تصنيف الكتب الإنكليزية باستخدام قواعد الربط
(Association Rules)

م.م. ايناس جاسم هادي الشريفي
كلية الآداب / الجامعة المستنصرية
Enas_tech83@yahoo.com

التقديم: ٥٨ في ٢٠١٧/٣/٦
القبول: ١٢١ في ٢٠١٧/٣/٢٧

الملخص:

يكتسب المكتبات عناية خاصة من المؤسسات التعليمية المتنوعة، لما تشكله من محور أساس في اثراء الجوانب العلمية بمختلف الموارد المعرفية. ويبرز دور المكتبات في سرعة توصيل المعلومة المناسبة والسريعة للمتلقي. وعليه فإنَّ فهرسة وتصنيف الكتب والموارد المكتبية يعدُّ مسألة غاية في الأهمية لما تؤديه من دور في تحسين أداء المكتبات. إنَّ عملية تصنيف الكتب ووضعها ضمن الفئة الصحيحة يعدُّ من العمليات الحرجة ولا سيما في حالة تصنيف الكتب الإنكليزية. يقدم هذا البحث أنموذجاً لتصنيف الكتب الإنكليزية الكترونياً باستخدام قواعد الربط (Association Rules). كمرحلة أولى من هذا البحث تم اجراء استبانة للعاملين في المكتبات، وذلك لغرض، الوقوف على التحديات والمصاعب المصاحبة لعملية تصنيف الكتب، وبينت نتائج الاستبانة انه هنالك عدداً من المعوقات تصاحب تصنيف الكتب الإنكليزية. بعدها تم تكوين قاعدة بيانات لعدد من الكتب، وذلك، لتنفيذ النموذج عليها. تم تنفيذ النموذج على قاعدة بيانات الكتب المذكورة. ويتم وفق النموذج المقترح إعطاء عنوان الكتاب فقط ومن ثم تقييم مدى قدرة النموذج على وضع الكتاب ضمن الفئة الصحيحة من خلال عنوان الكتاب المعطى. بعد تنفيذ النموذج تبيَّن أنَّ الخوارزمية قادرة على تصنيف الكتب بنحوٍ صحيح بنسبة (٧٦%)، أي بمعدل خطأ، يبلغ (٢٤%).

إنَّ نتائج هذا البحث تفتح المجال امام الدراسات لتنفيذ واختبار طرائق وأساليب برمجية أخرى من شأنها تحسين الأداء، وتقديم نموذج تصنيف الكتب الكترونياً بشكل كفوء، يخدم العاملين في المكتبات والمؤسسات التعليمية على حدٍّ سواء.

الكلمات المفتاحية: تصنيف الكتب، قواعد الربط، ديوي العشري

Classifiying English Books By Using Association Rules

Asist .instructor. Enas Jasim Hadi
Almustansiriyah University- College of Arts
Enas_tech83@yahoo.com

Abstract:

In general, libraries have taken a special interest from various educational institutions due to the key role that libraries play in terms of enriching the scientific aspects with different knowledge resources. The role of libraries is significantly can be seen in the fast information delivery for the recipients. Hence, the indexing and classifying of library resources is very important task because it contributes in enhancing the libraries performance. The classification of books (putting the book in the right class) is not an easy task, especially when the books are written in English. This research presents a model for books classification by using Association Rules techniques. The first stage of the research involved distributing a questionnaire for people who are working in libraries. The purpose of the questionnaire was to identify the difficulties that come with English books classification. The results of the questionnaire proved that there were number of obstacles in terms of English books classification process. After that we have collected an English books dataset to be employed in our model. Then the model has been implemented successfully and tested by given the title of the book. The results of the research proved that the model can put the book into the right class with ratio of (76%) and the error rate was (24%). That means it is possible to develop computer-based systems to classify the library resources. The results of the research open the way toward developing more sophisticated techniques and methods which can improve the performance of classifying the books electronically in an efficient manner that serves the librarians and users as well.

Keywords: Books Classification, Association Rules, Dewey Classification.

الفصل الأول/ الإطار العام للبحث:

١.١ مشكلة البحث:

لا يخفى على أحد الاهمية التي تشكلها الكتب في اثراء حياتنا العلمية والاجتماعية بالمعلومات. حيث كانت ومازالت الكتب تشكل العصب الأساس في الحصول على المعلومات. ونتيجة لذلك يتزايد الاهتمام بعلم المكتبات وبتصنيف الكتب. وتشكل عملية تصنيف الكتب احدى اهم محاور علم المكتبات، لما تؤديه من دور مهم في تسهيل عملية الوصول الى الكتاب المطلوب بأقل وقت وجهد. ونتيجة لذلك، ولتزايد اعداد الكتب وتنوعها، تتزايد التحديات المصاحبة لعملية تصنيف الكتب. ومن خلال الزيارات الميدانية التي اجراها الباحث لعدد من المكتبات تم ملاحظة أنَّ عملية تصنيف الكتب الانكليزية تتطلب من العاملين في مجال المكتبة مهارات معينة، لغرض تصنيف كتاب معين ووضعه ضمن الفئة المناسبة. وللوقوف على حيثيات الموضوع، وتحديات تصنيف الكتب الإنكليزية قام الباحث بتصميم وتوزيع استبانة على العاملين في المكتبة لبيان واقع حال تصنيف الكتب الإنكليزية (ملحق١)، حيث تم صياغة عدد من الأسئلة التي تبيّن فيما إذا كانت هنالك مصاعب تواجه المكتبي في مجال تصنيف الكتب الإنكليزية. وبعد تحليل نتائج الاستبيان تبيّن ما يأتي:

- يتطلب من العاملين في مجال تصنيف الكتب الإنكليزية معرفة جيدة باللغة الإنكليزية لغرض تصنيف الكتب وهذا الامر هو الذي يتطلب تواجد كادر متخصص باللغة الإنكليزية للإسهام بعملية التصنيف.
- مقارنةً مع الكتب العربية تعد عملية تصنيف الكتب الإنكليزية معقدة نوعا ما.
- افتقار المكتبات الى أنظمة معلومات محوسبة تقوم بعملية تصنيف الكتب الإنكليزية.

وتطرح مشكلة البحث التساؤل الآتي: هل بالإمكان انشاء نموذج معلومات محوسب قادر على تصنيف الكتب الإنكليزية، والمساهمة في تسهل عمل المكتبي؟

٢.١ أهداف البحث:

إنَّ الهدف الرئيس من هذا البحث هو تقديم نموذج لنظام معلومات لتصنيف الكتب الأجنبية الكترونيا بالاعتماد على قواعد الربط (Association Rules). فضلاً عن الهدف الرئيسي هنالك عدد من الأهداف الفرعية التي تسهم في تحقيق الهدف الرئيس وكما يأتي:

- تحديد العلاقات التي تربط بيانات الكتب الأجنبية.
- تطبيق قواعد الربط لتصنيف الكتب الأجنبية.
- قياس أداء النموذج المقدم وتقييم، النتائج المستخلصة.

٣.١ أهمية البحث:

يهتم اخصائيو المعلومات والمكتبات بتنظيم وخزن واسترجاع وحفظ المعلومات وآلية التعامل معها من خلال تسهيل عمل مؤسسات المعلومات والمكتبات والمراكز البحثية. وتتجلى أهمية هذا البحث في أنه يقدم خوارزمية قادرة على تصنيف الكتب الإنكليزية والمصنفات البحثية المبنية على نظام تصنيف ديوي العشريّ. وتعدُّ هذه الخوارزمية بوابة أمام تطوير خوارزميات أكثر تطوراً، مبنية على أساليب الذكاء الاصطناعيّ، ومن الممكن تطويرها بالاعتماد على:

- المستخلصات (Abstracts).
- قائمة رؤوس الموضوعات (Subjects Headings List).
- الكلمات المفتاحية الدالة (Keywords).

٤.١ المنهجية:

في هذا البحث تم اعتماد المنهج التجريبيّ التحليليّ. حيث تم جمع بيانات لعدد من الكتب الأجنبية، وتم ترتيبها، وتجميعها على وفق نظام ديوي العشريّ. بعد ذلك تم تطبيق أنظمة قواعد الربط لغرض تطوير نموذج تصنيف البيانات، اعتماداً على نظام ديوي العشريّ. وأخيرا تم تحليل أداء النظام، وتقييم النتائج للوقوف على اهم نقاط النموذج.

٥.١ الدراسات السابقة:

١.طريقة تصنيف الكتب بالاعتماد على نظرية الرياضيات الضبابية (Ping & Tai-shan, 2011).

ركزت هذه الدراسة في تقديم طريقة جديدة لتصنيف الكتب الكترونياً، وذلك، لتسهيل عملية تصنيف الكتب وتسريعها، فضلا عن تقديم الدقة والكفاية المناسبتين. تم اعتماد طريقة الرياضيات الضبابية (Fuzzy Mathematics) لتصنيف الكتب، وهي احدى طرائق الذكاء الاصطناعيّ. اعتمدت الطريقة على أنَّ اليات الذكاء الاصطناعي تسهم الى حدٍّ كبير في تسهيل المهام، وتحسين الأداء. اثبتت نتائج الدراسة أنه يمكن لهذه الطريقة من تصنيف الكتب الكترونياً، ممّا يشجع بناء أنظمة تعتمد على هذه الطرائق.

٢.نظام مكتبي ذكي متعدد الأطراف لتصنيف الكتب (Pengfei, Liangxian, & Junxia, 2012).

تم في هذه الدراسة تقديم نظام لتصنيف الكتب اعتمادا على مبادئ الذكاء الاصطناعي، وأنَّ الهدف من النظام المقترح هو تصميم نظام يشبه الإنسان الآليّ لإدارة المكتبات. تم تطبيق النظام على الكتب التابعة للمكتبة الصينية العامة. كما أنَّ الخوارزمية التي تم تصميمها كانت قادرة على إيجاد

موقع الكتاب بالاعتماد على بطاقة شريط الرموز (Barcode) التابع للكتاب. اثبتت نتائج الدراسة أنَّ الخوارزمية المقدمة تسهم بنحوٍ كبير في تقليل الوقت والجهد المطلوبين، لغرض تصنيف الكتب.

٣.تصنيف موارد المكتبة باستخدام منهج تعليم (PU) (Shirude & Kolhe, 2016).

تم بموجب هذه الدراسة تقديم مقترح نظام تصنيف محتويات المكتبات الكترونيا. كما وتم اعتماد آليات الذكاء الاصطناعيّ، لغرض القيام بعملية التصنيف. المهمة الأساسية للنظام المقترح هي تصفية الموارد وتقسيمها على فئات محددة. تضمنت نتائج الدراسة أنَّ النظام المقترح قام بتحسين عملية التصنيف لموارد المكتبة، وذلك، من خلال تقييمه من العاملين في المكتبات. كما وتم تقديم عدد من المقارنات للطرائق الإحصائية المتبعة في عملية التصنيف.

الفصل الثاني/ الجانب النظري:

١.٢ مقدمة تاريخية عن التصنيف:

قديما قال العرب (أن اخبرتك أنَّ اول العلم بالمعرفة هو تصنيفها) إلا أنَّ تاريخ التصنيف يبدأ مع ترتيب الافكار للبشر، وبدأت هذه المحاولات منذ ان بدأ الانسان يميز الكون، ويجد له تفسيرات، واخذ يكّون طريقة او نظام، ومثل هذا التصنيف يفيد الفلاسفة والمنطقيين، لأنها تقسيم منطقي للمعرفة، ولقد كان هناك كثير من الانظمة والافكار لتنظيم المعرفة البشرية وفرزها وترتيبها، وكان هناك حوالي (١٦١) نظاما، وتبدأ من فلسفة افلاطون بكتابه الجمهورية (٤٢٨ ـ ٤٧ ق.م)، ولقد قام افلاطون بأولى محاولاته لتصنيف موضوعات المعرفة، وقسمها على قسمين: الاول العالم المحسوس وهو الطبيعيات، والآخر عالم المعقول وهو الرياضيات، وبعدها جاءت محاولات ارسطو (٢٤٠– ٣١٠) ق.م، وجاءت محاولات اخرى، مثل الاشوريين واليونان والمحاولات العربية الاسلامية، ولكن كلّ هذه الانظمة والمحاولات تمثل تنظيماً وتطبيقاً للمعرفة، وتجميعاً للأفكار، ولكن تطبيقها في المكتبات لترتيب الكتب كان مستحيلاً. ومن هنا بدأت محاولات المكتبات بتطبيق انظمة مختلفة. أي أنَّ فكرة التصنيف موجودة منذ القدم، ولكن الترتيب كان كيفما اتفق.

(Sayers, 1944)

التصنيف كما عرفته مارجريت مان (Mann, 1949): (التصنيف في ابسط بيان له هو وضع الاشياء المتشابهة معاً وهو ترتيب الاشياء وفق التشابه والاختلاف ويعتبر تصنيف الكتب تصنيفا للمعرفة مع التعديلات اللازمة التي يمليها الشكل المادي للكتاب)

وعرفه ريتشاردسون (Richardson, 1912): (وضع الاشياء المتشابه معاً)، ويعرف قاموس مصطلحات جمعية المكتبات (American library association) (Sheehy, Keckeissen, McIlvaine, & Winchell, 1976) (ان التصنيف عبارة عن سلسلة او مجموعة من الاقسام مرتبة بنظام معين طبقا لمبدأ معين او غرض محدد).

والتصنيف من وجهه نظر الباحث هو تصنيف المصنفات المعرفية وترتيبها بصورة تسهل عمل المكتبي والقارئ في الوقت نفسه بالاعتماد على أحد انظمة التصنيف مثل تصنيف ديوي العشري.

٢.٢ فوائد التصنيف في المكتبات. (الحديدي، ١٩٦٩):

١. يساعد القراء على الوصول الى ما يريدونه من مواد بسرعة وسهولة عن الموضوع المراد.

٢. يضع حدودا واضحة لمختلف اصول المعرفة وفروعها ومن ثمَّ يمنع اختلاط وتداخل مواد المكتبة بعضها بعضٍ.

٣. يعدُّ الاساس في تصنيف المجموعات المكتبية

٤. يوفر الوقت عند استخراج الكتب وإعادتها إلى مكانها بدقة.

٥. كثير من القراء يبحثون عن حاجاتهم دون مساعدة المكتبي.

٦. اختيار الوثائق المبني على التصنيف يضمن التقنين والتوازن بين المجموعات.

٧. قد يكون التصنيف أساساً لتنظيم التعاون بين المكتبات في الترتيب الموضوعيّ المتخصص في ميدان التعاون الدولي، وهنا يؤدي التصنيف دور اللغة العالمية.

٣.٢ معيار التصنيف الجيد (فهمي، ١٩٨٤):

يوجد هناك عدة معايير ينبغي توافرها في نظام التصنيف الجيد، ومنها:

١. الشمول للموضوعات (Subjects Comprehension): ينبغي أن يكون نظام التصنيف الجيد متصفاً بالشمول والاتساع، وقابلاً للتوسع المستقبليّ، والجداول المكانية والمدد الزمنية، والسلالات العرقية، أو اللغوية والادبية.

٢. الترتيب المنطقيّ للمجموعات المكتبية: ينبغي أن تكون انظمة التصنيف المكتبية الناجحة أن ترتب الموضوعات وفقا لدرجة تجاورها، مثلاً ترتب كتب الدين مجاورة لكتب الفلسفة.

٣. ترتيب الكتب والاقسام حسب الاسس المنطقية: ان الترتيب المنطقي للرتب والاقسام ووحدة الموضوع وتكاملها، وعدم تشتيته لأغراض اضطرارية، تعطي نموذجاً جيداً لنظام التصنيف الجيد. على سبيل المثال في الطبعة الرابعة عشر من نظام ديوي العشريّ عولج موضوع النسبية في ثلاثة اماكن:

- مع علم الفلك.
- مع الجاذبية.
- مع علم النفس.

وإذا طبقنا خاصية التسلسل المنطقي لفرضنا تواجد كل ما كتب عن النسبية في مكان واحد

٤. كفاية المصطلحات المستخدمة في النظام: يشترط في نظام التصنيف الجيد أن تكون المصطلحات محكمة ومؤدية للغرض الذي اعدت من أجله، ويشترط عدم توافر شك حول معنى

المصطلح المستخدم في الجداول أو الكشاف حتى يكون مضبوطاً، ففي الطبعة الرابعة عشر من تصنيف ديوي العشريّ، يوجد أكثر من مكان لتصنيف كلمة (Dogs) الكلاب باللغة الإنكليزية:

أولاً: صنفت أولاً تحت الرقم (٦٣٦.٧) (Domestic Animals) وهنا يتم معاملة الكلاب كحيوانات منزلية.

ثانياً: وصنفت تحت الرقم (٦٢١.٨٣٧) يتعلق بمصطلح مختلف موضوعياً ومتشابهاً من حيث الحروف الهجائية، وحيث يقصد هنا السقاطة واللسان الخاص بالتروس. وفي الوقت نفسه لا نجد في الكشاف رقم تصنيف يعبر عن مصطلح استخدام نماذج للكلاب على شكل دمى أطفال.

٥.مرونة الرمز (The symbol flexibility): إنَّ مرونة الرمز تضيف بعداً جديداً للاستخدام العمليّ للتصنيف، فإنَّ لطول رقم التصنيف اهمية كبيرة، فإنَّه لقصره مزية في توضيح كلّ الفروق الضرورية في التجميع فضلاً عن الى سهولة التذكر. ويجدر بنا أن نعرف أنَّ قصر الرمز يزيد من فرص تعميمه وتذكره، ويظهر هذا بوضوح في التصنيف العشريّ، إذ يتكون من ارتباطات بين عشرة ارقام من صفر إلى تسعة بحد أدني ثلاثة ارقام.

٦.النجاح العملي للكشاف: إنَّ أية خطة رتبت بها موضوعات المعرفة البشرية ترتيباً مصنفاً بحاجة الى كشاف بسبب أنَّه كلما زاد عدد الموضوعات في الكشاف زادت العلاقات الموجودة بين هذه الموضوعات ومن ثمَّ زادت فائدة الكشاف ويستخدم لغرض كشف المكان الصحيح للمصطلح في الخطة عند تصنيف الكتب.

٧.تقوم انظمة التصنيف بترتيب الكتب على الرفوف، بحيث يعطي للكتاب مكاناً محدداً نسبياً بمعنى أنَّ كلّ كتاب على الرف يكون دائماً في مكانه بالنسبة للكتب المتواجدة إلى جانبيه، وذلك على الرغم من مكان الكتاب على الرف قد يتغير بإضافة كتب جديدة، أو سحب كتب، أو تحريكها من مكان لآخر على الرف.

٨.يعدُّ الرمز الفائدة الثانية لنظام التصنيف، وهي سهولة ارجاع الكتاب الى مكانه على الرفوف المكتبية، ولا يمكن ان يتم هذا الا بوصل الرمز الذي يوضع على كعب الكتاب، او في بطاقات الفهارس، والذي من الممكن عدَّه اختزالا لخطة التصنيف.

مثال (اوديت بدران، ١٩٧٦)

جاذبية الاجسام الصلبة تأخذ رقم التصنيف (531.5) يتم تحليل هذا الرقم كالاتي

٥٠٠ العلوم البحتة.

٥٣٠ الفيزياء.

٥٣١ الميكانيكا.

531.1 جاذبية الاجسام الصلبة.

والاساس في هذه التقييمات واحد، حيث يمكن القول إنَّ أفضل رمز هو الرمز الذي يعطي أفضل علاقة تذكر.

٩. ارقام الكتب: تستخدم ارقام الكتب لتمييز الكتب او المطبوعات التي تحمل ارقام التصنيف نفسها.

٤.٢ نظم التصنيف المستخدمة في المكتبات حسب التسلسل الزمني لإصدارها (اتيم، ١٩٨٧):

١. نظام التصنيف العشريّ ـ وضعه ملفل ديوي عام ١٨٧٦.

٢. نظام التصنيف التوسعيّ ـ وضعه شارل كتر عام ١٨٩١.

٣. نظام مكتبة الكونغرس ـ وضع عام ١٩٠٢.

٤. نظام التصنيف الموضوعيّ ـ وضعه جيمس براون عام ١٩٠٦.

٥. نظام التصنيف التوضيحيّ ـ وضعه رانجاناثان عام ١٩٣٣.

٦. نظام التصنيف البيبلوغرافيّ ـ وضعه هنري بليس عام ١٩٣٥.

٧. نظام التصنيف العشريّ العالميّ ـ وضعه إعتماداً على تقسيمات ديوي للمعرفة هنري لافونتين وبول اوتليت باللغة الفرنسية عام ١٩٠٥ لمصلحة المعهد الدولي للبيبلوغرافيا (والذي يدعى الآن الاتحاد الدولي للتوثيق) (FID).

فضلاً عن نظم التصنيف الرئيسة التي وضعت لتصنيف المكتبات تصنّف الكتب والوثائق وفقا لطرق اخرى في التصنيف:

١. التصنيف العرضيّ (Accidental Classification) وفي نظام التصنيف العرضيّ تصنف الكتب والوثائق، أو المقالات، وفقاً لحادثة معينة، أو زمن معين، أو بلد معين، أو موضوع خاص.

٢. التصنيف الاصطناعي (Artificial Classification): يعد هذا التصنيف اساسا للترتيب وقد يكون هجائياً، أو لغوياً، أو رقمياً، أو زمنياً، أو حسب الشكل، أو الحجم، أو وفق خاصيات اخرى يحددها القائم بعملية التصنيف امثلة:

- ترتيب للغة داخل الوثيقة الواحدة.
- الرقمي للأسطوانات والشرائح والافلام والمواد التي لا يجوز العبث بها يدويا.
- احتمال التلف (المخطوطات القديمة او اوراق البردي)
- التصنيف حسب اهتمامات القراء ان الوثائق والكتب تكون مفيدة بسبب المواضيع التي تتضمنها تلك الوثائق الا في الآداب يكون الشكل اهم من موضوعه. (محمود الاخرس، ١٩٦٥)

على الرغم من كثرة نظم التصنيف فإنَّه اغلب المكتبات، ومن ضمنها المكتبات العراقية تميل لاستخدام تصنيف ديوي العشريّ، وذلك لعدة أسباب: ومنها

١. يمثل المواد المكتبية في عشرة اقسام من (٠٠٠– ٩٩٩)، ويمكن تفريع كل منها الى فروع عشرية الى ما لا نهاية.
٢. يتمتع النظام بمرونة الى الحد الذي توسع فيه الارقام بأسلوب خطي، لتغطي المظاهر الخاصة.
٣. يمتاز بالسهولة، حيث يمكن للمصنف أن يقوم بعملية التصنيف بعد فهمه للنظام.
٤. سرعة إدراك القارئ للتسلسل الموضوعيّ فيه.
٥. يصبح مألوفا لدى القارئ نتيجة لتطبيقه في المكتبات المدرسية والمكتبات العامة.
٦. يتبع النظام مبدا التدرج من العام الى الخاص، كما أنه مبني على مفاهيم فلسفية.

قسم ديوي المعرفة البشرية الى ثلاثة عشرة قسماً تبعا لتسلسل حياة الانسان:

١. بدأ الانسان في التفكير بذاته وماهية وجوده على الارض فكانت الفلسفة.
٢. قاده تفكيره الى ان وجوده على الارض راجع الى الخالق، فكانت الديانات.
٣. بسبب طبيعية الانسان الاجتماعية لابد له من العيش كجماعات من خلال الاسرة او العشيرة القبيلة، فكانت العلوم الاجتماعية.
٤. ادت الحياة الاجتماعية الى ضرورة الاتصال والتواصل، فكانت اللغة.
٥. اخذ الانسان في ملاحظة الحياة من حوله وبدأ التفكير في أسرار هذه الحياة، فكانت العلوم النظرية.
٦. لبعض جوانب النظرية السابقة تطبيقات تخدم الانسان، فكانت العلوم التطبيقية.
٧. كان الانسان مضطرا للبقاء في سكنه عندما لم تسنح له الظروف للصيد، أو الزراعة، أو الرعي، فكان يقضي اوقات فراغه في الرعي والرسم والنقش على جدران الكهوف فكانت الفنون الجميلة.
٨. تطورت الفنون الجميلة، بحيث أصبح الانسان يعبّر عن أحاسيسه باللغة فكانت، الآداب.
٩. تقدمت حياة الجماعة وظهرت في اماكن اخرى متباعدة، فكانت الحاجة الى الرحلات، ومعرفة اساليب الحياة لدى كل منها فكان التاريخ.
١٠. هناك حالات تجمع بين بعض، أو كل الميادين السابقة، فكانت المعارف العامة.

ولكن ملفل ديوي استخدم الاصول الرقمية (٠– ٩) ليدل على تلك الاقسام واشترط ان لا يقل اصل كل رقم عن ثلاثة أعداد مما جعل الاصول (٠٠٠– ٩٩٩)، وعليه أصبحت الاقسام العلمية الرئيسة العشرة كالآتي، بعد أن قرر أن تبدأ بالمعارف العامة (Chan, 1996)

- ٠٠٠–٠٩٩ المعارف العامة.
- ١٠٠–١٩٩ الفلسفة.
- ٢٠٠–٢٩٩ الديانات.

- ٣٠٠-٣٩٩ العلوم الاجتماعية.
- ٤٠٠-٤٩٩ اللغات.
- ٥٠٠-٥٩٩ العلوم النظرية (البحتة).
- ٦٠٠-٦٩٩ العلوم التطبيقية (التكنولوجيا).
- ٧٠٠-٧٩٩ الفنون الجميلة.
- ٨٠٠-٨٩٩ الآداب.
- ٩٠٠-٩٩٩ التاريخ والجغرافيا.

٥.٢ مشكلة تصنيف الكتب الاجنبية في المكتبات العراقية:

بعد الدراسة الميدانية التي قام بها الباحث، ودراسة واقع الحال في قسم فهرسة وتصنيف الكتب الانكليزية لاحظ الباحث ان اغلب الكتب والمصنفات البحثية التي تم تصنيفها تمت من خلال الاستعانة بأحد خريجي اقسام اللغة الانكليزية، بسبب عدم استطاعة المكتبيين على انشاء رقم تصنيف معتمد وصحيح للمطبوعات الانكليزية هذا من جانب، أما من جانب آخر فيكون خريج اقسام اللغة الانكليزية والترجمة على غير علم ودراية بمهنة فهرسة وتصنيف الكتب، فقبل أن يقوم المكتبي بعملية التصنيف عليه أن يكون ملما بنظام التصنيف المعتمد في المكتبة، وهو نظام تصنيف ديوي العشري واقسامه وتفريعاته الرئيسة والمساعدة والكشافات:

وعليه، على المصنف غير المكتبي الالمام بما يأتي ليكون قادراً على أداء عملية التصنيف:

١. ملماً بنظام تصنيف ديوي العشري.

٢. أن يدرس الخلاصات الثلاث.

٣. أن يقوم بحفظ الخلاصة الاولى المتضمنة عشرة اقسام.

٤. أن يعرف ماهية الجداول الاصلية.

٥. أن يكون ملما بالجداول المساعدة (٧) والكشافات.

٦. أن ينتبه على التسلسل الهرمي ودرجات العلاقة القائمة بين فروع المواضيع.

٧. على المصنف ان يكون واعياً الى المعارف العامة والآداب (٠٠٠) والآداب (٨٠٠) حيث إنَّ عملية التفريع تتم على اساس الشكل وليس الموضوع.

٨. وبعد أن يقوم المصنف بفهم الخطوات السابقة، يكون قد حدد موضوع الوثيقة، ونظام المعرفة الذي عولج الموضوع من خلاله، أي يكون مستعدا لتصنيف الوثيقة (حدد مكانها في النظام).

٩. على المصنف ملاحظة أن هناك وثائق تغطي موضوعين أو أكثر، أو قد تكون الوثائق تعالج موضوعين مترابطين أو مستقلين.

٦.٢ قواعد الربط (Association Rules Mining (ARM)

تعد قواعد الربط (ARM) من اهم تقنيات التنقيب عن البيانات لما تؤديه من دور مهم وفعال في إيجاد علاقات منطقية بين مكونات البيانات. إنَّ طبيعة عمل قواعد الربط تتلخص في إيجاد علاقات خفية وارتباطات تظهر بنحوٍ متكرر بين عناصر البيانات المختلفة ولا سيما ضمن قواعد البيانات الكبيرة (Big Database). تتبع قواعد الربط مبدأ (إذا كان – فإنه) حيث جملة اذا تمثل العنصر الشرطي، في حين جملة (فأنه) تعني النتيجة، ويمثل الشرط والنتيجة مجموعة العناصر التي تتم معالجتها باستخدام قواعد الربط (Han, Pei, & Kamber, 2011). وبتعبير اخر، حالما يتم إيجاد علاقة معينة يمكن بسهولة تتبع مجموعة العلاقات التي تمثل البيانات (Krishna Kumar, Amrita, & Priya, 2013).

تعود بداية استخدام قواعد الربط الى العام ١٩٩٣ من قبل (Agrawal, Imieliński, & Swami, 1993)، حيث تم بموجب هذه الدراسة تم تقديم نموذج علائقي وهو (تحليل سلة السوق) (Market basket analysis)، وهو من اشهر الأمثلة شائعة الاستخدام في تمثيل البيانات المتكررة وإيجاد العلاقات فيما بينها وتحليلها، حيث يتمثل بتحليل عمليات البيع والشراء في الأسواق، ومن ثمَّ إيجاد علاقات بين العناصر الداخلة في العملية التسويقية. إنَّ إيجاد هذه العلاقات وتعريفها يسهم بشكل كبير في تطوير استراتيجيات تسويقية جديدة من خلال تحديد العناصر الأكثر تكرارا، ومن ثمَّ زيادة الإنتاجية. فضلاً عن المثال المذكور آنفاً حول علاقات الربط، هنالك عدد كبير من الأمثلة التي تم تطبيق قواعد الربط بها مما أدى الى اتساع رقعة استخدام هذه الطرق في تطوير التطبيقات المختلفة في مختلف مجالات تكنولوجيا المعلومات، على سبيل المثال امنية البيانات، الاتصالات، نظم اتخاذ القرار، أنظمة المبيعات، وغيرها (Han et al., 2011).

بنحوٍ عام، تعمل قواعد الربط على إيجاد تفاصيل معينة وظيفتها الأساسية إيجاد خصائص مشتركة ومؤشرات فيما بين العناصر المشتركة (Zaki, Meira Jr, & Meira, 2014). وطبقا الى النموذج القياسي لقواعد الربط، والمقدم من قبل (Agrawal et al., 1993)، ولكي تكون القواعد مهمة، وذات قيمة، هنالك أمران يظهران مدى أهمية وموثوقية قواعد الربط. الامر الأول: هو مقياس الدعم (Support)، والذي يطلق عليه أيضا (الدعم النسبيّ) والذي يقيس مدى تعددية ظهور عنصرين محددين سويا في عملية واحدة وبالتالي حساب عدد العمليات التي تمثل هذا الظهور للعناصر المشتركة. والامر الثاني: هو مقياس الثقة (Confidence) والذي يمثل النسبة بين العمليات التي تضم عنصرين محددين الى العمليات التي تحوي العنصر الأول وبمعنى اخر هو تمثيل لتقدير الاحتمالية (Probability). ومن الجدير بالذكر ان مقياس الثقة حساس جدا

لاتجاه العملية (بمعنى اخر غير متماثل)، حيث ان احتمال العملية (X, Y) يختلف تماما عن العملية (Y, X) (Han et al., 2011). ونحوٍ عام، تتلخص إجراءات إيجاد قواعد الربط بما يأتي:

١. إيجاد مجموعة العناصر المتكررة ضمن عمليات معينة مع تحديد عتبة للتكرارات لتحديد أدنى مقياس للدعم.

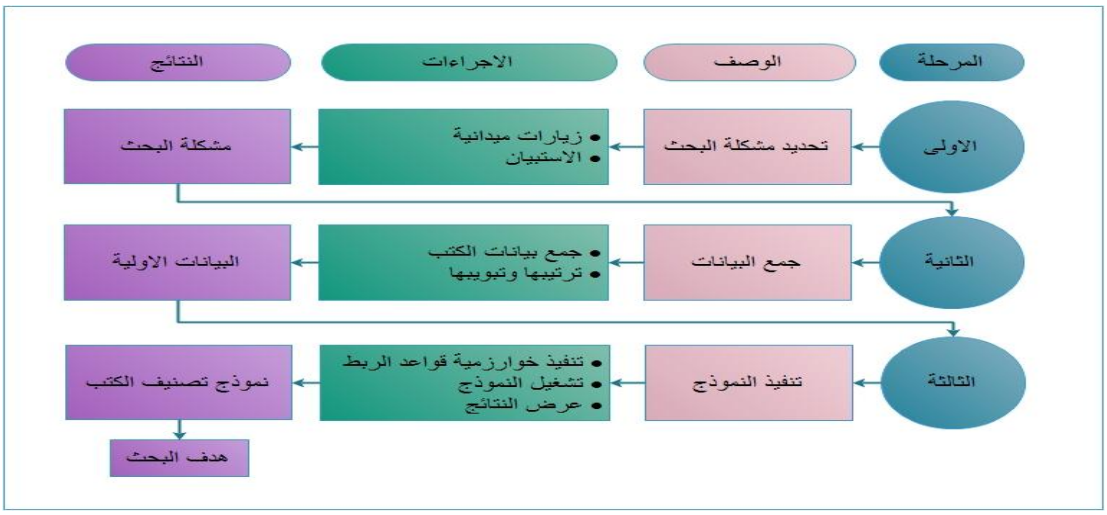
٢. توليد قواعد لربط العناصر من بين مجموعة العناصر المتكررة، حيث أنَّه العلاقة الجيدة هي التي تحقق أدنى مستوى للثقة والدعم التي تم تحديدها مسبقا. أنَّ من اهم الأمور الذي يجب اخذها بالحسبان هو أنَّ عملية إيجاد العناصر المترابطة تعدُّ عملية صعبة ومعقدة نوعا ما، وذلك نتيجة تطبيق هذا المبدأ على كم كبير من البيانات، وإيجاد تكراراتها.

في بحثنا الحالي، الهدف الأساس من استخدام قواعد الربط هو لإيجاد العلاقات المهمة التي تربط خصائص البيانات (بيانات الكتاب)، والتي تحدد تصنيفه. بمعنى اخر إيجاد البيانات المتكررة للكتب التي تقع ضمن صنف معين (تخصص معين)، والتي تسهم الى حدٍّ كبير في تصنيف الكتب ووضعها ضمن فئة معينة اعتماداً على تصنيف ديوي العشريّ.

الفصل الثالث/ الجانب العمليّ:

١.٣ هيكل البحث:

بنحوٍ عام يتكون أي بحث علمي من مجموعة من الخطوات التي تشكل بمجملها جميع مراحل تطور البحث. وتكون هذه الخطوات متسلسلة منطقيا بحسب تسلسل زمنيّ لكلِّ مرحلة من مراحل البحث هنالك مجموعة من المدخلات والمعالجات والمخرجات. وبتكامل هذه المراحل ومكوناتها الفرعية يتكامل البحث، ويتشكل هيكله واطاره العام. شكل (١) يوضح الإطار العام للبحث الحالي ومراحله الثلاثة التي سيتم تفصيلها في الفقرات القادمة.



شكل (1): الإطار العام للبحث

٢.٣ تحديد مشكلة البحث:

كما ذكر في مشكلة البحث، انه تم اجراء زيارات ميدانية وتوزيع استبانة على عينة من العاملين في المكتبات والمختصين بعلم المعلومات والمكتبات، ملحق (١). كما تم اجراء الاستبانة في المكتبات الآتية، وذلك لكونها تقع ضمن نطاق جغرافي متقارب مما يسهل زيارتها:

١. مكتبة كلية الآداب/ الجامعة المستنصرية.

٢. المكتبة المركزية/ جامعة بغداد.

٣. مكتبة كلية التربية/ ابن رشد/ جامعة بغداد.

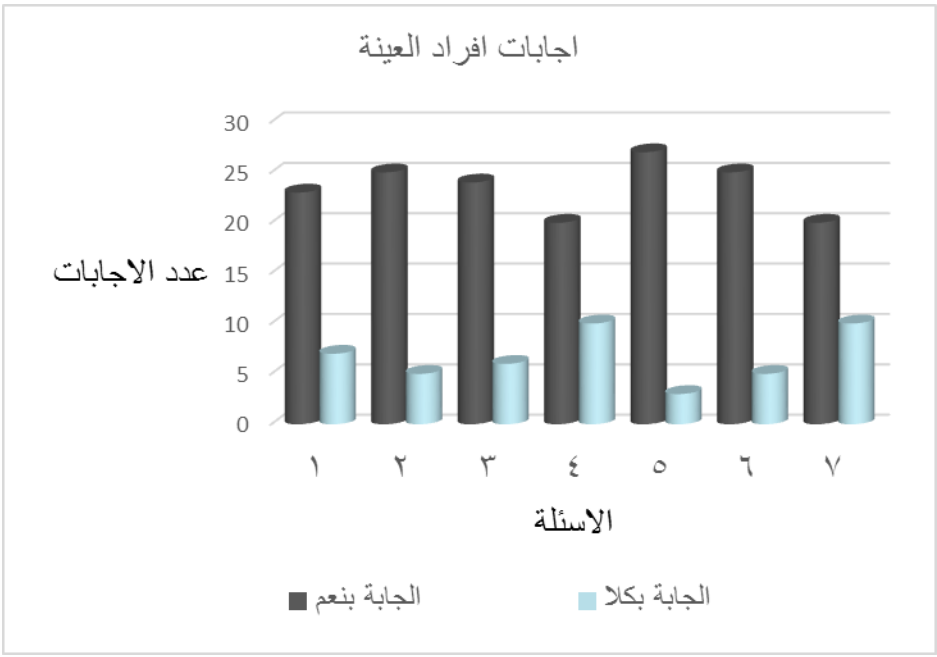
٤. مكتبة الكلية التقنية الإدارية/ الجامعة التقنية الوسطى.

حيث أنَّ حجم عينة الاستبانة تتكون من (٣٠) استمارة، وتوزعت إجابات افراد العينة كما هو موضح في جدول (١). والذي يبيّن قيم الإجابات عن أسئلة الاستبانة.

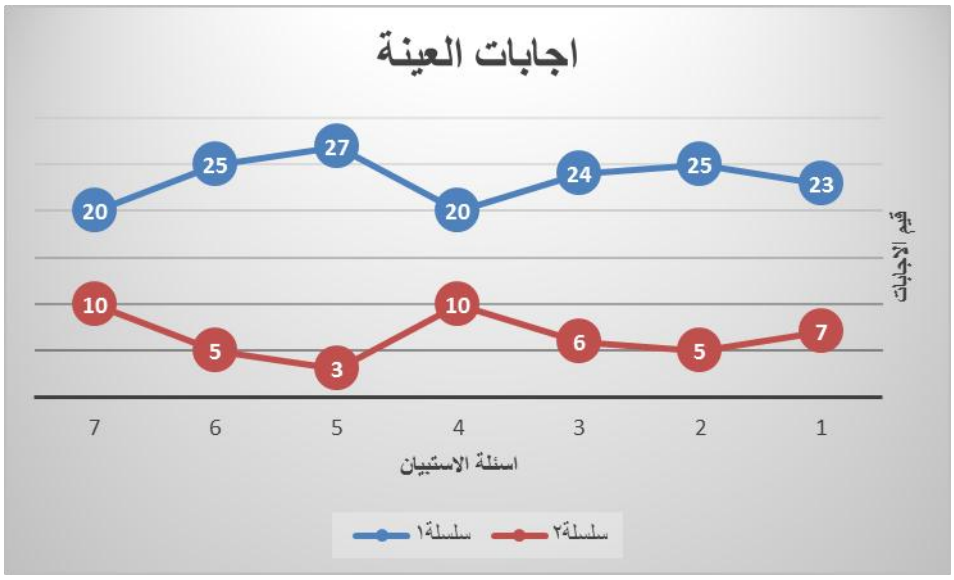
جدول (1): توزيع اجابات العينة حول مشكلة البحث

ت	السؤال	الجابة بنعم		الجابة بكلا	
		عدد الإجابات	النسبة	عدد الاجابات	النسبة
	هل يواجه العاملون في المكتبات صعوبة في تصنيف الكتب الإنكليزية؟	٢٣	٧٦.٧	٧	٢٣.٣
	هل يتطلب تصنيف الكتب الإنكليزية مهارات لغوية خاصة لغرض القيام بمهمة تصنيف الكتب؟	٢٥	٨٣.٣	٥	١٦.٧
	هل تحتاج المكتبة الى كادر متخصص لتصنيف الكتب الإنكليزية؟	٢٤	٨٠	٦	٢٠
	تعد عملية تصنيف الكتب الإنكليزية معقدة نوعا ما بالمقارنة مع تصنيف الكتب العربية؟	٢٠	٦٦.٧	١٠	٣٣.٣
	هل تشجع بناء نظم معلومات برمجية تقوم بعملية تصنيف الكتب الإنكليزية؟	٢٧	٩٠	٣	١٠
	هل تعتقد ان نظم المعلومات المحوسبة تساهم في تسهيل عمل المكتبي من حيث تصنيف الكتب الإنكليزية؟	٢٥	٨٣.٣	٥	١٦.٧
	تشكل اللغة الإنكليزية عائقا كبيرا امام العاملين في المكتبات لغرض تصنيف الكتب الإنكليزية؟	٢٠	٦٦.٧	١٠	٣٣.٣

والواضح من الجدول أعلاه أنَّ هنالك اجماعاً من أفراد العينة على أنَّ عملية فهرسة الكتب الإنكليزية وتصنيفها لا تتسم بالسهولة إذا ما قورنت بالكتب العربية، شكل (٢)، شكل (٣). فضلاً عن ذلك، ومن ملاحظة قيم الجدول المذكور في أعلاه وللأسئلة (٥، ٦)، واستناداً الى القيم التي أجاب عنها افراد العينة والخاصة بالحاجة الى نظم معلومات محوسبة تساعد على فهرسة الكتب وتصنيفها، نلاحظ ان هنالك تأييداً وبنسبة كبيرة لتطوير هكذا طرق من شأنها المساهمة في التقليل من صعوبات عملية فهرسة الكتب الإنكليزية وتصنيفها.



شكل (2): مخطط اجابات العينة



شكل (3): التباين في اجابات افراد العينة

إنَّ كلَّ ما ذكر أعلاه يبيّن أنَّ هنالك مشكلة تواجه العاملين في فهرسة الكتب الإنكليزية وتصنيفها، وبهذا تكتمل المرحلة الأولى من مراحل البحث التي تهدف الى تحديد مشكلة البحث.

٣.٣ جمع البيانات:

تعدُّ عملية جمع البيانات من اهم مراحل تنفيذ البحوث العلمية، لما لها من أثر مباشر في نتائج البحث. في هذا البحث تم جمع بيانات لغرض إجراء عملية تصنيف الكترونية مبنية على مبادئ الذكاء الاصطناعي للحاسوب. وبما أننا نستخدم قواعد الربط، لذا يجب أن يتمّ تبويب

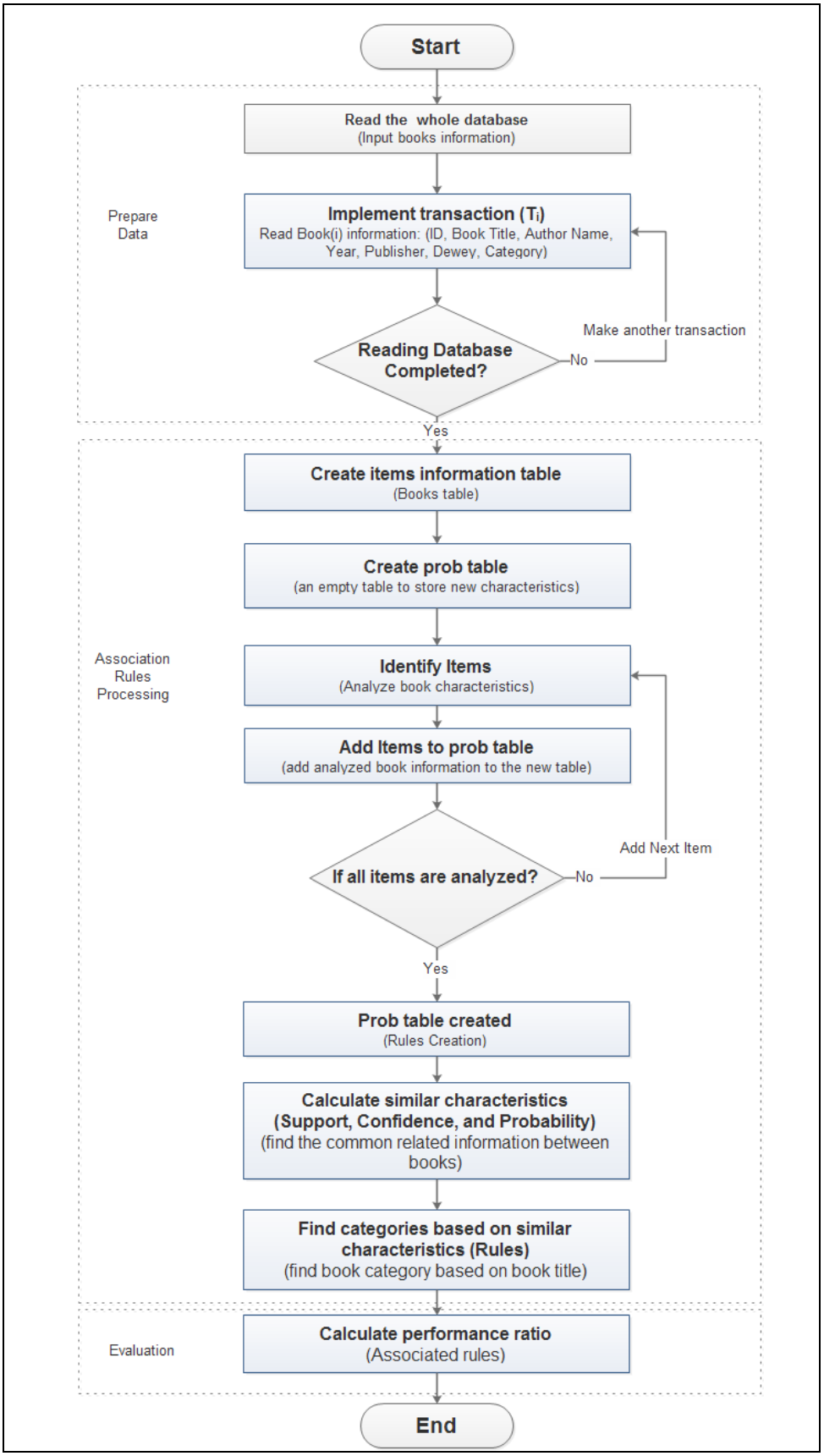
البيانات بشكل جدول. تم جمع بيانات الكتب التي من مكتبة كلية الآداب/ الجامعة المستنصرية. وتكونت عينة الكتب من (٥٠) كتاب باللغة الإنكليزية، ملحق (٢). كما وتم تمثيل كل كتاب من الكتب التي تم ادراجها ضمن العينة بمجموعة من الحقول في الجدول، تمثلت بما يأتي (عنوان الكتاب، اسم المؤلف، السنة، الناشر، الفئة، تصنيف ديوي العشري). يمثل جدول البيانات التي تم تجميعها مدخلات خوارزمية قواعد الربط التي سيتم تنفيذها.

٤.٣ تنفيذ النموذج المقترح والنتائج:

كما ذكر آنفاً أنَّ البحث الحالي يهدف إلى تقديم نموذج لتصنيف الكتب الإنكليزية باستخدام قواعد الربط (Association Rules). وعليه، تم بناء نموذج يوضح خوارزمية قواعد الربط ،وكيفية تطبيقها على قاعدة بيانات الكتب التي تم جمعها، كما هو موضح في شكل (٤) الذي يبين الهيكل العام والخوارزمية للنموذج المقترح والمصمم ضمن هذا البحث. كما وتم برمجة إجراءات قواعد الربط (Association Rules) باستخدام لغة البرمجة (Visual Basic.Net –MS 2012)، وهي احدى أشهر لغات البرمجة المستخدمة في بناء وتنفيذ الأنظمة والبرامج، شكل (٥). في بداية الامر تم اجراء بعض التعديلات على البيانات التي تم جمعها، وذلك بتحويلها الى أصناف (مجاميع)، وكما هو موضح في الجدول (٢).

جدول (2): اصناف الكتب التي تم جمعها

ت	اسم الصنف	عدد الكتب
١	Applied science and technology	١
٢	Arts and recreation	١
٣	Generalities	٥
٤	Geography and history	٨
٥	Languages	٤
٦	Literature	١٦
٧	Natural science and mathematics	٢
٨	Philosophy and psychology	٨
٩	Religion	٢
١٠	Social science	٣
المجموع	10	٥٠



شكل (٤): نموذج تصنيف الكتب الانكليزية باستخدام خوارزمية قواعد الربط (Association Rules)

وحيث إنَّ قواعد الربط تقوم بإيجاد علاقات لكلّ عنصر من عناصر قاعدة البيانات، تم تكوين قواعد وفق المبدأ الآتي:

لكل قاعدة بيانات مجموعة من العناصر، ولكل عنصر مجموعة من البيانات

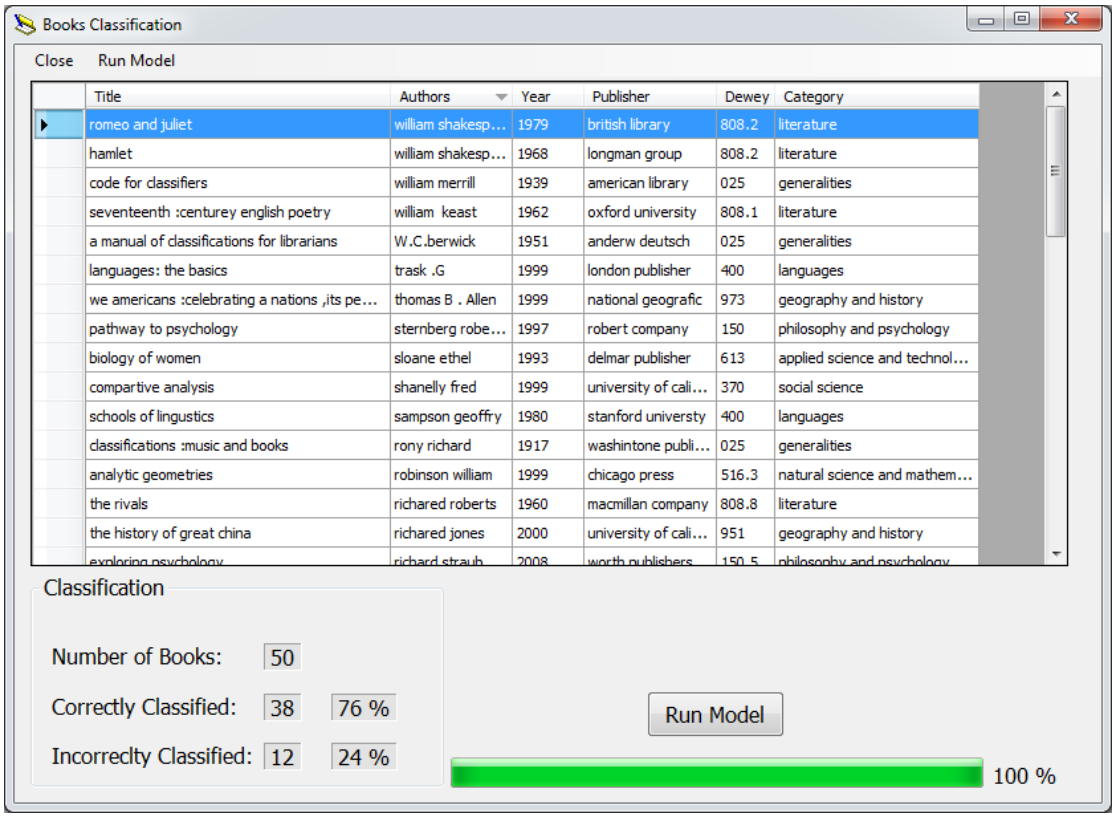
ووفقا للقاعدة أعلاه فإن القاعدة الأساسية التي تم اعتمادها في بحثنا هي كيفية تصنيف كتاب معين عن طريق إعطاء العنوان فقط. وبما أنَّ كل كتاب (عنصر) من الكتب التي جمعها يتكون من الخصائص التالية:

عنوان الكتاب، اسم المؤلف، السنة، سنة الطبع، اسم الناشر، تصنيف ديوي العشري، الصنف

بنحوٍ عام تعمل قواعد الربط على إيجاد علاقات معينة ضمن خصائص البيانات المستخدمة. وتم الاستفادة من هذا المبدأ في إيجاد علاقات معينة تربط الكتب التي تقع تحت صنف معين، من خلال تحليل بيانات كل كتاب، وهذا ما تقوم به قواعد الربط. حيث تعمل وفقاً لمبدأ الذكاء الاصطناعيّ الذي من شأنه القيام بعمليات معقدة، ومتكررة، لتحليل البيانات وتفسيرها، وإعطاء النتائج.

بعد تنفيذ عمليات قواعد الربط على عينة الكتب التي تم جمعها تم الحصول على النتائج الآتية:

- استطاعت خوارزمية الربط إيجاد علاقات تربط الكائنات التي تقع ضمن نفس الصنف وهو ما يشجع على بناء أنظمة معلومات قادرة على تصنيف الكتب اعتمادا على مبادئ تكنولوجيا المعلومات والذكاء الاصطناعي.
- ان النموذج المقترح والمعتمد على قواعد الربط تمكن من تصنيف الكتب الإنكليزية الى مجاميعها بنسبة (٧٦%) أي بمعدل ثلاثة ارباع العينة. أي أنَّ الخوارزمية قادرة على وضع كتاب معين ضمن صنفه بعد إعطاء العنوان فقط بنسبة (٧٦%).
- إنَّ من بين الخصائص العائدة لكلِّ كتاب، كان الأثر الأكبر من نصيب تصنيف ديوي العشريّ، حيث يسهم بنحوٍ كبير في تسهيل عملية التصنيف.



شكل (5): واجهة تنفيذ النموذج المقترح

٥.٣ الاستنتاجات:

خرج الباحث بعد اتمام البحث بجملة من الاستنتاجات، وهي كما يأتي:

١. هنالك صعوبة واضحة في عملية تصنيف الكتب الإنكليزية، وهي صعوبة مرتبطة بمحددات اللغة للعاملين في اقسام فهرسة الكتب في المكتبات وتصنيفها.

٢. يمكن الاستفادة من أساليب التنقيب عن البيانات والذكاء الاصطناعي في بناء خوارزميات متخصصة في علمية تصنيف الكتب نظرا لنجاح هذه التقنيات في كثير من فروع العلم والمعرفة، والاستفادة منها في اثراء العمل وتحسينه.

٣. يمكن لتقنيات قواعد الربط إيجاد علاقات تربط الكتب الواقعة تحت التصنيف نفسه، عن طريق تحليل دقيق لخصائص كل كتاب.

٤. إنَّ النموذج المقدم يستطيع تصنيف الكتب الى اصنافها الأساسية، من خلال عنوان الكتاب وبنسبة نجاح قدرها (٧٦%) من حجم العينة.

٥. ان النسبة أعلاه تعدُّ جيدة وتشكل نقطة بداية مقنعة، يمكن الاستفادة منها في بناء أنظمة تصنيف الكترونية تسهم في تسهيل مهمة تصنيف الكتب الإنكليزية للعاملين في اقسام فرسة الكتب وتصنيفها في المكتبات العراقية.

٦.٣ التوصيات:

استناداً لما قدمه البحث من نتائج، تم صياغة جملة من التوصيات وكما يأتي:

١. العمل على تحسين نتائج الخوارزمية من خلال تجربة خصائص أخرى للكتاب، مثل الكلمات المفتاحية والمستخلصات.

٢. استخدام الطرائق الأخرى للذكاء الاصطناعيّ، كالشبكات العصبية، وتقييم نتائجها.

٣. العمل على بناء نموذج مشابه لتصنيف الكتب العربية.

٤. الاهتمام بتطوير نظم المعلومات المبنية على أدوات وأساليب الذكاء الاصطناعيّ، لتحسين جودة ونوعية نظم المعلومات.

المصادر:

١. إتيم محمود أحمد. التصنيف بين النظرية والتطبيق. – عمان: الدار العربية للموسوعات، ١٩٨٧. – ص٥١.

٢. اوديت بدران. التصنيف في المكتبات: دليل ارشادي لتصنيف ديوي العشري. – بغداد: الجامعة المستنصرية، ١٩٧٦. – ص٥٣.

٣. خالد الحديدي. فلسفة علم تصنيف الكتب كمدخل لفلسفة العلوم. – مصر: مكتبة النهضة، ١٩٦٩. – ص١٤٨

٤. فؤاد إسماعيل فهمي. تصنيف ديوي العشري بين النظرية والتطبيق. السعودية: جامعة الملك عبد العزيز، ١٩٨٤– ص١٢.

٥. محمود الاخرس. التصنيف. – عمان: الاخرس، ١٩٦٥. – ص٤١.

6. Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. Paper presented at the Acm sigmod record.

7. Chan, L. M. (1996). Dewey Decimal Classification: a practical guide: Forest Press.

8. Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques: Elsevier.

9. Krishna Kumar, A., Amrita, D., & Priya, N. S. (2013). Mining Association Rules between Sets of Items in Large Databases. International Journal of Science and Modern Engineering (IJISME), ISSN, 2319-6386.

10.Mann, M. (1949). Introduction to Cataloguing and Classification of Books (2nd ed.). Chicago: ALA.

11.Pengfei, G., Liangxian, D., & Junxia, Q. (2012). A Library Book Intelligence Classification System based on Multi-agent. Physics Procedia, 24, 2187-2193.

12.Ping, W., & Tai-shan, Y. (2011). A method for books classification based on fuzzy mathematics theory. Paper presented at the Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference on.

13.Richardson, E. C. (1912). Classification: Theoritical and Practical. New York: Scribner.

14.Sayers, W. (1944). Manual of classification for librarians and bibliographers. 33.

15.Sheehy, E. P., Keckeissen, R. G., McIlvaine, E., & Winchell, C. M. (1976). Guide to reference books (Vol. 976): American Library Association Chicago, IL.

16.Shirude, S. B., & Kolhe, S. R. (2016, 16-18 March 2016). Classifying library resources in Library Recommender Agent using PU learning approach. Paper presented at the 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE).

17.Zaki, M. J., Meira Jr, W., & Meira, W. (2014). Data mining and analysis: fundamental concepts and algorithms: Cambridge University Press.

18.

References:

1) Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. Paper presented at the Acm sigmod record.
2) Chan, L. M. (1996). Dewey Decimal Classification: a practical guide: Forest Press.
3) Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques: Elsevier.
4) Krishna Kumar, A., Amrita, D., & Priya, N. S. (2013). Mining Association Rules between Sets of Items in Large Databases. International Journal of Science and Modern Engineering (IJISME), ISSN, 2319-6386.
5) Mann, M. (1949). Introduction to Cataloguing and Classification of Books (2nd ed.). Chicago: ALA.
6) Pengfei, G., Liangxian, D., & Junxia, Q. (2012). A Library Book Intelligence Classification System based on Multi-agent. Physics Procedia, 24, 2187-2193.
7) Ping, W., & Tai-shan, Y. (2011). A method for books classification based on fuzzy mathematics theory. Paper presented at the Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference on.
8) Richardson, E. C. (1912). Classification: Theoritical and Practical. New York: Scribner.
9) Sayers, W. (1944). Manual of classification for librarians and bibliographers. 33.
10) Sheehy, E. P., Keckeissen, R. G., McIlvaine, E., & Winchell, C. M. (1976). Guide to reference books (Vol. 976): American Library Association Chicago, IL.
11) Shirude, S. B., & Kolhe, S. R. (2016, 16-18 March 2016). Classifying library resources in Library Recommender Agent using PU learning approach. Paper presented at the 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE).
12) Zaki, M. J., Meira Jr, W., & Meira, W. (2014). Data mining and analysis: fundamental concepts and algorithms: Cambridge University Press.