تطبيق العنقدة المتعددة المستويات على نص القرآن الكريم

محمد سعيد دسوقي
المعهد العالي المعهد العالي للعلوم التطبيقية والتكنولوجيا، ص. ب. 31983 دمشق سورية
said.desouki@hiast.edu.sy

الخلاصة: نهدف في هذا البحث إلى تطبيق تقنية عنقدة الوثائق المتعددة المستويات Multi-Level Document Clustering، إحدى التقنيات الهامة في التنقيب في النصوص Text Mining، على نص القرآن الكريم. قمنا باستخراج الكلمات المفتاحية من الآيات القرآنية وحساب التشابهات فيما بينها اعتماداً على هذه الكلمات، وتصنيفها ضمن مجموعات بحسب كلماتها المفتاحية وتشابهاتها، ثم قمنا بحساب التشابهات بين المجموعات وإنشاء مستويات أعلى إلى أن حصلنا على شجرة تصنيف هرمية تعبِّر عن توزع الآيات القرآنية وتجميعها بحسب كلماتها.

الكلمات الجوهرية. التنقيب في النصوص, العنقدة المتعددة المستويات, المعالجة الآلية لنص القرآن الكريم.

1. المقدمة

تزايد في الآونة الأخيرة الاهتمام بتقنيات التنقيب في النصوص واستكشاف المعارف واستخراجها منها. وشرع المتخصصون في هذا المجال في وضع طرق وآليات وخوارزميات تعالج النصوص آلياً وتستخرج منها قواعد أو روابط بين الكلمات أو أجزاء الكلام أو تصنيفات وملخصات. يرتبط موضوع التنقيب في النصوص ارتباطاً وثيقاً بالمعالجة الآلية للغة، إذ لا بد لأي تناول للنصوص ومعالجة لها من استخدام قواعد وأنظمة التحليل الصرفي والنحوي والدلالي [رونين وجيمس، 2007]. وإذا أردنا التنقيب في النصوص فأين سنجد نصاً أعظم مكانةً وأصدق حديثاً وأحق بالدراسة والتحليل والمعالجة والتناول والتمحيص من نص القرآن الكريم. كلام الله الذي يستمد العظمة من الله العظيم. ومهما كتب الكاتبون ونشر الناشرون وملأ المبدعون من صفحات على الإنترنت وغيرها فأين نجد نصاً نثق به ونصدق قائله ونسبته له مثل القرآن الكريم، النص الحكيم والحقيقة المطلقة.

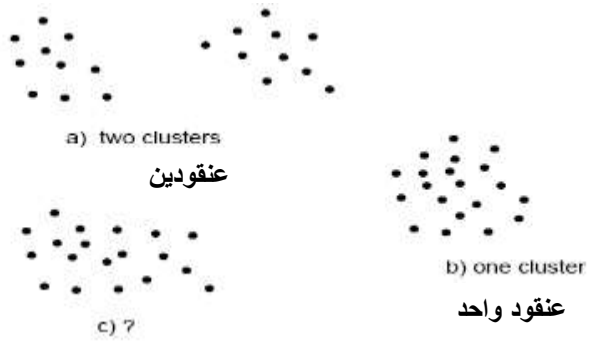
تقنيات التنقيب في النصوص كثيرة ومتنوعة فمنها مثلاً الربط بين الكلمات والمقاطع في النصوص، ومنها استخراج الأسماء والتواريخ واستخدامها لفهرسة النصوص، ومنها تصنيف النصوص ضمن موضوعات محددة مسبقاً، ومنها عنقدة الوثائق المتعددة المستويات ووضعها ضمن أقسام وموضوعات حسب محتواها [نيكولاس وإدوارد، 2007]. وهذه التقنية الأخيرة هي التي اخترنا العمل عليها وتطبيقها على نص القرآن الكريم.

لا يفترض هذا البحث الانطلاق من مشكلة بحاجة إلى حل، فالقرآن الكريم كما وصل إلينا بالتواتر جمعاً عن جمعٍ عن رسول الله صلى الله عليه وسلم عن جبريل عليه السلام عن رب العزة جل جلاله مصنف ضمن سور في كل منها مجموعة من الآيات، وتجميع الآيات في السور وترتيبها ضمنها وترتيب السور وتسميتها أمور توقيفية عن رسول الله صلى الله عليه وسلم ولا مجال للرأي والاجتهاد فيها. ولا شك أن هذا التجميع والترتيب هو الأكمل والأفضل لتحقيق مقاصد القرآن الكريم التي أرادها الله عز وجل. لكننا نحاول من خلال هذا البحث تطبيق فكرة العنقدة على نص القرآن الكريم، من

خلال تجميع الآيات ضمن السورة الواحدة وتصنيفها بحسب موضوعها أو تجميع السور ومقارنتها ببعضها وتصنيفها، أو تجميع الآيات ضمن القرآن كاملاً وتصنيفها، بهدف الدراسة والتفسير والتدبر والوصول إلى تصنيف موضوعي يسهم في دعم الإعجاز الدلالي للقرآن الكريم.

2. ما هي العنقدة؟

تعتمد فكرة العنقدة Clustering على استخراج الكلمات المفتاحية من النصوص وحساب التشابهات فيما بينها اعتماداً على هذه الكلمات، وتصنيفها ضمن مجموعات تبعاً لمحتواها تسمى عناقيد Clusters (انظر الشكل 1). يمكن أن تجري العنقدة على عدة مستويات فتسمى العنقدة المتعددة المستويات وينتج عنها شجرة تصنيف هرمية تعبِّر عن محتوى النصوص المعالجة. فالعنقدة هي التصنيف ضمن أصناف غير معروفة وغير محددة مسبقاً. تعتمد تقنية العنقدة المتعددة المستويات على استخراج الكلمات المفتاحية من النصوص، ثم حساب التشابهات فيما بين هذه النصوص اعتماداً على الكلمات المفتاحية المستخرجة، ثم تصنيف النصوص ضمن مجموعات تبعاً لنتائج التشابه. يجري بعد ذلك حساب التشابهات بين هذه المجموعات لضم بعض المجموعات إلى بعضها في مستوى آخر من التصنيف، وهكذا حتى نصل إلى مجموعة واحدة هي رأس الشجرة أو الهرمية التي تعبِّر عن تصنيف النصوص حسب محتواها [نيكولاس وإدوارد، 2007].



الشكل 1. عدة مجموعات من النقاط في فضاء ثنائي البعد

نلاحظ أنّ المجموعة الأولى a تبدو قابلة للتقسيم بشكل طبيعي لا يقبل الجدل إلى عنقودين. أما المجموعة الثانيةb فهي تبدو مستحيلة التقسيم فهي عنقود واحد مترابط. أما المجموعة الثالثة cفهي تشكل مشكلة جوهرية. حيث لا يوجد فصل واضح بين النقاط فلا يمكن اعتبارها عنقود واحد ولا يمكن اعتبارها عنقودين لأننا لا نعلم الحد الفاصل بين العنقودين اللذين نرغب بتكوينهما. تستخدم عملية العنقدة إجرائيات تحاول إيجاد تقسيم معقول للعينات غير المصنفة (كما في c).

أي أننا انطلاقاً من مجموعة من النصوص نصل إلى شجرة تصنيف هرمية مؤلفة من مجموعة من المجلدات تعبِّر عن محتوى هذه النصوص وتُسهِّل الاستفادة منها وتحليلها. للمسألة متغيرات كثيرة يمكن التحكم بها. إذ يجب تعريف التشابه وطرق قياس المسافات بين النصوص وطرق استخراج الكلمات المفتاحية منها، وطرق تجميع الوثائق في مجموعات. انظر [دسوقي والعبدو، 2012] للوقوف على تفاصيل العنقدة المتعددة المستويات والمعادلات المستخدمة.

3. الأعمال السابقة

يمكن أن نذكر هنا أعمال الدكتور محمد زكي خضر في قواعد بيانات القرآن الكريم [محمد زكي وأكرم، 2011] والتي أخرج منها "المعجم المفهرس للتراكيب المتشابهة لفظاً في القرآن الكريم" الذي يتضمن أكثر من 44000 عبارة مختلفة. إلا أنَّ البحث الأكثر صلة ببحثنا هذا هو بحث السيدة نجلاء ثابت عن العنقدة الهرمية للقرآن الكريم [نجلاء، 2005]. استخدمت مؤلفة البحث نسخة الكترونية من القرآن الكريم مكتوبة بأحرف لاتينية تجنباً لمشاكل معالجة اللغة العربية، ثم قامت ببناء محلل صرفي خاص بها، ثم بغية تقليل تعقيد المسألة أخذت فقط أكثر 500 كلمة وروداً وطبقت العنقدة على أطول 24 سورة فقط. يعتبر البحث أن كل سورة هي كيان مستقل وأن العنقدة هي لإيجاد علاقات السور مع بعضها. نتج عن هذا البحث شجرة تصنيف تجمعت فيها كل مجموعة من السور مع بعضها في حوالي خمسة مستويات. بعد تجميع السور في مجموعات ومجموعات أعلى نحصل في المستوى الأعلى على مجموعتين تتضمن الأولى إحدى عشرة سورة (البقرة، آل عمران، المائدة، النحل، الحج، الزمر، النساء، التوبة، الأنفال، الأحزاب، النور)، وتتضمن الثانية ثلاثة عشرة سورة (الأنعام، يونس، غافر، الأعراف، هود، النمل، القصص، الإسراء، الكهف، الأنبياء، يوسف، طه، الشعراء). تفسر المؤلفة هذا التقسيم بأن المجموعة الأولى باستثناء سورتي النحل والزمر[1] هي سور مدنية تحدد الشريعة الإسلامية بتفاصيلها الاجتماعية والاقتصادية والقانونية وتوضح العبادات والأخلاق والمعاملات. أما المجموعة الثانية فهي سور مكية تؤكد وحدانية الله عز وجل ونبوة الرسول محمد صلى الله عليه وسلم وقيام الساعة وتعد المؤمنين بالجنة وتتوعد العصاة بالعقاب والكافرين بالنار وتذكر بقصص الأنبياء والأمم السابقة. لكن كيف نفسر الاستثناءات؟ وماذا بشأن باقي سور القرآن الكريم المدنية والمكية وانتمائها إلى إحدى المجموعتين؟

حاول هيرمان مويصل في بحث آخر [مويصل، 2009] التصدي للمسألة ومعالجة مشكلة السور الصغيرة التي واجهتها نجلاء ثابت فاستخدم بعض المفاهيم المعقدة في التوزيع الإحصائي ووصل إلى حد أدنى لطول السورة التي يجب إدخالها في عملية العنقدة هو 300 كلمة. كما حدَّ الكلمات المفتاحية بتسع كلمات هي: الله، لا، رب، قال، كان، يوم، ناس، يومئذ، شر. حصل مويصل على تصنيف هرمي مختلف للسور بحسب الكلمات التي اختارها. أما البحث الأهم في هذا المجال فقد قام به عبد الباقي محمد شرف ضمن أطروحة الدكتوراه [عبد الباقي، 2012]، حيث استفاد ضمن محاولاته لتصنيف سور القرآن الكريم، باستخدام مفاهيم تعلم الآلة وأداة Weka، من بعض المعارف اللغوية والمعارف المتعلقة بمجال تصنيف السور إلى مكية ومدنية. حصل عبد الباقي على ثمان مجموعات مختلفة من السور في مستوى واحد، أربع منها تحتوي على سور مكية أو مدنية فقط ومجموعتان تغلب عليهما السور المكية ومجموعتان فيهما سور مكية ومدنية، تحتوي إحداهما سورة البقرة المدنية مع سورة الأعراف المكية بسبب تشابهما في المحتوى خاصةً فيما يتعلق بقصة سيدنا موسى عليه السلام.

4. تطبيق العنقدة على نص القرآن الكريم

نود الإشارة أولاً إلى أننا استخدمنا نسخةً إلكترونية موثوقة من القرآن الكريم مكتوبة وفق الرسم العثماني دون تشكيل، وفَّرها لنا مشكوراً الزميل الدكتور أشرف قطنة[2]، وقد أوردنا الآيات ضمن أمثلة هذا البحث منها.

[1] يجب أن نستثني أيضاً سورة الحج حسب تصنيف كتب علوم القرآن.

[2] الدكتور أشرف قطنة مؤلف كتاب رسم المصحف والإعجاز العددي، دراسة نقدية في كتب الإعجاز العددي في القرآن الكريم، دار المنار، 1420 هـ.

عند تطبيق العنقدة أو التصنيف على القرآن الكريم تظهر متغيرات جديدة، إذ يمكن تصنيف القرآن كاملاً باعتبار كل آية منه كياناً مستقلاً، أو تصنيف سوره باعتبار أن كل سورة هي كيان مستقل بذاته، أو تصنيف الآيات ضمن السورة الواحدة. وفي جميع الحالات سنحصل على تصنيف هرمي للآيات أو السور بحسب محتواها. سنستعرض مراحل عملنا بتسلسلها الزمني وتطور تعقيدها، حيث بدأنا بمعالجة نصية بسيطة باستخدام برامج Perl بسيطة، جرت الإضافة عليها وتطويرها شيئاً فشيئاً. ثم تبين لنا ضرورة بناء أداة برمجية متكاملة تسمح من خلال موسطاتها بالتبديل بين عدد كبير من خيارات ومتغيرات مسألة العنقدة وإجراء عدد كبير من التجارب المختلفة. بنينا هذه الأداة بلغة Java وحرصنا أن تكون أداة عامة يمكن استخدامها لعنقدة أي مجموعة من النصوص، ثم طبقناها على نصوص القرآن الكريم.

أ. تطبيق العنقدة البسيطة على نص القرآن الكريم كاملاً

المرحلة الأولى: قمنا في البداية بإجراء مجموعة من التجارب فبدأنا بتعداد الكلمات في الآيات لإحصاء التكرارات، ثم البحث عن الآيات التي تحقق أكبر تشابه من حيث الكلمات على أساس تطابق تام في الكلمات دون استخدام تحليل صرفي، أي دون إعادة الكلمات إلى جذورها، ففي هذه الحالة لا تتطابق كلمات مثل "كتب" و"كتاب" و"كاتب" إلا إذا كان لها نفس الرسم العثماني. في هذه المرحلة الأولى لم يجرِ تجاهل أو حذف أي من الكلمات (يجري عادةً حذف مفردات الربط على الأقل والتي تسمى أيضاً الكلمات الزائدة Stopwords وتتضمن حروف الجر والأسماء الموصولة والضمائر ...)، وهنا يمكن أن نتوقف قليلاً ونبرر هذا الخيار السهل! إذ كيف يمكننا تجاهل مفردات من القرآن الكريم أو اعتبارها كلمات زائدة مهما كان السبب؟

جرى في هذه المرحلة اعتبار كل آية هي نص مستقل، فقمنا بتحديد أزواج الآيات (مثنى مثنى) التي تحقق أكبر تشابه باستخدام خوارزمية بدائية تمر على الآيات واحدةً فواحدة وتقارنها مع ما تبقى من آيات لم نصل إليها بعد. واعتمدنا على طريقة مبسطة في حساب التشابه باستخدام العلاقة: $\sum_{i=1}^{n} x_i \cdot y_i$ حسب فضاء الأشعة (X,Y) الذي يمكن استخدامه لتمثيل النصوص من خلال الكلمات المركبة لها وتواتر هذه الكلمات ضمنها.

في هذه المعالجة البدائية لم تكن النتائج معبرة جداً، فنتيجة لعدم حذف مفردات الربط ونتيجة لورود كلمات كثيرة غير مميزة للآيات مثل لفظ الجلالة "الله"، كان للآيات الطويلة النصيب الأكبر من التطابقات فورد لدينا مثلاً تطابق الآيتين:

(يأيها الذين ءامنوا إذا تداينتم بدين إلى أجل مسمى فاكتبوه وليكتب بينكم كاتب بالعدل ولا يأب كاتب أن يكتب كما علمه الله فليكتب وليملل الذي عليه الحق وليتق الله ربه ولا يبخس منه شيا فإن كان الذي عليه الحق سفيها أو ضعيفا أو لا يستطيع أن يمل هو فليملل وليه بالعدل واستشهدوا شهيدين من رجالكم فإن لم يكونا رجلين فرجل وامرأتان ممن ترضون من الشهداء أن تضل إحديهما فتذكر إحديهما الأخرى ولا يأب الشهداء إذا ما دعوا ولا تسموا أن تكتبوه صغيرا أو كبيرا إلى أجله ذلكم أقسط عند الله وأقوم للشهدة وأدنى ألا ترتابوا إلا أن تكون تجرة حاضرة تديرونها بينكم فليس عليكم جناح ألا تكتبوها وأشهدوا إذا تبايعتم ولا يضار كاتب ولا شهيد وإن تفعلوا فإنه فسوق بكم واتقوا الله ويعلمكم الله والله بكل شيء عليم) [البقرة: 282] . (الطلق مرتان فإمساك بمعروف أو تسريح بإحسن ولا يحل لكم أن تأخذوا مما ءاتيتموهن شيا إلا أن يخافا ألا يقيما حدود الله فإن خفتم ألا يقيما حدود الله فلا جناح عليهما فيما افتدت به تلك حدود الله فلا تعتدوها ومن يتعد حدود الله فأولئك هم الظلمون) [البقرة: 229]

المرحلة الثانية: قمنا في هذه المرحلة باستخراج مفردات الربط أو الكلمات الزائدة، حيث جرى تعداد الكلمات كلها وترتيبها تنازلياً بحسب تكراراتها ووضعها في ملف ثم جرى استعراضها يدوياً لحذف الكلمات التي لا تعتبر مفردات ربط. نورد فيما

يلي مقطعاً من ملف مفردات الربط: "من في ما إن لا الذين على إلا ولا وما أن إلى لهم ومن ثم لكم به بما أو ذلك له الذي هو هم وإن كل فيها عن إذا عليهم هذا لم وهو فإن إذ والذين عليكم فلا إنا منهم بعد إنه عليه حتى وهم وإذا أولئك إني أم ولقد بل فيه قد عند قبل إنما ولكن ولو مما منكم فلما ألا لمن دون فمن فما منه فإذا منها كذلك لنا".

مع أنَّنا استبعدنا مفردات الربط عند تحديد أزواج الآيات (مثنى مثنى) التي تحقق أكبر تشابه، إلا أنَّ النتائج ما زالت غير معبرة لأن كلمات مكررة كثيراً مثل لفظ الجلالة "الله" ما زالت تعطي وزناً كبيراً لتطابق الآيات خاصةً في الآيات الطويلة التي تكررت فيها كثيراً. فتطابقت مثلاً آية المداينة التي ذكرناها سابقاً وهي أطول آية في القرآن الكريم (الآية 282 من سورة البقرة) مع الآية الأخيرة من سورة المزمل:

- (إن ربك يعلم أنك تقوم أدنى من ثلثي اليل ونصفه وثلثه وطائفة من الذين معك والله يقدر اليل والنهار علم أن لن تحصوه فتاب عليكم فاقرءوا ما تيسر من القرءان علم أن سيكون منكم مرضى وءاخرون يضربون في الأرض يبتغون من فضل الله وءاخرون يقتلون في سبيل الله فاقرءوا ما تيسر منه وأقيموا الصلوة وءاتوا الزكوة وأقرضوا الله قرضا حسنا وما تقدموا لأنفسكم من خير تجدوه عند الله هو خيرا وأعظم أجرا واستغفروا الله إن الله غفور رحيم). [المزمل:20]

المرحلة الثالثة: جرى في هذه المرحلة إعداد برنامج لتجميع كل الآيات التي لها المسافة نفسها عن آية معينة (ليس فقط آية واحدة هي أول آية كما كان سابقاً). فلم يعد التجميع مثنى مثنى، بل أصبح مع كل آية مجموعة الآيات التي تشبهها ولها نفس المسافة عنها وهي المسافة الأصغر (أي التشابه الأكبر).

كانت هناك بعض النتائج المعقولة في هذه المرحلة كما في المجموعة التالية من الآيات:

- (وإذ نجينكم من ءال فرعون يسومونكم سوء العذاب يذبحون أبناءكم ويستحيون نساءكم وفي ذلكم بلاء من ربكم عظيم) [البقرة: 49]
- (وإذ أنجينكم من ءال فرعون يسومونكم سوء العذاب يقتلون أبناءكم ويستحيون نساءكم وفي ذلكم بلاء من ربكم عظيم) [الأعراف: 141]
- (وإذ قال موسى لقومه اذكروا نعمة الله عليكم إذ أنجيكم من ءال فرعون يسومونكم سوء العذاب ويذبحون أبناءكم ويستحيون نساءكم وفي ذلكم بلاء من ربكم عظيم) [إبراهيم: 6]

إلا أن تأثيرات الكلمات التي لا تؤدي دوراً في تمييز الآيات ظلت كبيرة، كما لاحظنا تجميع بعض الآيات في مجموعة واحدة بسبب اشتراكها في أجزاء مختلفة كما في المجموعة التالية:

- (وإذ تأذن ربك ليبعثن عليهم إلى يوم القيمة من يسومهم سوء العذاب إن ربك لسريع العقاب وإنه لغفور رحيم) [الأعراف: 167]
- (ثم إن ربك للذين هاجروا من بعد ما فتنوا ثم جهدوا وصبروا إن ربك من بعدها لغفور رحيم) [النحل: 110]
- (ثم إن ربك للذين عملوا السوء بجهلة ثم تابوا من بعد ذلك وأصلحوا إن ربك من بعدها لغفور رحيم) [النحل: 119]

المرحلة الرابعة: جرى في هذه المرحلة الاحتفاظ بالكلمات المفتاحية التي تساهم في مطابقة كل آية مع باقي الآيات. ثم جرى تجميع الآيات المتشابهة بعد وضع كل منها في ملف يحمل نفس الاسم في مجلدات أسماؤها هي الكلمات المفتاحية المشتركة. فحصلنا على نتائج من الشكل التالي:

الشكل 2. عنقدة على مستوى واحد للقرآن الكريم كاملاً.

ب. تطبيق العنقدة المتعددة المستويات على نص القرآن الكريم كاملاً

جرى في هذه المرحلة تطبيق الأعمال البرمجية السابقة نفسها لكن مع ضم المجلدات في كل مستوى إلى مستوى أعلى للوصول إلى هرمية كاملة. بلغ عمق الشجرة أكثر من 130. الجذر الرئيسي هو لفظ الجلالة "الله"، لأنه الأكثر تكراراً واشتراكاً بين الآيات. وردت في العقد العلوية كلمات مثل "قل"، "آمنوا"، "كان"، "السموات"، "الأرض"، "الكتاب"، "الناس"، أما الأوراق فهي الآيات الكاملة.

ت. تطبيق العنقدة المتعددة المستويات على جزء واحد

جرى في هذه المرحلة تطبيق البرامج السابقة على أجزاء مختلفة من القرآن الكريم. فعند التطبيق على الجزء الأول مثلاً بلغ عمق الشجرة عشرة مستويات. الجذر الرئيسي هو لفظ الجلالة "الله". ووردت في العقد العلوية كلمات مثل "قل"، "يعملون"، "كان"، "إبراهم"، "اسمعيل"، "الكتاب"، "موسى"، أما الأوراق فهي الآيات الكاملة.

ث. تطبيق العنقدة المتعددة المستويات على سورة الشعراء

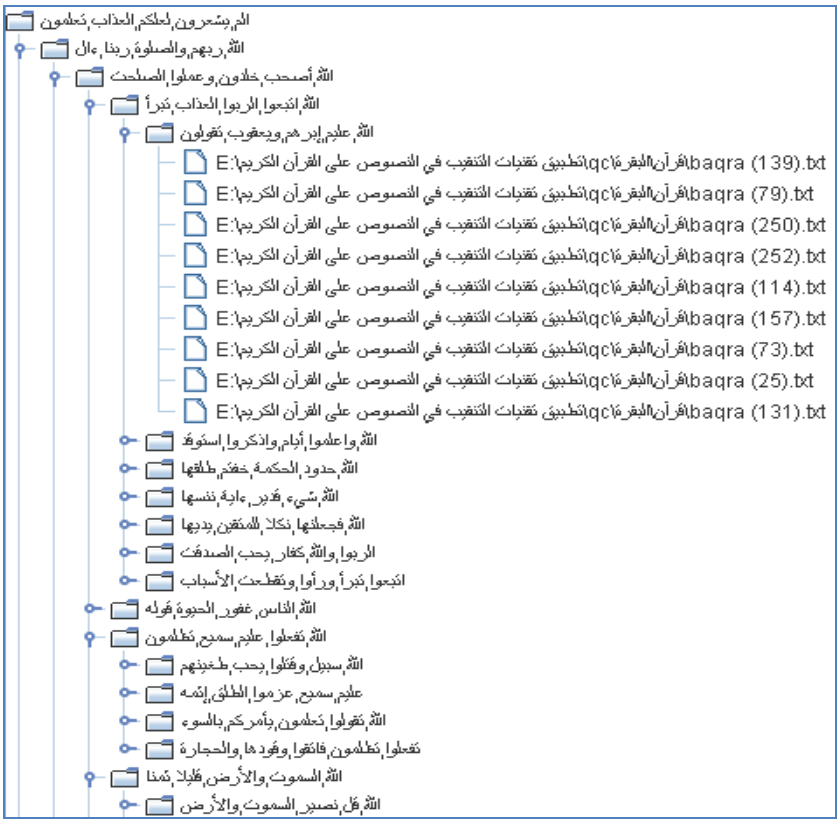
الشكل 3. تطبيق العنقدة المتعددة المستويات على سورة الشعراء.

ج. بناء أداة العنقدة

بعد المجموعة السابقة من الاختبارات تبين لنا ضرورة بناء أداة خاصة بالعنقدة، فحاولنا بناء أداة على درجة عالية من المرونة تسمح بالتحكم بمتغيرات المسألة عند تطبيقها، وتسمح بإجراء تجارب عديدة باستخدام ظروف وسياقات مختلفة [دسوقي والعبدو، 2012].

ح. تطبيق أداة العنقدة على نص سورة البقرة

الحالة الأولى: في الحالة الأولى لم نستخدم المحلل الصرفي بل جرت مطابقة الكلمات كما هي. إلا أن المشكلتين السابقتين (تأثيرات الكلمات التي لا تؤدي دوراً في تمييز الآيات، وتجميع بعض الآيات في مجموعة واحدة بسبب اشتراكها في أجزاء مختلفة) ما زالتا موجودتين. في الشكل التالي نموذج من النتائج، حيث نمثل كل عنقود بالكلمات المفتاحية الخمسة الأكثر تكراراً في مكوناته:



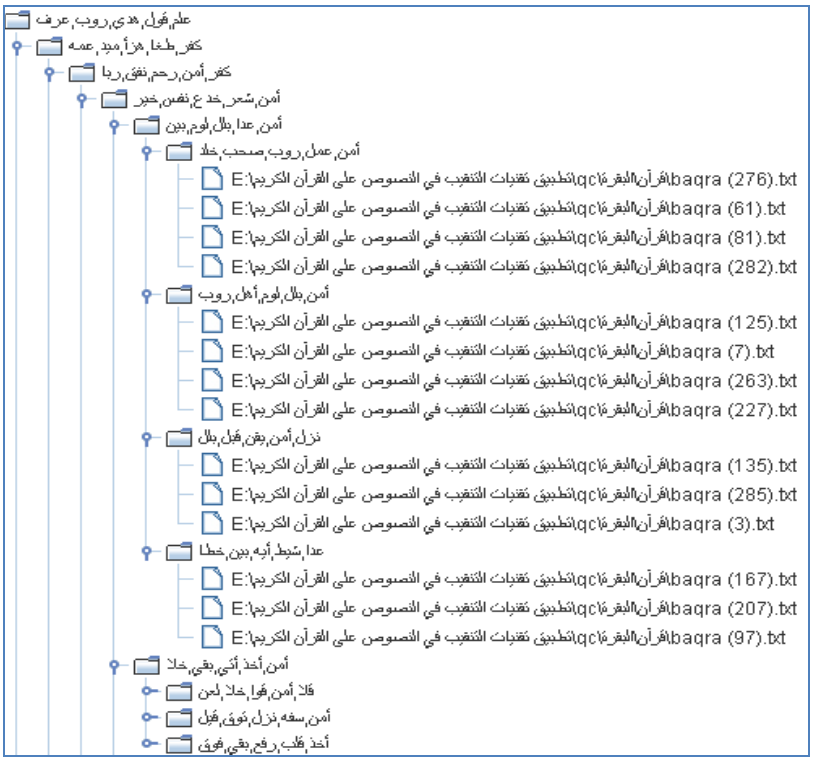
الشكل 4. تطبيق أداة العنقدة على سورة البقرة دون استخدام المحلل صرفي.

نلاحظ الآيات التالية ضمن عنقود واحد:

- (أم تقولون إن إبرهم وإسمعيل وإسحق ويعقوب والأسباط كانوا هودا أو نصرى قل ءأنتم أعلم أم الله ومن أظلم ممن كتم شهدة عنده من الله وما الله بغفل عما تعملون) [البقرة: 139]
- (وقالوا لن تمسنا النار إلا أياما معدودة قل أتخذتم عند الله عهدا فلن يخلف الله عهده أم تقولون على الله ما لا تعلمون) [البقرة: 79]

- (فهزموهم بإذن الله وقتل داود جالوت وءاتيه الله الملك والحكمة وعلمه مما يشاء ولولا دفع الله الناس بعضهم ببعض لفسدت الأرض ولكن الله ذو فضل على العلمين) [البقرة: 250]
- (تلك الرسل فضلنا بعضهم على بعض منهم من كلم الله ورفع بعضهم درجت وءاتينا عيسى ابن مريم البينت وأيدنه بروح القدس ولو شاء الله ما اقتتل الذين من بعدهم من بعد ما جاءتهم البينت ولكن اختلفوا فمنهم من ءامن ومنهم من كفر ولو شاء الله ما اقتتلوا ولكن الله يفعل ما يريد) [البقرة: 252]
- (ولله المشرق والمغرب فأينما تولوا فثم وجه الله إن الله وسع عليم) [البقرة: 114]
- (إن الصفا والمروة من شعائر الله فمن حج البيت أو اعتمر فلا جناح عليه أن يطوف بهما ومن تطوع خيرا فإن الله شاكر عليم) [البقرة: 157]
- (ثم قست قلوبكم من بعد ذلك فهي كالحجارة أو أشد قسوة وإن من الحجارة لما يتفجر منه الأنهر وإن منها لما يشقق فيخرج منه الماء وإن منها لما يهبط من خشية الله وما الله بغفل عما تعملون) [البقرة: 73]
- (إن الله لا يستحي أن يضرب مثلا ما بعوضة فما فوقها فأما الذين ءامنوا فيعلمون أنه الحق من ربهم وأما الذين كفروا فيقولون ماذا أراد الله بهذا مثلا يضل به كثيرا ويهدي به كثيرا وما يضل به إلا الفسقين) [البقرة: 25]
- (ووصى بها إبرهم بنيه ويعقوب يبني إن الله اصطفى لكم الدين فلا تموتن إلا وأنتم مسلمون) [البقرة: 131]

الحالة الثانية: في الحالة الثانية جرى استخدام المحلل الصرفي فكان عمق الشجرة أكبر وعدد المجموعات الفرعية أكثر، وكان التجميع منطقياً أكثر، إلا أنه ما تزال هناك عيوب كما في الشكل 5، حيث نرى أن الآيات الثلاث الأولى تجميعها منطقي أما الآية الرابعة فقد ضمت إليها بسبب اشتراكها معها في الجذر "أمن".

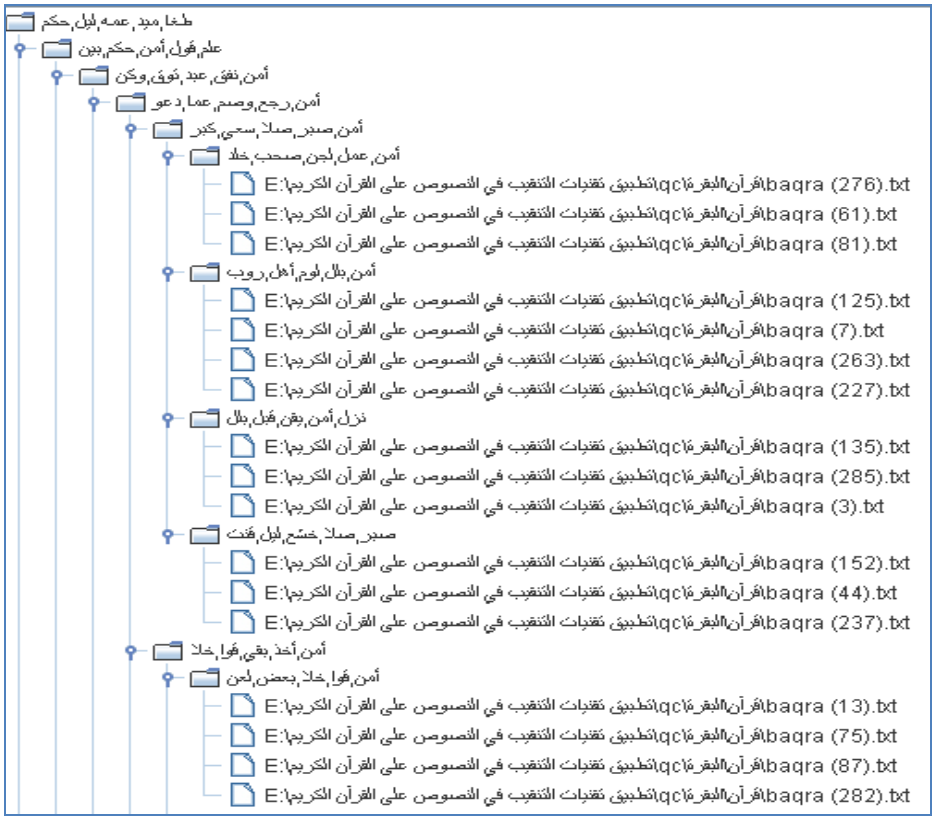


الشكل 5. تطبيق أداة العنقدة على سورة البقرة مع استخدام المحلل الصرفي.

لاحظ العنقود الأول الذي يحتوي الآيات:

- (إن الذين ءامنوا وعملوا الصلحت وأقاموا الصلوة وءاتوا الزكوة لهم أجرهم عند ربهم ولا خوف عليهم ولا هم يحزنون) [البقرة: 276]
- (إن الذين ءامنوا والذين هادوا والنصرى والصبين من ءامن بالله واليوم الأخر وعمل صلحا فلهم أجرهم عند ربهم ولا خوف عليهم ولا هم يحزنون) [البقرة: 61]
- (والذين ءامنوا وعملوا الصلحت أولئك أصحب الجنة هم فيها خلدون) [البقرة: 81]
- (وإن كنتم على سفر ولم تجدوا كاتبا فرهن مقبوضة فإن أمن بعضكم بعضا فليؤد الذي اؤتمن أمنته وليتق الله ربه ولا تكتموا الشهدة ومن يكتمها فإنه ءاثم قلبه والله بما تعملون عليم) [البقرة: 282]

بعد ذلك تمكنًا باختيار عتبتي تقليم دنيا وعليا مناسبتين، من الوصول إلى مجموعة تضم فقط الآيات الثلاثة الأولى السابقة(انظر الشكل 6)، لاحظ أيضاً أن المجموعة الأخيرة (الآيات 167،207،97) في نفس التصنيف والتي تتعلق بالعداوة والشيطان والخطأ والبينة، قد تغيرت فأصبح لدينا مجموعة جديدة من الآيات (152،44،237) تتعلق بالصبر والصلاة والخشوع والليل والقنوت.



الشكل 6. تطبيق أداة العنقدة على سورة البقرة مع استخدام المحلل الصرفي واستخدام عتبات تقليم مناسبة.

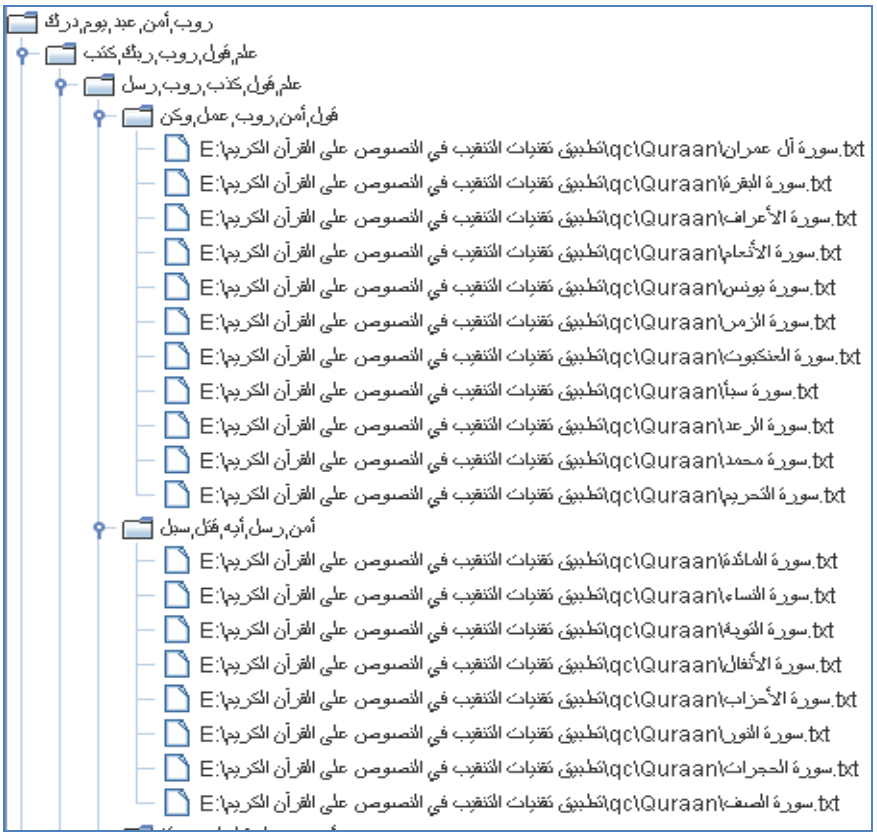
لاحظ كيف جرى تغيير العناقيد عن الشكل السابق والحصول على نتائج أكثر دقة. العنقود الأول مثلاً يضم الآيات:

- (إن الذين ءامنوا وعملوا الصلحت وأقاموا الصلوة وءاتوا الزكوة لهم أجرهم عند ربهم ولا خوف عليهم ولا هم يحزنون) [البقرة: 276]

- (إن الذين ءامنوا والذين هادوا والنصرى والصبين من ءامن بالله واليوم الأخر وعمل صلحا فلهم أجرهم عند ربهم ولا خوف عليهم ولا هم يحزنون) [البقرة: 61]
- (والذين ءامنوا وعملوا الصلحت أولئك أصحب الجنة هم فيها خلدون) [البقرة: 81]

خ. تطبيق أداة العنقدة على سور القرآن الكريم

في هذه الحالة اعتبرنا أن كل سورة من القرآن الكريم هي نص مستقل، ثم جرت مقارنة كل سورة مع السور الأخرى مع استخدام المحلل الصرفي. لاحظنا تجميع سور مثل الفاتحة مع القدر، والمجادلة مع المنافقون، والشمس مع الليل، والبينة مع الانشقاق، والتين مع العصر، والناس مع النصر، والانفطار مع القيامة والعاديات، والحشر مع الفتح والطلاق، والتغابن مع الحديد والممتحنة والبروج، يُظهر الشكل 7 نموذجاً من النتائج.



الشكل 7. تطبيق أداة العنقدة على سور القرآن الكريم.

5. الختام

قمنا في هذا البحث بتطبيق إحدى تقنيات التنقيب في النصوص وهي تقنية العنقدة على القرآن الكريم. طبيعة العمل في التنقيب في النصوص أنك لا تبحث عن شيء محدد أو أنك لا تعرف عن أي شيءٍ تبحث، وربما من هنا جاء تعبير "التنقيب"، وإنما تتوقع الوصول إلى علاقات أو روابط أو معلومات جديدة لم تكن معروفة مسبقاً. توقعنا في بداية العمل الوصول إلى تصنيف موضوعي للقرآن الكريم يسهم في دعم الإعجاز الدلالي للقرآن الكريم، لكن يبدو أننا ما نزال بعيدين عن ذلك لعدة أسباب. أهم هذه الأسباب أننا اعتمدنا حصراً على مبنى الكلمات ولم نعتمد على المعنى ولم نتطرق حتى

للمرادفات. أما الخيارات والتباديل المختلفة التي طبقناها فهي كلها تتعلق بالمبنى الصرفي. وقد يكون من الضروري إضافة مراحل أخرى من المعالجة تتطرق إلى المعنى والدلالة مثل نظام تحديد أنماط الكلمات وأنظمة التحليل النحوي والتحليل الدلالي، أو على الأقل المرادفات. لاحظنا من جهة أخرى أن بعض الكلمات في آية معينة يكون لها أهمية أكبر بكثير من كلمات أخرى، بل قد تكون هي محور موضوع الآية، مما يجعل من الضروري إعطاؤها ثقلاً أكبر من غيرها عند حساب التشابهات. فقد لاحظنا أن كثيراً من التشابهات تنتج عن كلمات مساعدة لا تُمثِّل الموضوع الأساسي للآيات. نتيجة أخرى مهمة تعيد لنا شيئاً من معنوياتنا وتصب في الهدف الأساسي الذي أجرينا البحث من أجله وهو إعجاز القرآن الكريم، هي أن القرآن الكريم لعظيم قدره وسمو معانيه وعمقها وصلاحه للعصور كلها، لا يمكن كشف أسراره بمعالجات سطحية بسيطة تنجح مع النصوص الأخرى التي يكتبها البشر. بل يمكننا في طرح أعمق أن نقول إن البرامج الحاسوبية، ومنها بحثنا هذا، تكشف لنا يوماً بعد يوم معجزة الخالق في العقل البشري. فكثيرٌ من الذين تدبروا القرآن الكريم من المتقدمين والمتأخرين فسَّروا الآيات ببعضها أو ربطوا بينها، ومنهم من صنفها ضمن مجموعات حسب معناها، ومعجزة العقل البشري هنا في معرفة الكلمة الهامة التي تشكل محور الآية ومعرفة معناها وارتباطها الدلالي مع كلمات أخرى وآيات أخرى.

6. المصادر والمراجع

[دسوقي والعبدو، 2012]

Desouki M. S., AL-Abdo A. A., Experiments in Mining Arabic Texts, International Journal for Communications and Computer Engineering, Vol. 2, No. 1, 2012.

[رونين وجيمس، 2007]

Ronen Feldman && James Sanger, The Text Mining Handbook, Cambridge University Press, 2007.

[عبد الباقي، 2012]

Abdul Baquee Muhammad, Annotation of Conceptual Co-reference and Text Mining the Qur'an, Ph.D. Thesis, Leeds University, 2012.

[محمد زكي وأكرم، 2011]

محمد زكي خضر وأكرم محمد زكي، دراسة احصائية لكلمات القرآن الكريم، المؤتمر الثالث للغة العربية وآدابها، الجامعة الإسلامية العالمية بماليزيا،2011.

[مويصل، 2009]

Moisl H., Sura Length and lexical probability estimation in cluster analysis of the Qur'an. ACM Transactions on Asian language information processing 8, 4 (Dec. 2009), 1–19.

[نجلاء، 2005]

Naglaa Thabet, Understanding the thematic structure of the Quran: an exploratory multivariate approach, Proceedings of the ACL Association for Computational Linguistics Student Research Workshop,, Michigan, 2005.

[نيكولاس وإدوارد، 2007]

Nicholas O. Andrews and Edward A. Fox, Recent Developments in Document Clustering, Department of Computer Science, Virginia Tech, Blacksburg, 2007.

7. جدول الالفاظ

English	عربي
Multi-Level Document Clustering	عنقدة الوثائق المتعددة المستويات
Text Mining	التنقيب في النصوص
Cluster	عنقود
Stopwords	الكلمات الزائدة

8. الخلاصة باللغة الانجليزية

Applying Multi-Level Documents Clustering to Quran Text

Desouki Mohammad Said

Higher Institute for Applied Sciences and Technology, P.B. 31983 Damascus, Syria

said.desouki@hiast.edu.sy

Abstract

The Objective of this paper is to apply Multi-Level Document Clustering, one of the important text mining techniques, to the holly Quran text. We started by extracting the Quranic Ayats keywords and measuring their similarity using this keywords, and then categorizing them into groups based on their keywords and similarities. We, then, performed similarity measures between the groups to create higher levels until constructing a hierarchical categorization tree describing the distribution and grouping of Quranic Ayat based on their words.

Keywords: Quran Text, Multi-Level Clustering, Hierarchical Categorization.