسيد المرئي للخرائط و ذلك بتحقيق التكامل بين جميع المعطيات المتوفرة.

إنشاء و إدارة قواعد المعطيات و البيانات بإجراء مسح لمكامن أخطار الانزلاقات الأرضية. إعداد منهجية فعالة مجدية عن طريق إستخدام معطيات الإستشعار عن بعد، نظام تحديد المواقع (GPS) ، ونظام المعلومات الجغرافية (GIS). إن البساطة والتفرد في تنسيقات الملفات تسمح بإنتاج المعلومات من بيانات متفرقة و بالإستمرارية السلسة في المعالجة الآلية.

تشغيل 9.3 ArcGIS®

إن إستغلال نظم المعلومات الجغرافية في حزمة 9.3 ArcGIS يتطلب إستخدام التطبيقات التالية : ArcMap يستخدم بكافة المهام الكرتوغرافية من ادراج و تحليل وتحديث للخرائط.

ArcCatalog تطبيق لإدارة البيانات يستخدم من أجل تطوير قواعد البيانات المكانية وعرض البيانات الوصفية. ArcToolbox يتيح تنفيذ مهام التحويل، والتجهيز (الجيوبروسيسينج). إضافة إلى ArcScene، ArcReader، و ArcGlobe، يمكن لجميع هذه التطبيقات تنفيذ مهام مختلف نظم المعلومات الجغرافية، من تحليل جغرافي، وإدارة وتحديث للبيانات.



شكل 01: الكيانات المختلفة في ArcGIS®.

4-3 أهداف الدراسة

تهدف الدراسة إلى وضع خارطة مخاطر تصنف الجهة الشرقية للوطن إلى مناطق مخاطر طبيعية ضعيفة، متوسطة، أو عالية تأخذ بعين الاعتبار الإرتباط المباشر بالتوزيع السكاني، البنية التحتية، و المراكز الصناعية بإستخدام GIS. تتميز الخارطة التي تجمع مختلف عوامل المخاطر الطبيعية مع العوامل البشرية بأنها تساعد على تحديد المناطق ذات الإحتمالية الواسعة لتعرضها للإنزلاق و حدوث أضرار مادية و بشرية، و بالتالي المساعدة على التخفيف من هذه المخاطر و الإختيار الأفضل للمواقع من أجل انجاز المشاريع مختلفة التوجهات.

4 استخدام نظم المعلومات الجغرافية في مشاريع البحث في علوم الأرض و الكون

يعرف كوليت (1992) نظام المعلومات الجغرافية على انه بيئة رقمية تتيح إدارة و إستغلال المعطيات ذات الطابع الفضائي. مما يسمح بإستخدام الأبعاد الموضوعية المكانية و الزمانية للمعلومات من أجل أغراض التحليل و الوصف و التنبؤ للواقع المنظور .

تعد ظاهرة الإنزلاقات الأرضية إحدى أخطر الظواهر التي تهدد أراضي الجمهورية الجزائرية، حيث تأخذ أشكالا متعددة فنجد الانهيارات، سقوط الصخور و الحجارة، التدفقات الطينية... تشير التقديرات الى أن حوالي نصف مشاريع البنية التحتية في المنطقة الشمالية الشرقية للجزائر تعاني من هذه الظاهرة بدرجات متفاوتة و تحت عدة أشكال، و تظهر آثارها السلبية من خلال تدهور الوضع العام للمنشآت العمومية و الخاصة الشيء الذي ضاعف معاناة الأهالي في مقر الولاية و العديد من القرى الموزعة جغرافيا على امتداد مشارفها الشمالية. سارعت الجهات البحثية و الأكاديمية في الجامعة الجزائرية الى توجيه مشاريع بحثها للمساهمة في الجهود الرامية الى السيطرة على هذه الظاهرة المدمرة لكل تقدم في وتيرة تطبيق مشاريع التنمية المستدامة المسجلة ضمن المخططات التنموية كالمخططات الخماسية 1999-2004/2004-2009/2009-2014

والتي تتضمن مشاريع الحفاظ على البيئة و كل ما يتعلق بها من اعادة تأهيل الأراضي و معالجة مشاكلها من انجراف و عدم استقرار، في هذا الإطار أنجزت المخابر البحثية المنضوية تحت لواء وزارة الجامعات و البحث العلمي العديد من الدراسات البحثية من أجل تثبيت الأراضي و حصر الصالحة منها للزراعة، تنمية الغابات و المساحات الخضراء كعامل اساسي في إستقرار السفوح.

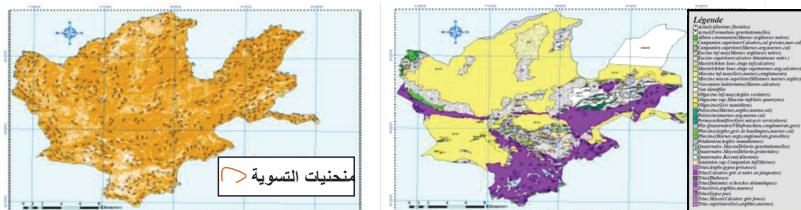
5 الإجراءات المنهجية للدراسة:

1-5 مصادر جمع المعلومات:

المكتبات العامة و الخاصة و تقارير الإدارات و المؤسسات الرسمية: بدأنا الدراسة ببحث مكتبي مستفيض قصد توفير المعلومات الكامنة في الكتب و الدوريات، و الموسوعات و دوائر المعارف و المقالات و الرسائل و الأطروحات، و البيانات و التقارير و المطبوعات الحكومية.

الاطالس و الخرائط الجغرافية: تعد الخرائط الطبوغرافية و الجيولوجية و الرسوم البيانية و الخرائط الكنتورية من بين أهم المصادر الأساسية اللازمة لإثراء الدراسة.

ان التجسيد المرئي (Visualisation) على شكل خرائط له دور كبير في تخزين و توصيل المعلومات الجغرافية، لذا رقممت الخرائط الجيولوجية للوحات أرقام 55، 77، 78، تحت مقياس 1/50000 المسوحة ضوئيا و المعلمة جغرافيا في نظام UTM حيث تم تحويل شعبة الخطوط المفتوحة لكل طبقة جيولوجية إلى مضلعات مغلقة، و نت ثم قمنا بإنشاء قاعدة بيانات تحتوي على خصائص كل وحدة جيولوجية (شكل 08 و 09). يمكن ل GIS أن يعطي شكلا دقيقا لسطح الأرض و ذلك عن طريق ادخال الخرائط الكنتورية للمنطقة و نقاط الارتفاع في شكل احداثيات ثلاثية، ثم يتم استنتاج شكل سطح الأرض من قبل البرنامج على شكل نموذج ارتفاع رقمي. DEM



مفتاح الخريطة	
Actuel (Alluvions fluviales)	Paléocène (Marnes-arg. marno-cal)
Actuel (Formations gravitationnelles)	Pernocarbonifère (Grès micacés versicolores)
Albien-cénomannien (Marnes argileuses noires)	Plio-Quaternaire (Villafranchien, conglomérats, grès)
Campanien-supérieur (Calcaires, cal gréseux, mar-cal)	Pliocène (Argiles, grès de boudingues, marnes-cal)
Campanien-supérieur (Marnes-arg. marnes-cal)	Pliocène (Marnes arg. conglomérats, gravellites)
Eocène inf-moy (Marnes argileuses noires)	Priabonien (Argiles numidiennes)
Eocène-supérieur (calcaires bitumineux noirs)	Quaternaire-Moyen (Déluvio-gravitationnelles)
Maestrichtien Sous-étage inf (calcaires)	Quaternaire-Moyen (Déluvio-proluviales)
Maestrichtien Sous-étage sup (marnes-arg. calcaires)	Quaternaire-Recent (Alluvions)
Miocène inf-moy (Grès, marnes, conglomérats)	Santonien sup-Campanien inf (Marnes)
Miocène moyen-supérieur (Siltstones marnes argiles)	Trias (Argilo-gypso-gréseuses)
Néocomien indéterminé (Marno-calcaires)	Trias (Calcaires gris et noirs en plaquettes)
Non identifiée	Trias (Diabases)
Oligocène inf-moy (Argiles verdâtres)	Trias (Dolomites et breches dolomitiques)
Oligocène sup-Miocène inf (Grès quartzeux)	Trias (Grès, argillites, marnes)
Oligocène (Grès numidiens)	Trias (Gypse pur)
Paléocène (Marnes argiles, marno-cal)	Trias-Moyen (Calcaires gris foncé)
	Trias-supérieur (Grès, argillites, marnes)

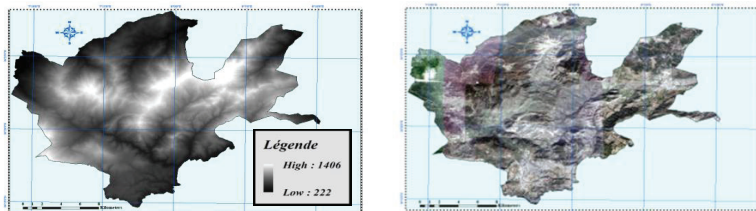
شكل 02: الخريطة الليثولوجية الرقمية للصخور مستمدة من الخرائط الجيولوجية رقم 55، 77، 78. الخريطة الطبوغرافية المستمدة من النموذج الرقمي للتضاريس kriging (على اليسار)

(في الاسفل) مفتاح الخريطة الجيولوجية.

صور الأقمار الصناعية:

استغلت الدراسة المرئيات الفضائية للقمر الصناعي Landsat 7 + ETM، بتميز 90 م يعود تاريخ التقاطها إلى عام 2010 م وذلك بهدف التعرف على توزيع المساحات المزروعة ، حجم التجمعات المأهولة بالسكان، الطرقات، و الحوادث....

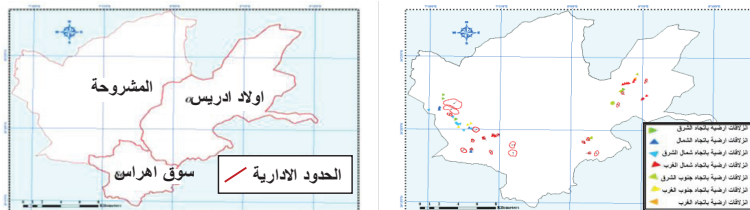
حملنا هذه المرئيات بإتباع الطريقة التالية : فتح الرابط على موقع Landsat View Images ← النافذة Collection ← Landsat Archive ← تحديد المرئية ← تحدد Path & Row لمنطقة شمال شرق الجزائر ← GO ← أيكونة Add ← أيكونة Download ← Arcgis.



شكل 03 : الصور العمودية للقمر الصناعي Landsat 7 و النموذج الرقمي للأرضية.

2-5 الدراسة الحقلية

تم الإعتماد على الدراسة الميدانية لإستيفاء البيانات الأولية بشكل مباشر من الميدان حيث قمنا بإنجاز بطاقة جرد حقلية بإتباع نظم تحديد المواقع بإستعمال أجهزة GPS المتصلة بعدد من الأقمار الصناعية التي تعطي دقة عالية للإحداثيات. من ثم تم إستكمال بياناتها عن طريق إنجاز خارطة توزيع مناطق الإنهيارات الأرضية الضرورية لإشتقاق خارطة الحساسية. يتطلب جرد الإنهيارات الأرضية إنجاز عملية تعامد للخطوط الكنتورية مع صور الأقمار الصناعية.



شكل 4 :خريطة جرد توزيع الانهيارات الارضية في منطقة الدراسة.

3-5 أسلوب الدراسة:

لجاناً في هذه الدراسة إلى استعمال أسلوب الحصر الشامل حيث قمنا بحصر جميع المراكز العمرانية في منطقة الدراسة وتحديد أكثرها تعرضاً لمخاطر الإنزلاقات باستغلال صور الأقمار.

نمذجة البيانات Data Modeling

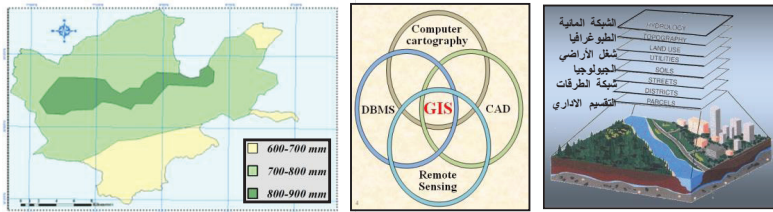
النمذجة هي عملية محاكاة للواقع عن طريق بناء نموذج، يمكننا من خلاله فهم ظاهرة معينة. هذا النموذج عبارة عن مجموعة من الخطوات و القواعد المكانية الخاصة بنظمنا للمعلومات الجغرافية.

البيانات المناخية:

إعتمد البحث على بيانات الأمطار خلال للفترة الممتدة ما بين عامي 1979 م و2009 لأربع محطات مناخية و هي محطة سوق اهراس، محطة المشروحة، محطة عين الدالية و محطة بوشقوف، علما أن هذه المحطات تقع على إمتداد مطقة الدراسة.

بيانات جريان المياه السطحية:

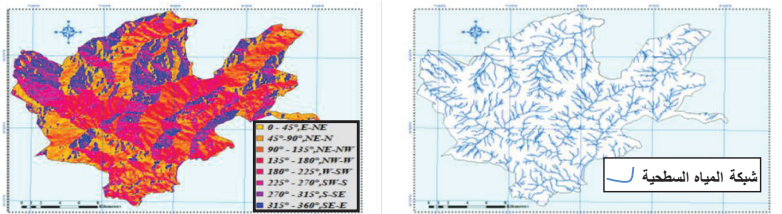
إعتمدت الدراسة على تقارير قياس كميات السيول لكل من حوض سبيوس الذي يصب في الشمال، حوض مجردة الذي يصب في الشرق، و المحطة السيلية لسد عين الدالية.



شكل 5: الطبقات الكاذبة لسطح الأرض، التقنيات المرتبطة ب GIS، خريطة توزيعات التهاطل.

الادخال Data Input

يمكن ل GIS أن يحول جميع البيانات رقمية و أنوماتيكية الى ملفات رقمية Digitalizing باستخدام التكنولوجيا الضوئية مثل المسح الضوئي Scanning ، كما يمكن تحويل البيانات من صورة CAD الى صورة GIS باستخدام Data Conversion tools .



شكل 6: الخريطة الحالية لشبكة المجاري المائية داخل الحدود الإدارية لمنطقة الدراسة مستمدة من خارطة إتجاهات المنحدرات.

الجدول 1: قائمة الموضوعات و الملفات في نظام المعلومات الجغرافية لمنطقة الدراسة.

اسم الملف	نوع الملف	موضوعات	نوع الخرائط
Topo. 1	File System Raster	Carte topographique de Souk Ahras	الخرائط الأساسية
Topo. 2	File System Raster	Carte topo. d'Oued Mougras	
Topo. 4	File System Raster	Carte topographique de Bou Chegouf	
Géol. 1	File System Raster	Carte géologique de Souk Ahras	
Géol. 2	File System Raster	Carte géologique d'Oued Mougras	
Géol. 3	File System Raster	Carte géologique de Bou chegoug	
Extra† 1 /14	File System Raster	Orthophoto1 à .Orthophoto14	الصور المتعامدة
MNT	Grid	Modèle numérique de terrain	خرائط مشتقة من الدرجة الأولى
MNA	Grid	Modèle numérique d'altitude	
Contour	Grid	Carte des courbes de niveau	
C. pente	Grid	Carte des pentes	
C. exposition	Grid	Carte d'exposition	
c. géol.	Shape	Carte géologique numérisée	خرائط موضوعية
c. du climat.	Shape	Carte de climat	
C. des facteurs S.	Shape	Carte des facteurs structuraux	
c. de pendage	Shape	Carte de pendage	
c. des routes	Shape	Carte du réseau routier	
c. d'agglomération	Shape	Carte d'agglomération	
c. des glissements	Shape	Carte des glissements repérés	
c. de vulnérabilité	Shape	Carte vulnérabilité	
c. d'aléa	Shape	Carte d'aléa	
C. de risque	Shape	Carte de risque	

4-5 منهج الدراسة:

إتبعنا هذه الدراسة المناهج الآتية:

المنهج الوصفي التحليلي: يهدف هذا المنهج المستعمل في تقدير الحساسية إلى وصف الآثار السلبية للإنزلاقات الماضية، وتحليل الأضرار المادية والبشرية المنجزة عنها.

المنهج الكمي: يعد هذا المنهج الدقيق أحد أهم المناهج الجغرافية الحديثة نظراً لما يتمتع به من مزايا في تحليل الارتباط لربط العلاقة بين المتغيرات التابعة والمتغيرات المستقلة، إستعملنا فيه:

معادلة خطوط تساوي المطر $Pav = (A1(p1+p2)+A2(p2+p3)+.../2(A1+A2)+...)$.
معادلة معدل التساقط $P = Pa/A$.

معادلة حجم التساقط: حجم التساقط = مساحة الحوض × معدل التساقط.

معادلات الخصائص التضاريسية وتشتمل على: معادلة معدل الانحدار، معادلة نسبة التضاريس، معامل الإنمناج، معادلة نسبة التقطع.

معادلة نسبة التشعب = عدد المجاري من مرتبة م/ عدد المجاري من المرتبة الأعلى منها.

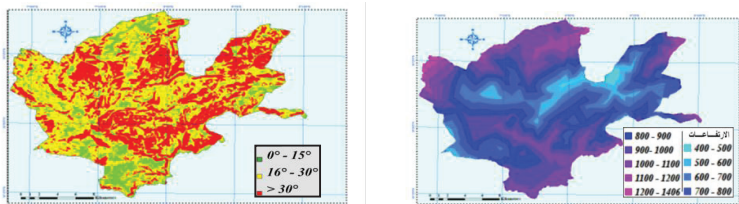
معادلة تكرار المجاري = عدد المجاري بجميع رتبها في الحوض/مساحة الحوض(كم²)

معادلة كثافة التصريف = إجمالي أطوال مجاري الأودية بمنطقة الحوض/ مساحة الحوض.

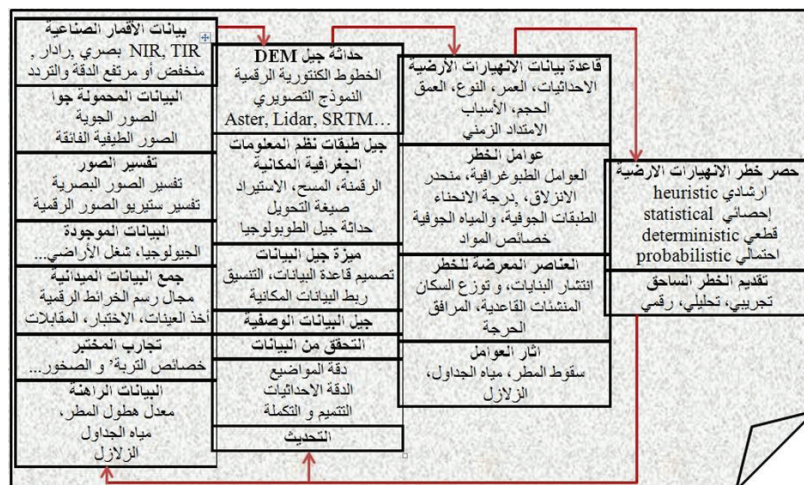
5-5 الطرائق المتبعة

بعد جمع عديد الخرائط من عدة مصادر حيث شملت الخرائط الأساسية، المشتقة أو المعالجة مثل: خارطة النطاقات الزلزالية CRAAG/CGS. خارطة الارتفاعات الطبوغرافية digitaldatabase. خارطة انتشار شبكة المياه السطحية في منطقة الدراسة. خارطة الكثافة السكانية لسنة 2008. خارطة البنية التحتية لسنة 2010 المستخرجة بتطبيقات arcview على صور الأقمار الصناعية. خارطة المنحدرات حيث اشتقت من النموذج الرقمي للارتفاع بتطبيقات "اشتقاق المنحدر" وحدة من ArcView. و لبعديد من لخرائط الأخرى كما هو ممثل في الجدول 1

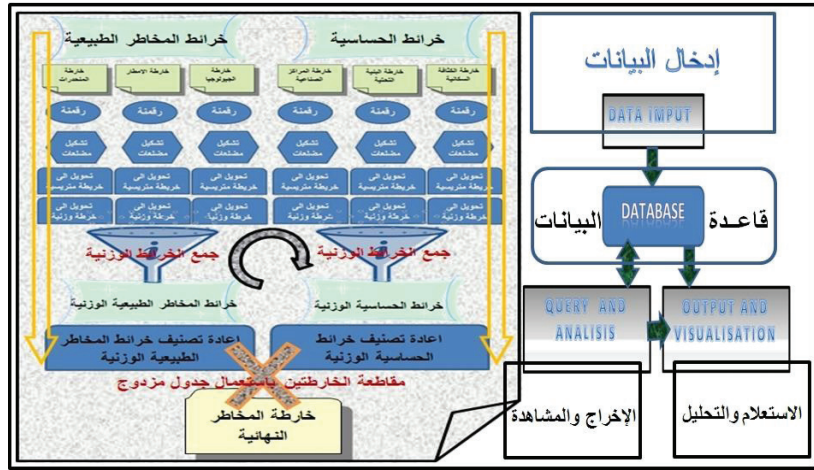
أستعمل الأسلوب الكارتوجرافي لتحويل البيانات والأشكال الإحصائية وتمثيلها ببيانيا، حيث أستخدم التحليل الحاسوبي و المعالجة الرقمية للمربيات الفضائية ببرنامج ArcGIS وبرنامج إيرداس، برنامج SPSS، و البرامج الإحصائية: Image 9.0 و Excel.



شكل 7: النموذج الرقمي للارتفاعات، و خارطة المرتفعات لمنطقة الدراسة.



شكل 8 : منهجية تتابع GIS و كيفية معالجة العلومات.



شكل 9 : منهجية تراكم نظام المعلومات الجغرافية لخريطة المخاطر لمنطقة الدراسة.

بإتباع المخطط المفصل لمنهجية تتابع الدراسة، قمنا بإدخال البيانات ل GIS و ذلك من خلال الرقمنة باستعمال برمجيات R2V، و من ثم تحليل و معالجة البيانات الرقمية في برنامج ILWIS بعد عملية الرقمنة تم إنشاء خرائط المضلعات (Polygon Maps) و تحويل هذه الخرائط إلى خرائط مترية (Raster Maps) ثم إستنتاج الخرائط الوزنية (Weight Maps). صنفت الخرائط وفق جداول تحوي قيم مقدرة من خلال المعرفة الجيولوجية بمنطقة الدراسة. من خلال معالجة البيانات في GIS تم الحصول على الخرائط الوزنية للحساسية و الأخطار.

6 خارطة الحساسية لمنطقة الدراسة

للحصول على خارطة الحساسية الوزنية إستخدمنا تقنيات الاستشعار عن بعد remote Sensing تتلخص هذه التقنيات في تجميع بيانات عن سطح الأرض باستخدام صور الأقمار من أجل تصنيف المعالم الجغرافية الموجودة بها مثل إستخدامات الأراضي و الطرقات... بإستعمال النماذج المتعددة المتغيرات تم جمع الخرائط الوزنية للكثافة السكانية والبنية التحتية والمراكز الصناعية وفق صفوف تصاعدية تمثل كل منها درجة من درجات المخاطر.

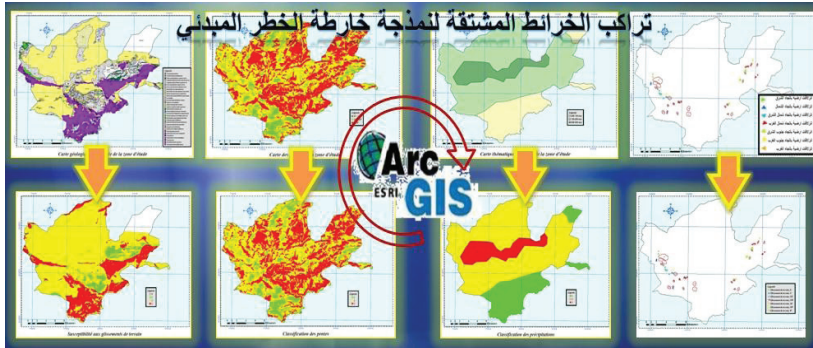


شكل 10: خارطة توزيع المدن، خارطة شبكة الطرقات، و خارطة الحساسية الناتجة.

7 خارطة المخاطر الطبيعية لمنطقة الدراسة

لتحديد المجالات الجغرافية لظهور الانهيارات الأرضية، تم الجمع بين خارطة الظروف الجيولوجية، ظروف المنحدرات، وخارطة توزيع الانهيارات الأرضية، وهطول الأمطار.

شرط الانحدار: تبين ملاحظتنا المستقاة من الدراسة الميدانية أن الانهيارات الأرضية أمر شائع عندما يتجاوز المنحدر 30 درجة (عالية)، ويمكننا تحديد المجالات التي يكون فيها خطر الانهيار الأرضي متوقع في خيارات "درجة <30 < درجة المنحدر ≤ 15" (معتدلة) كما يمكن أيضا تمثيل منطقة المخاطر الطبيعية المنخفضة، <15 درجة المنحدر.



شكل 11 : مبدأ نمذجة خارطة المخاطر لمنطقة الدراسة.

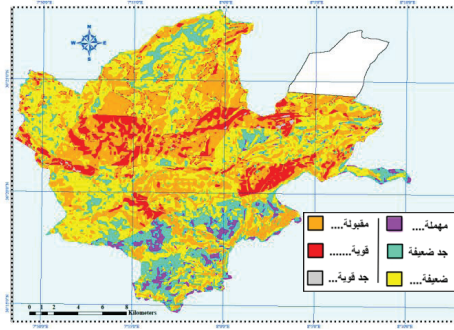
شرط جيولوجية الصخور و الأترية: خلال مرحلة الإستطلاعات الجيولوجية ؛ لاحظنا أن غالبية الانهيارات الأرضية تتركز على فضاءات بروز صخور العصر الترياسي. نلاحظ نفس الشيء في جانبي السفوح الحادة للغضار نظرا للطبيعة الطينية والمتكسرة لهذه الصخور.

إستعملنا أيضا نموذج خارطة جرد مواقع الانهيارات الأرضية في منطقة الدراسة المعروض سافا.

المنافشة والنتائج

تبين خارطة الأخطار الطبيعية الأولية أن الصفوف عالية المخاطر تتواجد بالمناطق الشمالية للولاية وهي الأكثر وعورة، أمطاراً، تركيزاً للإرتفاعات الطبوغرافية، و تركزا للإنزلاقات الأرضية.

إن معظم الأضرار المتعلقة بهذه الأخطار تجلت في هذه المنطقة بالذات و أدت الى خسائر مادية جليلة من إنكسار الطرق و تهمد البيوت و تشققها...



شكل 1: خارطة الخطر المبدئي للانهيارات الأرضية داخل الحدود الإدارية لمنطقة الدراسة.

8 إستنتاج خارطة الأخطار النهائية

لتحديد قيم ناتج تتطابق الخارطتين بتنسيق (shap) وفق صفوف محددة مسبقاً تم استخدام جدولاً مزدوجاً لإعادة تصنيف خارطتي الخطر المبدئي والحساسية. من خلال تطبيق مقترحات الجدول في نمذجة الخطر والضعف، أنتجنا خارطة جديدة تجمع تراكب صفوف الخارطتين، ألا وهي خارطة المخاطر الجامعة لمخاطر الانهيارات الأرضية.

جدول 02 : الجدول ثنائي الأبعاد المستخدم لجمع خارطتي المخاطر الطبيعية والحساسية.

	حساسية منخفضة جداً	حساسية منخفضة	حساسية عالية	حساسية عالية جداً
مخاطر منخفضة جداً	مخاطر منخفضة	مخاطر منخفضة	مخاطر منخفضة	مخاطر منخفضة
مخاطر طبيعية منخفضة	مخاطر منخفضة	مخاطر منخفضة	مخاطر منخفضة	مخاطر منخفضة
مخاطر طبيعية معتدلة	مخاطر منخفضة	مخاطر منخفضة	مخاطر منخفضة	مخاطر منخفضة
مخاطر طبيعية معتدلة	مخاطر منخفضة	مخاطر منخفضة	مخاطر منخفضة	مخاطر معتدلة
مخاطر طبيعية عالية	مخاطر منخفضة	مخاطر معتدلة	مخاطر عالية	مخاطر عالية
مخاطر طبيعية عالية جداً	مخاطر منخفضة	مخاطر معتدلة	مخاطر عالية	مخاطر عالية

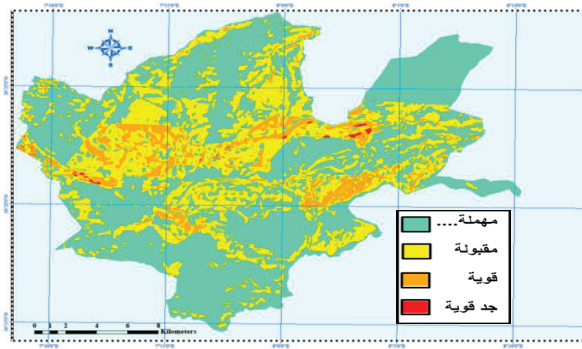
لإستخلاص خارطة المخاطر الجامعة لمنطقة الدراسة تم استخدام المعادلة:

$$\text{two-Dim_table} [\text{map1}, \text{map2}] = \text{Output_map}.$$

حيث two-Dim_table هو الجدول المزدوج، map1 تمثل خارطة المخاطر الطبيعية، map2: خارطة الحساسية، بينما Output_map هي خارطة الأخطار النهائية.

تتكون خارطة الأخطار النهائية من أشكال ومضلعات تحدد، تقييم، وتحصي مدى الضعف وتعطي توزيع المناطق المعرضة لخطر الانهيارات الأرضية، التي تنص على أن جميع الأشخاص، الطرق، والمنازل المتواجدة على المسار المفترض للإنزلاق هي عرضة لخطر الانهيارات الأرضية.

تعتمد درجة الخطورة أساساً على حجم الأرض المنزلة، وكمية المياه المخزنة في هذه الأرضية. تم تصنيف هذه القيم إلى أربع فئات مخاطر، ولكل منها نفس العدد من البكسل.



شكل 12 : خارطة خطر الانهيارات الارضية في منطقة الدراسة.

الخارطة المعروضة أعلاه تبين أن المناطق التي يتركز فيها الخطر موجودة أساسا في شمال و في وسط منطقة الدراسة، هذه الأقاليم تتميز بغزارة أمطارها، كثافة نسيجها العمراني ، وشبكة الطرقات، والتي تتوافق مع النواحي الأكثر حساسية. تم تحديد النسبة المئوية للخطورة لمختلف المناطق الإدارية لمنطقة الدراسة من خلال إجراء عملية المقاطعة بين خارطة الأخطار النهائية مع خارطة المناطق الإدارية، حيث نلاحظ أن منطقة المشروحة تضم على أعلى نسبة من الخطر.

9 الخلاصة:

مكن هذا العمل البحثي من إنتاج خريطة أخطار إنزلاق الأراضي في المنطقة الشمالية الشرقية للجزائر اعتمادا على خارطتي الخطر الجيولوجي الأولي و الحساسية. يتكون نظام المعلومات الجغرافية لمنطقة الدراسة و المتمركز أساسا على النموذج الرقمي للتضاريس من وثائق مختلفة، كالخرائط الورقية وصور الأقمار الصناعية و المتعمدة...

إتبع المنهج الرئيسي للدراسة الخطوات التالية:

مسح ورقمنة الخرائط الجيولوجية الكلاسيكية، أتاح إصدار الخريطة الليتولوجية للصخور، للعوامل الهيكلية، للإتجاهات والغوص ، والخريطة الجيولوجية الرقمية.

إشتقاق نموذج الإرتفاع الرقمي وخريطة المنحدرات و فئات المنحدرات ابتداء من النموذج الرقمي للأرضية و الصور المتعمدة للأقمار الصناعية.

وضع خريطة التوزيع المتوسط لإرتفاعات الأمطار السنوية، الذي يكشف بشكل واضح السمات الأساسية لتوزيع مياه الأمطار.

تمثيل إحدائيات، إتجاه وشدة الإنزلاقات الموجودة إنطلاقا من رقمنة خريطة توزيع الإنهيارات.

رسم خرائط رقمية للمناطق الحساسة لهذه الظواهر .

دمج جميع هذه الموضوعات في نفس النظام الجغرافي ArcView ®.

تصميم نموذج لتحديد مناطق الخطر الأولي يجمع بين مختلف المتغيرات المأخوذة بعين الاعتبار مزج ومطابقة المناطق المعرضة والمناطق الحساسة لخطر الإنهيار.

النتيجة النهائية تجلت في خارطة أخطار الإنهيارات الأرضية لمنطقة شمال شرق الجزائر و التي تعد مثالا جيدا لمزايا إستخدام نظم المعلومات الجغرافية، والجمع بين مصفوفات من طبقات متعددة من المواضيع.

10 الخاتمة

أثبتت هذه الدراسة الميدانية المعتمدة على ما هو موجود من معلومات متعلقة بالتضاريس، نوعية التربة، ومعدل التساقطات المطرية... قدرة الشراكة بين الجامعة و المجتمع الإقتصادي و الإجتماعي على تسخير النظم الحديثة في خدمة مشاريع واقعية تساهم في رفع المعانات عن المواطنين القاطنين في المناطق الخطرة، الشيء الذي شجعنا على المضي قوما بتسجيل مشاريع بحث أخرى تحت نفس السياق. إن إنجاز خارطة المخاطر الطبيعية لشمال شرق الجزائر كنموذج تنبئي يمكنه أن يتنبأ بأن منطقة ما معرضة للانزلاقات، بعد أداة جد مهمة في بين أيدي ممتهني مجالات تهيئة الإقليم.

11 المراجع

- [1] Carrera A., Guzzetti M., Cardinali M., and Reichenbach P. (2000). Use of GIS Technology in the prediction of landslide hazard. *Natural Hazards*, 20: 1170-135.
- [2] Darawcheh R., Sbenati R., (2000). The July 551 AD Beirut earthquake, eastern Mediterranean region. *Journal of earthquake engineering* Vo. No 4 403-414.
- [3] Giardini D., Jimenez J., and Grunthal G., (2002). European-Mediterraneaseismic hazard map ter Resour. Res. 30: 1153-1171.
- [4] Van Westen. C. J. Terlien M. T. J. (1995). An approach towards deterministic landslide hazard analysis in GIS. (Colombia). *Earth surfaces and landforms*, 21: 853-868.
- [5] Van Westen. C. J., 1993. Application of geographical information system to landslide zonation, ITC publication No. 15, ITC, Enschede.
- [6] Aleotti, P. and Chowdury, R. (1999) Landslide hazard assessment: summary review and new perspectives. *Bull. Eng. Geol. Env.*, Pages 21- 44
- [7] Baeza, C. and Corominas, J. (2001) Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surface Processes and Landforms*, 26:12 1251-1263.
- [8] Crosta GB, Agliardi F (2002) How to obtain alert velocity thresholds for large rockslides. *Phys Chem Earth Parts A/B/ C* 27(36):1557-1565
- [9] Baldelli, P.3 et al « Landsliding-susceptibility numerical mapping et the messina strait crossingsite Italythe 7th international symposium onland1, 153-158.
- [10] Argyris C. et Sch.n D.A., « Apprentissage organisationnel. Théorie, méthode et pratique », De Boeck Université, Bruxelles, Paris, 1996 .
- [11] ARMONIA, « Multirisik and emergency management », rapport de recherche, 2005.
- [12] Beck E. et Glatron S., « Vulnérabilité socio-spatiale aux risques majeurs: l'approche du géographe », actes du colloque « Vulnérabilités sociétales, risques et environnement. Comprendre et évaluer », Université Toulouse – le Mirail, 14, 15 et 16 mai 2008
- [13] Béguin M. et Pumain D., « La représentation des données géographiques », Armand

Colin, 1994

- [14] Brehm P.J. et al, « Enterprise risk analysis for property and liability insurance companies », Guy Carpenter & Company, LLC, 2007

Digital geographic fields in earth sciences and their relationship to development of information "GIS applications in completing the maps of natural hazards model"

Abstract. Aspire various national research laboratories to control the use of computer technologies and software and databases in the development of its research. Among these mention the software technology of geographic information systems, which are used in this study in order to map the risk of landslides of the North-eastern Algeria. Limit following a map distributed by the first in Algeria distributed areas most vulnerable to this type of risk. Indicate that these incidents are repeated every year with the solutions of the rainy season in the mountain regions located in the north of the state Tagaste.

This study shows how important GIS join an indispensable technique in the collection, management and matching a huge number of complex geological data in a digital environment in order to reach information and documents submitted in the form of a modern and simple. The use of following programs for the conduct of studies on the ground lead to raise the quality of decisions and the effectiveness of future plans and the resulting success of the schemes in the areas of sustainable development to create the region.

Keywords: GIS, maps of natural hazards, ArcGis 9.3, northeastern Algeria, sustainable development programs, applications and software.



محرك بحث عربي كامل التوزيع بالاعتماد على المُعالجة اللغوية الآلية

محمد عابدين¹، تغريد عنبر²

¹كلية الحاسبات والمعلومات - جامعة عين شمس - القاهرة - جمهورية مصر العربية
²تقنيات الحاسبات واللغة (كولتيك) - 6 أكتوبر - جمهورية مصر العربية

Taghride@coltec.net ، mabdeen@alumni.uottawa.ca

الخلاصة: أصبحت الشبكة العنكبوتية العالمية مقصداً مُفضلاً للباحثين عن المعلومات في جميع أنحاء العالم. إن كم المعلومات الهائل على الشبكة العنكبوتية - بما يتضمنه من بلايين الصفحات الإلكترونية التي تغطي شتى الموضوعات - جعل المعلومة على بُعد نقرة إصبع. ويمثل المحتوى العربي الإلكتروني جزءاً هاماً من الشبكة العنكبوتية، كذلك تأتي اللغة العربية في المرتبة الخامسة من حيث الاستخدام على مستوى العالم، أضف إلى ذلك، العدد الكبير لمستخدمي الإنترنت باللغة العربية والذي يتزايد بمعدلات هائلة. في ظل هذه الحقائق يصبح من الضروري تيسير التعامل مع المحتوى العربي الإلكتروني بالنسبة لمتصفح الشبكة العربية. يوجد على الساحة عدد من محركات البحث العالمية مثل جوجل، إم إس إن، ياهو و غيرها إلا أن الطبيعة الخاصة للغة العربية لم تؤخذ في الاعتبار عند تصميم هذه المحركات.

في هذا البحث نستعرض ما أنجزناه لتطوير نموذج أولي لمحرك بحث عربي. تم تصميمه بحيث يكون كامل التوزيع بهدف توسيع نطاق البحث وخدمة عدد كبير من المستخدمين نتوقع تعاملهم مع هذا المحرك. نستعرض في هذا البحث ما تم إنجازه بالنسبة لمكونات هذا المحرك. جدير بالذكر أن محرك البحث هذا يعتمد اعتماداً أساسياً على المعالجة اللغوية الآلية التي تبدأ مع عملية الفهرسة وتنتهي بنهاية عمل وحدة الاستعلام. هذه الميزة التي ينفرد بها هذا المحرك تجعله مختلفاً بصورة ملحوظة عن المحركات الأخرى المتاحة.

الكلمات الجوهرية: محرك بحث - معالجة لغوية آلية - استعادة المعلومات - شبكة الإنترنت العربية - النظم المؤرعة

1 المقدمة

لقد أصبحت الشبكة العنكبوتية العالمية المستودع المعلوماتي الأكبر على الإطلاق و المجاني في ذات الوقت. للباحثين عن المعلومات عبر العالم تعتبر الشبكة العنكبوتية هي الاختيار رقم واحد للبحث عن المعلومات بداية من التشخيصات الطبية مروراً بأحدث أدوات الاتصال النقالة و حتى نتائج بطولة كأس العالم لكرة القدم.

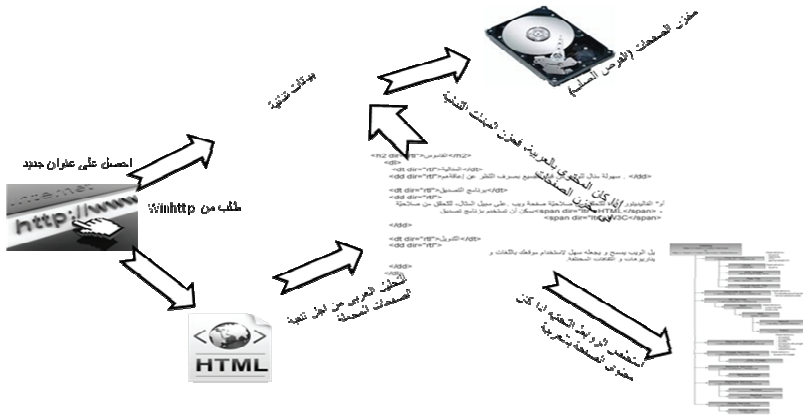
تعتبر محركات البحث هي التقنية الفاعلة التي تمكن من استعادة و فهرسة و ترتيب و تصنيف و عرض معظم المعلومات المتعلقة باستعلام المستخدم. تجمع محركات البحث المعلومات المتاحة على الخوادم الإلكترونية العنكبوتية من جميع أنحاء العالم و تقوم بالمعالجة المكثفة لها لنقل الكمية

العنكبوتى عليها ومن ثمّ توسيع نطاق الصفحات المطلوب زيارتها. عناوين المواقع الالكترونية من القائمة تتم زيارتها بتكرار بناءً على مجموعة من القواعد سيتم شرحها تفصيلاً فيما بعد في هذا البحث العلمي. شكل 2 بالأفـل يُظهر تخطيطاً لكيفية عمل اللـمـام العنكبوتى العربى.

إن تصميم اللـمـام العنكبوتى يضم أربعة مكوّنات :

- 1- برمجة و تتابع شبكى
- 2- ترميز (تشفير) الصفحة
- 3- استخراج الرابط
- 4- إدارة الذاكرة

شكل 2 بالأفـل يصوّر مكوّنات اللـمـام العنكبوتى الذى قمنا بتطويره خلال العمل على مشرونا.



شكل 2 . تخطيطاً لكيفية عمل اللـمـام العنكبوتى العربى

لتصميم وتنفيذ لـمـام عنكبوتى فعّال كفاء كامل التوزيع متعدد المتابعية، واجهنا العديد من التحديات وواجهناها بتطبيق طرق متنوعة تم تفصيلها في عمل نشرناه حديثاً [4]. هناك مسألة هامة تختص بتصميم اللـمـام العنكبوتى العربى ألا وهي سياسة تحديث الصفحة. تحدّد سياسة تحديث الصفحة الميعاد المناسب لتحديث نسخة الصفحة التي تم تحميلها بالفعل من قبل. قمنا كذلك بتقديم تحسينات لطريقة "طول عمر الصفحة" لملائمة طبيعة الشبكة العنكبوتية العربية [8] [9] [10]. ولقد طبقنا تقنية معالجة لغوية لخلق محتوى الصفحة للحد الأدنى الممكن. هذه التقنية أثبتت فاعليتها وكفاءتها في استخدام الذاكرة. إن حفظ 90 % من استخدام الذاكرة أصبح في متناول يدنا بتطويرنا لمثل هذه التقنية [3].

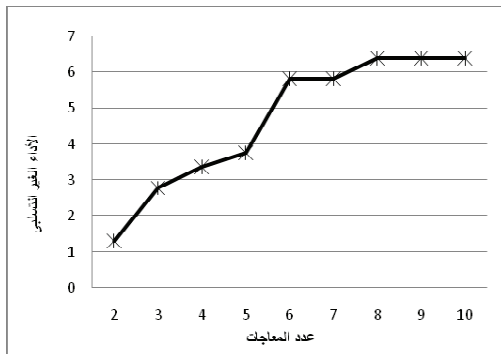
(أ) الأداء المتوازي للـمـام العنكبوتى

طورنا لـمـاماً عنكبوتياً متوازياً مع اللـمـام العنكبوتى السابق بناؤه بجوهره وتم تقييم أدائه. وحيث أن عدد الماكينات المتاحة كان محدوداً، فقد قمنا بتطوير تقنية محاكية لمحاكاة عدد اعتباطي من الماكينات ورمزناه بـ N. لدراسة تأثير التوازي الخاص بالخوارزميات المطوّرة على سرعة عمل اللـمـام العنكبوتى، قمنا بتشغيل الخوارزميات المطوّرة مسبقاً على مجموعة مكونة من 2000

صفحة عربية على الشبكة، ثم قمنا بقياس الوقت المتوسط (بالثانية) المنصرم في المراحل الثلاث التالية.

- 1- مرحلة قراءة الصفحة و تقسيمها إلى قطع متساوية أو شبه متساوية
- 2- مرحلة المعالجة اللغوية للصفحة المطبقة من كل معالج بيانات .
- 3- مرحلة الدمج المطبقة بواسطة العقدة الرئيسية

إن أداء اللام العنكبوتي المتوازي هو المعكوس الزمني (1/ الزمن). هذا الأداء تم حسابه لحجم صفحة يساوي 500 كيلو بايت؛ النتائج موضحة في الشكل التالي . هذا الشكل يوضح الأداء غير الانتسابي في مقابل عدد المعالجات.



شكل 3. الأداء في مقابل عدد المعالجات

تم حساب الأداء المُطَبَّع الخاص باللام العنكبوتي المتوازي (الأداء عند بعض ماكينات N / الأداء عند $N=2$). النتائج موضحة في الشكل 4 بالأسفل.



شكل 4 . الأداء الانتسابي في مقابل عدد المعالجات

يوضح شكل 4 أنه عند تطبيق خوارزمتنا على حجم الصفحة المساوي لـ 500 كيلو بايت فإننا نحصل على أداء أفضل بمقدار خمس مرات عند استخدامنا لـ N أربع مرات أكبر . أمّا عند تطبيقنا لخوارزمتنا على حجم صفحة أصغر من 500 كيلو بايت فقد حصلنا على أداء أفضل بمقدار ثلاث

مرات كحد أقصى عند استخدامنا لنفس ماكينات N . إن هذا يعني أن الزيادة في الأداء الناتجة من زيادة الموارد تعتبر ذات دلالة على المحتوى الإلكتروني الكبير. لذا فإن الخوارزمية المتوازنة المقترحة تصبح فعالة جداً عند تطبيقها على الصفحات الكبيرة. المزيد من المناقشة التفصيلية و الدراسة في [5].

3 وحدة رتبة الصفحة

تم تطوير خوارزمية رتبة بايج بواسطة لاري بايج وسيرجي برين وتم تنفيذها واستخدامها بواسطة محرك البحث الخاص بجوجل [1]. الغرض منها هو ترتيب الأهمية النسبية لكل عنصر داخل مجموعة الوثائق ذات الارتباط التشعبي (مثل الشبكة العنكبوتية العالمية) من خلال هذه المجموعة

رتبة الصفحة يعتبر نظاماً عالمياً للترتيب والتصنيف لكل الصفحات الإلكترونية، بصرف النظر عن محتواها وبالإعتماد فقط على موقعها في البناء التصوري على الشبكة العنكبوتية حيث إنها ترتب الصفحات بموضوعية وبطريقة ميكانيكية.

يتم تقسيم الترتيب على روابطه الموجهة بالتساوي للمساهمة في ترتيب الصفحات التي تشير إليها. مشكلة واحدة هي أن صفحة i ربما لا يوجد لها ارتباط خارجي. لحل هذه المشكلة يعطي جوجل كلا من الخلايا في العمود i القيمة $1/N$.

مشكلة أخرى هي أنه في حالة ما إذا أشارت بعض الصفحات لبعضها البعض وليس لصفحات أخرى فإننا أمام بالوعة، حيث تذهب كل الأهمية إلى صفحات بالوعة. يعرف جوجل $d = 0.85$ على أنها الاحتمالية التي سيبثها المتصفح باعتبارها ارتباطاً خارجياً حقيقياً. و لذا فإن $1 - d$ تعتبر احتمالية أن المتصفح سيبث ارتباطاً خارجياً صناعياً. قيمة رتبة الصفحة لأي صفحة p_i تكون متوقعة كالتالي :

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

حيث أن $p_1, p_2, \dots, p_N$ تعتبر الصفحات محل الاعتبار، $M(p_i)$ هي مجموعة الصفحات التي تشير كرايط إلى p_i ، $L(p_j)$ هو عدد الروابط الصادرة على الصفحة p_j ، N هي العدد الكلي للصفحات، d هي عامل الإغراق .

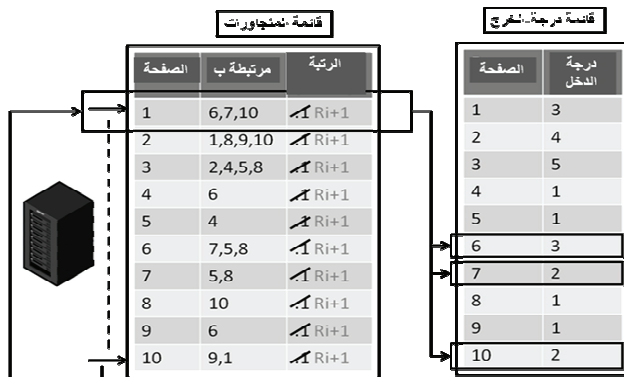
لقد قمنا بتطوير خوارزمية لرتبة الصفحة كاملة التوزيع تعتمد على خوارزمية الصفحة المتبعة من طرف جوجل. هذه هي الخطوة الاستهلاكية نحو تطوير محرك لرتبة الصفحة كامل التوزيع لمحرك البحث العربي القائم تحت التطوير. تم اختبار الخوارزمية الحالية المطورة لرتبة الصفحة لتعمل على ما يصل إلى ست ماكينات ثنائية النواة مع تدرج جيد كما سيتم مناقشته لاحقاً.

يقوم محرك الترتيب بالتفقيب عن الصفحات الإلكترونية التي تم تحميلها من اللام العنكبوتى الكامل التوزيع، ويؤلف قائمتين: قائمة المحاذاة (الجوار) و قائمة درجة-الخروج. يتم إنشاء ثلاثة جداول باستخدام هاتين القائمتين الحالية والتالية والبائدة التي تحمل درجة-الخروج لكل صفحة . خطوات الخوارزمية هي كالتالي :

- 1- املأ الجدول البائد باستخدام قائمة الجوار التي تم إنشاؤها استهلالاً
- 2- ابذر القائمة الحالية بقيم الرتبة الأولية للصفحة التي هي $1/N$ حيث أن N تمثل رقم الصفحات في الرسم.
- 3- كرر الصفحات ككل في الرسم التوضيحي بحساب رتبة الصفحة وباستخدام هذه المعادلة

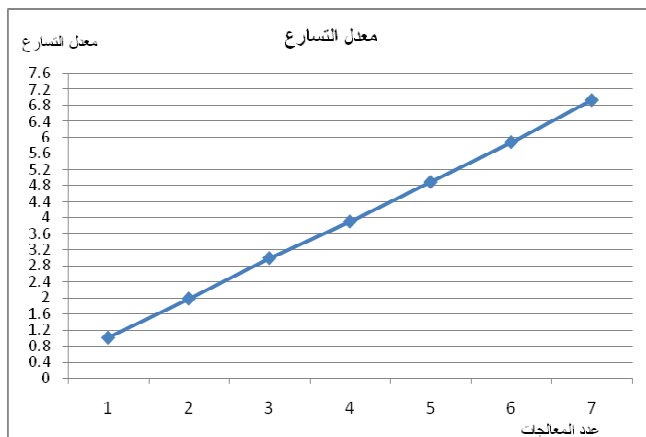
$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in N(p_i)} \frac{PR(p_j)}{L(p_j)}$$

- 1- وزع رتبة الصفحة من القائمة الحالية للقائمة التالية
- 2- عدّ للخلف لخطوة التكرار (1) أو أنه في حالة التقارب
- شكل رقم 5 بالأسفل يوضّح نموذج تشغيل للخوارزمية الخاصة بعشر صفحات
-



شكل 5. نموذج لخوارزمية رتبة الصفحة الموزعة لعشر صفحات مترابطة.

إنّ تدرّج الخوارزمية المطوّرة للترتيب يظهر واضحاً في شكل 6 بالأسفل. أرقام متعددة من الماكينات (N) يتم استخدامها لترتيب 140,000 صفحة عربية على الشبكة العنكبوتية من الشكل يتضح أن تدرجات الخوارزمية ستخبر بعدد الماكينات المختبر (سبع ماكينات ثنائية المعالج)



شكل 6. تقييم للأداء الخاص بالوقت المستغرق لترتيب 140,000 صفحة إلكترونية

وما يحدث هو أن هناك خوارزمية للترتيب أكثر مناسبة للصفحات الالكترونية على الشبكة العنكبوتية ذات المحتوى العربي تأخذ المحتويات في الحسبان. هذه الخوارزمية الخاصة بترتيب الصفحات تستخدم وحدة المعالجة اللغوية لعد الكلمات المنتمية لنفس الوحدة اللغوية. إن هذا يخدم في تعديل قيمة رتبة الصفحة والترتيب العام، وفي هذه الحالة يتم التعبير عنها بـ :

$$FR(p_i) = a * PR(p_i) + b. MR(p_i)$$

حيث FR هي الترتيب النهائي، PR هو ترتيب الصفحة، MR هو الترتيب الظاهري، a و b ثابتان حيث أن $b = (1-a)$. و هذا العمل مازال جاريا.

4 تقنيات الاستعلام من منظور عربي

تستخدم محركات البحث عند التعامل مع استعلام معين إحدى تقنيتين: البحث المتطابق، والبحث العشوائي.

البحث "المتطابق" يسترجع من النصوص المخزنة تلك التي تتطابق مفرداتها كتابيا مع مفردات الاستعلام دون زيادة أو نقص. وأما البحث "العشوائي" فيسترجع من النصوص المخزنة المفردات المتطابقة كتابيا مع مفردات الاستعلام، بالإضافة إلى المفردات الأخرى المسبوقة بحروف عن يمين (سوابق) أو عن يسار (لواحق) مفردات الاستعلام: إذا افترضنا أن مفردة الاستعلام هي "book"، فإن البحث "المتطابق" يقتصر على البحث عن "book" فقط، أما البحث العشوائي فيشمل "Book" و "books". تلائم هاتان التقنيتان اللغات الأوربية، لأن نظامها اللغوي مبسط للغاية إذا قيس بنظام اللغة العربية

إذا طبقت كلا التقنيتين على نصوص عربية، جاءت نتائج البحث غير مرضية؛ تقنية البحث المتطابق تسفر عن نتائج قاصرة قصورا شديدا: إذا كان الاستعلام هو "وزير التربية" فإن نتائج البحث لن تتضمن "وزير التربية، لوزير التربية، كوزير التربية..."، أما تقنية البحث العشوائي فقد يترتب عليها دخول كلمات ليس بينها وبين كلمة الاستعلام صلة: إذا كانت كلمات الاستعلام هي "جريد النخل"، فإن نتائج البحث يمكن أن تتضمن "تجريد، جريدة..." وكذلك الحال بالنسبة لاستعلام مثل: "سكين"، قد تتضمن نتائج البحث "مسكين - سكين". وهكذا يمكن أن تتضمن نتيجة البحث عددا غير قليل من الوثائق التي لا صلة بينها وبين موضوع الاستعلام... في نفس الوقت قد تفقد نتائج البحث مفردات ذات صلة مباشرة بالاستعلام كما في حال إذا كان جمع مفردة الاستعلام من جموع التكسير. لتلافي هذا القصور، أدخلت بعض محركات البحث تحليلا صرفيا سطحيا عند التعامل مع النصوص العربية. هذا الحل، وإن حسن نتائج البحث أحيانا، فإنه يحيد عن الصواب في أحيان كثيرة.

هناك ظاهرة أخرى هي التنوع الكبير في مستوى كتابة المحتوى الرقمي وبالذات على الشبكة العنكبوتية؛ يتميز مستوى الكتابة في بعض الصفحات بالدقة من حيث الالتزام بالقواعد الكتابية التقليدية، وتحرر بعض الصفحات من بعض القواعد الكتابية ولكنها تظل في إطار القبول اللغوي: مثل إسقاط همزة القطع (إبراهيم ← إبراهيم) أو عدم الالتزام بكتابة الباء اللينة في آخر الكلمة (يحتفي ← يحتفي). هذا الأسلوب الكتابي يسود في بعض المناطق الجغرافية مثل مصر. معظم الصفحات العربية على الشبكة - وبالذات الصفحات الصحفية و صفحات الشبكات الاجتماعية وغيرها - تعج بالأخطاء الكتابية بجميع أنواعها. من النماذج الفعلية لهذه الأخطاء "ذلة لسان، الجرائم اللانسانية، السعودي، المساول الحكومي، الإستقرار، حديقت الحى، الإستفتاء..." ولا يخفى مدى خطورة الأخطاء الكتابية في المحتوى الرقمي على عملية البحث؛ إن البحث بأشكاله الحالية يغطي فقط الكلمات الصحيحة دون الكلمات الخطأ مما يعني تجاهل وثائق عديدة في عملية البحث، وبالتالي عدم شمولية النتائج.

في ظل هذه الأوضاع، رأينا أن تطوير أي محرك بحث عربي لا يمكن أن يكون ناجحا ومميزا إلا إذا استحدثت تقنيات تلائم طبيعة اللغة العربية وواقعها.

لقد طورنا نوعين من التقنيات: تقنيات خاصة بمستوى الكتابة وتقنيات خاصة بعملية البحث.

4.1 تقنيات مستوى الكتابة

أخذاً في الاعتبار لواقع المحتوى الرقمي العربي من حيث مستوى الكتابة أُنحْنَا للمستخدم اختياريين بالإضافة إلى البحث الأولي "default":

- المستوى الكتابي الدقيق كتابي متميز مثل مواقع نصوص التراث
- المستوى الكتابي المقبول وينحصر استخدامه في المواقع التي تهتم بالمستوى الكتابي وفي نفس الوقت تتخفف من قيود همزة القطع والياء اللينة آخر الكلمة. عند تطبيق هذا الاختيار، فإن نتائج البحث عن استعلامات مثل "البنك المركزي"، الإقبال على الانتخابات، ستتضمن الوثائق التي تحتوي على "البنك المركزي، الإقبال على الانتخابات".
- البحث الأولي يأخذ في الاعتبار الأخطاء الكتابية الشائعة عند تطبيق هذا الاختيار، فإن نتائج البحث عن استعلامات مثل "قرى، السعودية، الإقبال على الانتخابات"، ستتضمن الوثائق التي تحتوي على الكلمات "قرى، السعودية، الإقبال على الانتخابات" بالإضافة إلى "قرى، السعودية، الإقبال، الإقبال". رويتنا أن مستخدم محرك البحث يهدف إلى الحصول على معلومات كاملة عن استعلامه، وليس من الإنصاف حرمانه من هدفه بسبب أخطاء كتابية لا ذنب له فيها داخل النصوص موضوع البحث.

4.2 تقنيات البحث

بالإضافة إلى البحث المتطابق المستخدم في جميع محركات البحث، يسمح النظام بالبحث وفق أكثر من خيار:

- البحث العشوائي الذكي وهو تطوير للبحث العشوائي المستخدم في كثير من محركات البحث؛ نتائج البحث وفق هذه التقنية تتضمن جميع الصفحات التي تحتوي على مفردة الاستعلام بالإضافة إلى جميع الصفحات الأخرى التي تحتوي على مفردة الاستعلام بما يلتصق بها من سوابق ولواحق مقبولة لغوياً. وفي نفس الوقت تستبعد جميع المفردات المسبوقة بحروف تخلق كلمة جديدة ليس بينها صلة وبين مفردات الاستعلام. الاستعلام عن "مال عام" يمكن أن تتضمن نتائجه صفحات بها "المال، للمال، والمال..." وفي نفس الوقت لن تتضمن "عمال، جمال، رمال..." وتسمى هذه التقنية أيضاً بالبحث بالزوائد.
 - البحث التصريفي المترابط حيث تتضمن نتائج البحث بالإضافة إلى نتائج البحث بالزوائد، التصريفات وثيقة الصلة بمفردة الاستعلام: الاستعلام عن "مال عام" يمكن أن يتضمن أيضاً "الأموال العامة..."
 - البحث وفقاً لـ "وحدة المفاهيم الصرفية" حيث يضاف إلى نتائج البحث الناتجة عن التقنيتين السابقتين جميع المفردات ذات الصلة بالاستعلام من حيث المضمون والأصل الصرفي: حرب ← حروب، محارب، حربي... وفي نفس الوقت لن تتضمن نتائج البحث الكلمتين "حاربة، محراب" رغم أن جميع المفردات مشتقة من جذور واحدة هي "ح ر ب". والسبب هو أن المضامين الذي تعبر عنها "حاربة ومحراب" تختلف عن مضمون "حرب، حروب...". ميزة البحث وفقاً لهذه التقنية أنها تقدم نتائج بحث شاملة ودقيقة في نفس الوقت.
- جدير بالذكر أن جميع هذه التقنيات تمثل العمود الفقري لمحرك البحث فهي أساس عمليات الفهرسة والبحث.

رغم أن تقنيات المعالجة اللغوية الآلية هي أساس عملية الفهرسة، فإنها لم تؤثر على حجم الفهرس أو على سرعة الفهرسة؛ فهرسة صفحات حجمها 3 جيجا أنتجت فهرسا حجمه حوالي جيجا واحدة بسرعة قدرها 468484 كلمة / الدقيقة ، علما بأنه لم تطبق بعد تقنيات لتحسين الأداء وعلى رأسها بناء قاعدة بيانات خاصة بالنظام تتجاوب مع بنية الفهرس ومتطلباته كبديل لقاعدة البيانات المستخدمة حاليا وهي الـ SQL Server و هي قاعدة بيانات عامة.

مما لا شك فيه أن التقنيات اللغوية المطبقة تواجه بعض المشاكل وتتطلب مزيدا من المراجعة والتطوير وهو ما سنقوم به في المرحلة القادمة.

و الشكل التالي يضم نموذج من الاستعلامات ونتائج البحث عنها.

COLTEC

Search Engine Prototype

صراعات سياسية - حزبية

Writing mode: ☐ Strict ☐ Relaxed ☒ Common-Error

Search technique: ☐ Affixation ☒ Inflectional ☐ Morpho-Conceptual

☐ Exact Search ☒ Enable Search Options for Excluded Words

Search

Search for صراعات سياسية - حزبية with Writing mode = Common-Error and Search technique = Inflectional

318 result(s)

1 2 3 4 5 6 7 8 9 10 ... >>	View	الحكم .. البيت رى الرؤد .. هيريس 0865 عدد هيريس البيت دلوحي بنابيك وكس و كزاية و صراعة حرة و صراعية امر بن سني كفاك .. ناقصها ايه البيت عشاق نوب كل من الرؤد في التمسبوتول http://imgsrc.dnry.com/logo_2004_200577
View	الإدارة التنفيذية ، وواجهة العملاء ، وإنتاج العملاء ، ولا يمكن علم أكثر معروف الصراع الذي تار في تلك الأيام والذي طاق أي الرئيس نفسه ، والذي لم يكن ليخرج في	
View	تطاول الصراع بين الألقاب المتعارضة وهو غري ساذ وهو الأمر الذي يتناول مع الوقت الصراع بين مسلم ومسيحي وليس من ضرر ومصرى وهذا الألقاب في تفسير القاموس كل في صالح	
View	التأثير من التفسيرات العامة لا تملك أن تحلها مرة أخرى إلى السير في التاريخ وقد أسس السياسي من المبروكي تعاليم السياسة لا تملك أن تخرج له مخرجا فسياسة لم يتأخذه رؤس	
View	و الأتني والحاكم والحكوم والعني والعنير تكامل وليس صراعاتها إلا أن الصراع أصل الاشتراك بين ما إله في فوضف و تمت ومنتبت 4 # posted by القلب	
View	بواسطة رمضان منذ ساعة واحدة 295,009 47,244 الصراعات السياسية الأقسام الفرعية: إمكانية التفسيرية ساحة الأخبار بجمع أصنافا	
View	شعرب عريف جازنا لعل تلك "الطوائف" الشعبية، وإليك الصفحة التي تعرف أفراد كل "فرع" أو "مذهب" معتبرات وصوب فليخربوا أو ما يوقونها إليه، ليس أقل في تغير وجوان تطويعم	
View	الماضية الحزبية الصراعات الصراعات السياسية 2005 الصراعات الحزبية 2005 الصراعات الحزبية 2005 الصراعات الحزبية 2005 الصراعات الحزبية 2005 الصراعات الحزبية 2005	
View	تفسيرات لبريتانيون محدودة لاستخدام شباب العالم فئات وصورة المصيرين في ذوق الصراعات الحزبية الاسترقاق إضافة إلى أمريكا الجنوبية (بوليفيا، بيرو، إلخ) هري	
View	تعدلتها التفسيرية القاطرة ضحا أو التفسيرية، بل كل تلوغ فئات الرئيس والتجديد والذي في سياسات الخشاع والفتك على الذي الذي يتأخذه العرب في صراعاتها وفي كل نمو وكس وريب	
View	تذكر إلى القديس ... رقة السياسية الفلسطينية عربيا "الحزب" كتب "مذهب جبهة جبهة السياسية" ككثيرة : فلسطين حرك شعار راحة الأجداد وولنته الأجداد، كان يعني لتطويعه	
View	تطاول الصراع بين الألقاب المتعارضة وهو غري ساذ وهو الأمر الذي يتناول مع الوقت الصراع بين مسلم ومسيحي وليس من ضرر ومصرى وهذا الألقاب في تفسير القاموس كل في صالح	
1 2 3 4 5 6 7 8 9 10 ... >>		

شكل 6. نتائج البحث لمفردات (صراعات سياسية حزبية) بطريقة (البحث التصريفي المترابط)



شكل 7. نتائج البحث لمفردات (صراعات سياسية-حزبية) بطريقة (البحث بالزوائد)

5 النتائج

اللمام العنكبوتي المطور في هذا العمل يتم استخدامه لمتابعة الشبكة العربية العنكبوتية . امتلك هنا اللمام العنكبوتي 300 خيطاً، وتم تشغيله لمدة شهر على اتصال بالانترنت 2 ميجا بايت / الثانية.

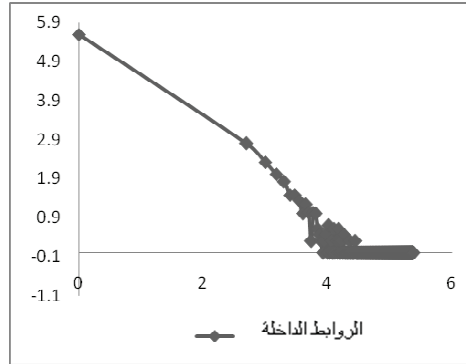
جدول 1. أسماء النطاقات الإلكترونية و عددها ونسبتها المئوية

اسم امتداد النطاق	عدد المواقع الإلكترونية	النسبة المئوية (%)
com	813795	81.3795
net	97719	9.7719
org	28895	2.8895
sa	14848	1.4848
ws	8157	0.8157
info	6657	0.6657
cc	5124	0.5124
eg	2329	0.2329
tv	1890	0.189
me	1719	0.1719
ae	1705	0.1705
kw	1662	0.1662

العدد الكلي لعناوين المواقع التي تم تحميلها تدور حول مليون صفحة بحجم كلي 45 جيجا بايت للبيانات. تم عمل تحليل لمجموعة عناوين المواقع الإلكترونية لإلقاء الضوء على الشبكة العنكبوتية العربية. لقد استخرجنا الموقع الذي ينتمي لامتداد اسم نطاق بعينه والذي يمثل المجال الأخير في النطاق (مثل com, net, org, eg, ... إلخ) . الجدول رقم "1" أعلاه يوضح أسماء النطاقات الإلكترونية المستخرجة مع العدد الكلي للمواقع الإلكترونية و النسب المئوية المناظرة .

و من الجدير بالملاحظة أن الامتداد الأكثر شهرة هو ".com" والذي يشكل ما يزيد عن 80 % من العدد الكلي للامتدادات. من الملاحظ أيضاً أن الامتدادات الخاصة بالدول بشكل عام (مثل eg, .sa, .ae, .kw) لا تشكل أكثر من قليل من النسبة المئوية و تملك المملكة العربية السعودية فيها نصيب الأسد و مصر فقط 0.23 بالمائة.

على الصعيد الآخر ، قمنا بدراسة اتجاهات الروابط الخاصة بالصفحات العربية الإلكترونية على الشبكة العنكبوتية التي تم تحميلها. لقد قمنا بتعيين عدد العقد لدرجة n المعطاة في مقابل قيم درجة-الدخل (شكل 8) . توضح العلاقة توزيعاً من نوع قانون الأس للصيغة $1/d^n$ مع قيمة ألفا لتساوي 2.1 تقريباً. هذه النتيجة تعتبر متوافقة مع نظيرتها للشبكة العنكبوتية ككل . هذا يعني أن الشبكة العنكبوتية العربية تتحى منحاً مشابهة لباقي الشبكة العنكبوتية الدولية .



شكل 8 . توزيع درجة-الدخل للمواقع الإلكترونية العربية التي تم تتبعها باللمام العنكبوتي

6 الخاتمة

في هذا البحث العلمي قمنا بتقديم النتائج الخاصة بعملنا عند تطوير النموذج الأولي لمحرك البحث العربي المعتمد على المعالجة اللغوية الآلية. وقمنا بالتركيز على وحدتين حيث أنهما اللذان انتبهنا من تطويرهما بقدر معقول من التفصيل. اللامم العنكبوتي الكامل التوزيع هو الأكثر نضجاً و تطوراً حالياً . لقد قمنا باستخدام هذه الوحدة لتحميل واحد مليون صفحة عربية إلكترونية على الشبكة العنكبوتية. ما زال العمل جارياً على هذه الوحدة لتحميل الشبكة العربية العنكبوتية (القابلة للتحميل). كما قمنا كذلك ببعض الدراسات حول الجزء الهام من الشبكة العربية العنكبوتية الذي يجب علينا حالياً تحليله ألا وهو المحتوى. نعمل كذلك على المزيد من المداخل الأخرى التنقيبية في الشبكة العربية العنكبوتية بغرض التصنيف. أما فيما يخص خوارزمية ترتيب الصفحة فإن عملنا إلى هذه النقطة ما زال مهتماً بتطوير خوارزمية ترتيب الصفحة وذلك بالاعتماد على تلك المقترحة من قبل العالم بيج والتي لها أداء جيد جداً مع تزايد عدد المعالجات. لقد قمنا باختبار

خوارزمتنا الخاصة بترتيب الصفحة ووجدناها ذات أداء جيد حتى سبع ماكينات دون أن نلاحظ أي تدهور واضح في الأداء .
 العمل المتوقع والمنتظر فيما يخص خوارزمية ترتيب الصفحة هو تطوير خوارزمية يتم تحسينها بواسطة إدخال المعالجة اللغوية.
 لم نناقش تفصيلاً عملنا للتطوير الكامل التوزيع لوحدة الفهرسة حيث أن المفهرس الكامل التوزيع الذي نعمل عليه حالياً ذو وحدة معالجة مفردة من مؤسسة كولتيك [12] (شريكنا الصناعي في هذا المشروع) وهو يعمل على الماكينات التي تحوى الصفحات التي تم تحميلها بواسطة اللام العنكبوتى. العمل الحالي على وحدة الفهرسة يشمل مُفهرساً تعاونياً كامل التوزيع.

المراجع

- [1] Sergey Brin , Lawrence Page, The anatomy of a large-scale hypertextual Web search engine, Proceedings of the seventh international conference on World Wide Web 7, p.107-117, April 1998, Brisbane, Australia
- [2] A. Arasu, J. Cho, H. Garcia-Molina, A. Paepcke, S. Raghavan, "Searching the Web," ACM Transactions on Internet Technology, 1: August 2001.
- [3] D. Ezzat, M. Abdeen, M. F. Tolba, "A Memory Efficient Approach for Crawling Language Specific Web: The Arabic Web as a Case Study," In Proceedings of the International Conference on Information Management and Engineering, April 3 - 5, 2009.
- [4] M. Abdeen and M. F. Tolba, "Challenges and design issues of an Arabic web crawler", the Sixth International Conference on Computer Engineering and Systems ICCES'10, Dec. 2010, Cairo, Egypt.
- [5] M. Abdeen , Doaa E. Mohamed, Mohamed F. Tolba, "An Enhanced Performance Distributed Arabic Web Crawler", the International Journal of Intelligent Computing and Information Science , Ain Shams University , Egypt , June 2010.
- [6] J. Cho, and H. Garcia-Molina, "Parallel crawlers," In Proceedings of the 11th World Wide Web, pages 124–135, 2002.
- [7] Allen Heydon and Mark Najork, "Mercator: A Scalable, Extensible Web Crawler". The Journal of World Wide Web, Springer Link, 1999.
- [8] V. Shkapenyuk and T. Suel, "Design and implementation of a high-performance distributed web crawler," In Proceedings of the 18th International Conference on Data Engineering, 2002.
- [9] J. Cho, and H.Garcia-Molina, "Effective page refresh policies for Web crawlers," ACM Transactions on Database Systems, 28(4): December 2003.
- [10] G. Singh Manku, A. Jain, and A. Das Sarma, "Detecting NearDuplicates for Web Crawling," World Wide Web, May 8–12, 2007.
- [11] Debora Donato, Luigi Laura, Stefano Leonardi, Stefano Millozzi, "Simulating the Webgraph: A Comparative Analysis of Models," Computing in Science and Engineering, vol. 6, no. 6, pp. 84-89, Nov./Dec. 2004.
- [12] Coltec Computer and Language Technology, www.coltec.net

Title: An NLP Based Fully Distributed Arabic Search Engine

Abstract: The world-wide-web has become the favorite destination of information seekers across the globe. With its massive amount of information that includes billions of web pages, information for just about any topic is a click-of-a-finger away.

The Arabic web represents an important portion of the world wide web. With Arabic as the 5th most spoken language in the world, with the increasing number of Arabic internet users at exponential rates, and with the special nature of the Arabic language, it is becoming important to introduce a search engine that suits better searching the Arabic web. Many search engines exist such as Google, Msn, and Yahoo. However, these search engines are not designed with proper consideration of the Arabic language and its special nature.

This paper presents some of our work while developing a prototype of an Arabic search engine. This search engine is designed to be fully distributed to be able to scale up and serve the huge number of users expected to use this engine. The paper presents the work done on the modules of the search engine, namely, the crawler, the page-rank, the query, and the indexing components. At the heart of the search engine is an NLP- based module that integrates with both Indexing and query components, which makes this engine significantly different from other ones available.

Keywords - component; Search Engine; NLP; Information retrieval; Arabic Web; Distributed Systems.

المصطلحات

الشبكة العنكبوتية	World Wide Web (Web)
المُعالجة اللغوية الآلية	NLP
النطاق الإلكتروني	Domain Name
الأداء الأنتماسي	Normalized performance
وحدة الفهرسة	Indexer
اللمام	Crawler
وحدة الاستعلام	Query Engine
مرتب الصفحات	Page Rank
محرك البحث	Search Engine
المعالج اللغوي	Linguistic processor
درجة الدخل	In-Degree
درجة الخرج	Out-degree
ثنائية المعالج	Dual-Core
طول العمر	Longivity
وحدة البرمجة	Module
المستوى الكتابي الدقيق	Strict Writing Mode
المستوى الكتابي المقبول	Relaxed Writing Mode
الأخطاء الكتابية الشائعة	Common Error Mode
البحث العشوائي الذكي	Intelligent Wildcard
البحث بالزوائد	Affixional
البحث التصريفي المترابط	Inflectional

مشروع نموذجي لحوسبة الصناعة العربية الرقمية في البحث والتعليم

أحمد الفرشيشي¹، روضة مروان¹، مرح الطاهر¹، أمل الميلي²

¹ جامعة تونس، المعهد العالي للتصرف بتونس، تونس

² جامعة بنسلفانيا، فيلادلفيا، و م أ

ahmed.ferchichi@planet.tn¹, raoudha.marouane@isg.rnu.tn¹,

marah.altaher@gmail.com¹, amel@themilis.net²

الخلاصة. إستجابة للنسق السريع الذي تتطور به المعارف الرقمية، ومساهمة في تطوير الصناعة العربية الرقمية، نقترح عبر هذه المقالة مشروع إنجاز بوابة رقمية لحوسبة الصناعة الرقمية العربية، نركز على العمل الجماعي العنكبوتي المشترك وتقانة المعلومات والإتصالات. لتفعيل البوابة المقترحة، نناقش المقالة العديد من الأنشطة النموذجية الفرعية وتبرز فوائد إنجازها.

الكلمات الجوهرية: الشبابة، المحتوى الرقمي، الصناعة الرقمية، العمل الجماعي، اللغة العربية، البحث والتعليم، بوابة الزيأتين.

1 مقدمة

تعتبر الشبابة [1] اليوم وسيلة عمل أصبح من الصعب الاستغناء عنها في مختلف الأنشطة الحياتية، حيث نراها تتجه بخطى سريعة لتصبح أكبر مكتبة للمعارف الإنسانية. إذا اهتمنا بأشكال المعرفة التي يحتاجها المتعلم، نرى أنه يمكن تلخيصها عبر خمسة لغات: لغة طبيعية يستعملها المتعلم للتواصل مع الآخرين (كاللغة العربية مثلاً)، لغة تخصصية يستعملها المتعلم لخدمة المجتمع الذي يعيش فيه (كالطب مثلاً)، لغة بحثية يستعملها المتعلم لتحصيل المعرفة عندما تكون هذه اللغة سائدة (كاللغة الإنجليزية اليوم مثلاً)، لغة مدنية يستعملها المتعلم للتفاهم مع الآخرين (كميثاق عيش مشترك بين الناس)، ولغة وسيلة يستعملها المتعلم لتسهيل أعماله (كتقانات المعلومات والاتصالات اليوم مثلاً). في هذا الإطار، تفرز الشبابة المشاهد التالية:

مشهد 1

تتطور المعرفة ولا تتطور مناهج البحث والتعليم بالنسق المناسب.

مشهد 2

يتطور المحتوى الرقمي باللغة الإنجليزية ولا يتطور المحتوى الرقمي باللغة العربية بالنسق المناسب.

مشهد 3

تتطور وسائل التكنولوجيا والاتصال ولا تتطور طرق البحث والتعليم بالنسب المناسبة.

تشير المشاهد السابقة إلى أن الشبكة تمثل بالنسبة للبحث والتعليم هدفاً ووسيلة، فهي هدف لأنها تتضمن المعرفة التي يمكن أن يستهدفها البحث والتعليم، وهي وسيلة لأنها تتضمن تقانات المعلومات والاتصالات التي يمكن أن يستعين بها البحث والتعليم. هذا، ويرتبط التعليم بالمعرفة وفق علاقة يساعد التعليم فيها المتعلم على تحصيل المعرفة، ويساعد البحث فيها المتعلم على اكتشاف المعرفة، وتساعد المعرفة فيها المتعلم على حل المسائل. تشير المشاهد العامة المشار إليها سابقاً، إلى اختلال العلاقة بين البحث والتعليم في العديد من المجالات المعرفية. نلاحظ على سبيل المثال في مجال الحوسبة الآلية [2]، كيف أن المعرفة تتطور بنسق كبير وأن التعليم يعجز على التطور بالنسق المقابل المقبول. نلاحظ أيضاً في مجال اللغة العربية، كيف أن المعرفة تنسم بالجمود وأن التعليم يعجز على ترسيخ اللغة كرمز ثقافي هام. من ناحية أخرى، تشكو طرق التعليم من صعوبة إدارة العلاقة بين مختلف متغيرات التعليم (خصائص المتعلمين، طبيعة البيئات التعليمية، أهداف التعليم، محتوى التعليم، طرق التدريس، طرق التقييم...). وللتدليل على صعوبة إدارة متغيرات التعليم، نذكر الصعوبات الهائلة التي تعترض جمعية (أ س م) للحوسبة الآلية في كل محاولة تحيين / تطوير للتوصيات المتعلقة بمناهج تعليم الحوسبة الآلية [2]. تبعاً لذلك، نستنتج أنه يصعب كثيراً على الأفراد والجماعات الصغيرة أن تتواصل مع النسق الذي تنمو به المعارف والنسق الذي يستوجب أن تتطور به مناهج التعليم، فيترتب عن ذلك العديد من الأسئلة:

سؤال 1

كيف يمكن تطوير البحث والتعليم بالتوازي مع تطور المعرفة الرقمية؟

سؤال 2

كيف يمكن تطوير المحتوى الرقمي باللغة العربية بالتوازي مع تطور المحتوى الرقمي باللغة الإنجليزية؟

سؤال 3

كيف يمكن تطوير البحث والتعليم بالتوازي مع تطور تقانات المعلومات والاتصالات؟

للإجابة عن الأسئلة السابقة، ندعو للتخلي عن أساليب البحث والتعليم التقليدية، واعتماد أساليب عمل جديدة. تبعاً لذلك، نقترح إنجاز بوابة رقمية نموذجية تشير لها بـ **بوابة الزياتين**، من أجل استعمالها كوسيلة لحوسبة الأنشطة المتعلقة بالصناعة العربية الرقمي في مجال البحث والتعليم. يركز المقترح على:

- تجربة سابقة في مجال الحوسبة الآلية، تجدها العديد من المؤلفات والملفات [3]،
- تجارب دولية سابقة مثيلة، لعل أهمها مشروع ويكي الجامعة [4]،
- عينة من أمثلة التصنيع الرقمية التي يمكن إنجازها عبر البوابة المقترحة (أنظر الجزء 6).

تتبع المقالة الخطوات التالية: (1) في الجزء الثاني، نعرض نظام العمل الذي يحدد هدف البوابة وطريقة العمل المقترحة لتصنيع المحتوى الرقمي باللغة العربية؛ (2) في الجزء الثالث، نعرض واجهة البوابة التي تمكن لمختلف الفاعلين أن يفقدوا مشاريع تصنيع جديدة أو ينخرطوا في مشاريع عمل جارية؛ (3) في الجزء الرابع، نجلب الإنتباه إلى تنوع عمليات التصنيع الرقمي وضرورة توفير الأدوات العنكبوتية ذات العلاقة؛ (4) في الجزء الخامس، نحدد أنواع المنتجات الرقمية الموجب تصنيعها عبر البوابة خدمة للبحث والتعليم؛ (5) في الجزء السادس، نناقش مجموعة من أنشطة التصنيع الرقمي التي يمكن أن تتجز بنجاعة عبر البوابة؛ (6) في الجزء السابع، نحسب الحمولة التي تترتب عن إنتاج بعض مناهج البحث والتعليم للتدليل على أهمية العمل الجماعي العنكبوتي المشترك؛ (7) في الجزء الثامن، نقترح مخططا عاما لإنجاز البوابة وبعض المنتجات الرقمية النموذجية.

2 ميثاق التصنيع الرقمي

نظرا أننا نأمل أن تحظى بوابة الزيتتين (موضوع هذه الورقة) بإجماع واسع في أوساط البحث والتعليم، ونظرا لطبيعة المؤلفات الرقمية، نتوقع أن تلعب البوابة المقترحة أدورا عديدة، فتكون: مرصدا للمستجدات، مخبرا للبحوث، جامعة للتعليم، دارا للنشر، منبرا للنقاش، ومكتبة للمؤلفات. تبعا لذلك، يستوجب أن تحظى البوابة بميثاق خاص ينظم العلاقة بين مختلف الفاعلين المهتمين بالتصنيع الرقمي، نعرضه عبر الرسم 01.

3 واجهة التصنيع الرقمي

إنطلاقا من نص الميثاق المبين سابقا، نقترح أن تكون واجهة بوابة الزيتتين على الصورة المبينة في الرسم 02.

7. تنشر البوابة منتجاتها على شكل موسوعات رقمية، بعد عمليات التقييم والمراجعة الضرورية شكلاً ومضموناً، من قبل الخبراء المختصين، وتشارك في المؤتمرات الدولية ذات العلاقة للتعريف بمنتجاتها الرقمية، وللتعريف بالخدمات التي يمكن أن يقدمها الخبراء لمؤسسات البحث والتعليم في العالم العربي وخارجه. 8. تستعمل البوابة تقانات المعلومات والاتصالات ولا سيما وسائل العمل الإلكتروني المشترك لإنجاز مهامها. 9. تفتح البوابة نوافذ لـ الشراكة مع جهات البحث والتعليم الغير ناطقة باللغة العربية. 10. تعتبر البوابة المقاييس والمعايير الدولية المعتمدة في تحرير الكتب والرسائل الجامعية والمقالات والكيونات التعليمية، من أجل تسهيل: رصدها، اكتشافها، تجميعها، تبويبها، تقييمها، ترجمتها، تأليفها، إخراجها، نشرها، إشهارها، وإسترجاعها.	1. تعتبر بوابة الزياتين مؤسسة غير ربحية. 2. تهدف البوابة إلى المساهمة في تطوير الصناعة العربية للمحتوى الرقمي. وتشمل مهامها الرصد والبحث والتعليم والنشر والتوزيع. 3. تنجز البوابة هدفها على شكل مشاريع مشتركة بين الخبراء، حيث يعتمد كل باحث مشروع سياسة العمل التالية: يحجز الباحث فضاء رقمياً للمشروع، يدعو الخبراء للإنخراط في المشروع، ويلتزم باستشارة المرجعيات العلمية ذات العلاقة، ويطور المنخرطون منتجاتهم بصفة تفاعلية وفورية. 4. ينخرط في البوابة كل حامل لشهادة الماجستير أو الدكتوراه، يهتم بممارسة البحث والتعليم باللغة العربية، يؤمن بالعمل الجماعي المشترك، ويلتزم بخصائص المشاريع التي ينخرط فيها. 5. تشرف على إنجاز هدف البوابة هيئة إشراف وتشرف على كل مشروع هيئة تنفيذية من خبراء البوابة. 6. تتكون موارد البوابة من دعم الجهات المؤمنة بدورها، وكذلك من مقابل الأعمال التي ينجزها الخبراء، وتستعمل البوابة مواردها المالية المتوفرة لتسريع إنجاز مشاريعها.
رسم 01: ميثاق عمل بوابة الزياتين للصناعة العربية للمحتوى الرقمي في البحث والتعليم	

إختيار لغة التواصل إختيار مجال البحث والتعليم الترحيب بالزائرين تقديم ملخص عن خدمات البوابة الدعوة للتسجيل	الترحيب بالزائرين																
عرض مشاريع البحث والتعليم الجارية الدعوة للتسجيل في المشاريع الجارية الدعوة لبعث مشاريع مستقبلية حجز فضاءات للعمل تجهيز فضاءات العمل <table border="1" data-bbox="224 384 909 588"> <tr> <th colspan="2">أنواع الانخراط</th></tr> <tr> <td>باعت</td><td>يقترح مشاريع عمل ويدعو للانخراط فيها</td></tr> <tr> <td>مشرف</td><td>يسير مشاريع عمل</td></tr> <tr> <td>باحث</td><td>يشارك في إنجاز مشاريع عمل</td></tr> <tr> <td>مدرس</td><td>يدرس في نطاق مناهج تعليم</td></tr> <tr> <td>متعلم</td><td>يتابع مناهج تعليم</td></tr> <tr> <td>داعم</td><td>يدعم مشاريع العمل</td></tr> <tr> <td>زائر</td><td>يتابع مشاريع البوابة</td></tr> </table>	أنواع الانخراط		باعت	يقترح مشاريع عمل ويدعو للانخراط فيها	مشرف	يسير مشاريع عمل	باحث	يشارك في إنجاز مشاريع عمل	مدرس	يدرس في نطاق مناهج تعليم	متعلم	يتابع مناهج تعليم	داعم	يدعم مشاريع العمل	زائر	يتابع مشاريع البوابة	التسجيل في مشاريع البحث والتعليم
أنواع الانخراط																	
باعت	يقترح مشاريع عمل ويدعو للانخراط فيها																
مشرف	يسير مشاريع عمل																
باحث	يشارك في إنجاز مشاريع عمل																
مدرس	يدرس في نطاق مناهج تعليم																
متعلم	يتابع مناهج تعليم																
داعم	يدعم مشاريع العمل																
زائر	يتابع مشاريع البوابة																
عرض حالة تطور المشاريع تحيين خارطات المشاريع	متابعة مشاريع البحث والتعليم																
<table border="1" data-bbox="224 683 922 1251"> <tr> <td>تتضمن بطاقات تسجيل المهتمين باستعمال البوابة، وفضاءات عملهم وفق المشاريع التي انخرطوا فيها.</td><td>مكتبة الباحثين</td></tr> <tr> <td>تتضمن بطاقات المشاريع الماضية والجارية والمستقبلية.</td><td>مكتبة المشاريع</td></tr> <tr> <td>تتضمن الوثائق القارة التي تحتاجها المشاريع والوثائق الصادرة عن إنجاز المشاريع.</td><td>مكتبة الوثائق</td></tr> <tr> <td>تتضمن تقانات المعلومات والاتصالات التي يستحقها المنخرطون في المشاريع لتساعدهم على إنجاز مهامهم.</td><td>مكتبة التقانات</td></tr> <tr> <td>تتضمن منتديات النقاشات التي تبيح لمناقشة نقاط معينة تثار خلال إنجاز مشاريع البحث والتعليم.</td><td>مكتبة المنتديات</td></tr> </table>	تتضمن بطاقات تسجيل المهتمين باستعمال البوابة، وفضاءات عملهم وفق المشاريع التي انخرطوا فيها.	مكتبة الباحثين	تتضمن بطاقات المشاريع الماضية والجارية والمستقبلية.	مكتبة المشاريع	تتضمن الوثائق القارة التي تحتاجها المشاريع والوثائق الصادرة عن إنجاز المشاريع.	مكتبة الوثائق	تتضمن تقانات المعلومات والاتصالات التي يستحقها المنخرطون في المشاريع لتساعدهم على إنجاز مهامهم.	مكتبة التقانات	تتضمن منتديات النقاشات التي تبيح لمناقشة نقاط معينة تثار خلال إنجاز مشاريع البحث والتعليم.	مكتبة المنتديات	فضاءات مشاريع البحث والتعليم						
تتضمن بطاقات تسجيل المهتمين باستعمال البوابة، وفضاءات عملهم وفق المشاريع التي انخرطوا فيها.	مكتبة الباحثين																
تتضمن بطاقات المشاريع الماضية والجارية والمستقبلية.	مكتبة المشاريع																
تتضمن الوثائق القارة التي تحتاجها المشاريع والوثائق الصادرة عن إنجاز المشاريع.	مكتبة الوثائق																
تتضمن تقانات المعلومات والاتصالات التي يستحقها المنخرطون في المشاريع لتساعدهم على إنجاز مهامهم.	مكتبة التقانات																
تتضمن منتديات النقاشات التي تبيح لمناقشة نقاط معينة تثار خلال إنجاز مشاريع البحث والتعليم.	مكتبة المنتديات																
<table border="1" data-bbox="224 1289 922 1410"> <tr> <td>تتضمن الموسوعات الرقمية التي تنمها مختلف المشاريع</td><td>مكتبة الموسوعات</td></tr> </table>	تتضمن الموسوعات الرقمية التي تنمها مختلف المشاريع	مكتبة الموسوعات	منتجات مشاريع البحث والتعليم														
تتضمن الموسوعات الرقمية التي تنمها مختلف المشاريع	مكتبة الموسوعات																
رسم 02: واجهة بوابة الزائرين للصناعة الرقمية في البحث والتعليم باللغة العربية																	

4. مخبر التصنيع الرقمي

تتمثل مهمة مخبر التصنيع الرقمي في توفير بيانات العمل الضرورية من أجل أن تحقق البوابة أهدافها المرسومة في الميثاق. نظرا لطبيعة أنشطة ومنتجات بوابة الزياتين، يستوجب أن توفر البوابة لمستخدميها تقانات المعلومات والاتصالات المتعلقة بالعمليات التالية: الرصد، الاكتشاف، الاسترجاع، التجميع، التبيب، التقييم، الترجمة، التأليف، الإخراج، النشر، الإشهار، والتوزيع. يستوجب على التقانات المستعملة أن تأخذ بعين الاعتبار المقاييس والمعايير المعتمدة في تحرير المؤلفات (كتب، مجلات، رسائل جامعية، مقالات، كينونات تعليمية)، من أجل تسهيل معالجتها من طرف البرمجيات ذات العلاقة (محركات بحث، بيئات تعليم، ...).

5. مخبر الموسوعات الرقمية

يتمثل مخبر الموسوعات الرقمية، في تنظيم منتجات البحث والتعليم المطورة عبر بوابة الزياتين، على شكل موسوعات معرفية. ندعو في مرحلة أولى بأن يهتم العمل الموسوعي في مختلف المجالات المعرفية بأنواع الموسوعات التالية: موسوعة الخبراء، موسوعة البحوث، موسوعة الترجمات المتميزة، موسوعة المصطلحات العلمية، موسوعة البرمجيات الحاسوبية، وموسوعة مناهج التعليم. تجدر الإشارة إلى أن الموسوعات الرقمية على الشبكة تتميز بسرعة التطوير والتحسين والمراجعة الفورية التفاعلية مقارنة بالموسوعات الورقية التقليدية، الأمر الذي يبقى مرهونا فقط بدعوة الباحثين المهتمين للإنخراط في المشاريع الفرعية المبرمجة للغرض.

6. أنشطة مستقبلية

نعتبر في هذا الجزء، بعض مواضيع البحث والتعليم التي يمكن أن تنجز بواسطة بوابة الزياتين، نعرضها للأغراض التالية:

- تأكيد أهمية إنجاز مثل تلك المواضيع على شكل مشاريع مشتركة تنجز بواسطة البوابة،
- إعداد عينة من مواضيع البحث والتعليم يمكن استعمالها لتشجيع تمويل البوابة والمشاركة في مشاريعها.

1.6 تطوير موسوعة الخبراء

يتمثل هذا النشاط في إدارة موسوعة رقمية تفاعلية للخبراء المهتمين بممارسة البحث والتعليم باللغة العربية. يمكن تحقيق هذا النشاط بفسح المجال للباحثين للتسجيل عن بعد. تجدر الملاحظة إلى أن الجهات العربية الرسمية تجد صعوبات للتواصل مع الخبراء العرب المتواجدين داخل الوطن العربي وخارجه.

2.6 تطوير موسوعة للمصطلحات العلمية

يتمثل هذا النشاط في إدارة موسوعة رقمية تفاعلية للخبراء المهتمين بممارسة البحث والتعليم باللغة العربية. يمكن رصد المصطلحات من خلال متابعة نداءات المشاركة التي تنشرها المؤتمرات. يجدر أيضا بفرق العمل المهمة بشأن المصطلحات أن تكون على اتصال بالمرجعيات العلمية العربية وغيرها لتستعين بها خلال عملية ترجمة المصطلح، حيث يجب أن تتوفر شروط خاصة، مثل: إطلاع المترجم على قواعد الترجمة، إنماء المترجم لحقل المصطلح، إطلاع المترجم على تعريف المصطلح باللغة الإنجليزية من مصادر متفق على قيمتها العلمية. يشمل هذا النشاط العمليات التالية: التعرف على مصادر المصطلحات، تجميع المصطلحات، وضع خطة عمل لرصد المصطلحات الجديدة، ترجمة المصطلحات، تقييم الترجمات المقترحة، إشهار موسوعة المصطلحات لدى الأوساط الأكاديمية والمهنية. يمكن تحقيق هذا النشاط جزئيا بصفة تفاعلية عن بعد عبر بوابة الزياتين. تجدر الإشارة إلى أن موسوعات المصطلحات العلمية باللغة العربية نجدها اليوم غير مستعملة وغير محيئة، هذا إن سمع الجمهور بوجودها أصلا. يمكن لهذا النشاط أن يعيد لها الاعتبار ويمكن من استعمالها عن بعد علاوة على تحيينها بصفة تفاعلية.

3.6 تطوير موسوعة لمواضيع البحوث

يتمثل هذا النشاط في إدارة موسوعة للبحوث تتميز بأنها: (1) تخدم مشكلات حياتية تنموية، (2) تتجز عبر العمل الجماعي المشترك لضمان جودة نتائجها وسرعة إنجازها بما يضاها المشاريع الدولية. يمكن تحقيق هذا النشاط بفسح المجال للباحثين أن يقترحوا مواضيع بحث عن بعد. تجدر الملاحظة إلى أن هذه الموسوعة سوف تغطي عجز التأخير الذي تشهده مؤسسات البحث والتعليم في العالم العربي سواء من ناحية نوعية مواضيع البحث أو من ناحية توفير المؤثرين.

4.6 تطوير موسوعة لمناهج التعليم

يتمثل هذا النشاط في إدارة موسوعة لمناهج التعليم، تتناسب مع المتطلبات المهنية المستحدثة، والخدمات التي توفرها تقانات المعلومات والاتصالات. إذا اهتمنا بقطاع التعليم الجامعي على سبيل المثال، نلاحظ أن النظام التعليمي الجامعي الأوروبي الموحد الجديد (الذي هو أصلا مستوحا من منظومة التعليم الجامعي الأمريكية) يقسم الشهادات الجامعية إلى ثلاثة أنواع [5]: شهاد الإجازة وشهاد الماجستير وشهاد الدكتوراه، إذ تدوم الإجازة ثلاثة سنوات وتتطلب 30 وحدة تعليمية سداسية ويوم الماجستير سنتين ويتطلب 20 وحدة تعليمية سداسية وتدوم الدكتوراه ثلاثة سنوات وتتطلب 30 وحدة تعليمية سداسية. إضافة إلى ذلك، تتطلب الوحدة التعليمية ما يقارب 45 ساعة تعليمية حضورية و 90 ساعة تعليمية غير حضورية. إذا اعتبرنا أن الوحدات التعليمية متساوية من حيث الأرصدة و المدة الزمنية المخصصة لها، وإذا اعتبرنا أن كل وحدة تعليمية تشمل 10 مباحث (مفردات)، نستنتج أن الإجازة تتضمن 300 مبحثا وأن الماجستير يتضمن 200 مبحثا وأن الدكتوراه تتضمن 300 مبحثا. يشير الحساب السابق إلى أن كل مرشح لشهادة دكتوراه، يستوجب عليه أن يجابه 80 وحدة تعليمية أو 800 مفردة تعليمية. من ناحية الشكل، لا بأس أن تعتمد الموسوعة بالنسبة للمناهج الجامعية النظام الأوروبي، غير أنه من ناحية المضمون، ينبغي إيجاد حلول ناجعة لمجابهة المسائل التقليدية المتعلقة أساسا بفرز المفردات التعليمية والربط بينها وترتيبها [6] [7]. تبع لما ذكر، لا يمكن للأفراد والجماعات الصغيرة أن تعمل بصفة منعزلة لتطوير مناهج

البحث والتعليم. يكفي أن نعمم التصور السابق لجميع المراحل العمرية وللتنخصصات التي يمكن أن تحتاجها المجتمعات العربية، لنفهم أنه لا حل خارج العمل الجماعي المشترك على نطاق واسع، وبوابة الزياتين يمكن أن تكون إطاراً أيضاً لمثل هذا العمل.

نقترح أن تكون مناهج التعليم الجامعية دقيقة التخصص، ترتبط ببحوث تشغل أوساط البحث، وتنجز وفق الخطة التالية: (1) تحديد أهداف مناهج التعليم، إنطلاقاً من مشاريع البحوث، (2) فرز المفردات التعليمية من مجالات البحوث، (3) تقييم المتعلمين بواسطة نشر البحوث، و(4) تحيين مناهج التعليم تبعاً لتطور مجالات البحوث.

نقترح أيضاً التفكير في تطوير مناهج التعليم انطلاقاً من اكتشاف المعرفة المختزلة في البرمجيات الحاسوبية، عبر المراحل التالية: (1) يختار الخبراء تطبيقاً برمجية، (2) يحللون البرمجية، (3) يستخلصون المصطلحات، (4) يكتبون التعيينات، (5) يطورون الدروس، (6) يطورون حالات الدراسة، (7) يحددون البحوث ذات العلاقة، (8) يدونون أدب المصطلحات، و(9) ينظمون المؤتمرات العلمية ذات العلاقة.

5.6 تطوير موسوعة للترجمات المتميزة

يستعمل الفرد أساساً لغة المجموعة البشرية (قبيلة، شعب، أمة) التي ينتمي إليها، ثم يستعمل لغات المجموعات البشرية الأخرى عندما يحتاج للتواصل معها، مثلاً يستعمل الترجمة كوسيلة لنقل المعرفة. بالنسبة لمشاريع الترجمة، نقترح إعطاء الأولوية للمشاريع التي تستجيب أكثر لمتطلبات البحث والتعليم في العالم العربي.

يجدر الذكر إلى أنه في نطاق المشاريع الطلابية، توصلنا بفضل مجهودات الطلبة إلى ترجمة محاضرات الفائزين بجائزة تورينق للحوسبة الآلية (المساوية لجائزة نوبل) التي بدأ منحها منذ سنة 1966. تتضمن تلك المحاضرات تجارب ونصائح مفيدة، في عصر تراكمت واختلطت فيه المعرفة. نستخلص من هذا العمل ثلاثة فوائد: (1) تنشيط اللغة العربية عبر الممارسة العلمية في الأوساط الجامعية، (2) تشريك الطلبة المتقدمين للمساهمة في مجهود الترجمة من اللغات الأخرى للغة العربية، (3) إمكانية استعمال المحاضرات كنقطة انطلاق للمصطلحات العلمية، نظراً لما يتسم به أصحاب تلك المحاضرات من تجارب ومن إسهامات علمية: بالنسبة لمجال الحوسبة الآلية، إذا ضربنا عدد المحاضرات (55) بعدد المفردات العلمية في المحاضرات، نحصل على ما يقارب 550 مصطلحاً علمياً يمكن أن تنطلق منها موسوعة مصطلحات الحوسبة الآلية. زيادة على ذلك، يمكن الإتصال بأصحاب تلك المحاضرات من أجل الحصول على معلومات إضافية عبر استمارات ترسل إليهم أو ندوات افتراضية يدعون إليها، خصوصاً وأن الأغلبية الساحقة منهم مازالت على قيد الحياة.

6.6 تطوير مكتبة لبرمجيات البحث والتعليم

نظرا لما يتطلبه التصنيع الرقمي من أدوات حاسوبية، يمكن بعث مشاريع تطوير أدوات برمجية لمساعدة التصنيع الرقمي المتعلقة بـ: الرصد، الإكتشاف، الإسترجاع، التجميع، التوبيخ، التقييم، الترجمة، التأليف، الإخراج، النشر، الإشهار، التوزيع، توليد تمارين الدروس ومواضيع وإصلاح الإمتحانات، تقييم البحوث، تطوير مناهج الشهادت والمقررات والمفردات الدراسية.

7.6 تطوير بيئة لتعليم اللغة العربية

يواجه تعليم اللغة العربية تحديات كثيرة ومتنوعة، حيث أننا نلاحظ: (1) فشلا في إبراز اللغة كرمز ثقافي هام، (2) تدنيا في استعمال اللغة كأداة اتصال وتواصل على نطاق واسع، (3) صعوبة في تحسين نجاعة التعليم، (4) عجزا في تطوير مشاريع عمل مشتركة حول تطوير اللغة وتعليمها. (5) تأخرا في استفادة اللغة بوسائل التقانات والاتصالات. دأبت جهات عديدة للاهتمام بتعليم اللغة العربية، غير أن الجهود المبذولة تنسم بالتشتت، وبالتالي تفتقد القوة الكافية التي تسمح لنتائجها بالانتشار والاستثمار. توفر بوابة الزياتين بالمقابل المناخ المناسب للمشاورات والاستشارات والنقاشات حول تعليم اللغة العربية. في هذا الإطار، يمكن أن يستثمر التعليم البرامج الحاسوبية لتعليم المهارات اللغوية وفق الوضعيتين التاليتين:

متابعة المتعلمين للبرامج الحاسوبية وهي تؤدي المهارات اللغوية:

وفق هذه الوضعية: يختار المتعلم نصا لغويا ثم يتابع البرامج الحاسوبية في: قراءته، أو كتابته، أو تلوينه، أو تفكيكه، أو تركيبه، أو تشكيله، أو رفع تشكيله، أو تصريفه، أو إعرابه، أو ترجمته.

متابعة البرامج الحاسوبية للمتعلمين وهم يحاكون البرامج الحاسوبية في تأدية المهارات اللغوية:

وفق هذه الوضعية، يختار المتعلم نصا لغويا ثم تدخل البرامج الحاسوبية في مساعدته على محاكاتها مع الاستعداد لتقديم المساعدة كلما دعت الحاجة.

من المفيد الإشارة إلى أن مختلف الأنشطة المذكورة سابقا (قراءة، كتابة، تلوين، تفكيك، تركيب، تشكيل، رفع تشكيل، تصريف، إعراب، ترجمة، ...) يستوجب تنفيذها بالاستعانة بالحواسيب وفق طريقة تطويرية، تتجود فيها أهداف التعليم بصفة حلزونية، حيث في كل دورة، يمكن مراجعة موضوع سابق أو التعمق في موضوع سابق، أو إدماج موضوع جديد.

8.6 تطوير بيئة لتعليم هندسة البرمجيات

تمثل هندسة البرمجيات منذ السبعينات الحل الأمثل لمجابهة المعادلة الصعبة التي تربط بين جودة البرمجيات وتكلفتها وآجال إنجازها. ونظرا للنسق السريع الذي تتطور به المعارف في المجال، وتتغير به مناهج التعليم تبعا لذلك، يمكن أن تستعمل بوابة الزياتين لتطوير العديد من المناهج المختصة في هندسة البرمجيات من خلال تظافر جهود الباحثين المهتمين بالمجال.

1.8.6 مشكلة بحوث هندسة البرمجيات

تسعى البحوث في هندسة البرمجيات عموماً لتوفير السبل المعرفية التي تسمح للمشاريع البرمجية بأن تحقق المعادلة المنشودة بين متطلبات جودة البرمجيات من ناحية وضغوط الأجل والتكلفة المرتبطة بالإنتاج من ناحية أخرى. هذا ولا تزال طبيعة البرمجيات وكثرة تنوعها تصعب مهمة البحوث في المجال [8].

2.8.6 مشكلة مناهج تعليم هندسة البرمجيات

تشتمل مناهج تعليم هندسة البرمجيات مثل سائر المناهج الأخرى سواء في المعلوماتية أو غيرها من تشتمت المعارف في وحدات تعليمية كثيراً ما تتداخل مباحثها ويصعب ترتيبها والتنسيق بينها، وتشتمل مناهج تعليم هندسة البرمجيات بالأساس من غياب التجربة التطبيقية لدى المدرسين والتوازن بين الجوانب النظرية والجوانب التطبيقية [8].

3.8.6 تركيز التعليم على مناقشة المشاريع البرمجية

نقترح أن تركز خطط تعليم هندسة البرمجيات على الرؤيا المتمثلة في أن هندسة البرمجيات تحقق في الواقع صلة ثنائية بين المسائل البرمجية التي يمكن أن تطرح والحلول البرمجية التي يمكن أن تقابلها، بمعنى أننا نقترح أن يركز تطوير المناهج التعليمية على حالات دراسة فعالية تناقش بصفة تدريجية خصائص البرامج الحاسوبية (مجال التطبيق، طريقة التنفيذ، أسلوب المعمارية، الحجم، الجودة...) ومحتوياتها (لغات تمثيل، خوارزميات معالجة، تراكيب بيانات...) وعمليات تطويرها (توصيف، تصميم، تصديق...) [9].

4.8.6 تنويع مسارات التكوين

تفرّع مجال هندسة البرمجيات بسرعة فائقة خلال العقدين الأخيرين، بعد أن استقل عن علم الحاسب الآلي في السبعينات [8]. تبعاً لذلك، يمكن لمناهج التعليم أن تتأقلم مع هذا الواقع باقتراح تنويع مسارات التكوين كما نشرح ذلك في الرسم 03 مع مثال للمفردات المعرفية التي يمكن أن يعتمدها المنهج [7] [10].

بالنسبة لشهادة الإجازة	بالنسبة لشهادة الماجستير	بالنسبة لشهادة الدكتوراه
مقرر 1: منطق رياضي مقرر 2: رياضيات محددة مقرر 3: نظرية الحساب الآلي مقرر 4: خوارزميات وتراكيب بيانات مقرر 5: لغات البرمجة مقرر 6: أساليب البرمجة مقرر 7: توصيف البرمجيات مقرر 8: عبارة البرمجيات مقرر 9: تصميم البرمجيات مقرر 10: برمجة الحواسيب مقرر 11: إدارة المشاريع البرمجية مقرر 12: بيئات هندسة البرمجيات	مقرر 1: مقرر 2: مقرر 3: مقرر 4: مقرر 5: مقرر 6: تحدد مقررات الشهادة وفق التخصص	مقرر 1: مقرر 2: مقرر 3: مقرر 4: مقرر 5: مقرر 6: تحدد مقررات الشهادة وفق التخصص

7 حمولة المنتجات الرقمية

لنبرهن من جديد عن أهمية العمل الجماعي المشترك بواسطة بوابة الزياتين، نحسب عبر الرسم 04، عدد الدروس الرقمية الموجب تطويرها في نطاق ثلاثة أنواع من مناهج التعليم، وهي تباعا: منهج لتعليم هندسة البرمجيات موجه للطلبة الجامعيين، منهج لتعليم اللغة العربية موجه للناطقين باللغة العربية، منهج في اللغة العربية موجه لغير الناطقين باللغة العربية. يمثل بصفة جلية، عدد الدروس الموجب تطويرها في نطاق كل منهج استحالة توصيفها وتصميمها وتطويرها وتدريسها وتقييمها وتحسينها وتقييمها من طرف الأفراد والجماعات الصغيرة دون المساس بقيمتها [10] [7].

منهج الإجازة	منهج الماجستير	منهج الدكتوراه
30	20	30
450	300	450

نعتبر تخصيص 8 وحدات تعليم (من أصل 20) لتأطير رسائل الماجستير في هندسة البرمجيات (على أساس وحدتين في كل فصل دراسي)، وتخصيص 18 وحدة تعليم (من أصل 30) لتأطير رسائل الدكتوراه في هندسة البرمجيات (على أساس ثلاثة وحدات في كل فصل). الأمر الذي يعني أنني نكتفي بإنتاج 12 وحدة تعليمية (أو 180 درسا) للماجستير و 12 وحدة تعليمية (أو 180 درسا) لشهادات الدكتوراه أيضا

عدد الكتب الرقمية			عدد الدروس الرقمية		
منهج الإجازة	منهج الماجستير	منهج الدكتوراه	منهج الإجازة	منهج الماجستير	منهج الدكتوراه
30	12	12	450	180	180

54 كتاب رقمي = 810 درس رقمي

رسم 04 أ: حساب الدروس الرقمية لتعليم هندسة البرمجيات (إجازة+ماجستير+دكتوراه)

عدد الكتب الرقمية منهج التعليم الابتدائي (6 سنوات)	عدد الكتب الرقمية منهج التعليم الإعدادي (3 سنوات)	عدد الكتب الرقمية منهج التعليم الثانوي (4 سنوات)
6 كتب رقمية 180 درس رقمي (30x6)	3 كتب رقمية 90 درس رقمي (30x3)	4 كتب رقمية 120 درس رقمي (30x4)

13 كتاب رقمي = 390 درس رقمي

رسم 04 ب: حساب الدروس الرقمية لتعليم اللغة العربية للناطقين بها (إبتدائي+إعدادي+ثانوي)

عدد الكتب/الدروس الرقمية منهج مستوى المتقدمة (1 سنة)	عدد الكتب/الدروس الرقمية منهج المستوى المتوسط (1 سنة)	عدد الكتب/الدروس الرقمية منهج المستوى المتقدم (1 سنة)
1 كتاب رقمي = 30 درس رقمي	1 كتاب رقمي = 30 درس رقمي	1 كتاب رقمي = 30 درس رقمي

3 كتب رقمية = 90 درس رقمي

رسم 04 ت: حساب الدروس الرقمية لتعليم اللغة العربية لغير الناطقين بها (أول+ثاني+ثالث)

8 المقارنة بين ويكي جامعة وبوابة الزياتين

إنطلق مشروع ويكي الجامعة كأحد مشاريع مؤسسة ويكيميديا سنة 2006 [7]، بهدف استعمال الشبكة بصفة جماعية من أجل: (1) التشجيع على إنتاج المواد التعليمية واستضافتها على شكل وسائط متعددة في مختلف الاختصاصات، ثم تمكين المتعلمين من متابعتها في غرف وقاعات تدريس؛ (2) التشجيع على تأليف الكتب والموسوعات ونشرها في مواقع متخصصة مثل موقع ويكي الكتب وويكيبديا [4]. تبعاً لما سبق قوله، نلاحظ تقارباً في الأهداف العامة بين ويكي الجامعة وبوابة الزياتين من حيث اهتمامهما بالبحث والتعليم والتعامل معهما عبر المشاريع المشتركة بواسطة الشبكة. ولكن نجلب انتباه القارئ إلى أن ويكي الجامعة لا يحقق في نسخته الجارية المتطلبات التي سبق أن حددناها سابقاً في المقالة، علاوة على ما قد تقتضيه الصناعة العربية الرقمية من خصوصيات لا سيما في التعامل مع اللغة العربية.

يمكن لمشروع الزياتين جمع العديد من المشاريع العربية في بوابة واحدة، فهناك عدة اقتراحات وتوصيات ومشاريع لإثراء المحتوى الرقمي العربي من بينها تلك التوصيات التي صدرت عن ورشة العمل التي انعقدت في سوريا بالمعهد العالي للعلوم التطبيقية والتكنولوجية بمشاركة العديد من الدول العربية والمنظمة العربية للتربية والثقافة والعلوم، والتي أوصت إلى ضرورة إثراء المحتوى الرقمي العربي وبعث مشاريع في هذا المجال [12]. كما أن هناك مشاريع متعلقة بالترجمة ومن بين مشاريع الترجمة العربية الرائدة، مشروع الألف كتاب في مصر، والمشروع القومي للترجمة في المجلس الأعلى للثقافة المصري، ومشاريع وزارة الإعلام الكويتية، والمنظمة العربية للترجمة في بيروت، كما أن جائزة خادم الحرمين الشريفين الملك عبد الله بن عبد العزيز العالمية للترجمة، ومقرها مكتبة الملك عبد العزيز العامة بالرياض، التي انطلقت في أكتوبر 2006، تعد أحد مشاريع الترجمة المهمة والتي تمنح سنوياً للأعمال المتميزة والجهود البارزة في مجال الترجمة، هذا بالإضافة إلى دور جامعة الدول العربية ومؤسساتها وأجهزتها في مجال الترجمة، وكذلك مشاريع الترجمة التي تقوم بها المؤسسات الأهلية ودور النشر العربية وكذلك الجهود الفردية، كما انطلق أخيراً مشروع «كلمة» أحد مشاريع هيئة أبوظبي للثقافة والتراث، والذي سيقوم بترجمة مائة كتاب في العام ويتميز بزيادة عدد الكتب العلمية المترجمة، وكذلك مشروع مؤسسة محمد بن راشد آل مكتوم في ترجمة خمسين كتاباً من أبرز الكتب العالمية خلال عام واحد، والتي تعد من المشاريع الثقافية الضخمة والرائدة في العالم العربي للترجمة إلى اللغة العربية من مختلف اللغات الأجنبية، إلى غير ذلك من مشاريع للترجمة في معظم الدول العربية، تهدف لتنشيط حركة الترجمة، وإثراء المكتبة العربية بالكتب المترجمة في مختلف مجالات الفكر والمعرفة [13].

بوابة الزياتين ستمكن مستعملها بتنسيق الجهود بين مختلف الأطراف لبعث مشاريع مشتركة لتحقيق الأهداف المنشودة متجنباً تشتت الجهود والخبرات وتكرار المشاريع ذات الهدف الواحد.

هناك عدة أطراف يمكن أن تساهم علمياً و/أو مادياً في مشروع بوابة الزياتين وهي: (1) المنظمة العربية للتربية والثقافة والعلوم لأنها تهتم بمجال تطوير المحتوى الرقمي العربي كما أن لها مشروع النهوض باللغة العربية التي أقرتها جامعة الدول العربية بقيمة الدوحة. وكذلك مشروع تطوير المناهج في العالم العربي [14]، (2) المؤسسات والجامعات القطرية المهمة بالموضوع مثل مؤسسة قطر للتربية والعلوم وتنمية المجتمع [15]، (3) المؤسسات والجامعات السعودية، (4) مؤسسة الفكر العربي ببلدان [16]، (5) المركز العربي للتعريب والترجمة والتأليف والنشر بدمشق

[17], مكتب تنسيق التعريب في الرباط [18], المنظمة الإسلامية الدولية للتربية والثقافة والعلوم [19].

9 خطة إنجاز المشروع

نبين عبر الرسم 5، أنشطة وعناصر تكلفة مشروع بوابة الزيتين للصناعة العربية الرقمية في البحث والتعليم.

قبل المشروع	تحديد المشروع تحديد فريق البحث دراسة أدبيات المشروع تجميع تقانات المعلومات والاتصالات ذات العلاقة تطوير موقع المشروع على الشبكة تحديد كراسات لجنة من المنتجات الرقمية	
السنة الأولى للمشروع	الفصل 1 الندوة الأولى للمشروع	تطوير النسخة الأولى للبوابة تقييم النسخة الأولى للبوابة تحرير التقرير الأول للمشروع
	الفصل 2 الندوة الثانية للمشروع	بداية تطوير عينة المنتجات الرقمية بواسطة البوابة تطوير النسخة الثانية للبوابة تقييم النسخة الثانية للبوابة تحرير التقرير الثاني للمشروع
	الفصل 3 الندوة الثالثة للمشروع	مواصلة تطوير عينة المنتجات الرقمية بواسطة البوابة تطوير النسخة الثالثة للبوابة تقييم النسخة الثالثة للبوابة تحرير التقرير الثالث للمشروع
السنة الثانية للمشروع	الفصل 4 الندوة الرابعة للمشروع	مواصلة تطوير عينة المنتجات الرقمية بواسطة البوابة تطوير النسخة الرابعة للبوابة تقييم النسخة الرابعة للبوابة تحرير التقرير الرابع للمشروع
	الفصل 5 الندوة الخامسة للمشروع	مواصلة تطوير عينة المنتجات الرقمية بواسطة البوابة تطوير النسخة الخامسة للبوابة تقييم النسخة الخامسة للبوابة تحرير التقرير الخامس للمشروع
السنة الثالثة للمشروع	الفصل 6 الندوة السادسة للمشروع	تجريب البوابة في مؤسسات بحث وتعليم تقييم تجريب البوابة في مؤسسات البحث والتعليم تحرير التقرير السادس للمشروع تحرير التقرير النهائي للمشروع
		تسويق البوابة على نطاق واسع بحث مشاريع تستثمر البوابة
الرسم 5 أ: أنشطة مشروع الزيتين للصناعة العربية الرقمية في البحث والتعليم.		

01	إدارة المشروع
02	إنجاز بوابة الزياتين
03	تطوير أدوات برمجية جديدة
04	تدريب فريق البحث على منهجية العمل
05	تطوير منتجات رقمية بواسطة بوابة الزياتين
06	مناخلة / تنظيم: مؤتمرات علمية، زيارات علمية
07	متفرقات
الرسم 5 ب: عناصر تكلفة مشروع الزياتين للصناعة العربية الرقمية في البحث والتعليم	

10 الخاتمة

نظرا لبروز الشائكة كمكتبة تتجمع فيها المعارف، صار لزاما علينا المشاركة في تطويرها، من أجل المحافظة على موقع للحضارة الإسلامية العربية ضمن الحضارة الإنسانية. تبعا لذلك، اقترحنا فكرة إنجاز بوابة رقمية (سميناها بوابة الزياتين) وسط الشائكة، تمثل في مجال البحث والتعليم فضاء تفكير وإبداع يجمع بين المراسد العلمية والمكتبات الرقمية والجامعات الافتراضية والمجلات العلمية ومنابر النقاش، عبر مشاريع عمل جماعية مشتركة. حسبنا الحمولة الفكرية لبعض مناهج البحث والتعليم، للتدليل على المجهود الفكري المطلوب خصوصا عندما تنتج المناهج في ظروف عمل صعبة تتسم بخلل في الموارد البشرية والمالية والتشريعات الإدارية. ناقشنا أيضا عينة من المنتجات الرقمية النموذجية التي طالما شكلت تحديا للمشاريع التقليدية. يتضمن مشروع البوابة مخطط إنجاز نأمل أن يهتدي للممولين المؤمنين.

المراجع

- [1] <http://ar.wikipedia.org/wiki/%D8%A5%D9%86%D8%AA%D8%B1%D9%86%D8%AA>
- [2] ACM/AIS/ IEEE-CS Joint Task Force for Computing Curricula. **Computing Curricula 2005**. Association for Computing Machinery, Association for Information Systems, and Computer Society, 2005
- [3] <http://www.phillips-publishing.com>
- [4] <http://www.wikiversity.org>
- [5] http://ec.europa.eu/education/higher-education/doc1290_en.htm
- [6] أحمد الفرشيشي، روضة مروان، علي الملي، لمياء العابد، مرجح الطاهر، نبيل عباس، هاني عمار، اسكندر الزغلامي. **كنديبديا بين التوصيف والتفعيل: مشروع موسوعة الكندي الرقمية العربية للحوسبة الآلية**. الندوة الدولية السادسة لعلوم وهندسة الحاسوب، تونس، 2010.
- [7] مرجح الطاهر، روضة مروان، لمياء العابد، أحمد الفرشيشي. **منهجية الكندي لتطوير المناهج التعليمية الجامعية في الحوسبة الآلية: حالة دراسة لمناهج الإجازة والماجستير والدكتوراه في هندسة البرمجيات**. الندوة الدولية السادسة لعلوم وهندسة الحاسوب، تونس، 2010.
- [8] IEEE/ACM Joint Task Force on Computing Curricula. **Software Engineering 2004**,

Curriculum Guidelines for Undergraduate Degree Programs in Software**Engineering.** IEEE Computer Society Press and ACM Press, August 2004.

(http://www.computer.org/curriculum or

http://www.acm.org/education/curricula.html).

- [9] أحمد الفرشيشي. تعليم البرمجة الحاسوبية بالخوارزميات والصلات الثنائي. بحث، مقترح للنشر لدى مؤسسة فيليبس للنشر، 2010.
- [10] مرح الطاهر. مقترحات جديدة لتصميم المناهج الجامعية لتعليم المعلوماتية وتطبيقها في هندسة البرمجيات. رسالة بحث ماجستير، المعهد العالي للتصرف بتونس، جامعة تونس، مقترح للنشر لدى مؤسسة فيليبس للنشر، 2010.
- [11] http://ar.wikipedia.org/wiki/%D9%88%D9%8A%D9%83%D9%8A_%D8%A7%D9%84%D8%AC%D8%A7%D9%85%D8%B9%D8%A9
- [12] <http://www.sana.sy/ara/15/2010/10/19/314295.htm>
- [13] صفات سلامة. مشاريع الترجمة العربية.. ضرورة تنسيق الجهود. مقال في جريدة الشرق الأوسط. 10 ديسمبر 2007. <http://www.aawsat.com/leader.asp?section=3&article=449159&issueno=10604>
- [14] http://www.alecso.org.tn/index2.php?option=com_content&task=view&id=1601&Itemid=1&pop=1&page=0
- [15] <http://www.qf.org.qa/output/page40.asp>
- [16] <http://www.arabthought.org>
- [17] <http://www.acatap.htmlplanet.com/arabic-h.htm>
- [18] <http://www.arabization.org.ma>
- [19] http://www.isesco.org.ma/arabe/history/history.php?idd=TDD_REF

الملخص باللغة الانجليزية

Title: The Zayatin Port: A Project for the Computerization of Digital Arab Industry in Research and Education

Abstract. In response to the rapid development of the digital knowledge, and as a contribution to the development of the Arab digital industry, we propose in this article a project for realizing a digital port for computing the Arab digital industry, to be based on group networking and information and communication technology. In order to activate the proposed port, the article discusses a number of subsidiary model activities and underlines the benefits of its achievement.

Keywords: World Wide Web, Digital Content, Digital Industry, Group Work, Arabic Language, Research and Education, Zayatin Port.

المصطلحات

World Wide Web	الشابكة
Digital Content	المحتوى الرقمي
Digital Industry	الصناعة الرقمية
Port	البوابة
Software Engineering	هندسة البرمجيات
Association for Computing Machinery	رابطة مكائن الحوسبة
Computer Science	علم الحاسوب
Computing Science	علم الحوسبة الآلية

