توظيف خوارزمية الأسبقية في التنقيب عن أعلام المحدثين بطرق الحديث النبوي

آلاء فيصل المختار[1]، حسن مظفر الرزو[2]
[1] وزارة الكهرباء، الموصل، جمهورية العراق
[2] هيئة المستشارين، الموصل، جمهورية العراق
[1]afs712000@yahoo.com , [2]halrizzo@gmail.com

الخلاصة: تعد مسألة التنقيب عن أعلام نقلة الحديث النبوي الشريف من المسائل المهمة التي تمنح الباحث الإسلامي في ميدان علوم الحديث النبوي فرصة سبر رواة الحديث وطرق نقل الرواية للتأكد من سلامة المرويات التي تحفل كتب السنن النبوية والآثار. وبالرغم من الجهد الكبير الذي قام به أئمة الجرح والتعديل في تصنيف أمهات الكتب التي تعنى برجال الحديث إلا أننا أصبحنا أمام تحد كبير في التعامل مع هذه النصوص بطريقة آلية وبعيداً عن النص المطبوع والذي يتطلب معرفة عميقة وجهداً كبيراً لتتبع رواة الحديث ضمن سلاسل الإسناد. ولتحقيق هذه الغاية استهدف البحث توظيف خوارزمية الأسبقية على نص سنن الدارمي بوصفها مرحلة أولية، يمكن بعد ضمان نجاحها السعي نحو توسيع دائرة توظيفها لتشمل نخبة من أمهات كتب السنة النبوية الشريفة.وقد أثمرت عملية التنقيب المعلوماتي التي وظفت خلالها خوارزمية الأسبقية عن اقتناص أسماء المحدثين مبوبة على عدة مستويات (بحسب ورود اسم العلم ضمن سلاسل الإسناد: أحادي، ثنائي، ثلاثي، رباعي، خماسي،والاسم الكامل مع اللقب). ولترسيخ قناعتنا بصحة نتائج عملية التنقيب في تحديد هوية المحدث، اعتمدت دالة الموثوقية Confidence والتي يمكن أن تعد خطوة مهمة على طريق توفير المادة العلمية لنقد طريق الرواية.

الكلمات الجوهرية : تنقيب البيانات، تنقيب النصوص، خوارزمية الأسبقية، دالة الموثوقية، تحليل محوسب لأسماء أعلام الحديث النبوي الشريف، خوارزميات النصوص.

1. المقدمة

ساهم التطور في العلم و الاقتصاد وتكنولوجيا المعلومات و الاتصالات بدور كبير في زيادة كمية البيانات الرقمية بالآونة الأخيرة بحيث لم تعد وسائل التحليل التقليدية والإحصائية قادرة على التعامل مع الكميات الهائلة المطروحة من البيانات. وقد ظهر التنقيب عن البيانات (Data Mining) منذ أواخر الثمانيات قدرته الغاشمة بوصفه أحد الحلول الناجحة لتحليل محتوى البيانات الرقمية , وذلك بتحويلها من مجرد معلومات متراكمة وغير مفهومة (بيانات) إلى معلومات (تمتلك قيمة معرفية) يمكن استغلالها و الاستفادة منها [ديين، 2005].بصورة عامة تعتمد عملية التنقيب عن البيانات على مبادئ جوهرية تسعى الى تقسيم موارد المعرفة الرقمية الى ثلاثة مفاصل أساسية:

1. البيانات Data: هي عبارة عن الحقائق والأرقام والنصوص التي يمكن أن تعالج من قبل الحاسب.
2. المعلومات Information: النماذج والعلاقات بين تلك البيانات والتي تشكل مادة معرفية مفيدة.
3. المعرفة Knowledge: تعد خلاصة المحتوى الرقمي للمعلومات والتي يمكن استنباطها من الأنماط السائدة في المعلومات، والقواعد الارشادية التي تستمد من مادتها [اندرياس، 2005].

استخدمت تقنية تنقيب النصوص Text Miningلاستخراج المعرفة المفيدة في الكميات الكبيرة من النصوص غير المنتظمة Non Structured وذلك من خلال تحويل النص الحر إلى نص يمكن تحليله من خلال إيجاد حلول للوصول إلى المعرفة في الكم الهائل من النصوص الموجودة على الحاسبات وشبكة الانترنيت باستخدام تقنيات الفهم الآلي، والبحث عن المعلومات، والفهم الآلي للغات الطبيعية. تتم هذه العملية من خلال معالجة أولية للنص باستخراج الكلمات والمفاهيم وبعدها إيجاد العلاقات بين المفاهيم وتمثيل النص في قواعد ربط وتصنيفه وإمكانية عرضه للمستخدم بطريقة سهلة الفهم [لانا، 2009].

ويبقى التحدي الكبير أمام العالم العربي في تفعيل دور اللغة العربية وإمكانية استعمالها في الميادين العلمية والتقنية ومواكبة التطور في ميادين تكنولوجيا المعلومات في كافة أنشطة المجتمع. وكان لنصوص التراث الإسلامي بمختلف تجلياتها المعرفية فرصة خصبة للولوج في ميدان التنقيب المعلوماتي بعد ان أولاها الباحثون وطلبة العلم عناية بالغة، شملت تحقيق مخطوطاتها، وإعادة دراسة نصوصها، والتأليف في كثير من مسائلها الشائكة، ثم بلغت آثار تقنيات المعلومات تراثنا العريق فباشرت بتحويله إلى نصوص رقمية استوطنت مواقع الانترنيت، أو الأقراص المدمجة التي انتشرت بكثافة في سوق المعرفة الإسلامية المعاصرة. بيد أن العقبة ذاتها قد برزت مع ازدياد حجم الفيض المعلوماتي لمادة النصوص الإسلامية الرقمية، فلم نعد قادرين على مطالعتها أو دراستها بأناة لكي نحسن فهم مضامينها، وتوظيف مادتها المعرفية على مفردات حياتنا اليومية. من أجل هذا سنحاول دراسة بعض النتائج التي يمكن الحصول عليها من تطبيق آلية التنقيب المعلوماتي على نصوص الحديث المنتخبة، لنطرق أبوابا جديدة يمكن أن تمنحنا فرصة لننهل معارف جديدة من نصوص لم تعد بمتناول ملكتنا العلمية، رغم أنها تشخص على رفوف مكتباتنا العامة والشخصية على حد سواء. وسيسهم هذا العمل في فتح أبواب واسعة أمام العاملين في ميدان الحوسبة الذكية لإنشاء خوارزميات برمجية قادرة على التعامل مع النصوص العربية، كما أنه سيدعم في الوقت نفسه العاملين في ميادين التحليل اللغوي والدلالي، والعاملين في ميدان نظرية المعرفة لتعميق فهمهم بدلالة مختلف أنواع المفردات التي توظف في كتب الحديث بجميع تجلياتها المعرفية.

2. عينة البحث

حرصنا على تطبيق خوارزمية إيجاد أسماء الأعلام على كتاب مسند الدارمي المعروف بـ (سنن الدارمي)[1] . وقد اشتهرت سنن الدارمي عند المحدثين وأطلق عليه اصطلاح (المسند) على خلاف اصطلاحهم. قال العراقي: "اشتهر تسميته بالمسند كما سمى البخاري كتابه بالمسند لكون أحاديثه مسندة "وتبوأ هذا السفر الجليل من الأحاديث النبوية مكانة مهمة بين كتب السنة النبوية المشهورة حتى قال بعض أهل العلم : "كتاب الدارمي أحرى وأليق بجعله سادساً للكتب لأن رجاله أقل ضعفاً، ووجود الأحاديث المنكرة والشاذة نادرة فيه، وله أسانيد عالية، وثلاثياته أكثر من ثلاثيات البخاري[2].

[1]. هو الإمام الحافظ شيخ الإسلام بسمرقند، أبو محمد عبد الله بن عبد الرحمن بن الفضل بن بهرام بن عبد الصمد التميمي الدارمي السمرقندي. كان مولده سنة إحدى وثمانين ومائة (181) للهجرة. نسبه: الدارمي، نسبة إلى دارِم. توفي رحمه الله سنة خمس وخمسين ومائتين (255) للهجرة يوم التروية، ودفن يوم عرفة يوم الجمعة. وهو ابن أربع وسبعين سنة.

[2]. أنظر بواسطة السنة النبوية الشريفة.www.el3b.com/islam/Islamic-school / sunnah /s3-11.html

مورست عملية التنقيب المعلوماتي لطرق اسناد الحديث النبوي والآثار الواردة فيه عن طريق توظيف سلسلة من الخوارزميات الحاسوبية لاستخلاص أجزاء محددة من أسماء نقلة الحديث النبوي، مع تمييز مراتب حضورها في سلسلة الإسناد.

3. الأسس المفاهيمية والمعلوماتية لآلية التنقيب المعلوماتي

تعد آلية التنقيب المعلوماتي جزءاً من مجموعة الآليات المحوسبة التي تسعى الى استنباط الأنساق المعرفية، والعلاقات المقيمة بين المفردات السائدة في النصوص. وتتألف هذه الآلية من مجموعة من معالجات الحوسبة الذكية مع توظيف خوارزميات برامج اللغة الطبيعية بحيث تكون قادرة على توليد عدد كاف من الأنساق المعرفية السائدة في النص، مع تحديد طبيعة العلاقات التي تربط فيما بينها [الرزو،2006]. وتسعى هذه الآلية الى تقطير محتوى النصوص عبر عملية الكشف عن المباني والمعاني المستبطنة داخل النصوص، والتي تتسم بتعقيد محتواها، أو تلك التي تتميز بحجمها الكبير الذي تعجز الطرق التقليدية في استيعاب المادة المعرفية المطروحة فيها. ويكمن الهدف الأساس لتحويل النص التقليدي إلى وصف دلالي محوسب لمعالجة المحتوى الرقمي وتحويله إلى نسق يسهل التعامل معه بواسطة تقنيات الحوسبة الذكية، وبما يضمن التحليل العميق للمفردات الموجودة فيه، والعلاقات الدلالية الموجودة فيما بينها [الرزو،2008].

متى استكملت عملية تقطيع النص وترميزه Tokenization عن طريق إزالة الفراغات المقيمة بين المفردات، وإزالة النقاط الميزة لنهاية العبارات، وغيرها من الإشارات والهيئات المستخدمة داخل تركيبة النص، تصبح مفردات النص متوافقة مع الأنساق المعلوماتية التي يسهل التعامل معها بواسطة تقنيات الحوسبة الذكية [فيلدمان ومن معه، 1998].

وتستخدم خوارزميات ذكية لتمييز المراتب المختلفة للألفاظ فتنتقى الأسماء، والأماكن، والصفات، عبر توظيف قواعد لغوية . دلالية Rule-based Approach. بعد ذلك تمارس عملية التبويب الموضوعي، والتنقيب عن الأنساق المفاهيمية التي تربط الكلمات فتنتج لنا اصطلاحات ومعاني يمكن استثمارها في تعميق فهمنا بمحتويات النص ومضامينه المعرفية [فينسنت، 2005]. بصورة عامة تتوفر مجموعة من الخوارزميات التي تستخدم في معالجة البيانات غير المهيكلة ومن هذه الخوارزميات التي تستخدم في تنقيب البيانات

خوارزمية الأسبقية A Priori Algorithmus

استخدمت خوارزمية الأسبقية في تنقيب البيانات لاشتقاق قواعد التجميع association rule صممت للعمل مع قواعد البيانات التي تحتوي على معاملات [ماركوس، 2005]. الناتج من تنفيذ الخوارزمية هو مجموعة من القواعد التي تشير إلى عدد المرات التي ترد العناصر في مجموعات البيانات [سانجبيف ومن معه، 2010]. تستخدم هذه الخوارزمية للعثور على الارتباطات القائمة بين مجموعات مختلفة من البيانات حيث تمتلك كل مجموعة عدداً من البنود Items. ويطلق عليها اصطلاح المعاملات ransction. تحاول الخوارزمية العثور على مجاميع فرعية مشتركة بين ما لا يقل عن الحد الأدنى من عدد(c) من مجموع العناصر وتعتمد فيها المعالجة من الأسفل الى أعلى "Bottom Up" حيث يتم توسيع المجاميع الفرعية المتكررة بعنصر واحد في كل مرة ويسمى بالجيل المرشح Candidate Generation. ثم تقارن مجموعة العناصر مع البيانات بصورة مستمرة لحين عدم العثور على توسعات أخرى ناجحة، حينئذ تتوقف فاعلية الخوارزمية [اكسندونك ومن معه، 2007]. ويمكن مراجعة آلية عمل هذه الخوارزمية من خلال سلسلة الايعازات البرمجية أدناه.

A Priori Algorithmus Ck: Candidate itemset of size k Lk: frequent itemset of size k L1= {frequent items}; for(k= 1; Lk!=∅; k++) do begin Ck+1= candidates generated from Lk; for each transaction t in database do increment the count of all candidates in Ck+1 that are contained in t Lk+1= candidates in Ck+1with min_support End Return ∪k until no more frequent item sets found

استفسارات الجبل الثلجي Iceberg Queries

سميت باستفسارات الجبل الثلجي نسبة إلى الكمية الكبيرة من البيانات المسترجعة من جملة استفسار صغيرة. ويمكن تشبيه العديد من استفسارات تنقيب البيانات بجبل ثلجي لأنها تمارس عملها على كمية كبيرة من البيانات[جينيوك ومن معه، 2000]. وتنفذ هذه الاستفسارات بواسطة مجموعة من الوظائف وعلى عدد من الصفات / الخصائص Attribute لإيجاد القيم الإجمالية للموضوع الذي تمارس عليه عملية الاستفسار، بناءاً على القيم المحددة كما في الإيعاز التالي [ميني ومن معه، 1998].

HAVING agr(R.B)=CONSTANT Select R.A1 , R.A2 ,..... $R.A_k$, agr(R.B) FROM Relation R GROUP by R.A1 , R.A2

4. نهج المعالجة المحوسبة والخوارزميات المستخدمة

وظفت خلال هذا البحث مجموعة من الخوارزميات التي تضمن توفير فرصة كافية لإيجاد مفردات النص ذات الصلة بطريق الإسناد الذي تناولته الدراسة، مع ممارسة عملية التنقيب عن طبيعة الدلالات الكامنة وراء استخدامها في عمليات إنشاء المفاهيم الواردة فيه. تألفت الخوارزميات المستخدمة في برنامجنا للتنقيب المعلوماتي من المراحل الآتية :[جيوي ومن معه، 2006]

– تنقية البيانات Data Cleaning : وهي مرحلة عزل البيانات التي تحتوي على تشويش أو شوائب Noise من مادة النص الذي تناولناه بالتحليل. ونفذت بمجموعة إجراءات ببرامج معالج النصوص MS–word عن طريق إزالة الفراغات المقيمة بين المفردات والنقاط المميزة لنهاية العبارات وغيرها من الإشارات والهيئات المستخدمة داخل تركيبة النص مثل الفواصل والأرقام والعلامات المستخدمة في الطبع.

– نقل البيانات Data Transformation : وهي عملية نقل البيانات التي تم اختيارها إلى شكل ملائم لإجراءات البحث والاسترجاع.ونفذت بلغة البرمجة C Language حيث قامت بتحويل النص إلى قاعدة بيانات تألفت من حقل واحد، لكي يسهل التعامل مع مفرداتها.

– استخراج البيانات Data Mining :في هذه المرحلة تم تطبيق أسلوب ذكي لاستخراج نماذج مفيدة قدر الإمكان وشملت تنفيذ الخوارزميات التالية :

1– خوارزمية الأسبقية Prior Algorithm : وتم تنفيذها بواسطة برنامج صمم بلغة قواعد البيانات MS–Access لتكوين اصطلاح اسم المحدث المفرد الوارد بعد صيغ الرواية المعتمدة في جدول(3) من خلال إيجاد الاسم والبحث عنه في داخل الملف، وحساب عدد مرات تكراره في الملف وخزنه في ملف يحتوي مجموعة الأسماء الملتقطة من أسانيد الحديث النبوي، مع عدد مرات تكرارها.

2- استفسارات الجبل الثلجي Iceberg_Query : وتم تنفيذها بواسطة جمل SQL كتبت بلغة قواعد البيانات MS-Access ووظفت في اقتناص الأسماء الثنائية والأسماء الثلاثية والرباعية والخماسية واعتمد مبدأ تمييز اسم المحدث الثنائي بوصفه الكيان الوارد بين صيغتي تحديث متتاليتين مضافاً عليه شرط ورود صيغة (بن/ابن) الفاصلة بين اسم المحدث ووالده، مع تكرار الحالة مع الاسم المناظر إلى الجد لتكوين الاسم الثلاثي أو الرباعي والخماسي، كما في جمل SQL المدرجة3، وروعي في كتابة جمل الاستفسار والتنقير عن الأسماء التي تبدأ بكلمة عبد مثل عبد الله، وتحديد موقعها، ومن ثم دمجها بالأسماء الاعتيادية لتكوين قاعدة موحدة مع ممارسة عملية التجميع Group لإيجاد تكرار الأسماء في قواعد البيانات المكونة سابقا وعدد مرات ورودها في عموم مسند الدارمي لتكوين الاسم الثنائي والاسم الثلاثي.

3- استفسارات الجبل الثلجي IceBerg_Query لإيجاد أسانيد الرواة المنتهية باللقب حيث استخدمت الخوارزمية السابقة لإيجاد أسانيد الرواة مع إضافة شرط لجملة Select Query Langage SQL بان تنتهي الكلمة الأخيرة على حرف (ي . النسبة) وقد اشترط أن لا يكون حضورها ضمن كلمات مثل (في، اخبرني...). - تمثيل المعرفة Knowledge Representation : وهي المرحلة الأخيرة من مراحل اكتشاف المعرفة في قواعد البيانات والتي تساعد المستفيد في فهم و وتفسير نتائج استخراج البيانات، يتم في هذه المرحلة تطبيق أسلوب ذكي لاستخراج نماذج مفيدة قدر الإمكان واستخدمت قواعد التجميع Association Rule وتطبيق دالة الموثوقية عليها Confidence Rule. وقد روعي في عمل هذه الخوارزميات جميعاً مبدأ المحافظة على ترتيب ورود المفردات داخل النص لكي تكون المعلومات أسهل تناولاً عند البحث عن الدلالات، أو تحديد نمط العلاقات القائمة فيما بينها، وبين بقية المفردات السائدة في نصوص مسند الدارمي.

5. تقييم حصيلة المعالجة المعلوماتية وبيان نتائجها

أثمرت المعالجة التي ارتكزت إلى المبادئ التي نوقشت أعلاه عن مجموعة ثرية من المعلومات التي أودعت بقواعد البيانات كمثال عليها الشكل(1) والتي استخدمت في استفسارات الجبل الثلجي لإيجاد أسانيد الرواة وخزنت محصلة النتائج في الجداول (1،3،2،4).

book1_1 لحقل	book2_1 لحقل	book3_1 لحقل	book4_1 لحقل	book5_1 لحقل	book6_1 لحقل	book7_1 لحقل	book8_1 لحقل
حدثنا	خالد	بن	مخلد	حدثنا	مالك	عن	طلحة
حدثنا	سليمان	بن	حرب	حدثنا	حماد	بن	سلمة
حدثنا	سعيد	بن	زيد	عن	شعيب	بن	الحبحاب
حدثنا	حجاج	بن	منهال	حدثنا	حماد	بن	سلمة
حدثنا	حماد	بن	سلمة	عن	حبيب	بن	أبي
حدثنا	مسلم	بن	إبراهيم	حدثنا	وهيب	حدثنا	أيوب
حدثنا	سليمان	بن	حرب	حدثنا	شعبة	عن	مغيرة
حدثنا	سليمان	بن	حرب	عن	حماد	بن	سلمة
حدثنا	الفضيل	بن	مرزوق	حدثنا	عدي	بن	ثابت
حدثنا	حميد	بن	الأسود	عن	عيسى	قال	سمعت
حدثنا	محمد	بن	الصلت	حدثنا	هشيم	أخبرنا	مغيرة
حدثنا	سعيد	بن	منصور	حدثنا	أبو	عوانة	عن
حدثنا	يحيى	بن	يمان	عن	المنهال	بن	خليفة
حدثنا	محمد	بن	عيسى	حدثنا	معمر	حدثنا	خصيف
حدثنا	يحيى	بن	حمزة	عن	سليمان	بن	داود
حدثنا	يعقوب	بن	إبراهيم	حدثنا	يحيى	بن	سعيد
حدثنا	هاشم	بن	القاسم	حدثنا	شعبة	قال	قتادة

الشكل(1) نموذج لقاعدة البيانات المستخدمة في استفسارات الجبل الثلجي

[1] " Selectكتاب سنن الدارمي."الحقل2" ,كتاب سنن الدارمي."الحقل3" ,كتاب سنن الدارمي."الحقل4] ,الحقل2] &""&[الحقل3] &""&[الحقل4 AS [الاسم الثنائي

FROMكتاب سنن الدارمي) WHEREالحقل3"="بن) AND ("الحقل4"<>عبد) AND ("الحقل1"<>عبد("

يظهر الجدول (1) نموذجاً من أسماء الأعلام كاملة مع اللقب والتي استخرجت بخوارزمية الجبل الثلجي. ويبدو واضحاً من بيانات الجدول أن خوارزمية الجبل الثلجي قد نجحت في اقتناص أسماء المحدثين مع تثبيت الألقاب والتي تعد خطوة جوهرية للفصل بين أسماء الأعلام الواردة ضمن نصوص نقلة الحديث النبوي.

الجدول (1) نموذج من أسماء الأعلام كاملة مع اللقب.

الاسم الاول	الاسم الثاني	الاسم الثالث	الاسم الرابع	الاسم الخامس	الاسم السادس	اللقب	اسماء الاعلام مع اللقب
عبد	الرحمن	بن	محمد	بن	المظفر	الداوودي	عبد الرحمن بن محمد بن المظفر الداوودي
عبد	الله	بن	أحمد	بن	حمويه	السرخسي	عبد الله بن أحمد بن حمويه السرخسي
عبد	الله	بن	زيد	بن	عاصم	المازني	عبد الله بن زيد بن عاصم المازني
عبد	الله	بن	عمرو	بن	عوف	المزني	عبد الله بن عمرو بن عوف المزني
عبد	الله	بن	عدي	بن	الحمراء	الزهري	عبد الله بن عدي بن الحمراء الزهري
عبد	الله	بن	نافع	بن	نضلة	العدوي	عبد الله بن نافع بن نضلة العدوي

بينما احتوى الجدول (2) على تكرارات الأسماء الأحادية والثنائية والثلاثية والرباعية (بحسب ورودها في النص) مع حساب النسبة المئوية لحضورها باستخدام خوارزمية الأسبقية. وقد وقع اختيارنا على اسم محمد بوصفه نموذجاً للنتائج التي حصلنا عليها بتوظيف الخوارزمية على أسماء الرواة تيمناً باسم الرسول الكريم. وتظهر بيانات الجدول حضور مجموعة منتخبة من هذه الأسماء في ضوء التكرار الكلي داخل حدود الكتاب المنتخب مع النسبة المئوية التي يمكن أن نميزها ضمن حضور أسماء الأعلام التي تبتدئ باسم العلم "محمد" حيث كلما ازدادت النسبة المئوية لحضور الاسم فإنها تؤشر إلى مستوى التمييز من خلال مراجعة مكونات الاسم (الأول، الثاني، الثالث،..... الخ)

الجدول (2) نموذج من تكرارات الأسماء مع النسب المئوية لحضورها في النص.

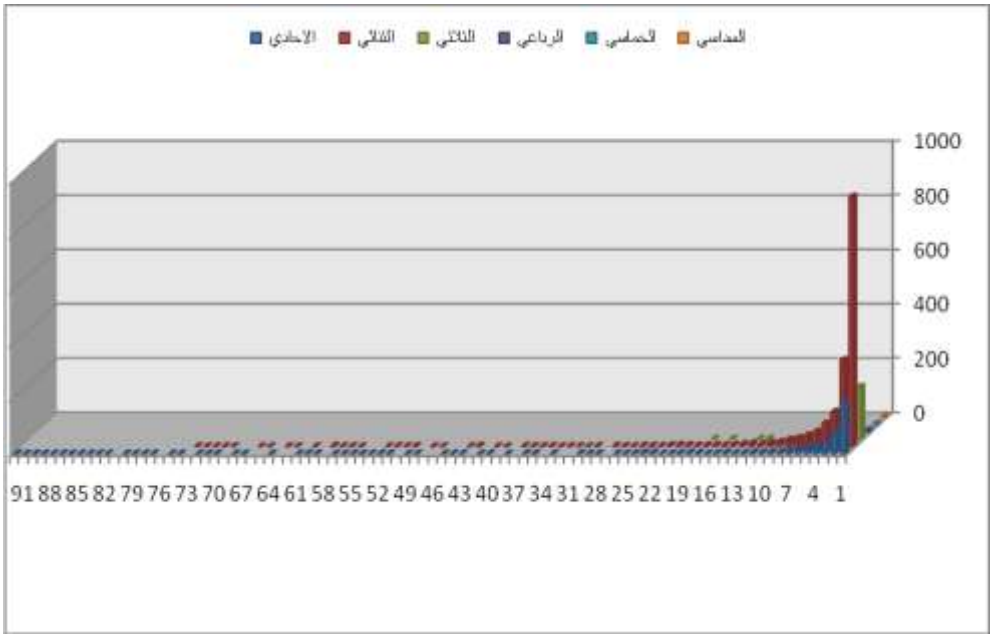
	الاسم	التكرار الكلي	النسبة المئوية
الأحادي	محمد	1091	0.476
الثنائي	محمد بن إبراهيم	18	0.008
	محمد بن أبي	21	0.009
	محمد بن أحمد	36	0.016
	محمد بن إدريس	1	0
	محمد بن إسحاق	64	0.028
	محمد بن أسعد	4	0.002
	محمد بن إسماعيل	3	0.001
	محمد بن اسحاق	1	0
	محمد بن الحسن	3	0.001
	محمد بن الزحاف	1	0
	محمد بن الصلت	17	0.007
	محمد بن الطفيل	1	0
	محمد بن العلاء	16	0.007
	محمد بن الفرج	1	0
	محمد بن الفضل	11	0.005
	محمد بن الفضيل	1	0

	الاسم	التكرار الكلي	النسبة المئوية
الثلاثي	محمد بن إبراهيم بن الحارث	3	0.001
	محمد بن إسماعيل بن أبي	2	0
	محمد بن الزحاف بن أبي	1	0
	محمد بن جبير بن مطعم	6	0.003
	محمد بن جعفر بن الزبير	3	0.001
	محمد بن حمزة بن عمرو	1	0
	محمد بن شعيب بن شابور	2	0.001
	محمد بن عباد بن جعفر	2	0.001
	محمد بن عبد الرحمن بن أبي	5	0.002
	محمد بن عبد الرحمن بن ثوبان	4	0.002
	محمد بن عبد الرحمن بن يزيد	1	0
	محمد بن عبد الله بن أبي	3	0.001
	محمد بن عبد الله بن الحسن	1	0
	محمد بن عبد الله بن الزبير	1	0
	محمد بن عبد الله بن زيد	3	0.001
	محمد بن عبد الله بن مسلم	1	0
	محمد بن عبد الله بن نمير	2	0.001
	محمد بن عبد الله بن نوفل	1	0

	الاسم	التكرار الكلي	النسبة المئوية
	محمد بن القاسم	8	0.003
	محمد بن المبارك	22	0.01
	محمد بن المصفى	2	0.001
	محمد بن المنتش	5	0.002
	محمد بن المنكدر	9	0.004
	محمد بن المنهال	6	0.003
	محمد بن الوليد	1	0
	محمد بن بشار	6	0.003
	محمد بن بشر	4	0.002
	محمد بن بكر	2	0.001
	محمد بن جبير	7	0.003
	محمد بن جعفر	6	0.003
	محمد بن حاتم	3	0.001
	محمد بن زيد	5	0.002
	محمد بن سالم	20	0.009

	الاسم	التكرار الكلي	النسبة المئوية
	محمد بن عكرمة بن عبد	1	0
	محمد بن علي بن الحسين	3	0.001
	محمد بن عمار بن ياسر	1	0
	محمد بن عمرو بن الحسن	2	0.001
	محمد بن عمرو بن حزم	17	0.007
	محمد بن عمرو بن عطاء	3	0.001
	محمد بن كثير بن أبي	4	0.002
	محمد بن كعب بن مالك	1	0
	محمد بن يحيى بن حبان	12	0.005
	الاسم	التكرار الكلي	النسبة المئوية
	محمد بن علي بن الحسين بن علي	3	0.00130776
الرباعي	محمد بن عمرو بن الحسن بن علي	2	0.00087184
	مجموع تكرارت اسم محمد	2294	

ويظهر الشكل (2) تكرار التكرارات للأعلام التي وردت في نصوص الكتاب، والذي يتوافق مع قانون زيف (Ziffs law) الذي ينص على وجود علاقة عكسية مابين حجم الحضور للكلمة داخل النص ومستوى تكرارها.



الشكل(2) حساب تكرار التكرارات لأسماء الأعلام

ولغرض توفير مناخ مناسب يمكن أن يستثمره العاملون في حقل محدد (الحديث النبوي) وللوقوف على كثير من الأسس التي تعتمد في ذلك الحقل المعرفي فقد وجهنا سلسلة عملية التنقيب للمعلومات لتحديد مستويات رواية الحديث وفق معايير أئمة مصطلح الحديث والتي يمكن أن تلقي حول صحة هذه الروايات واتصال إسنادها وموثوقية مواردها لدى التلاميذ عن شيوخهم الذين رووا عنهم لكي تتطابق مع المبادىء المعتمدة وانتقيت أنواع الرواية بعد أن بوبت وفق أنواع الرواية، ومختلف أنماط هذه الروايات في مختلف الأبواب انظر الجدول (3).

الجدول (3) أعداد صيغ الرواية وحسب نوعية الصيغة المستخدمة

نوع الرواية	صيغ الرواية	التكرار
الاخبار	أخبر	6
	أخبرنا	2547
	أخبرني	134
	المجموع	2687
الاعلام	أعلمني	0
الانباء	أنا	117
البلاغ	بلغنا	8
	بلغني	23
	المجموع	31
التحديث	حدثتني	17
	حدثنا	3849
	حدثني	416
	حدثنيه	2
	حدثه	34
	حدثوا	1
	المجموع	4319
الرواية	رواه	2
	روايته	4
	روى	4
	روية	0
	رويت	0
	روبناه	0
	المجموع	10
السماع	سمع	122
	سمعت	399
	المجموع	531
العنعنة	عن	9409
الاقوال	قال	5500
	قالت	260
	قالوا	108
	المجموع	5868

قاعدة الارتباط Association Rule

تعد قواعد الارتباط احد أدوات اكتشاف المعرفة Knowledge discovery of Data mining – KDD لما تتمتع به من قدرة على تصفح كميات هائلة من البيانات، و تسمح (في الوقت ذاته) بالتقاط طيف واسع من العلاقات المحتملة التي يمكن أن تستنبط بناءً على بعض الصفات الموجودة قبالة حضور صفات أخرى [محمد جي ومن معه]. تنقسم الخوارزمية إلى جزئيين الأول يعنى بإيجاد مجموعة العناصر item sets التي تتجاوز عدد مرات حضورها القيم المحددة في قاعدة البيانات وتسمى بالعناصر المكررة. والثانية تعنى بتوليد قاعدة الارتباط من خلال معالجة العناصر المكررة ذات مستوى الثقة المنخفض mini confidence. [سوتيرس، 2006]

دالة الوثوقية Confidence هو مقياس لقوة قواعد الارتباط [بيويك،2008].

An association rule is composed of two item sets:

1. Antecedent or Left-Hand Side (LHS) **2. Consequent or Right-Hand Side (RHS)**

Support :support (LHS RHS) The percentage of transactions that contain all of items.

Confidence : support (LHS RHS) / support(LHS)

In an indication of the strength of the rule P(RHS LHS)

An algorithm for finding all rules with min support and min confidence

Step1: find all frequent item sets with minsup

Step2: generate all rules from step1.Step2:for each frequent item set I with support (I)Divid I into LHS_I (AND RHS_I CONFIDENCE=Support(I)/Support(LHS_I)

واحتوى الجدول رقم (4) على حساب نتائج قاعدة الارتباط لورود الأسماء الثنائية والثلاثية والرباعية في المسند بحساب النسبة المئوي

جدول(4) حساب نتائج قاعدة الارتباط

نوع الارتباط	مستوى الثقة بهوية المحدث من خلال عدد تكرار حضور كل حالة														
	0	10%	20%	30%	40%	50%	60%	70%	80%	90%	1	1.30%	1.70%	3.60%	14.40%
الاسم الأحادي:الاسم الثنائي	733	311	153	93	32	79	19	31	12	10	270		1		
الاسم الأحادي:الاسم الثلاثي	188	32	10	7	2	8	3	1			21				
الاسم الأحادي:الاسم الرباعي	15	31	35	19	7	48	1	8	6	2	102				
الاسم الثنائي:الاسم الثلاثي	5	3		1										1	
الاسم الثنائي:الاسم الرباعي	1				1	1					5	1			1
الاسم الثلاثي:الاسم الرباعي	1					1					8				

(ارتباط الاسم الأحادي : الاسم الثنائي) : يمثل قسمة عدد مرات ورود الاسم الثنائي على الاسم المفرد مثلا: تكرار الاسم الأحادي محمد (1091) وتكرار الاسم الثنائي (محمد بن ابراهيم 18) بقسمة 1091/18 وتقريب الناتج لمرتبة واحدة باستخدام دالة round والحصول على النتيجة و إيجاد عدد مرات تكرار ناتج كل قسمة ولكل الأسماء الواردة في النص

والذي يمثل مجموع نسب الثقة في تمييز حضور الاسم بمفرده (الأحادي) مقارنة مع وروده في أسانيد أخرى بصيغة ثنائية (اسم الراوي+اسم أبيه).

(ارتباط الاسم الأحادي : الاسم الثلاثي) : يمثل قسمة عدد مرات ورود الاسم الثلاثي على الاسم المفرد مثلا تكرار الاسم الأحادي محمد (1091) وتكرار الاسم الثلاثي (محمد بن إبراهيم بن الحارث 3) بقسمة 1091/3 وتقريب الناتج لمرتبة واحدة باستخدام دالة round والحصول على النتيجة و إيجاد عدد مرات تكرار ناتج كل قسمة ولكل الأسماء الواردة في النص والذي يمثل مجموع نسب الثقة في تمييز حضور الاسم بمفرده (الأحادي) مقارنة مع وروده في أسانيد أخرى بصيغة ثلاثية (اسم الراوي+اسم أبيه+اسم جده).

(ارتباط الاسم الثنائي : الاسم الثلاثي) : يمثل قسمة عدد مرات ورود الاسم الثلاثي على الاسم الثنائي مثلا تكرار الاسم الثنائي (محمد بن ابراهيم 18) وتكرار الاسم الثلاثي (محمد بن إبراهيم بن الحارث 3) بقسمة 18/3 وتقريب الناتج لمرتبة واحدة باستخدام دالة round والحصول على النتيجة و إيجاد عدد مرات تكرار ناتج كل قسمة ولكل الأسماء الواردة بالنص والذي يمثل مجموع نسب الثقة في تمييز حضور الاسم الثنائي (اسم الراوي+اسم أبيه) مقارنة مع وروده في أسانيد أخرى بصيغة ثلاثية (اسم الراوي+اسم أبيه+اسم جده)وهكذا بالنسبة للأسماء الرباعية وحساب تكراراتها في النص. أما بالنسبة للنسب المئوية التي وردت في الجدول المذكور فإنها تؤشر إلى نسب التكرار الأصلي في حدود النص الذي تناولته الخوارزمية بمعالجات التنقيب المعلوماتي.

5. الختام

مورست آليات عمليات التنقيب المعلوماتي على نص كتاب الحديث المعروف بسنن الترمذي. وقد وقع الاختيار على انتقاء اسماء الاعلام الاحادية والثنائية والثلاثية الرباعية والخماسية والاسم الكامل مع اللقب بوصفها معياراً يمكن من خلاله المقارنة مع كل نقلة اسماء الرواية والحديث لبقية كتب الحديث.ولزيادة مستوى الموثوقية استخدمت خوارزمية قاعدة الارتباط لالتقاط كل العلاقات المحتملة التي تشرح بعض الصفات الموجودة اعتمادا على وجود الصفات الأخرى ولقياس الموثوقية من ورود الاسماء عولجت نتائج التنقيب المعلوماتي بواسطة دالة الوثوقية.ظهر لنا من خلال هذا الجهد المتواضع وجود أكثر من إمكانية لتمييز عناصر الحصيلة المعرفية لدى رواة الحديث مع إمكانية تطوير هذا العمل على أساس جهد مؤسسي يمكن من خلاله تناول طيف واسع من كتب الحديث والرواية لبلوغ قاعدة معلوماتية شاملة لهذا العلم الجديد بواسطة أدوات الحوسبة الذكية التي شاع استخدامها في هذه الأيام.

6. المصادر والمراجع

[الرزو،2006]

الرزو، حسن مظفر، مقارنة محوسبة بين نهج برنارد لويس ودانيال بايبس في معالجة مسائل إستشراقية معاصرة، مقبول للنشر في مجلة إسلامية المعرفة، المعهد العالمي للفكر الإسلامي، ماليزيا، 2006.

[الرزو،2008]

الرزو، حسن مظفر، التنقيب المعلوماتي على موارد الشريعة الإسلامية، مجلة أحوال المعرفة، المجلد الثالث عشر، العدد الواحد والخمسون، ابريل 2008، الصفحات 38–43.

[اندرياس، 2005]

Andreas , H. ,"**A Brief Survey of Text Mining**" , KDE Group , Kassel,hotho@cs.uni. 2005.

[بيويك،2002]

Baoeic , A. ," Data-**Mining**",http://www.authorstream.com/Presentation/Dionigi-48230-data-mining-Introduction-Knowledge-Discovery-KDD-Overview-Counting-Co-.Occurrences-as-Education-ppt-powerpoint, (2002)

[ديين، 2005]

Dean W., " **An Evaluation of High-end Data Mining Tools for Fraud Detection**", Abbott , Elder Research , San Diego. CA 92122, (**2005**).

[فيلدمان ومن معه، 1998]

Feldman,R., M. Fresko, H. Hirsh, Y. Aumann, O.Liphstat, Y. Schler,& M. Rajman ,"**Knowledge Management: A Text Mining Approach**" , Proc. of the 2nd Int. Conf. on Practical Aspects of Knowledge Management (PAKM98) Basel, Switzerland, 29-30 Oct..(1998).

[جيوي ومن معه، 2006]

Jiawei H. & Micheline K., "**Data mining : Concepts and techniques second edition**" , University of Illinois at a , Urbana_Champaign @ 2006 by Elsevier Inc. (2006)

[جينيوك ومن معه، 2000]

Jinuk B. & Sukho L. , "**Partitioning Algorithms for Computing of Averge ICEBERG Queries**" , seoul national university , korea ,© Springer-Verlag berlin Heidelberg. (2000).

[لانا، 2009]

Ian H. Witten, "**Text mining**" , Computer Science, University of Waikato, Hamilton, New .Zealand , email ihw@cs.waikato.ac.n,.(2009)

[ماركوس، 2005]

Markus H ,"**The A priori Algorithm _ atutorial**" , CMA Australian National university , .Markus.Hegland@anu.edu.au,. (2005)

[ميني ومن معه، 1998]

Z Min F. , narayanan S. , Hector G. , Rajeev.M , Jeffrey D. ulman ,"**Computing ICEBERG Queries Efficiently** " , Proceedings of the 24th VLDB Conference New York, USA..(1998).

[محمد جي ومن معه]

Mohammed J., Zaki & Ching-Jui H.," **CHARM: An Efficient Algorithm for Closed Association Rule Mining** ",http://www.cs.rpi.edu/_zaki.

[سانجييف ومن معه، 2010]

Sanjeev R. , Priyanka G," **Implementing Improved Algorithms over A Priori data mining Association rule algorithm**",International Journal of computer Scince and technology , JCST VOL.3.(2010)

[سوتيرس، 2006]

Sotiris K. , Dimitris K., "**Association Rules Mining: A Recent Overview**", Educational Software Development Laboratory. International Transactions on Computer Science and Engineering, Vol.32 (1), pp. 71-82, (2006).

[جفينسنت، 2005]

Vincent , K ," **Text Mining Methods for Event Recognition in Stories**", Tech Report MI-05-2, Knowledge Media Institute KMI, UK, April.(2005).

[اكسندونك ومن معه، 2007]

Xindong W. , Vipin K. , J. Ross Quinlan , Joydeep G. , Qiang Y. , Hiroshi M. , Geoffrey J. McLachlan , Angus N. , Bing Liu , Philip S. Yu , Zhi-Hua Z. , Michael S. , David J. Hand , Dan S," **Top 10 algorithms in data mining**", Published online: 4 December 2007.

7. جدول الالفاظ

اللغة العربية	اللغة الانجليزية
تنقيب البيانات	Data mining
اكتشاف المعرفة	Knowledge discovery of Data mining
تنقيب النصوص	Text Mining
خوارزمية	Algorithm
خوارزمية الاسبقية	Priori Algorithms
معاملات	Transaction
استفسارات الجبل الثلجي	Iceberge Queries
الصفات	Attribute
قاعدة الارتباط	Association Rule
الموثوقية	Confidence
تنقية البيانات	Data Clearing
نقل البيانات	Data transformation
تقييم النموذج	Pattern Evaluation
تمثيل المعرفة	Knowledge Representation

8. الخلاصة باللغة الانجليزية

Usage of Priori Algorithm in Mining the Names of AL_Muhadtheen

Alaa Faisal Al_Mukhtar[1], Hassan Muzaffar Al_reso[2]
[1]General of Electricity Distribution Nineveh province, Governorate , IRAQ.
[2]Ninevah Governorate , IRAQ.
[1]afs712000@yahoo.com,[2]halrizzo@gmail.com

Abstract: Searching for masters narrators of prophetic tradition is considered one of the most significant issues which grant the Islamic researcher in the prophetic tradition field an opportunity to explore al hadith's narrators and the ways of narration's transcription to make sure of the exactness of narrations that prophetic Sunna and traditions books flow with. In spite of the great efforts made by Imams of contest and amendment in classifying the major books that deal with al hadith's men but we became among a big challenge in tackling with these texts mechanically and away from the printed text demanding a deep knowledge and hard work to follow al hadith narrators among ascription series. In order to achieve this objective our research aimed to apply the priori algorithm in al-Darmi tradition's text as a primary stage, so that it would be possible after its success the work to expand its applying circle to include the elite of prophetic tradition books. As a result of using data research process by applying the priority algorithm we have got the names of narrators classified in many levels (according to the name mentioned in the ascription series: single name, dual name, triple name, quadric name, penta name and the full name with surname). With a view to strengthen our satisfaction in the correctness of the research process's results in defining the identity of the narrator, confidence-function was used, which can be regarded as important step to present the scientific subject to criticize the narration manner.

Keywords: Data Mining, Text mining, Association Rule Mining Algorithms, Apriori Algorithm , confidence.