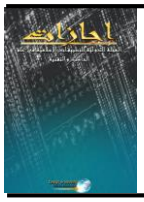


التكنولوجيا في خدمة البلاغة القرآنية

فواز شخير العنزي، ضياء الدين محمد أسعد أبوزينة
كلية علوم وهندسة الحاسوب، جامعة الكويت، الكويت
abuzeina@ku.edu.kw, fawaz.alanzi@ku.edu.kw

الخلاصة: يقدم هذا البحث دراسة علمية تتعلق بدور التكنولوجيا الحديثة في خدمة اللغة العربية والقرآن الكريم. تم التطبيق العملي لاحد المواضيع الهامة في مجال خدمة القرآن الكريم والعاملين عليه الا وهو ايجاد المتشابهات في القرآن الكريم. ولقد بذل علماء البلاغة القرآنية جهودا كبيرة لحصر الآيات المتشابهة بهدف بيان الاوجه البلاغية لهذه الآيات القرآنية. في هذا البحث نقدم دور التكنولوجيا في تسهيل عملية إيجاد المتشابهات بحيث يتم الاستغناء عن الطريقة التقليدية والانتقال الى استخدام ما توفره التكنولوجيا الحديثة في هذا المجال. تم استخدام طريقة فهرسة الدلالات الكامنة (LSI) وطريقة التقسيم (ME clustering) لإنجاز العمل المطلوب، كما تم استخدام مدونة (corpus) قرآنية تحتوي على 4781 آية. تكمن أهمية هذه الدراسة في استعراض الطرق الحديثة المستخدمة في أنظمة معالجة اللغات الطبيعية بشكل عام للإفادة منها في مجال حوسبة اللغة العربية.

الكلمات الجوهرية: القرآن الكريم، اللغة العربية، المتشابهات، فهرسة الدلالات الكامنة، البلاغة القرآنية.

1. المقدمة

تعتبر اللغة العربية واحدة من اللغات الحية اذ انها مرتبطة بالقرآن الكريم، كما انها اللغة المتحدث بها عند مئات الملايين من العرب والمسلمين. وكغيرها من اللغات الاخرى، فقد حظيت حوسبة اللغة العربية اهتماماً واضحاً في السنوات الأخيرة. نناقش في هذا البحث مسالة إيجاد المتشابهات في القرآن الكريم وما توفره التكنولوجيا الحديثة من إمكانيات لتحقيق هذا الهدف. ان هذا الكتاب العزيز وقد تحدى الله به الانس والجن ان يأتوا ولو بآية من مثله قد احتوى على أوجه بلاغية كانت معجزة للعرب من اهل البلاغة والفصاحة الذين لما يدركوا في حينه وجوه الاعجاز الأخرى كعلم الاجنة والافلاك والبحار. ان المتشابه في القران الكريم يشمل المتشابه المعنوي والمتشابه اللفظي. في هذا البحث نتناول المتشابه اللفظي. ولقد نشرت العديد من المؤلفات التي تتناول مسالة المتشابه اللفظي في القران الكريم بهدف بيان أوجه الاعجاز في الآيات المتشابه ولماذا حصل شيء من الاختلاف بين آيتين او أكثر من آيات القران الكريم، علماً ان التوجيه في بيان البلاغة هي مسألة اجتهاد بين العلماء. ومن الأمثلة على آيات متشابهات ما جاء في سورة الأعراف في الكلمات "ضلالة" و "سفاهة"، علما ان الرسم القرآني للآيات تم الحصول عليها من الموقع [1]:

سورة الأعراف الآية 61	قَالَ يَٰقَوۡمِ لَيۡسَ بِي ضَلَٰلَةٞ وَلَٰكِنِّي رَسُولٞ مِّن رَّبِّ ٱلۡعَٰلَمِينَ
سورة الأعراف الآية 67	قَالَ يَٰقَوۡمِ لَيۡسَ بِي سَفَاهَةٞ وَلَٰكِنِّي رَسُولٞ مِّن رَّبِّ ٱلۡعَٰلَمِينَ

في المثال السابق كانت المتشابهات في نفس السورة، وهذا لا يكون في جميع الحالات. فهناك الكثير من الآيات المتشابهة تقع في سور مختلفة، كما هو الحال في كلمة "فانفجرت" وكلمة "فانبجست" في المواضع القرآنية التالية:

سورة البقرة الآية 60	۞وَإِذِ ٱسۡتَسۡقَىٰ مُوسَىٰ لِقَوۡمِهِۦ فَقُلۡنَا ٱضۡرِب بِّعَصَاكَ ٱلۡحَجَرَۖ فَٱنفَجَرَتۡ مِنۡهُ ٱثۡنَتَا عَشۡرَةَ عَيۡنٗاۖ قَدۡ عَلِمَ كُلُّ أُنَاسٖ مَّشۡرَبَهُمۡۖ كُلُواْ وَٱشۡرَبُواْ مِن رِّزۡقِ ٱللَّهِ وَلَا تَعۡثَوۡاْ فِي ٱلۡأَرۡضِ مُفۡسِدِينَ ٦٠
سورة الأعراف الآية 160	وَقَطَّعۡنَٰهُمُ ٱثۡنَتَيۡ عَشۡرَةَ أَسۡبَاطًا أُمَمٗاۚ وَأَوۡحَيۡنَآ إِلَىٰ مُوسَىٰٓ إِذِ ٱسۡتَسۡقَىٰهُ قَوۡمُهُۥٓ أَنِ ٱضۡرِب بِّعَصَاكَ ٱلۡحَجَرَۖ فَٱنۢبَجَسَتۡ مِنۡهُ ٱثۡنَتَا عَشۡرَةَ عَيۡنٗاۖ قَدۡ عَلِمَ كُلُّ أُنَاسٖ مَّشۡرَبَهُمۡۚ وَظَلَّلۡنَا عَلَيۡهِمُ ٱلۡغَمَٰمَ وَأَنزَلۡنَا عَلَيۡهِمُ ٱلۡمَنَّ وَٱلسَّلۡوَىٰۖ كُلُواْ مِن طَيِّبَٰتِ مَا رَزَقۡنَٰكُمۡۚ وَمَا ظَلَمُونَا وَلَٰكِن كَانُوٓاْ أَنفُسَهُمۡ يَظۡلِمُونَ ١٦٠

ان الأمثلة السابقة تتعلق بالمتشابه الحاصل نتيجة إبدال كلمة بكلمة، وهناك ما يكون بإبدال حرف بحرف مثل "اولم يهد لهم" و "افلم يهد لهم". وقد اشار بدر الدين محمد بن عبد الله الزركشي في كتابه "البرهان في علوم القرآن" الى ان المتشابه اللفظي يقسم الى إلى ثمانية أقسام. كما نشير هنا الى ان الآيات المتكررة (التطابق التام) لا تدخل ضمن تصنيف الآيات المتشابه في القران الكريم مثل قوله تعالى:

سورة الرحمن الآية 13	فَبِأَيِّ ءَالَآءِ رَبِّكُمَا تُكَذِّبَانِ ١٣

ان إيجاد المتشابه اللفظي في القران الكريم يقوم على حصر الآيات المتشابهة من سورة الفاتحة الى سورة الناس. ان هذا الامر يحتاج الى جهد كبير جدا خصوصا في العصور السابقة في ظل غياب الوسائل المساعدة المتوفرة في يومنا هذا. ولقد اجتهد العلماء السابقون اجتهادا عظيما في هذا الامر حيث وجد عدد من الكتب التي تناولت هذه المسالة ومن أمثال ذلك كتاب "متشابه القران العظيم" للمؤلف أبو الحسين أحمد بن جعفر المنادي (المتوفى 336 هـ). وكتاب "درة التنزيل وغرة التأويل في بيان الآيات المتشابهات في كتاب الله العزيز" للخطيب الاسكافي (المتوفى 420 هـ). حيث تناول في كتابه 273 موضع من المتشابهات. وكذلك كتاب "ملاك التأويل القاطع بذوي الإلحاد والتعطيل في توجيه المتشابه اللفظ من آي التنزيل" للإمام الحافظ أحمد بن إبراهيم بن الزبير الغرناطي (المتوفى سنة 708 هـ). وهو عالم كبير من علماء الاندلس، وقد أشار الى جهد الاسكافي وأضاف عدد 104 موضعا جديدا من التشابه لم يذكرها الاسكافي. وهذا يشير الى مدى

صعوبة هذه المسألة وتداخل الآيات المتشابه مع بعضها البعض. ومن الكتب في هذا المجال ايضاً كتاب "متشابه القران" للمؤلف أبو الحسن علي بن حمزة الكسائي. وكتاب "البرهان في متشابه القرآن" للمؤلف محمود بن حمزة الكرماني.

وكذلك كتاب "كشف المعاني في المتشابه من المثاني" لمؤلفه بدر الدين ابن جماعة. بالإضافة الى كتاب "فتح الرحمن بكشف ما يلتبس من القرآن"، للمؤلف زكريا الأنصاري. الى غير ذلك من المؤلفات القديمة والحديثة التي تناولت هذه المسالة لأغراض بيان بلاغة وبديع القران الكريم. كما نشير هنا الى اننا في هذه المقدمة التاريخية قد نقلنا بعض المعلومات المختصرة من كتاب الدكتور إبراهيم بن عبد العزيز الزيد بعنوان "البلاغة القرآنية في الآيات المتشابهات من خلال كتاب «ملاك التأويل» لابن الزبير الغرناطي."

وحتى نوضح المقصود بالبلاغة القرآنية، والهدف من الحصول على الآيات المتشابهة نعرض المثال التالي بحيث حصل تشابه بين الآيتين في كلمة "بلداً" في سورة البقرة وكلمة "البلد" في سورة ابراهيم:

سورة البقرة الآية 126	وَإِذْ قَالَ إِبْرَٰهِـۧمُ رَبِّ ٱجْعَلْ هَٰذَا بَلَدًا ءَامِنًا وَٱرْزُقْ أَهْلَهُۥ مِنَ ٱلثَّمَرَٰتِ مَنْ ءَامَنَ مِنْهُم بِٱللَّهِ وَٱلْيَوْمِ ٱلْءَاخِرِ ۖ قَالَ وَمَن كَفَرَ فَأُمَتِّعُهُۥ قَلِيلًا ثُمَّ أَضْطَرُّهُۥٓ إِلَىٰ عَذَابِ ٱلنَّارِ ۖ وَبِئْسَ ٱلْمَصِيرُ ﴿١٢٦﴾
سورة ابراهيم الآية 35	وَإِذْ قَالَ إِبْرَٰهِيمُ رَبِّ ٱجْعَلْ هَٰذَا ٱلْبَلَدَ ءَامِنًا وَٱجْنُبْنِى وَبَنِىَّ أَن نَّعْبُدَ ٱلْأَصْنَامَ ﴿٣٥﴾
البلاغة القرآنية او التوجيه	يقول المفسرون ان لها عدة وجوه من ابرزها ان إبراهيم عليه السلام عندما وصل الى مكة وكانت ارض خلاء (البلد غير قائمة) فقال رب اجعل هذا بلدا امنا. ثم في زيارته التالية وكانت هناك بلد قائمة فقال رب اجعل هذا البلد امنا.

نناقش في هذا البحث كيفية الحصول على الآيات المتشابهة في القران الكريم من خلال البرامج المحوسبة المتوفرة حالياً تسهيلاً للوقت والجهد. علما ان العمل في هذا الاتجاه لا يخدم إيجاد المتشابه القرآني فحسب وانما في أي موضع من مواضع اللغة التي تحتاج الى إيجاد النصوص المتشابهة سواء في الحديث الشريف او التاريخ او حتى في الكشف عن الانتحال والسرقات الادبية. كما اننا لا نتحدث في هذا البحث عن طرق البحث عن كلمة او جملة في القران الكريم فهذا امر بسيط، بل نتحدث عن طريقة يقوم من خلالها النظام تلقائياً بإنتاج عدد من المجموعات القرآنية بحيث تحتوي كل مجموعة على مجموعة من الآيات المتشابهة. وبذلك فإننا يمكن ان نصف المسالة على ان لدينا ملفا نصيا ونريد ان نستخدم برنامجا معينا لتقسيم النصوص الى مجموعات متشابهة دون تدخل من قبل المستخدم. ولنا ان نتخيل صعوبة هذا العمل إذا أردنا القيام به بدون برنامج محوسب متخصص، فلا شك انه امر شاق حتى مع توفر القران الكريم بين أيدينا

بأشكال واحجام ورقية مختلفة وكذلك مع توفر برامج البحث عن كلمات معينة فان هذا لا يوفر لنا التلقائية التي نريدها لهذه العملية. وحتى لو افترضنا ان الشخص الذي يريد تحديد الآيات المتشابهة حافظا للقران الكريم فهذا لا يقلل من حجم المشكلة، حيث أشرنا الى ان كبار العلماء في هذا المجال قد اضافوا مواضع غفل عنها من سبقهم. يتركز عملنا في هذا الاتجاه على طريقة فهرسة الدلالات الكامنة (Latent Semantic Indexing) وهي طريقة معروفة في عدد من أنظمة استرجاع المعلومات (Information Retrieval) مثل محركات البحث، واكتشاف الموضوعات، وتلخيص النصوص. في القسم الثاني نناقش بعض الاعمال البحثية في هذا المجال، يليه في القسم الثالث الطريقة المقترحة. ونستعرض في القسم الرابع النتائج، ثم نختم هذه الدراسة في القسم الخامس.

2. الأعمال البحثية السابقة

حظيت اللغة العربية في الآونة أخيرة اهتماما ملحوظا نحو حوسبة اللغة وتعزيز الخوارزميات الخاصة بمعالجة المحتوى العربي على الانترنت خصوصا في ظل الزيادة الهائلة في هذه المعلومات. وتشمل هذه التطبيقات محركات البحث، الترجمة الالية، التشكيل الالي، تلخيص النصوص، التعرف على الكلام الخ. في هذا البحث نركز على طريقة فهرسة الدلالات الكامنة (Latent Semantic Indexing) والتي يتم اختصارها الىLSI ، كما وجدت ترجمة أخرى لهذه الطريقة بحيث تسمى نموذج التكشيف الدلالي الكامن. وبغض النظر عن المسمى الخاص بها في اللغة العربية فإننا نقول انها طريقة إحصائية يتم من خلالها انشاء مصفوفة (matrix) كبيرة تربط بين الوثائق النصية والكلمات الواردة في جميع الوثائق. كثيرة هي التطبيقات التي استخدمت هذه الطريقة في مجال حوسبة اللغة خصوصا اللغة الإنجليزية، حيث ان استخدامها في مجال اللغة العربية لا زال محدودا. هناك عدد من الميزات الهامة التي تحظى بها طريقة LSI واهمها قدرتها على إيجاد العلاقة بين النصوص المختلفة حتى في ظل غياب الكلمات المشتركة بين هذه النصوص وهذا ما يعرف بميزة .(semantic rich) والميزة الأخرى هي قدرتها على التعامل مع الابعاد الكبيرة بحيث يتم اختيار ابعاد قليلة لمتجهات النصوص (document vectors)بدون خسارة في المعلومات الخاصة بالنص وهذا ما يعرف بميزة .(dimension reduction technique) وبذلك فان طرق تقليل الابعاد تساعد في سرعة التنفيذ مع المحافظة على الخصائص الأساسية للبيانات.

نقدم في هذا القسم عددا من الاعمال البحثية التي استندت الى طريقة فهرسة الدلالات الكامنة (LSI) لخدمة التطبيقات اللغوية. ذكر في [2] ان هذه الطريقة استخدمت بشكل كبير في استرجاع المعلومات، البحث، التصنيف، والتصفية (filtering). تم استخدام هذه الطريقة في تلخيص النصوص (document summarization) كما تم الإشارة لذلك في [3]. ذكر في [4] ان استخدام هذه الطريقة تعطي نتائج أفضل عند استخدامها في مسألة تصنيف الوثائق. تمت الإشارة الى ان هذه الطريقة تستخدم في تحسين الأداء في محركات البحث حيث تمت الإشارة لذلك في [5]. استخدمت هذه الطريقة في فرز وتقسيم البرامج بحيث يتم تقسيم محتويات برنامج معين بحيث يتم استخلاص التوثيق (documentation) والنص البرمجي (source code) كل على حدة [6]. تم استخدام هذه الطريقة في تحديد الاعمال البحثية في حقل معين وذلك في الاعمال البحثية التي تشمل حقولا متعددة (7] (multidisciplinary field]. ويجدر الاشارة الى ان هناك العديد من الدراسات التي تستخدم طريقة فهرسة الدلالات الكامنة في تطبيقات مختلفة.

3. الطريقة المقترحة:

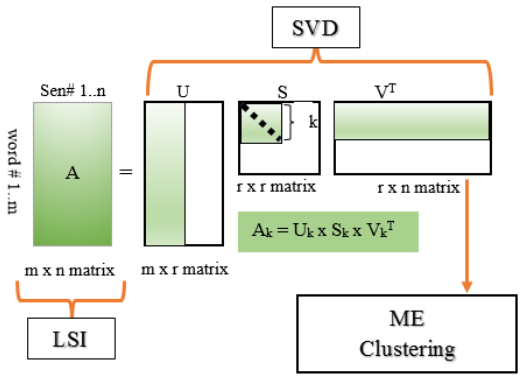
لوصف الطريقة المقترحة لابد من الإشارة الى ان طريقة فهرسة الدلالات الكامنة تستخدم طريقة رياضية خاصة بالجبر الخطي لتقسيم المصفوفة الناتجة الى ثلاث مصفوفات تحتوي على الخصائص المطلوبة وهذا ما يعرف بطريقة (singular value decomposition)او ما تعرف بـ .(SVD) ولابد من الإشارة الى ان أولى الخطوات في مجال الحصول على الآيات المتشابه هو الحصول على نسخة نصية من القران الكريم، وهذا ما قمنا به من خلال الحصول على هذه النسخة من موقع مصحف المدينة المنورة [8]. اما فيما يتعلق بالخوارزمية المقترحة لانتاج الآيات المتشابهة في القران الكريم فإنها تشمل الخطوات التالية:

أولا: انشاء مصفوفة كبيرة تحتوي على الارتباطات بين الوثائق والكلمات، وفي حالتنا فان الوثائق هي الآيات القرآنية والكلمات هي كلمات القران الكريم. وهذه الخطوة هي التطبيق العملي لطريقة.(LSI)

ثانيا: تطبيق التقسيم (انتاج ثلاثة مصفوفات) على المصفوفة الناتجة في الخطوة الاولى باستخدام.(SVD)

ثالثا: تطبيق خوارزمية التقسيم (ME clustering) باستخدام الخصائص الناتجة في الخطوة الثانية.

الشكل رقم 1 يوضح الخطوات اللازمة لانتاج الآيات القرآنية المتشابهة باستخدام الوسائل البرمجية الثلاثة التي تم الإشارة اليها في الخوارزمية المقترحة اعلاه (LSI+SVD+ME Clustering).



شكل 1: الطريقة المقترحة

ان الوصف الذي تم تقديمه في الخوارزمية السابقة هو وصف مختصر، ويمكن الرجوع الى المصادر المختلفة في هذا المجال للتعرف على آلية العمل لكل من الوسائل المستخدمة في الخطوات الثلاثة السابق ذكرها.

4. النتائج:

نعرض في هذا القسم التطبيق العملي للطريقة المقترحة أعلاه، وعرض لعينات من التقسيمات المستخرجة. وهي ما نسميه في بحثنا هذا بالآيات المتشابهة. لا بد ان نشير هنا الى ان المستخدم يحدد عدد المجموعات التي يريدها. والمقصود بعدد المجموعات هنا هو العدد الذي يرغب المستخدم بالحصول عليه عند تقسيم القرآن الكريم. فاذا اختار المستخدم عدد المجموعات 10 فان النظام سوف يقوم بعمل عشرة مجموعات ويسند الآيات القرآنية كلها على هذه المجموعات العشرة حسب درجة التشابه فيما بينها. وإذا لم يحدد المستخدم عدد المجموعات فان النظام سوف يقوم بتقسيم الآيات القرآنية حسب ما يراه مناسبا. اما بالنسبة لخطوات الحصول على النتائج فكانت على النحو التالي:

أولا: تطبيق طريقة (LSI) باستخدام لغة البرمجة بايثون.(Python)

ثانيا: تطبيق طريقة (SVD) باستخدام لغة البرمجة ماتلاب.(MATLAB)

ثالثا: تطبيق طريقة التقسيم (ME Clustering) باستخدام برنامج ويكا(WEKA)

فيما يلي نعرض النتائج التي حصلنا عليها بعد تطبيق الخطوات المطلوبة. نشير الى ان عدد آيات القران الكريم هو 6236 آية وقد تم اختيار 4781 آية فقط. ان الهدف من استبعاد عدد من الآيات هو الطاقة الاستيعابية للنظام (الذاكرة) إذ ان النظام كان يتوقف عن العمل عند ادراج جميع آيات القران الكريم. لذلك تم استبعاد (مضطرين لا مختارين) عدد من الآيات ذات الكلمات القليلة إذ أنها لا تؤثر كثيرا على الفكرة التي نرغب في ايصالها من خلال هذا العمل. على سبيل المثال، لا داعي لوجود بعض فواتح السور مثل "الم" او "كهيعص". وبما اننا نقوم بهذه الازالة بشكل برمجي، فقد تم اختيار الرقم 30 كمعيار لإزالة جميع الآيات التي تحتوي على عدد من الحروف يساوي او اقل من 30 حرفا (مع اعتبار المسافة الموجودة بين الكلمات لأنها من ناحية برمجية تعتبر حرفا مثل أي حرف من الأبجدية). كما نشير هنا الى ان النسخة القرآنية النصية التي عملنا عليها في هذا البحث هي نسخة غير مشكلة، مع فهمنا الى ان التشكيل جزء لا يتجزأ من النص القرآني.

بشكل فعلي، تم اختيار عدد المجموعات 1000 حتى نحصل على عدد كبير من المجموعات بحيث تكون الآيات داخل المجموعة الواحدة متشابهة الى حد كبير. فكلما تمت الزيادة على هذا الرقم كانت الآيات داخل المجموعات الناتجة متشابهة أكثر، وكلما قل هذا الرقم كانت الآيات اقل تشابها. ولذلك فقد أشرنا سابقاإلى أن هذا الرقم يتم تحديده من قبل المستخدم حسب الحاجة. وفيما يلي بعض المجموعات الناتجة من النظام:

رقم المجموعة	الآيات التي تم الحصول عليها
cluster901	"قال يا قوم ليس بي ضلالة ولكني رسول من رب العالمين"
cluster901	"قال يا قوم ليس بي سفاهة ولكني رسول من رب العالمين"
cluster901	"وقال موسى يا فرعون إني رسول من رب العالمين"

كما هو ملاحظ فان النظام استطاع ان يكشف عن التشابه في هذه الآيات ويقدمها الى المستخدم كمجموعة واحدة. نلاحظ ان هناك تشابه كبير بين الآية الأولى والآية الثانية وتشابه اقل مع الآية الثالثة. ويكون إلى هذه النقطة قد انتهى دور الوسائل التكنولوجية، ويترك الامر بعد ذلك لعلماء البلاغة القرآنية لتحديد فيما إذا كان هناك وجها بلاغيا في هذا التشابه ليتم بعد ذلك إعطاء التوجيه وبيان سبب حصول هذا الاختلاف. فيما يلي مجموعة أخرى من النتائج التي حصلنا عليها:

رقم المجموعة	الآيات التي تم الحصول عليها
cluster132	"فادخلوا أبواب جهنم خالدين فيها فلبئس مثوى المتكبرين"
cluster132	"قيل ادخلوا أبواب جهنم خالدين فيها فبئس مثوى المتكبرين"
cluster132	"ادخلوا أبواب جهنم خالدين فيها فبئس مثوى المتكبرين"

المجموعة التالية جمعت بين أربع آيات مع تشابه اقل وضوح مقارنة مع الأمثلة السابقة:

رقم المجموعة	الآيات التي تم الحصول عليها
cluster531	"وقالوا إن هي إلا حياتنا الدنيا وما نحن بمبعوثين "
cluster531	"إن هي إلا حياتنا الدنيا نموت ونحيا وما نحن بمبعوثين "
cluster531	"إلا موتتنا الأولى وما نحن بمعذبين "
cluster531	"إن هي إلا موتتنا الأولى وما نحن بمنشرين "

بالإضافة الى الأمثلة أعلاه فقد تم الحصول على عدد كبير من المجموعات المتشابهة. كما نود ان نشير في هذه الدراسة الى ان طريقة فهرسة الدلالات الكامنة تفيد أيضا في تصنيف الآيات القرآنية حسب الموضوع لأنها كما أشرنا في البداية لديها القدرة على تحديد العلاقة بين الجمل المختلفة حتى لو لم يكن هناك كلمات متشابهة. ومن الأمثلة على ذلك المجموعة التالية:

رقم المجموعة	الآيات التي تم الحصول عليها
cluster926	"وأوحينا إلى موسى أن ألق عصاك فإذا هي تلقف ما يأفكون "
cluster926	"وأدخل يدك في جيبك تخرج بيضاء من غير سوء في تسع آيات إلى فرعون وقومه إنهم كانوا قوما فاسقين "

نلاحظ في المثال السابق عدم وجود كلمات مشتركة بين الآيتين ومع ذلك فإنها صنفت في مجموعة واحدة لأنها تتكلم عن نفس الموضوع الذي يتعلق بقصة سيدنا موسى عليه السلام مع فرعون.

5. الخلاصة

في هذا البحث، تم توضيح دور التكنولوجيا في خدمة اللغة بشكل عام من خلال إيجاد الآيات المتشابهة في القران الكريم. كما تم عرض بعض النتائج التي تشير الى ان طريقة فهرسة الدلالات الكامنة يمكن ان تستخدم أيضا في تصنيف الموضوعات. وبذلك فان استخدام التقنيات الحديثة في أنظمة استرجاع المعلومات ومعالجة اللغات الطبيعية يسهل العمل ويعطي نتائج واعدة. الا ان هناك حاجة لمزيد من البحث لمقارنة نتائج هذه الدراسة مع اعمال تتعلق بالآيات القرآنية المتشابه.

6. المراجع

[1] The Holy Qur'an. (2015, August). Retrieved from http://www.sacred-texts.com/isl/quran/

[2] Kontostathis, April, and William M. Pottenger. "A framework for understanding Latent Semantic Indexing (LSI) performance." Information Processing & Management 42.1 (2006): 56-73.

[3] Lee, J. H., Park, S., Ahn, C. M., & Kim, D. (2009). Automatic generic document summarization based on non-negative matrix factorization.Information Processing & Management, 45(1), 20-34.

[4] Kantardzic, M. (2011). Data mining: concepts, models, methods, and algorithms. John Wiley & Sons.

[5] Osiński, Stanislaw, and Dawid Weiss. "A concept-driven algorithm for clustering search results." Intelligent Systems, IEEE 20.3 (2005): 48-54.

[6] Maletic, Jonathan, and Naveen Valluri. "Automatic software clustering via latent semantic analysis." Automated Software Engineering, 1999. 14th IEEE International Conference on.. IEEE, 1999.

[7] Kou, Gang, and Yi Peng. "An Application of Latent Semantic Analysis for Text Categorization." International Journal of Computers Communications & Control10.3 (2015): 357-369.

[8] Mus'haf al-Madinah. (2015, August). Retrieved from http://nashr.qurancomplex.gov.sa/site/

7. الخلاصة باللغة الانجليزية

Arabic Text Classification Using Linear Discriminant Analysis

Fawaz S. Al-Anzi, Dia AbuZeina
Department of Computer Engineering
Kuwait University
Kuwait City, Kuwait
fawaz.alanzi@ku.edu.kw, abuzeina@ku.edu.kw

Abstract
This work presents a potential role of technology to sever the Arabic language and the holy Quran. Finding similar Quranic sentences is one of the important topics in Quran related research. In fact, the language scholars have devoted significant effort to find the similar sentences for many purposes such as Qur'anic eloquence and its explanation. In this paper, we used latent semantic indexing (LSI) method along with expectation maximization (EM) clustering method to produce the groups of the similar sentences. We used a Quranic corpus that contains 4,781 sentences. The results show that the proposed method is useful to obtain the similar sentences in two levels, the similarity in words and the similarity in concepts (i.e semantic level). This work also shows the importance of LSI for information retrieval (IR) and natural language processing (NLP).
Keywords: Quran, Arabic language, latent semantic indexing (LSI).