

## تنقيب البيانات الضخمة في المكتبات الرقمية: دراسة في المفاهيم

## وآليات التطبيق

*Big data mining in digital libraries: a study of concepts and application mechanisms*

أسماء قرزيز\*

جامعة العربي التبسي- تبسة (الجزائر)، [asma.guerzize@univ-tebessa.dz](mailto:asma.guerzize@univ-tebessa.dz)

تاريخ النشر: 2021/06/15

تاريخ القبول: 2021/06/24

تاريخ الإرسال: 2021/06/16

## ملخص

تهدف هذه الدراسة لبناء منظومة معرفية حول تقنية تنقيب البيانات الضخمة باعتبارها أداة منهجية لتحليل واستكشاف بيئة المكتبات الرقمية حيث عرضت استخدامات هذه التقنية في تحليل هيكلية وبناء موقع المكتبات الرقمية ومحتواه المعلوماتي بالإضافة لتحليل الخدمات التي تتيحها ضمنه ، فقد جاءت هذه الورقة العلمية لتسليط الضوء على البيانات الضخمة من خلال توضيح مفهومها، خصائصها أسباب ظهورها بشكل عام، كما أبرزت كيفية استفادة البيانات الضخمة من علم التنقيب عن البيانات لاختزال البيانات غير النافعة داخل طوفان البيانات واستخراج الحكمة والمعرفة التقنية منها، فتم التعريف بتقنية التنقيب عن البيانات مع تقديم عرض مفصل ومبسط لمختلف تقنياتها، مع التعرّيج في الأخير على أهم أساليب تنقيب البيانات الضخمة المستخدمة في المكتبات الرقمية كخوارزميات العنقدة والتصنيف وتحليل الارتباط وتحليل السلاسل الزمنية والمجالات المطبقة عليها داخلها ، فضلا عن تقديم تجربتين عمليتين لمكتبتين رقميتين في هذا المجال أحدهما اجنبية والأخرى عربية ، وقد اعتمد في إنجاز هذه الدراسة على المنهج الوصفي، وخلصت الدراسة لمجموعة من النتائج وقدمت الاقتراحات من شأنها أن تعزز وتفعّل التوجه نحو استخدام تطبيقات تقنية تنقيب البيانات الضخمة في المكتبات الرقمية.

الكلمات المفتاحية: تنقيب البيانات؛ البيانات الضخمة؛ تنقيب البيانات الضخمة؛ المكتبة

الرقمية؛ تطبيقات البيانات الضخمة في المكتبات الرقمية.

## Abstract

*The purpose of this study is to build a knowledge system on the technique of large data mining as a methodological tool to analyze and explore the environment of digital libraries, where the uses of this technique were presented in the analysis of the structure and construction of the digital library site and its content and the analysis of the services it offers.*

*On the huge data by clarifying the concept, characteristics and reasons for their emergence in general, and highlighted how to benefit from the huge data from the science of data mining to reduce the data useless within the flood of data and the extraction of wisdom and technical knowledge m The data mining technology was introduced with a detailed and simplified presentation of its various functions, with the finalization of the most important data mining methods used in digital libraries such as cluster algorithms, classification, correlation analysis, time series analysis and applied fields, Two practical experiences in this field, one foreign and the other Arab.*

*The study was based on a descriptive approach. The study concluded with a set of results and presented some suggestions that would enhance and promote the use of large-scale data mining applications in digital libraries.*

**Keywords:** *Data mining; Big data; Massive data mining; Digital library; Massive data applications in digital libraries.*

## مقدمة

تعد المكتبات الرقمية من أهم وأحدث المكتبات في عصرنا الحالي فهي تساهم بشكل كبير في تطوير وتغيير المجتمعات العصرية جذريا من خلال ما تقدمه من معارف رقمية لجمهورها حيث تعمل على توفير آخر ما ينشر من مصادر المعلومات وتقديمها بسرعة فائقة وبطرق فعالة تسهل تداولها وتطبيقها والتفاعل معها. وقد ظهرت في الآونة الأخيرة ظاهرة الكم الهائل من البيانات الرقمية أو ما يسمى بالبيانات الكبيرة والتي تنبع من كل مكان وعن أي نشاط أو خدمة رقمية، و قد تأثرت المكتبات الرقمية بهذه الظاهرة نتيجة للطفرة الرقمية المتزايدة بها كما و نوعا بسبب استخدامها لتكنولوجيا المعلومات المختلفة و كذا باعتبارها الراعي و الحافظ الرسمي لمواد المعرفة الرقمية المتعددة و المختلفة و المتجددة حيث شكل ذلك تحديا رئيسيا لها، فالأساليب التقليدية لقواعد إدارة البيانات و المعلومات بها أصبحت عاجزة عن التحكم و السيطرة على

هذا الفيضان المعرفي من حيث رصده ، تجميعه ، إدارته ، تحليله وتخزينه هذا ما خلق ضجيجا وتشويشا لأدوارها و مهامها و أهدافها. مما دفع بمسيرها لتوجيه خططهم واستراتيجياتهم نحو الاستثمار في تقنيات استكشاف وتحليل البيانات الضخمة بها لاستخراج الحكمة المخترنة فيها ولبناء منظومة معرفية رقمية سليمة.

## 1. الإطار المنهجي للدراسة

### 1.1. إشكالية الدراسة

كل يوم يتم على مستوى المكتبات الرقمية مجموعة من الوظائف والخدمات المساهمة في جمع و رقمنة الانتاج الفكري العلمي وحفظه و اتاحته لتسهيل الوصول إليه عند الطلب ، وفي كل خطوة من هذه الخطوات يتم تجميع و تكديس كم هائل من البيانات الخام عن كل من الخدمات ، الوظائف و أخصائي المعلومات و استخداماتها من قبل المستخدمين بقواعد بيانات المكتبة الرقمية ، ولما كان الهدف الأسمى لوجود هذه المكتبات هو تحقيق رضى مستفيديها، و بما أن هذه البيانات تقدم صورة متكاملة عن المكتبة و كيفية تداولها داخلها فقد اتجهت هذه الأخيرة إلى استخدام تقنية تنقيب البيانات الضخمة لفهمها بغية الاستفادة منها بما تقتضيه الحاجة و الضرورة لتحقيق هذا الهدف. فكيف يمكن استثمار واستخدام آليات وأساليب تقنية تحليل البيانات الضخمة في المكتبات الرقمية؟

### 2.1. تساؤلات الدراسة

من الاشكالية السابقة تنبثق مجموعة التساؤلات التالية والتي تأتي الإجابة عنها من خلال عناصر الدراسة.

\* هل تعد تقنية تنقيب البيانات الضخمة اليوم مطلبا جديدا وخيارا استراتيجيا لتحليل والتحكم في بيانات ومعلومات المكتبات الرقمية؟

\* كيف يمكن أن تتعامل المكتبات الرقمية مع تقنيات تنقيب البيانات الضخمة؟

\* وهل ستساهم هذه التقنية في خلق قيمة مضافة لهيكل ومحتوى وخدمات المكتبات الرقمية؟

\* ماهي أهم المكتبات الرقمية العربية والأجنبية التي طبقت تقنية تنقيب البيانات الضخمة؟

### 3.1 أهداف الدراسة

نسعى من خلال الدراسة التي بين أيدينا إلى تحقيق جملة من الأهداف نذكرها في النقاط الآتية:

\* تكوين خلفية نظرية لمفهومي البيانات الضخمة وتنقيب البيانات.

\* تحليل وتوضيح مجالات استخدام أدوات التنقيب عن البيانات الضخمة في المكتبات الرقمية.

\* التحسيس بالإزامية تفعيل مشاريع واقعية لاستغلال تطبيقات وفوائد تنقيب البيانات الضخمة في المكتبات الرقمية.  
\* عرض تجربتين لمكتبتين رقميتين طبقنا تقنية تنقيب البيانات الضخمة.

#### 4.1. منهج البحث

اعتمدت هذه الدراسة على المنهج الوصفي كونه الطريق الأنسب لمثل هذه الدراسات النظرية، حيث أتاح لنا جمع البيانات والحقائق من الأدبيات والدراسات السابقة والمراجع العلمية ذات الصلة بمجال البحث، مما مكنتنا من فهم ماهية البيانات الضخمة، علاقة تقنية تنقيب البيانات بالبيانات الضخمة، ومجالات الإفادة من آليات تنقيب البيانات الضخمة في المكتبات الرقمية، وكذا عرض جهود تجربتين لمكتبتين رقميتين طبقنا أساليب تنقيب البيانات الضخمة.

#### 2. الإطار النظري للدراسة

##### 1.2. مدخل عام عن البيانات الضخمة (big data)

1.1.2. مفاهيم أساسية للبيانات الضخمة: يعد مصطلح البيانات الضخمة من المصطلحات الحديثة التي ظهرت كاتجاه جديد لوصف التدفق الهائل وكبير جدا للبيانات المنظمة وغير المنظمة، وقد استخدم هذا المفهوم لأول مرة بشكل علمي سنة 2001 من طرف لاني (Laney) حيث وصفها بأنها البيانات التي لا يمكن معالجتها بواسطة أدوات إدارة البيانات التقليدية، وأن هذه البيانات حتى تكون ضخمة يجب أن تتصف بثلاث خصائص تبدأ بحرف (V): وهي الحجم (Volume) والسرعة (Velocity) والتنوع (Variety).<sup>1</sup>

فالبيانات الضخمة مصطلح يستخدم لوصف مجموعات بيانات كبيرة للغاية أو على درجة من التعقيد أو التي تتطلب قدرا كبيرا من المعالجة السريعة لما يعرف في البيانات الضخمة بمشاكل الحجم/التنوع/السرعة. والتي من الصعب أو من المستحيل التعامل معها باستخدام قواعد البيانات والأدوات التقليدية. فقد بلغ حجم البيانات الرقمية اليوم 75 مرة أكثر من حبوب الرمال في العالم.<sup>2</sup>

وفيما يلي نستعرض بعض التعاريف التي وضعت لتوضيح هذا المفهوم، ونوردها على سبيل المثال لا الحصر:

تعريف كل من **Dollar** و **Gallagher** حيث عرفاها بأنها: "أصول المعلومات التي تتميز بالحجم الكبير والسرعة والتنوع ويتطلب تحليلها أساليب تقنية و تحليلية محددة لتحويلها إلى قيمة،

وتنتج البيانات الضخمة من عملية تراكم المعلومات السابقة و الحالية حول نشاط الأفراد في مختلف مجالات الحياة من أجل التنبؤ بالسلوكات بشكل منطقي أو الأخذ بالاحتياجات المستقبلية".

كما يعرفها المجلس الاقتصادي والاجتماعي للأمم المتحدة بأنها: "بيانات تتسم بضخامة كميتها وسرعتها الفائقة وشدة تنوعها بحيث تتطلب أشكالا فعالة من حيث التكلفة و مآطرة لفهمها على نحو أعمق واستخدامها على نحو أفضل في عمليات اتخاذ القرارات".<sup>3</sup>

كما جاء تعريفها بأنها: "الأشياء التي يمكن القيام بها على نطاق واسع لإنتاج بيانات واستخلاص رؤى جديدة أو خلق إشكال جديدة من القيمة بطرق تغير آليات الأسواق والمنظمات في حين نجد معهد ماكنزي العالمي سنة 2011 عرفها بأنها: "مجموعة من البيانات التي تتوافر بأحجام تفوق قدرة أدوات قواعد البيانات التقليدية في تجميع، تخزين، إدارة وتحليل تلك البيانات".<sup>4</sup>

أما المنظمة الدولية للمعايير (ISO) فتعرفها على أنها «مجموعة أو مجموعات من البيانات لها خصائصها الفريدة (مثل الحجم، السرعة، التنوع، التباين، صحة البيانات... إلخ)، لا يمكن معالجتها بكفاءة باستخدام التكنولوجيا الحالية والتقليدية لتحقيق الاستفادة منها".

ونجد شركة (IBM) تصف البيانات الضخمة بأنها: "تنشأ عن طريق كل شيء من حولنا وفي كل الأوقات وكل عملية رقمية وكل تبادل في وسائل التواصل الاجتماعي، تتناقلها الأنظمة، وأجهزة الاستشعار، والأجهزة النقالة فلها مصادر متعددة في السرعة والحجم والتنوع ولكي نستخرج منفعة معنوية من البيانات الضخمة نحتاج إلى معالجة مثالية، وقدرات تحليلية، ومهارات".<sup>5</sup>

إذن فالبيانات الضخمة تعبر عن كمية هائلة من البيانات المعقدة التي يفوق حجمها قدرة البرمجيات والآليات الحاسوبية التقليدية. ترصد برامجها وأجهزتها من خلال أنظمة استشعار وتطبيقات ما يحصل في البيئة المحيطة، وتنتج بيانات تخبرنا بما يحصل حولنا ومصادرها متعددة كمستشعرات الفضاء، الهواتف الذكية، التكنولوجيا القابلة للارتداء.<sup>6</sup>

#### 2.1.2. خصائص البيانات الضخمة: تتميز البيانات الضخمة بـ:

\* **كبر حجمها (Volume):** حيث أنه من المتوقع بحلول العام 2020 أن يحتوي الفضاء الإلكتروني على مايقرب من 40.000 زيتابايت من البيانات الجاهزة للتحليل واستخلاص المعلومات، من ملف نصي بسيط يقدر حجمه بعده كيلوبايتات مرورا بمقطع صوتي بالميجابايت أو فيديو بالجيجابايت إلى ملايين الهواتف الذكية التي تبث كميات ضخمة من البيانات إلى شبكات الهاتف كل ثانية. وبالتالي فحجم البيانات الضخمة كبير جدا ويحتاج إلى معالجات وأجهزة كبيرة قادرة على التعامل مع هذه البيانات<sup>7</sup>

\***سرعة نموها (Velocity):** ويقصد بها سرعة إنتاج واستخراج البيانات لتغطية الطلب عليها حيث تعتبر السرعة عنصرا حاسما في اتخاذ القرار بناء على هذه البيانات، وهو الوقت الذي نستغرقه من لحظة وصول هذه البيانات إلى لحظة الخروج بالقرار بناء عليها.

\***تعدد أنواعها (Volume):** فمع ازدياد أعداد مستخدمي الانترنت والهواتف الذكية وشبكات التواصل الاجتماعي المختلفة تغير شكل البيانات المؤلف من بيانات مهيكلية في قواعد بيانات إلى بيانات غير مهيكلية تتضمن عدد كبير من الصيغ مثل الصور ومقاطع الصوت والفيديو والرسائل القصيرة والوثائق بصيغها المتعددة مثل الصورة والصوت والفيديو والنص.<sup>8</sup>

\***متعددة الجودة والمصدقية (Veracity):** حيث انه ليست كل المعلومات والبيانات الواردة إلينا يمكن الاستفادة منها وتوظيفها في خدمة المؤسسة وصناعة القرار وبالتالي فانه يتم التخلص وأتلاف بعض البيانات data cleaning.

\***ذات قيمة كبيرة (value):** للاستفادة من البيانات الضخمة نحتاج إلى متخصصين يمتلكون الخبرات والمهارات الكافية للتعامل مع هذه البيانات وتحليلها التحليل المناسب وفي هذه الحالة تعتبر المعلومات ذات قيمة.

\***ذات قيمة متغيرة (variability):** بمعنى أن نفس المعلومات أو نفس البيانات يمكن إن تعنى عدة أشياء واستنادا إلى السياق الذي وردت فيه يمكن تحديد قيمتها وتحليلها تحليلا مناسبيا.  
\***متعددة المظاهر (visualization):** عند استخدام البيانات الضخمة يجب تحليلها وإظهارها بإشكال مختلفة تناسب مع طبيعة استخدامها وتأخذ أشكال متعددة مثل الإحصاءات والأرقام والأشكال الهندسية وغيرها.<sup>9</sup>

3.1.2. أسباب ظهور البيانات الضخمة: هناك أسباب عديدة أدت إلى ظهورها نجملها في:  
\* ظهور تقنيات انترنت الأشياء التي تتيح لجميع الأجهزة التواصل مع بعضها والترابط تقنيات الانترنت وإنتاج بيانات جديدة.

\* ظهور مجالات بحثية تنتج بيانات ضخمة جدا لا بد من تحليلها مثل علم علم الجينوم والمحاكاة الفيزيائية المعقدة والبحوث البيولوجية والبيئية وعلم الارصاد.  
\* ظهور الشبكات الاجتماعية التي ترسل كما ضخما من البيانات على مدار الساعة ومن مختلف الهيئات والجهات.

\* القوانين التي تحتم ضرورة بقاء هذه البيانات في قواعد البيانات لمتابعة المجرمين والمخربين والمتسللين.<sup>10</sup>

\*النمو المتسارع للنطاق العريض فائق السرعة للثابت والجوال بالإضافة للطاقة المتعاضمة للحوسبة السحابية.

\*النفوذ إلى تكنولوجيا البيانات الضخمة بطرق أسهل وأرخص.

\*تقدم وسائل الاتصال والتواصل بين الآلات (آلة إلى آلة).

\*تستخدم البيانات الضخمة لتوقع الجريمة من خلال تحليل البيانات للجرائم السابقة والحالية.

\*انخفاض تكاليف تخزين البيانات على غرار نظم الحوسبة السحابية.<sup>11</sup>

2.2 تنقيب البيانات وسيلة لتحليل البيانات الضخمة: تساعد تقنية تنقيب البيانات البحث في بحر ومناجم البيانات الضخمة وتحليلها لاكتشاف واستخلاص المعرفة منها وفي هذا الجزء سنعرض بالتفصيل كل المفاهيم التعريفية لها وطريقة عملها وأهم التقنيات التي تستعين بها لتحقيق ذلك.

1.2.2 مفهوم تنقيب البيانات (Data Mining): ظهر هذا المصطلح في منتصف التسعينات في الولايات المتحدة الأمريكية، وهو يجمع ما بين الإحصاء و تكنولوجيايات الإعلام (قواعد البيانات، الذكاء الاصطناعي، التعلم الآلي « machine Learning » ...).

وتوجد عدة تعريفات لهذا المفهوم منها: حيث يمكن تعريفها بأنها: " الاستكشاف الآلي أو المؤتمت لأنماط شائقة وغير جلوية مخفية في قاعدة بيانات معينة"<sup>12</sup>، أو أنها: " عبارة عن تحليلات لكمية كبيرة من البيانات بغرض إيجاد قواعد وأمثلة ونماذج التي يمكن أن تستخدم تقود وتدل أصحاب القرار، وتتنبأ بالسلوك المستقبلي"<sup>13</sup>، كما يمكن تعريفها كذلك بأنها: " تحليل لمجموعات كبيرة الحجم من البيانات المشاهدة للبحث عن علاقات محتملة وتلخيص للبيانات في أشكال جديدة لتكون مفهومة ومفيدة لمستخدمها"<sup>14</sup>.

من التعاريف السابقة نستنتج أن التنقيب في البيانات عبارة عن أداة لاكتشاف واستخراج المعلومات المفيدة والمثيرة للاهتمام بطريقة آلية من خلال تحليل مجموعة كبيرة جدا من البيانات بالاعتماد على النظم والأساليب الإحصائية والرياضية وعمليات الذكاء الاصطناعي وتقنيات التعليم الذاتي، حيث تساعد على استخلاص المعرفة المخفية والنماذج غير المتوقعة، إضافة إلى استكشاف قواعد جديدة موجودة في قواعد بيانات كبيرة لإنتاج معرفة تنبئية مستقبلية.

2.2.2. تقنيات التنقيب عن البيانات: هناك عدة اهداف للتنقيب في البيانات، نذكر أهمها:

\*التلخيص أو التكتيف (Summarization or Condensation): يشير التلخيص لأساليب فتتيت كتل البيانات الكبيرة إلى مقاييس موجزة، توفر وصفا عاما للمتغيرات وعلاقاتها. ومن الأمثلة على أساليب التلخيص نذكر: المتوسطات، والمجاميع، والإحصائيات الوصفية التي تتضمن

مقاييس النزعة المركزية مثل المتوسط الحسابي والوسيط والمنوال، ومقاييس التشتت مثل الانحراف المعياري. وعلى الرغم من أن مقاييس التلخيص تعطي صورة كبيرة عن بعض التفاصيل ذات العلاقة فإنها غالباً ما تهمل تفاصيل أخرى ذات أهمية كبيرة تتعلق بسلوك المستهلك خصوصاً.<sup>15</sup>

\***التصنيف (Classification):** هو تنظيم البيانات في أصناف و معطيات. وهو يستخدم مؤشرات الصنف المعطى لطلب المواضيع في مجموعة البيانات، وتستخدم طرق التصنيف عادة مجموعة التدريب حيث تكون جميع المواضيع مرتبطة مسبقاً مع مؤشرات اصناف معرفة، فتتعلم خوارزمية التصنيف من مجموعة التدريب وتبني نموذج يعمل على تصنيف المواضيع الجديدة. وبالتالي فإن التصنيف يساعد على تفسير أو التنبؤ بالسلوك المستقبلي للأصناف<sup>16</sup> أو الأفراد من خلال خصائص أخرى. ويمكن انجاز التصنيف بالاعتماد على الأساليب الإحصائية القديمة مثل الانحدار والتحليل التمييزي، أو بالاعتماد على أساليب حديثة نسبياً مثل قوى الارتباط والاستنتاج المستند إلى الحالة والشبكات العصبية.

وكأمثلة عن طرق التصنيف المستعملة كجزء من تطبيقات استكشاف المعرفة التي تتضمن تصنيف اتجاهات الأسواق المالية، والتحديد الآلي للأشياء المهمة في صورة كبيرة من قواعد البيانات.<sup>17</sup>

\* **التنبؤ (Prediction):** تتمثل الفكرة العامة للتنبؤ في استخدام عدد كبير من القيم السابقة لاعتبار قيم مستقبلية محتملة، حيث يقوم الشخص المعني بتخمين القيم المستقبلية اعتماداً على الأشكال التي تتكون من مجاميع البيانات والاشتقاق من العلاقات الاحتمالية بين المتغيرات المبحوث عنها للوصول إلى نتيجة تقديرية عن الأمر المتنبأ به<sup>18</sup>. ويوجد نوعان عامان من التنبؤ: ✓ واحد يستطيع إما محاولة التنبؤ ببعض قيم البيانات غير المتوفرة أو الأعمال رغبة الانتظار، أو تنبؤ بمؤشر صنف بعض البيانات.

✓ والآخر مرتبط مع التصنيف اعتماداً على مجموعة التدريب، يكون مؤشر الصنف تنبأً محدوس اعتماداً على قيم الصفات للموضوع وقيم الصفات للأصناف. والتنبؤ غالباً ما يشار للقيم العددية المفقودة، أو زيادة / نقصان الاتجاهات في وقت البيانات المناسب.<sup>19</sup> ومن الأدوات التقليدية المستخدمة في التنبؤ نذكر على سبيل المثال:



الانحدارات بأنواعها والتحليل التمييزي. أما الأساليب الجديدة فتشتمل على قواعد الارتباط وشجرة القرار والشبكات العصبية والخوارزميات الوراثية.<sup>20</sup>

\***العنقدة أو التجزئة (Clustering):** يتمثل التجميع العنقودي أو التجزئة إلى قطاعات في البحث عن مجموعات متجانسة في مجتمع من الأفراد ويشير التجميع العنقودي أو التجزئة إلى قطاعات إلى عملية تشكيل مجموعات أو قطاعات مؤلفة من أفراد أو أصحاب أسر، وذلك بالاستناد إلى معلومات متضمنة في مجاميع من المتغيرات التي تصفهم. والغرض من التجميع العنقودي المساعدة على تطوير برامج تسويقية مصممة على مقاسات الزبائن أنفسهم، والتي بالإمكان استخدامها لاستهداف أعضاء لكل قطاع من هذه القطاعات على أمل ترغيبهم في تكرار الشراء أو التحول إلى زبائن موالين.

وتتم أساليب التجميع العنقودي غالبا بمساعدة أساليب التحليل العنقودي الإحصائية والأساليب المستندة إلى شجرة القرار، والشبكات العصبية والخوارزميات الوراثية.

\***تحليل الارتباط (Rule Analysis):** يتمثل الارتباط في البحث عن علاقات أو ارتباطات موجودة بين عدة خصائص،<sup>21</sup> وتحليل الارتباط هو الاكتشاف لما يطلق عليه بصورة عامة قواعد الارتباط ( التي تتكون من مجموعتي السابقت والنتائج) وهي تدرس تكرار وجود أنماط ، ارتباطات ، تقاطعات أو هياكل عشوائية على رأس العناصر أو المواد في قواعد بيانات المعاملات ، قواعد البيانات العلائقية و مخازن المعلومات الأخرى. وهي تعتمد على:

✓ سماح ما يسمى اسناد، يحدد تكرار مجاميع العناصر.

✓ ويوجد سماح آخر يسمى الثقة، التي هي الاحتمالية المشروطة التي تجعل ظهور العنصر في المعاملة عندما يظهر العنصر الآخر، أنه يستخدم لتحديد قواعد الارتباط بالضبط تماما.

وبالتالي يكون هدف تنقيب قواعد الارتباط هو إيجاد جميع القواعد التي لها:

1. اسناد ≤ سماح minsup

2. الثقة ≤ سماح minsup

وللتوضيح ندرج المثال التالي: 98% من الزبائن الذين يشترون عجلات واكسسوارات للسيارة، هم أيضا يحصلون على خدمات ميكانيكية.<sup>22</sup>

\***تحليل السلاسل الزمنية:** هو من الأساليب التنبؤية للتنقيب عن البيانات المتميزة، لأنه يختص بقياس الزمن على وجه التحديد ويمكن تطبيقه على أي فترات زمنية متسلسلة بشرط ضبطها، كأن تكون على مدار عام كامل يبدأ بتاريخ محدد وينتهي بتاريخ محدد مع الأخذ في الاعتبار أيضا

أن يتم القياس باليوم والشهر حتى يتم قياس المتغير الزمني بشكل منتظم للتنبؤ بصورة منظمة أيضا.<sup>23</sup>

\*الكشف عن التغيرات أو الانحرافات (Change and deviation detection):

يرتكز على استكشاف التغيرات المهمة جدا في البيانات من خلال قياسات سابقة أو قيم معيارية.<sup>24</sup>  
3.2.2. طرق تنقيب البيانات:

1. التنقيب الوصفي (Descriptive Data Mining) : يعتمد على إعادة تنظيم البيانات، والتنقيب في أعماقها لاستخراج النماذج (Patterns) الموجودة فيها، كتشابه الزبائن الذي يسمح لك بإنشاء وصف بسيط عن مجموعة زبائن متشابهين، ولا يستوجب وجود هدف (Target) لمثل هذه البيانات.

2. التنقيب التنبؤي (Predictive Data Mining): يحاول إيجاد أفضل التنبؤات اعتمادا على المعطيات، كمعرفة المنتج الأفضل لزبون معين. باختصار يعتمد هذا التنقيب على استخدام المعلومات القديمة لتوقع ما سيحدث في المستقبل وتكون لدى مثل هذه البيانات هدف (Target).<sup>25</sup>

3.2. التنقيب عن البيانات الضخمة في مجال المكتبات الرقمية: لا تتوقف الاستفادة من البيانات الضخمة على المؤسسات والمشاريع التجارية بل تمتد إلى مجالات عديدة منها الطاقة والصحة والتعليم والبحث والمشاريع العلمية، وقد بدأت بوادرها تخوض غمار المكتبات الرقمية، وفيما يلي نعرض الاضافات التي قدمها هذه التقنية لهذا النوع من المكتبات كما ندرج نموذجين لتطبيقها بهما.

1.3.2. مجالات استخدام التنقيب عن البيانات في المكتبات الرقمية: تعد المكتبات الرقمية بيئة مناسبة للتعامل مع البيانات الضخمة ويمكن التعرف على مجالات الاستفادة منها داخلها من خلال ربطها بالتنقيب عن البيانات حيث يمكن أن نستفيد من تحليل البيانات الضخمة ومعالجتها<sup>26</sup>

\*تنمية المقتنيات واتخاذ القرارات: يتيح لنا استخدام خوارزميات تحليل الارتباط والتحليل العنقودي استنباط بعض القواعد التي ترشد وتقدم الرؤى الواضحة لمتخذي القرار حيال موقفهم من:

1. تحديد المصادر المطبوعة الأكثر استخداما وأهمية لرقمنتها: فسواء كانت المكتبة الرقمية متاحة من خلال الانترنت أو شبكة المكتبة داخل مؤسستها فإنها تستطيع تتبع وتسجيل زيارة

مستفيدها كاملة من خلال ملفات الولوج للمسفيد ذلك لأنها مبنية على قواعد بيانات علائقية واسعة لتخزين البيانات من مصادر معلومات مختلفة وربطها بالميتاداتا، وحفظ التسجيلات عن تحركات المستفيدين داخل صفحات المكتبة ومدى تفاعلهم واستخداماتهم لمصادر المعلومات. وبالتالي يصبح من السهل تحديد استخداماتهم داخل كل تخصص تغطيه المكتبة على حدة، ومن ثم إيجاد العلاقات بين استخداماتهم للمصادر ببعضها البعض للخروج بمثل هذه المؤشرات.

هذا علاوة على أن سياسة تنمية المقتنيات بالمكتبات الرقمية قائمة على الاتاحة والوصول وبالتالي كلما زادت درجة الاتاحة والوصول والاستخدام لمصادر معينة زادت من أهميتها وينطبق ذلك أيضا على تحديد الاشتراكات بقواعد البيانات والدوريات والكتب الإلكترونية.

2. تشخيص برامج التعاون في المصادر مع المؤسسات الأخرى ودرجة الحاجة إلى سد ثغرات مقتنياتهم: حيث يمكن الخروج بمجموعة من قواعد الارتباط التي تحدد أكثر المصادر متكررة الحدوث في استخداماتها من داخل مؤسسات أخرى ومدى ارتباطها وعلاقتها وتأثيرها على المصادر الموجودة بالمكتبة.

3. المساعدة في إدارة وتنظيم المحتوى الرقمي: من خلال الاهتمام بزوايا فاعلية الاستخدام التي يمكن توضيحها على النحو التالي:

\*بناء أنطولوجيا متخصصة موضوعيا: فمن تحليل ملفات الولوج باستخدام خوارزميات الارتباط للحصول على كثر المصادر استخداما وعلاقتها وتأثيرها على بعضها البعض على مستوى المكتبة ككل، يتم إعادة تقسيمها في عناقيد (مجموعات) باستخدام التحليل العنقودي للحصول على مجموعات مترابطة ومتداخلة داخل تخصص بعينه ويتم إتاحتها للمستفيدين للتصفح في شكل أنطولوجيا هرمية أو شجرية، وبالتالي يتم إعادة تنظيم مجموعات المكتبة بصورة أكثر فاعلية للمستفيد.

\*بناء أنطولوجيا عامة لكل مصادر المكتبة بشكل جديد مبنية على علاقات الاستخدام وتنقيب النصوص: حيث يتم إعادة هيكلية وتنظيم وإنشاء مجموعات من المصادر باستخدام خوارزميات التحليل العنقودي المعتمدة في ذلك على تحليل المستخلصات والكلمات الدالة المعبرة عن كل وثيقة، وبعد الخروج بهذه العناقيد وما تشتمل عليه من مصادر، تستخدم خوارزميات الارتباط للخروج بالعلاقات بين المصادر داخل العنقود الواحد( أما من خلال ملفات الولوج أو تحليل موضوعاتها) للخروج بشكل شجري عن علاقة المصادر ببعضها داخل العنقود الواحد الذي يرتبط هو الآخر بعناقيد أخرى. مما يسمح للمستخدم بتصفحها واستعراض المصادر ذات الصلة ببعضها البعض.<sup>27</sup>

\*تطوير خدمات المعلومات: حيث أتاحت هذه التقنية على إعطاء قيمة مضافة لخدمات

المعلومات من خلال تطبيق خوارزمياتها عليها حيث ساهمت في:

1. الرفع من كفاءة استخدام موقع الإلكتروني للمكتبة و ما يقدمه من خدمات: فباستخدام خوارزميات تحليل الارتباط و التنقيب التسلسلي يتم فحص مسارات المستفيدين لاكتشاف الطرق المتبعة للوصول لمختلف أنشطة المكتبة المتاحة عبر الموقع و يمكن من خلال ذلك الكشف عن أكثر الروابط و المسارات استخداما وأي منها يؤدي إلى بعضها البعض، والوصلات غير المفعلة أو التي لا يوجد إقبال عليها على الاطلاق، و هذا يعطي للمكتبة أفكارا واضحة عن أكثر المناطق حاجة داخل الموقع لتحسينها و أكثرها حيوية لنشر الرسائل المهمة مثلا أو وضع الإرشادات التوضيحية و خاصة في مواقع المكتبات الرقمية العلمية و المهنية، و بالتالي تساعد الأنماط المعرفية التنبؤية المكتشفة من مثل هذه التحليلات في تقديم المعلومات التي يرغبها المستفيدون و منه ضمان بقائهم فترة أطول داخل الموقع .

كما تساعد هذه التحليلات في فحص الأنماط المتكررة لمسارات وصول المستفيدين للمصادر الإلكترونية المتاحة عبر مواقع ناشريها أو مورديها للكشف عن المصادر غير المتاحة أو التي يصعب الوصول إليها وهذا من شأنه أن يساعد في إيجاد الحلول السريعة بدلا من تلقي الشكاوى من المستفيدين.<sup>28</sup>

2. كما مكنت خوارزميات " تحليل الارتباط " من:

\*تقديم خدمات استشارية للقراء من خلال دراسة معاملات الاستعارة السابقة لتحديد الأعمال ذات الصلة، و من ثم إتاحة هذه المعلومات من خلال الفهرس المتاح على الخط المباشر للسماح للمستفيدين برؤية الأعمال المشابهة لما تم اختياره من أعمال بناء على تاريخ تداول معاملات الاستعارة.

أو بناء ملفات شخصية لكل مستفيد استنادا إلى تاريخ تداول الاستعارة الخاصة به، مما يسهل إحاطته بكل ماهو جديد في ما يخص اهتماماته و تقديم مقترحات بالمواد و الموضوعات ذات الصلة بنتيجة تحليل الاستعارات مع المستفيدين الآخرين المشتركين معه في نفس مجال الاهتمام مما يساعد على تطوير ملفاتهم الشخصية دون تدخل من المستفيد نفسه، وقد يكون هذا الأمر مشكوكا فيه من الناحية الأخلاقية و القانونية دون الحصول على إذن مسبق من المستفيد، ويمكن الاستعاضة عن بناء الملفات الشخصية بتحليل بيانات عدد كبير من المستفيدين مع خذف بياناتهم الشخصية مما يمكن من الخروج بنتائج مماثلة.

\*تحسين كفاءة الفهرس المتاح على الخط المباشر: فمن خلال فحص ومتابعة إجراءات واستراتيجيات المستفيدين في تعاملاتهم مع الفهرس على مدار فترات زمنية مختلفة وما أسفر عنه ذلك من نتائج للخروج بمؤشرات حقيقية عن جودة أداء الفهرس في تلبية احتياجات المستفيدين، وبالتالي تحديد المناطق التي بها المشكلات وهذا يساعد المسؤولين على اتخاذ قرارات واعية حول تطوير أو تغيير واجهات نظام المكتبة الرقمية أو تحسينات في النظام عموماً. علاوة على الخروج بنتائج عامة عن أوجه القصور والنقص داخل المجموعات خاصة عندما تسفر نتائج البحث عن عدم وجود ما يطابق استراتيجية البحث.

وساعد استخدام أساليب " تحليل الانحراف " في التوقع والتنبؤ باتجاه الاختلاف والنتائج المتطرفة من استخدامات مصادر المكتبة من خلال تحليل واقع قاعدة بيانات الاستعارة لفترات زمنية محددة.

4. أما استخدام "تحليلات السلاسل الزمنية" فإنه يتيح الكشف عن التغيرات لقيم معينة على مدار الزمن كالكشف عن بعض الاستخدامات غير الصحيحة والخاطئة داخل المكتبة، أو نماذج الاحتمال من جانب المستفيدين خاصة دائمي التأخير في إعادة المواد المستعارة، مما يفتح الباب أمام أخصائي المكتبة ومتخذي القرار في إجراءات الفحص عن العوامل المؤثرة لحدوث مثل هذه الانحرافات.<sup>29</sup>

\*تنمية الموارد البشرية: حيث تمكن من تشخيص متكامل لسلوكيات الموظفين الذين يقدمون أداءً متفوقاً وطرق تحسين أدائهم وكذلك الطرق الأفضل التي تمكن من تكوين فرق عمل ناجحة في المكتبة الرقمية. ويمكن توضيح ذلك من خلال المثالين التاليين:

\* إجراءات التوريد: يتم التعامل مع الناشرين وموردين على المستوى المحلي أو العالمي، ويسجل النظام الفرعي للتوريد مثل هذه المعاملات بين المورد والموظف المختص، وتحليل هذه الأعمال والوقت المستغرق في أدائها تتضح الصورة أمام المديرين ومتخذي القرار حول سلوك الموظف في إنجاز الأعمال المنوطة بها ومدى درجة كفاءته في متابعة أوامر التوريد خاصة عند التعامل مع موردي قواعد البيانات والدوريات الإلكترونية والتي يتم فيها كثير من التفاوض حول الأسعار والخدمات المقدمة وكيفية سداد الاشتراكات وما إلى ذلك، وما ينتج عن ذلك من مراسلات وأوامر التوريد.

\* خدمات المعلومات: تستطيع النظم الآلية التي تسجل معاملات الاستعارة وكذلك قواعد البيانات الاستفسارات المرجعية أن تفرز أنماطاً معرفية للقياس بشكل كمي لدرجة التحميل

والضغط لمثل هذه الخدمات، مما قد يسلط الضوء بشكل واضح على إعادة توزيع موظفي المكتبة لإجراء خدمات معينة.

ولا تعد النماذج التنبؤية المقدمة من خلال هذه تنقيب البيانات الضخمة بديلا عن الطرق التقليدية لقياس أداء وكفاءة موظفي المكتبة، ولكنها تساعد في تقديم مبررات كمية ونوعية لتقسيم وتوزيع العمل داخل مناطق معينة بالمكتبة وتقديم طرق الدعم والمساعدة الحديثة والمستمرة للأخصائيين لزيادة قدراتهم ومهاراتهم.<sup>30</sup>

**\*الدعم والعلاقات الخارجية:** غالبا ما تكون معظم المكتبات الرقمية تابعة لمؤسسات أكبر لذلك لا بد من وجود قدر من التفاعل الجيد والمستمر بين المكتبة والجهة التابعة لها، ويسهل تحليل البيانات الضخمة في المكتبات الرقمية من تكوين الأنماط والنماذج التنبؤية نظرا لإتاحة جميع الأنشطة من خلال الويب، حيث تختص هذه المكتبات بقدرتها على زيادة سبيل التفاعل والتواصل مع المؤسسة الأم التابعة لها من خلال التنقيب داخل الاستشهادات المرجعة والاقبسات بالبحوث العلمية المنشورة من خلال موقعها الإلكتروني، وبتحليل مثل هذه البيانات تستطيع المكتبة قياس معاملات التأثير لأهمية البحوث ضمن مجال التخصصات البحثية دون غيرها من البحوث ومثل هذه الإمكانيات والقدرات تكون قيمة خاصة للمنظمات الكبيرة حيث يتم توزيع جهود البحث والتطوير في مواقع متعددة، وبالتالي يمكن الخروج ببعض التقارير التي تفيد متخذي القرار والمسؤولون بالمنظمة عن: أهمية هؤلاء الباحثين وموضوعاتهم المتناولة وخلق مجتمع من الباحثين لهم اهتمامات واتجاهات محددة هذا من شأنه أن يفيد في تطوير المؤسسة.

وعلاوة على ذلك تنظر المؤسسات التي تمول المكتبات الرقمية إلى آلية تنقيب النصوص على أنها وسيلة أفضل وأوفر ماديا لاسترجاع مصادر المعلومات بطريقة سهلة وأكثر دقة بدلا من الحلول المتعبة حليا كالفهرسة والميتاداتا، وذلك لأن هذه الآلية في تطور وتحسين مستمر، ومن خلال استخدامها يتم نقل المكتبة الرقمية من مجرد اقتراح النصوص التي قد تشتمل على الأجوبة إلى تقديم الأجوبة مباشرة من النص أو النصوص المناسبة، ويجدر الإشارة هنا إلى أن استخدام مثل هذه الأدوات يقلل من مخاطر استرجاع النصوص خارج نطاق السياق الصحيح، علاوة على تقديمها لبعض الملاحظات المهمة عن نوعية وجودة مصادر المعلومات المتوفرة بالمكتبة، لذلك من الأفضل ربط ما تم استخراج بروابط مباشر للنص نفسه حتى يظهر النص جنبا إلى جنب مع الأجوبة .

ومثل هذه التقنيات تقدم مكتسبا كبيرا إلى المؤسسات التي تشتمل على كميات هائلة من النصوص التقنية والفنية، حيث من شأنه تعزيز الوصول السريع والشامل للمواد المتفرقة وغير المفهرسة.

**\*تطوير تقارير النظم الآلية المتكاملة للمكتبات:** تستطيع تقنية النقيب عن البيانات الضخمة الإجابة عن مجموعة من التساؤلات التي لا تستطيع آليات التحليل التقليدية والخبرة البشرية الإجابة عنها، ويجب أن نضع في الاعتبار هنا شيئا مهما جدا وهو أن تقنية النقيب عن البيانات تنفذ داخل مستودعات البيانات، بينما النظم الآلية للمكتبات قائمة في الأساس على قواعد بيانات علائقية صممت من أجل رفع كفاءة معالجة مجموعة من المعاملات مثل (الفهرسة، التزويد، الاستعارة، خدمات البحث والاسترجاع... إلخ) وبالتالي فالغرض الوظيفي لكلا النظامين مختلف ولكن على الجانب الآخر ساعدت هذه التقنية على تسهيل الحصول على الأنماط والنماذج المعرفية وتجنب التعثر في أداء المعاملات المختلفة داخل النظم الآلية.<sup>31</sup>

**\*دور النقيب عن البيانات الضخمة في خدمة هيكل ومحتوى وخدمات المكتبات الرقمية:**

يمكن تقسيم تطبيقات نقيب البيانات الضخمة بالمكتبات الرقمية إلى ثلاثة عناصر رئيسية يتفرع منها أحد عشر عنصرا فرعيا وهي كالآتي:

1. هيكل المكتبة: ويضم ثلاثة عناصر هي (الموقع و الروابط التشعبية و الصفحات) ويستخدم النقيب عن البيانات الضخمة لهيكل المكتبة لاكتشاف المعرفة من هذا الهيكل التنظيمي و الروابط للصفحات، حيث ينقب عن الروابط التشعبية ( الفائقة) لصفحات المكتبة ووثائق الهيكل الداخلي لها علاوة على العناوين المباشرة لهذه الصفحات، لاستخلاص المعرفة من هذا الهيكل التنظيمي وروابطه، لتقديم معلومات جديدة و مفيدة ليست فقط على محتوى الوثائق ولكن بالربط مع صفحات أخرى، والهدف من ذلك اكتشاف و تطوير هيكل و نمط جديد لصفحات المكتبة لإعداد نموذج مفيد مبني على أساليب التصنيف و العنقدة و تحليل الارتباط بين الروابط و الصفحات ذات الصلة، لرفع جودة موقع المكتبة على الأنترنت و تحسين أساليب استرجاع المعلومات.

2. محتوى المكتبة: ويشتمل على ثلاثة عناصر أيضا هي (ملفات الاستخدام والبيانات النصية والميتاداتا) ويقصد بها استخراج المعلومات من وثائق المكتبات الرقمية التي غالبا ما تتاح على الويب وتمر عملية النقيب عن البيانات الكبيرة بالمكتبات الرقمية بالمراحل التالية:

\*تنظيم البيانات، \*الاستخراج التلقائي للبيانات، \*وصف خصائصها وسماتها، \*بعدها الاستخلاص، \*التصنيف التلقائي للوثائق، \*جمع وتنظيم المعلومات الموضوعية تلقائيا.

خدمات المكتبة: تضم خمسة عناصر هي (البحث والاسترجاع والتصفح ومقترحات القراءة وخدمات متخصصة لفئات معينة من المستخدمين) حيث من الممكن أن تقدم كثيرا من التطورات والمقترحات لخدمات المعلومات بالمكتبة كما سبق وذكرنا.<sup>32</sup>

4.2. تجارب التنقيب عن البيانات الضخمة في مجال المكتبات الرقمية: نستعرض فيما يلي تجربتين لاستخدام المكتبات الرقمية لتقنية التنقيب عن البيانات أحدهما أجنبية والأخرى عربية.

#### 1.4.2. تجربة اتحاد المكتبات الصربية (KOBSON):

أولا: التعريف بالمشروع: هي اتحاد للمكتبات الصربية، وتعد أكبر مكتبة رقمية تقدم مصادر التعلم الرقمية ب (صربيا) للطلاب و الباحثين و الأساتذة الصربيين، و تمكنهم أيضا من الوصول للدراسات الأجنبية من خلال قواعد بيانات النصوص الكاملة مثل: Science Direct , Springer , Black Well, Proquest ,Jstor.

ثانيا: هدفها: مساعدة المستخدمين الجدد من هذه المكتبة وكذلك الذين لديهم مشكلة في العثور على المصادر ذات الصلة من خلال تقديم حلول بالمقترحات من المصادر التي ربما تفيدهم وذلك بالتشابه مع مستفيدين آخرين لهم نفس الاهتمامات، معتمدين في ذلك على أساليب التنقيب عن البيانات وخاصة العنقدة والتصنيف.

ثالثا: المنهجية والتصميم: قامت منهجية هذا المشروع على تصميم نظم يسمى (REKOB)، ويمكن توضيح الخطوات التدريجية لعمل النظام كالآتي:

1. تجميع وتحليل بيانات البحث السابقة للمستخدمين وملفاتهم الشخصية، فمن خلال ملفات الاستخدام يستخرج المستخدمون الذين قاموا بعملية التحليل مرة واحدة على الأقل من المكتبة الرقمية. ونظرا لأنه تم الاستعانة ببيانات حقيقية عن المستخدمين كالرقم المعرف، والاسم والعنوان والقسم والنوع والبريد الإلكتروني، والبيانات الديمغرافية (السكن) وما إلى ذلك، ولحماية خصوصية المستخدمين تتم بعض التعديلات لعملية تنقيب البيانات حيث تستبدل الأسماء والمؤسسات برموز وبالتالي يكون الأشخاص المخولون فقط هم من يكونون على بيئة هوية المستخدمين وأقسامهم وكلياتهم.

يتم إعداد البيانات من خلال جدولين أحدهما للمستخدمين والآخر للخدمات، ويضم جدول المستخدمين بياناتهم الشخصية أما جدول الخدمات فيضم بيانات عن وقت وتاريخ الدخول للمكتبة الرقمية والملفات التي تم تحميلها، ثم يتم ربط جدول الخدمات والمستخدمين في جدول



جديد من خلال اسم المستفيد والخدمة المستفاد منها، ثم يحفظ كل ذلك داخل قاعدة بيانات Oracle لإدارة البيانات التي يتم تحويلها وتطبيعها لتناسب مع عملية التنقيب. 3. تبدأ عملية التنقيب عن البيانات باستخدام برنامج (Oracle DataMiner 10.2) وتحدد السمات والخصائص لعنقدة المستفيدين باستخدام خوارزمية (K-means) فعلى سبيل المثال تقسيم المستفيدين وفق الفئة العمرية في شكل بياني بالمستطيلات وهو عبارة عن مستطيلات متصلة تمثل كيفية التوزيع المتكرر للبيانات حيث تم تحليل عنصر العمر إلى خمسة مستطيلات ويمثل كل مستطيل فئة عمرية.

4. التصنيف من خلال تقسيم البيانات إلى مجموعتين، المجموعة المدبرة لاكتشاف أنماط المستفيدين ثم تجري عملية التقييم من خلال المجموعة المختبرة، والنتائج التي تم الحصول عليها من هذه المجموعة هي مؤشر لأداء أنماط المستفيدين الجدد، ويعني هذا أنه تم عنقدة المستفيدين وفق مجموعة من السمات والخصائص المختلفة، ثم يتم تصنيف هذه المجموعات بناء على استخراج القواعد من خلال خوارزمية (Naive Bayes) لاختبار هذه النتائج حيث أستخدم أسلوب الدعم (S) والثقة (C).

5. ولتصنيف المستفيدين الجدد، نجد أنه بمجرد إنشاء المجموعات/العناقيد ب (أنماط المستفيد) يتم مضاهاة المستفيد الجديد بالعنقود /المجموعة الأكثر تشابها لخصائصه، وأخيرا تنشأ توصية أو مقترح مناسب.

2.4.2. مشروع التنقيب عن البيانات في قاعدة بيانات الرسائل الجامعية بجامعة القاهرة:  
أولا: التعريف بالمشروع: هو مشروع للتنقيب المعلوماتي لرسائل الماجستير والدكتوراه في كليات الجامعة بجامعة القاهرة بهدف لتحليل بيانات الرسائل العلمية في القطاعات العلمية المختلفة بغية التعرف على الاتجاهات البحثية في الجامعات وتوجيه بحوث الرسائل العلمية نحو خدمة المجتمع وإتاحة التعاون البحثي بين القطاعات الأكاديمية المختلفة و الباحثين وتنمية التعاون الدولي في البحوث العلمية من خلال إتاحة ملخصات بحوث الرسائل العلمية بكليات الجامعة وبياناتها على شبكة المعلومات الدولية وهذا ضمن الخطة الاستراتيجية للجامعة حتى عام 2015. وقد كان التركيز في بداية المشروع على الرسائل العلمية باللغة الإنجليزية ولكن تم إضافة بعض الرسائل العلمية باللغة العربية ابتداء من عام 2007. ومن المنظر أن يساهم هذا المشروع بشكل كبير في دعم الخطة البحثية للجامعة وتطوير الدراسات العليا وتحديد صورة متكاملة حول الاتجاهات البحثية في المرحلة القادمة، علاوة على مساعدة الباحث في العثور على معلومات مفيدة عن الأطروحات على مستويات متعددة بما في ذلك المحتوى الموضوعي، المشرفون،

الباحثون، وعلى المستوى التنظيمي (الجامعة والقطاعات والكليات والأقسام) كما يساعد الباحث في اكتشاف بحوث مماثلة لاهتماماته.

ثانياً: المنهجية و التصميم: يستخدم المشروع خوارزميتين أساسيتين لتنقيب البيانات: الأولى مخصصة للبحث بالكلمات الدالة أو البحث المباشر والمحدد، والثانية خوارزمية شجرة الأقسام كما ويتبع أسلوبين من أساليب التنقيب وهما: تصنيف الرسائل العلمية حيث قسمت الرسائل العلمية وفق تصنيفات معينة تندرج تحت (الكلية، القطاع، القسم، الموضوع)، وعنقدة الرسائل العلمية حيث تم تقسيم الرسائل العلمية إلى مجموعة من العناقيد مستمدة في ذلك على خوارزميات التشابه والتي اعتمدت على التشابه في (العنوان، أو العنوان والمستخلص، أو العنوان والكلمات الدالة).

بالإضافة إلى استخدام تقنية التنقيب في النصوص، من خلال برنامج مفتوح المصدر، ولكن لم يتم الإشارة إلى أسماء خوارزميات العنقدة أو التصنيف التي تم استخدامها من الرسائل العلمية وقد تجاوز المليون كلمة وهذا القاموس تم الاستعانة به في كل عمليات التنقيب عن البيانات، وفي جميع مراحل ومستويات البحث مع الأخذ في الاعتبار استبعاد كلمات الوقف.<sup>33</sup>

## خاتمة

في الأخير يمكن القول أن تطبيق تقنية تحليل البيانات الضخمة يعد مطلباً وخياراً استراتيجياً وحتمية للمكتبات الرقمية خصوصاً وأنها تشهد غرقاً معرفياً رقمياً رهيباً والسبيل الأنسب للتحكم فيه هو بتوجه بؤر اهتماماتها نحو الاستثمار في الحلول التكنولوجية التي تتحها هذه التقنية، حيث يمكن استخدام بعض خوارزميات تنقيب البيانات كتحليل الارتباط، التصنيف، العنقدة وتحليل السلاسل الزمنية في استنباط رؤى ترشد مسيرتها أثناء اتخاذ قراراتهم حول تحسين وظائفها، وأنشطتها وتطوير خدماتها الرقمية بكل صورها وأشكالها وكذا ضمان تجديد وإتاحة وحفظ مصادرها، فهذه الأساليب تتيح دراسات الاستخدامية للمستخدمين وتحليل المجموعات الرقمية والدراسات المقارنة متعددة التخصصات هذا من شأنه أن يطور في أساليب البحث والاسترجاع واستخراج المعلومات والعلاقات بينهما للخروج بمختلف التوقعات والأنماط المعرفية غير المعروفة أو المتوقعة مسبقاً عن جميع الخطوات المكتبة الرقمية مما يوفر لها قيمة مضافة في الوقت الحقيقي.

إذن فتطبيق آليات تنقيب البيانات الضخمة في المكتبات الرقمية يساعد على استرجاع وتحليل معلومات الماضي الخاصة بجميع خطوات المكتبة والتي على أساسها يمكن بناء معلومات تفيد في صياغة قرارات حول سياسات واستراتيجيات حاضر المكتبة وكذا رصد والتنبؤ بمختلف الخطوات المستقبلية لها. كل ذلك يعزز مكانتها في محيطها وبين منافسيها، ويضمن رضى قرائها ومستفيديها.

### الاقتراحات:

- في الأخير يمكن عرض جملة من الاقتراحات التي تساعد على الاستثمار في تقنية التنقيب عن البيانات الضخمة في المكتبات الرقمية ونوجزها فيما يلي:
- صياغة خطة عمل فعالة لإنشاء واستخدام تقنية تنقيب البيانات الضخمة في مواقع ويب المكتبات الرقمية ودعمها ماديا ومعنويا.
- ضرورة التنويه ونشر الوعي لدى جميع مشرفي ومسيري المكتبات الرقمية بفوائد البيانات الضخمة وأدواتها من خلال عقد وتنظيم لقاءات ومؤتمرات دورية.
- تغيير أنماط التعليم والتكوين الأكاديمي في أقسام علم المكتبات والمعلومات بالجامعات العربية واعطاء الاهتمام الكافي لعلم البيانات وعلم التنقيب عن البيانات.
- الاستفادة من تجارب الآخرين في المؤسسات المماثلة أثناء التوجه لتجسيد تطبيقات تحليل البيانات الضخمة في المكتبات الرقمية على أرض الواقع.
- ضرورة توفير البنية التحتية في المكتبات الرقمية بما يواكب التطورات الحاصلة في مجال التكنولوجيا ويسهل عليها تبني مشاريع تنقيب البيانات الضخمة مع زيادة تدفق الانترنت فيها.

### قائمة المصادر والمراجع

1. البار، عدنان مصطفى. البيانات الضخمة ومجالات تطبيقها. [2018/04/10]. ص.02. متاح على الخط المباشر على الرابط: <https://www.kau.edu.sa/GetFile.aspx?id...fn...Adnan-Albar...>
2. البيانات الضخمة بالعربية = Big Data in Arabic: البيانات الضخمة الفرص والتحديات. [2018/03/25] متاح على الخط المباشر على الرابط: <https://bigdatainarabic.wordpress.com/2014/03/18/bigdatachancesandchallenges>
3. البيانات الضخمة ما أهميتها وما أهمية الاستثمار في تحليلها وكيف ستؤثر في حياتنا وقراراتنا؟ [2018/03/25]. متاح على الخط المباشر على الرابط: <http://www.urecten.com>
4. بوعناقة، سعاد. البيانات الضخمة في قطاع المكتبات: نقاط القوة والضعف الفرص والتحديات. بحث مقدم إلى المؤتمر 24 لجمعية المكتبات المتخصصة، مسقط، 6-8 مارس 2018.
5. الحمادي، علاء حسين. تنقيب البيانات = DATA MINING. عمان: إثراء، 2008.

## تنقيب البيانات الضخمة في المكتبات الرقمية: دراسة في المفاهيم وأليات التطبيق

-أسماء قرزیز-

6. حمدون، أم كلثوم صباحي محمد (وآخرون). استخدام تقنية تنقيب البيانات الضخمة في أمراض السرطان: بالتطبيق على مركز الخرطوم للعلاج بالأشعة الطب النووي. [2018/04/10]. 2001. متاح على الخط المباشر على الرابط: <https://2u.pw/byan7>
7. درويش، وسام محمود أحمد، تق. زايد، يسرية محمد عبد الحلیم. التنقيب عن البيانات في مجال المكتبات والمعلومات: المفاهيم، الأساليب، التطبيقات. القاهرة: الدار المصرية اللبنانية، 2013.
8. السالحي، جمال بن مطر. 2018. البيانات الضخمة ودورها في دعم اتخاذ القرارات والتخطيط الاستراتيجي. بحث مقدم إلى المؤتمر 24 لجمعية المكتبات المتخصصة، مسقط، 6-8 مارس 2018.
9. درويش، وسام محمود أحمد. استخدام تقنية التنقيب عن البيانات في تطوير المكتبات الرقمية العربية: دراسة تجريبية. الاتجاهات الحديثة في المكتبات والمعلومات. [2018/04/10]. 2013. مج 20، ع 39. متاح على الخط المباشر على الرابط: <https://platform.almanhal.com/Files/2/52016>
10. سعيد، منال حمدان. البيانات الضخمة في المكتبات الأكاديمية في سلطنة عمان: الواقع والتحديات. بحث مقدم إلى المؤتمر 24 لجمعية المكتبات المتخصصة، مسقط، 6-8 مارس 2018.
11. الشوابكة، يونس أحمد اسماعيل. الوعي بمفهوم البيانات الضخمة (Big data) لدى العاملين في المكتبات الأكاديمية: دراسة حالة المكتبة الجامعية الأردنية. بحث مقدم إلى المؤتمر 24 لجمعية المكتبات المتخصصة، مسقط، 6-8 مارس 2018.
12. فارس، شاشة. استخدام تقنيات التنقيب عن البيانات في الشبكات الاجتماعية. الملتقى الوطني حول شبكات التواصل الاجتماعي في الوسط الأكاديمي: فضاءات متطورة لإثراء العملية التعليمية التفاعلي والتشاركي. الجزائر، 04 فيفري 2014.
13. العاني، مزهر شعبان ( و آخرون). ذكاء الأعمال وتكنولوجيا المعلومات = Business Intelligence Information Technology. عمان: دار صفاء، 2012.
14. عثمان، سلوم. ماهي البيانات الضخمة (big data)؟ وما هو أسباب ضخامتها. [2018/04/10]. فيديو مدته 03:12 د. متاح على الخط المباشر على الرابط: <https://www.youtube.com/watch?v=So6rAvwwVKg>

### باللغة الأجنبية

15. Yu, Ping. 2011. Data mining in library reader management. In international conference on network computing and information security.
16. Neelamadhab, Padhy) and others ). The Survey of Data Mining Applications and Feature Scope. [14 /04/2018]. International Journal of Computer Science, Engineering and Information Technology (IJCSSEIT). 2012. Vol.2, No3.p.p. [en ligne] : <https://2u.pw/v4YFA>

### هوامش الدراسة

<sup>1</sup> الشوابكة، يونس أحمد اسماعيل. الوعي بمفهوم البيانات الضخمة (Big data) لدى العاملين في المكتبات الأكاديمية: دراسة حالة المكتبة الجامعية الأردنية. بحث مقدم إلى المؤتمر 24 لجمعية المكتبات المتخصصة، مسقط، 6-8 مارس 2018.

- <sup>2</sup> بوغناقة، سعاد. البيانات الضخمة في قطاع المكتبات: نقاط القوة والضعف والفرص والتحديات. بحث مقدم إلى المؤتمر 24 لجمعية المكتبات المتخصصة، مسقط، 6-8 مارس 2018.
- <sup>3</sup> سعيد، منال حمدان. البيانات الضخمة في المكتبات الأكاديمية في سلطنة عمان: الواقع والتحديات. بحث مقدم إلى المؤتمر 24 لجمعية المكتبات المتخصصة، مسقط، 6-8 مارس 2018.
- <sup>4</sup> Yu, Ping. 2011. Data mining in library reader management. In international conference on network computing and information security.
- <sup>5</sup> البيانات الضخمة ما أهميتها وما أهمية الاستثمار في تحليلها وكيف ستؤثر في حياتنا وقراراتنا؟ [2018/03/25]. متاح على الخط المباشر على الرابط: <http://www.urecten.com>
- <sup>6</sup> بوغناقة، سعاد. المرجع نفسه.
- <sup>7</sup> لبيانات الضخمة بالعربية = Big Data in Arabic: البيانات الضخمة الفرص والتحديات. [2018/03/25] متاح على الخط المباشر على الرابط: <https://bigdatainarabic.wordpress.com/2014/03/18/bigdatachancesandchallenges>
- <sup>8</sup> البار، عدنان مصطفى. البيانات الضخمة ومجالات تطبيقها. [2018/04/10]. ص. 02. متاح على الخط المباشر على الرابط: <https://www.kau.edu.sa/GetFile.aspx?id...fn...Adnan-Albar...>
- <sup>9</sup> السالحي، جمال بن مطر. 2018. البيانات الضخمة ودورها في دعم اتخاذ القرارات والتخطيط الاستراتيجي. بحث مقدم إلى المؤتمر 24 لجمعية المكتبات المتخصصة، مسقط، 6-8 مارس 2018.
- <sup>10</sup> عثمان، سلوم. ماهي البيانات الضخمة (big data)؟ وما هو أسباب ضخامتها. [2018/04/10]. فيديو مدته 03:12 د. متاح على الخط المباشر على الرابط: <https://www.youtube.com/watch?v=So6rAvwwVKg>
- <sup>11</sup> بوغناقة، سعاد. المرجع نفسه.
- <sup>12</sup> فارس، شاشة. استخدام تقنيات التنقيب عن البيانات في الشبكات الاجتماعية. الملتقى الوطني حول شبكات التواصل الاجتماعي في الوسط الأكاديمي: فضاءات متطورة لإثراء العملية التعليمية التفاعلي والتشاركي. الجزائر، 04 فيفري 2014.
- <sup>13</sup> حمدون، أم كلثوم صباحي محمد (وآخرون). استخدام تقنية تنقيب البيانات الضخمة في أمراض السرطان: بالتطبيق على مركز الخرطوم للعلاج بالأشعة الطب النووي. [2018/04/10]. متاح على الخط المباشر على الرابط: <https://2u.pw/byan7>
- <sup>14</sup> فارس، شاشة. المرجع نفسه.
- <sup>15</sup> الحمامي، علاء حسين. تنقيب البيانات = DATA MINING. عمان: إثراء، 2008. ص. 41.
- <sup>16</sup> فارس، شاشة. المرجع نفسه.
- <sup>17</sup> العاني، مزهر شعبان ( و آخرون). ذكاء الأعمال وتكنولوجيا المعلومات = Business Intelligence Information Technology. عمان: دارصفاء، 2012. ص. 65.
- <sup>18</sup> الحمامي، علاء حسين. المرجع نفسه. ص. 41-42.
- <sup>19</sup> فارس، شاشة. المرجع نفسه.
- <sup>20</sup> العاني، مزهر شعبان ( و آخرون). المرجع نفسه. ص. 65.
- <sup>21</sup> الحمامي، علاء حسين. المرجع نفسه. ص. 199-200.
- <sup>22</sup> فارس، شاشة. المرجع نفسه.
- <sup>23</sup> درويش، وسام محمود أحمد. استخدام تقنية التنقيب عن البيانات في تطوير المكتبات الرقمية العربية: دراسة تجريبية. الاتجاهات الحديثة في المكتبات والمعلومات. [2018/04/10]. 2013. مج 20، ع 39. متاح على الخط المباشر على الرابط: <https://platform.almanhal.com/Files/2/52016>

1. <sup>24</sup> Neelamadhab, Padhy ( And others ).The Survey of Data Mining Applications And Feature Scope. [14 /04/2018]. International Journal of Computer Science, Engineering and Information Technology (IJCSIT).2012. Vol.2, No3.p.p 02-16.[en ligne] : <https://2u.pw/v4YFA>

<sup>25</sup> یونس أحمد اسماعیل. المرجع نفسه.

<sup>26</sup> درویش، وسام محمود أحمد، تق. زاید، یسریة محمد عبد الحلیم. التنقیب عن البیانات فی مجال المكتبات والمعلومات:

المفاهیم، الأسالیب، التطبیقات. القاهرة: الدار المصریة اللبنانیة، 2013. ص.ص. 163-162-161.

<sup>27</sup> المرجع نفسه. ص.ص. 167-166.

<sup>28</sup> المرجع نفسه. ص.ص. 165-164.

<sup>29</sup> المرجع نفسه. ص. ص. 168-167.

<sup>30</sup> المرجع نفسه. ص 170. ص. 171. ص. 172.

<sup>31</sup> المرجع نفسه. ص. 181. ص. 182.

<sup>32</sup> المرجع نفسه. ص 215. ص 2.

<sup>33</sup> المرجع نفسه. ص 240. ص 243.