استخدام تقنية التنقيب عن البيانات فى تطوير المكتبات الرقمية العربية : دراسة تجريبية(*)

د. وسام محمود أحمد درويش
أخصائى تقنية المعلومات
المكتبة المركزية – جامعة حلوان
wessamdarwish@yahoo.com

تمهيد:

أصبح لدى الكثير من الهيئات العربية محتوى رقمي عربي من تقرير وإحصائيات ونشرات وبحوث ودراسات إلى غير ذلك من المواد العلمية، وقد يكون للمؤسسة الواحدة قاعدة بيانات أو أكثر؛ وربما يصل حجم هذه القواعد في بعض الأحيان إلى التيرابايت أي أكثر من ١,٠٠٠,٠٠٠,٠٠٠,٠٠٠ وداخل هذا الكم من البيانات ترقد الكثير من المعلومات المهمة الحقيقية التي لا يمكن العثور عليها وسط هذا الخضم من البيانات.

إضافة إلى ذلك، نجحت نظم إدارة قواعد البيانات Databases Management Systems (DBMS) في تنظيم المعلومات والتعامل معها داخل المؤسسات كافة ومن بينها المكتبات، إلا إنه مع الزيادة المستمرة والسريعة للمعلومات لم تعد هذه النظم كافية لعملية اتخاذ القرار وهذا هو التحدي الحقيقي، فإذا نظرنا إلى المعلومات التي يتم جمعها نجدها توجد على أنواع وأشكال وفئات موضوعية مختلفة ما بين معلومات تتعلق بالمعاملات التجارية، ومعلومات طبية وهندسية...إلخ، وتقارير نصية، ومواد سمعبصرية وبرامج على اختلاف أنواعها، إضافة إلى الصفحات والمواقع الإلكترونية المنتشرة على شبكة الإنترنت.

من هنا اهتم الباحثون المتخصصون في تجميع وتحليل البيانات ؛ بإيجاد تقنيات وآليات جديدة

(*)وسام محمود أحمد .استخدام تقنية التنقيب عن البيانات في تطوير المكتبات الرقمية العربية: دراسة تجريبية/ إشراف يسرية عبد الحليم زايد – القاهرة، ٢٠١٢. أطروحة دكتوراه. جامعة القاهرة، كلية الآداب. قسم المكتبات والوثائق والمعلومات.

يمكن من خلالها التنقيب داخل هذه البيانات حتى يمكن الحصول على فوائد أفضل كأن يتم تلخيص آلي للبيانات والمعلومات، واستخراج جوهر المعلومات المخزنة، واكتشاف أنماط جديدة من المعلومات لم تكن معروفة من قبل وذلك حتى يتسنى تجميع البيانات المتشعبة من القواعد المختلفة للمؤسسة الواحدة للحصول على أفضل النتائج لبحث واسترجاع المعلومات المفيدة للمستفدين والمؤسسة على حد سواء، كل ذلك ساعد على ظهور **تقنية التنقيب عون البيانوات** (DM) Mining Data لاكتشاف الأنماط المعرفية المختبئة.

والمكتبات بصفة عامة، والمكتبات الرقمية على وجه الخصوص توجد لديها العديد من البيانات، ومما لا شك فيه أن الخبرة المكتبية السابقة والمتراكمة على مدار السنوات وخاصة مهارات التكشيف والاستخلاص والتصنيف والفهرسة الموضوعية؛ تمكن أخصائيي المكتبات والمعلومات من تقدير احتياجات المستفيدين؛ وتقسيم وتنظيم وثائق المنظمة من وجهة نظر احتياجاتهم وتحيد أهم الخدمات التي تتعلق باهتماماتهم، إلا أن الزيادة في البيانات والمعلومات المتاحة عن المقتنيات والمستفيدين المجمعة والمكدسة تقف حائلاً دون مساعدة المستفيدين في الحصول على احتياجاتهم، لذا كان لابد من الاعتماد على آليات وتقنيات حديثة وقوية تستطيع التعامل مع هذا الكم الضخم من البيانات لتمكن أخصائي المكتبات من تقديم أفضل المساعدات للمستفيدين وقد تمثل ذلك في تقنية التنقيب عن البيانات.

وبالفعل كان لعلم المكتبات والمعلومات نصيب وافر في استخدام هذه التقنية الحديثة بداية من عام 1998 فيما يعرف بالتنقيب البيليوجرافي Bibliomining، أو تنقيب بيانات المكتبة Data Mining Library، وقد أفادت هذه التقنية الحديثة نسبي ا التي دخلت مجال المكتبات والمعلومات. بعد دخولها مجالات أخرى كثيرة مثل التسويق والإدارة والتربية ... الخ أفادت في تقديم أنماط جديدة من المعلومات لم تكن مكتشفة من قبل، كما أفادت في التنبؤ بأنماط مستقبلية من المعلومات قد تفيد المستفيد في عمليتي البحث والاسترجاع على حد السواء، إضافة إلى ذلك أفادت في تقديم رؤى جديدة للمجموعات والعمليات المكتبية المختلفة كما أفادت أيضًا في اتخاذ القرارات المتعلقة ببعض العمليات الفنية المطلوبة مثل الحفظ والرقمنة، علاوة على تقديم معلومات قيمة عن سلوك المستفيدين وتفاعلهم مع مجموعات المكتبة وخاصة في ظل البيئة الرقمية.

وقد لوحظ في السنوات الأخيرة الماضية حرص بعض المكتبات الرقمية على تضمين تقنية التنقيب عن البيانات داخل برمجياتها للاستفادة منها في تطوير أدائها على كافة الأوجه.

١- مشكلة الدراسة وأهميتها:

يمكن رصد الحاجة والأهمية لهذه الدراسة والأسباب التي دعت إليها في النقاط التالية:

١- إن كثيرًا من المكتبات ومراقق المعلومات بصفة عامة، والمكتبات الرقمية بصفة خاصة في الخارج قد طورت خدماتها باستخدام

تقنية التنقيب عن البيانات ، ومن أبرز الأمثلة على ذلك جهود (مركز المكتبات المحسب المتاح على الخط المباشر : (Online Computer Library Center المعروف بالحروف الاستهلالية (OCLC) في استخدام هذه التقنيات الحديثة بقواعد بياناتها وخاصة الفهرس العالمي WorldCat، وإعدادها لكثير من الدراسات والبحوث في هذا الصدد؛ لقياس مدى كفاءتها في تطوير مقتنيات المكتبة واتخاذ القرار.

٢- يقابل ذلك عدم الوعي الكافي بأهمية استخدام التنقيب عن البيانات، وإمكانية استفادة المكتبات وخاصة الرقمية منها في الدول العربية بصفة عامة ومصر بصفة خاصة رغم الفوائد الكثيرة التي يمكن تحقيقها من استخدام هذه التقنية.

٣- دراسات سلوك المستفيدين، تعد دراسات سلوك المستفيدين تجاه مجموعات المكتبة خاصة في بيئة الشبكات الرقمية معلومات قيمة لتطويرات مبتكرة في كيفية عرض وإتاحة المعلومات بالشكل الذي قد يرغبه المستفيدين ومن جهة أخرى قد يعزز ذلك قيمة مجموعات المكتبة للمستفيدين، وتمثل الدراسات الاستقصائية، وبيانات الإعارة، ومرات الولوج... إلخ من المصادر التي يمكن الاعتماد عليها في عملية التنقيب عن سلوك المستفيدين.

٤- إدارة مقتنيات المكتبة، فمكن خلال استخدام المكتبة لهذه التقنية فتح باب المشاركة في المصادر مع المكتبات الأخرى، فاستنباط أنماط جديدة لمجموعات المكتبة يوضح الأمر إمام المسئولين للحد من التكرار وسد الفجوات في مقتنيات المكتبة.

٥- الاستفادة من تقنية التنقيب عن البيانات في تحسين آليات البحث والاسترجاع بالمكتبات الرقمية: حيث يمكن استخدام بعض من أساليب تقنيات التنقيب عن البيانات ومستودعات البيانات (DW) Data Warehousing في تطوير آليات البحث والاسترجاع خاصة في المكتبات الرقمية العربية، وقد تمت العديد من الدراسات الأجنبية في هذا الموضوع، حيث تعتمد فكرة تطوير المكتبات الرقمية على استخدام هذه التقنيات لحل مشكلة تكامل المحتوى وذلك من خلال اكتشاف مصطلحات وأنماط معرفية ذات صلة، ووضعها في تصنيف هرمي باستخدام، على سبيل المثال وليس الحصر، تقنية متاجر البيانات Data Marts، وتقنية معالجة التحليل على الخط المباشر On-Line Analytical Processing : (OLAP) .

٦- يعد مجتمع المستفيدين من هذه الدراسة على وجه الخصوص، المجتمع المعلوماتي العربي وخاصًا أمناء المكتبات وأخصائيي المعلومات والمديرين والمسئولين عن اتخاذ القرار بالمكتبات ومراكز المعلومات بصفة عامة والمكتبة الرقمية بصفة خاصة، إلى جانب المهتمين بعمليات التنقيب عن البيانات من الباحثين وغيرهم.

٧- الزيادة الواضحة في حجم المحتوى الرقمي العربي في العقدين الآخرين، ومن ثم يمثل البحث والاسترجاع في هذا الكم الضخم من البيانات صعوبة شديدة وبالتالي تصبح نتائج البحث هي الأخرى غير دقيقة ؛ مما يدعو إلى ضرورة التفكير في استخدام تقنية التنقيب عن البيانات للاستفادة منها في تنظيم وتيسير سبل الاستفادة من هذا المحتوى، وتمثل ذلك في إنشاء مكتبة رقمية عربية تجريبية أسوة بالمكتبات الرقمية بالخارج التي استفادت من تقنية التنقيب عن البيانات في تطوير أدائها. ومنعًا للالتباس أو الوقوع تحت بعض القيود الإدارية والفنية في التعامل مع البيانات المطلوبة لتقنية التنقيب عن البيانات وقع الاختيار على القيام بتجربة بناء مكتبة رقمية عربية من البداية تعتمد آلياتها وأسلوب البحث والاسترجاع والتصفح على هذه التقنية بدلاً من الاستعانة بمكتبات رقمية موجودة بالفعل في مصر؛ حتى يتسنى للباحثة حرية التصرف والقدرة على استخدام أساليب وخوارزميات التنقيب عن البيانات وقياس واختبار النتائج والاستفادة من التغذية المرتدة من فئات المستفيدين من هذه المكتبة الرقمية التجريبية، والمجال الموضوعي لهذه المكتبة التجريبية هو الفنون الجميلة والتشكيلية واقتصر ذلك على الرسائل العلمية المجازة بالمكتبة المركزية لجامعة حلوان، ويرجع هذا الاختيار لأنه المكان الذي تعمل فيه الباحثة حيث يتوفر لديها كل الإمكانيات والتيسيرات من مادة علمية وأجهزة وشبكات الحاسب الآلي والمستفيدين لإجراء هذه التجربة.

٨- ندرة الدراسات العربية التي تناولت استخدام تقنية التنقيب عن البيانات في المكتبات ومراكز المعلومات العربية، كما إنه لا توجد أي دراسات علمية تناولت موضوع تطوير المكتبات الرقمية العربية تحديدًا باستخدام عملية التنقيب عن البيانات.

٢- أهداف الدراسة وتساؤلاتها:

تهدف هذه الدراسة لتحقيق ما يلي:

أولاً: تمييز مصطلح التنقيب عن البيانات عن المصطلحات الأخرى المتداخلة معه مثل التنقيب الببليوجرافي، وتنقيب بيانات المكتبة، والتنقيب في النصوص، وتتبع دخوله في مجال المكتبات والمعلومات.

ثانيًا: التعرف على أساليب وخوارزميات التنقيب عن البيانات المختلفة مع التركيز على أكثرها استخدامًا في مجال المكتبات والمعلومات.

ثالثًا: رصد المزايا والفوائد التي يمكن أن تحصل عليها المكتبات الرقمية العربية من خلال استخدام تقنية التنقيب عن البيانات ، وذلك عن طريق دراسة التجارب الأجنبية في هذا الصدد.

رابعًا: تصميم تجربة لاستخدام تقنية التنقيب عن البيانات في المكتبات الرقمية العربية على وجه الخصوص وذلك من خلال برنامج تجريبي لمكتبة رقمية عربية للرسائل العلمية المجازة في مجال الفنون

الجميلة والتشكيلية المتاحة بالمكتبة المركزية بجامعة حلوان.

وتحاول الدراسة الإجابة عن التساؤلات التالية:

١- ما المقصود بتقنية التنقيب عن البيانات وأساليبها وخوارزمياتها؟

٢- هل يمكن الاستفادة من هذه التقنية في مجال المكتبات والمعلومات، وما يمكن أن تقدمه هذه التقنية من جديد؟

٣- ما التطبيقات الأجنبية المستخدمة لهذه التقنية في مجال المكتبات والمعلومات؟

٤- كيف يمكن الاستفادة من هذه التقنية في تسهيل آليات البحث والاسترجاع من قواعد البيانات الببليوجرافية والنص الكامل؟

٥- كيف يمكن تطوير المكتبات الرقمية باستخدام هذه التقنية؟

٦- ما أهم الصعوبات التي واجهت هذه التقنية عند التعامل مع المحتوى الرقمي العربي؟

٣- حدود الدراسة ومجالها:

- **الحدود الموضوعية:**

تتناول الدراسة موضوع التنقيب عن البيانات من جوانبه المختلفة، مع التركيز على استخدام هذه التقنية في تطوير وتحسين عمليات البحث والاسترجاع والتصفح في المكتبات الرقمية العربية.

- **الحدود الزمنية:**

١- اهتمت الدراسة بموضوع التنقيب عن البيانات في مجال المكتبات والمعلومات منذ بداية دخوله المجال في نهاية التسعينيات من القرن العشرين وتحديدًا عام ١٩٩٨ وحتى وقت الانتهاء من الدراسة عام ٢٠١٢.

٢- أما فيما يتعلق بالدراسة التجريبية، فتم الاعتماد على الرسائل العلمية في الفترة الزمنية من ١٩٧٥ حتى ٢٠١١ وأجري الاختبار على هذا العمل في الفترة ما بين ٢٠١١/٩ حتى ٢٠١٢/٥.

- **الحدود المغوية:**

اقتصرت الدراسة التجريبية على الرسائل العلمية المجازة من جامعة حلوان بمصر **[باللغة العربية فقط]** في مجال الفنون الجميلة والتشكيلية لرصد الصعوبات التي يمكن أن تواجه هذه التقنية عند التعامل مع المحتوى الرقمي العربي.

- **الحدود المكانية:**

نظرًا لأن موضوع الدراسة يهتم باستخدام التنقيب عن البيانات خاصة في تطوير المكتبات الرقمية العربية؛ لذلك فقد شملت الدراسة التجارب الأجنبية دون التقيد بحدود مكانية خاصة أن ميدان البحث هنا (الإنترنت)، أما الدراسة التجريبية فقد تم تحديدها بجامعة حلوان بمصر.

٤- منهج الدراسة وأدواتها:

١/٤ منهج الدراسة:

اعتمدت هذه الدراسة بشكل أساسي على **(المنهج التجريبي)** الذي يعد من بين أبرز

استخداماته في مجال المكتبات والمعلومات اختبار الأساليب والتقنيات الجديدة اللازمة لتطوير المجموعات، أو الخدمات...إلخ بالمكتبات وذلك من خلال التجربة العملية لتطوير المكتبة الرقمية لقياس مدى إمكانية تطبيق خوارزميات التنقيب مع المحتوى الرقمي العربي فقط على وجه التحديد.

٢/٤ أدوات جمع البيانات:

نظرًا للحداثة النسبية لهذا الموضوع، وقلة بل ندرة الدراسات العربية التي تناولت استخدام التنقيب عن البيانات في مجال المكتبات والمعلومات؛ تمت الاستعانة بمجموعة من الأدوات البرمجية وغير البرمجية المختلفة والمتعددة لتجميع المادة العلمية للدراسة؛ بغرض التحقق من مدى كفاية ما تم تغطية من دراسات ووثائق حول هذا الموضوع وتطبيقاته بالخارج حتى تستطيع الباحثة استخدام خوارزميات التنقيب على بيانات الرسائل العلمية بطريقة صحيحة ومنهجية سبق تجريبها لضمان جودة وصحة النتائج، وانقسمت هذه الأدوات على النحو الآتي:

أولاً: الجانب النظري:

ويهتم هذا الشق بجمع كل ما يتعلق بإثراء المفاهيم وأساسيات الفهم والاستيعاب لتقنية التنقيب عن البيانات واستخدامها في مجال المكتبات والمعلومات، ولتحقيق ذلك تم الاعتماد على الأدوات التالية:

(أ) **الدورات التدريبية:** وتمثل السبيل الأساسي الذي اعتمدت عليه الباحثة في استيقاء واستيعاب المفاهيم والمعلومات الأساسية حول موضوع التنقيب عن البيانات ودراسة أساليبها وخوارزمياتها، وقد تم الحصول على دورات تدريبية خاصة ركزت على أساليب (العنقدة - التصنيف تحليل الارتباط السلاسل الزمنية) حيث تعتبر هذه من أهم أساليب التنقيب وشيوعًا واستخدامًا في مجال المكتبات والمعلومات، وتم الاستعانة في ذلك بمجموعة من برامج التنقيب في تطبيق ما تم دراسته؛ وسيتم ذكر هذه البرامج لاحقا في الجانب التطبيقي.

هذا بالإضافة إلى إنه لدراسة التنقيب عن البيانات وأساليبها وخوارزمياتها على وجه الخصوص فكان لابد من الاعتماد على مجموعة من الكتب العلمية المتخصصة في التنقيب عن البيانات وخاصة تلك التي عرضت المهام والأساليب المختلفة وكيفية بناء الخوارزميات بشكل من التفصيل.

(ب) **تجميع الإنتاج الفكري حول موضوع التنقيب عن البيانات واستخداماتها بالمكتبات بشكل عام والمكتبات الرقمية بشكل خاص:** نظرًا للحداثة النسبية لموضوع استخدام المكتبات لتقنية التنقيب عن البيانات فكان لابد من وجود أداة تساعد في توجيه وإرشاد الباحثة في معرفة وتحديد مجالات الموضوع المختلفة ومن ثم أهم الوثائق التي تناولت الموضوع وكذلك التطبيقات الفعلية لهذه التقنية بالمكتبات بصفة عامة والمكتبات الرقمية بصفة خاصة، والاستفادة منها كخبرات سابقة توضح كيفية الاستخدام

الأمثل لهذه التقنية علاوة على إثراء الجانب التجريبي للدراسة لحصر المادة النظرية عن استخدام المكتبات للتنقيب عن البيانات، فقد كانت الركيزة الأساسية هي المقالات العلمية وأعمال المؤتمرات حيث لم يستقر الموضوع بعد في كتب علمية.

ومن الجدير بالذكر في هذا الصدد، أن دراسة سمات الإنتاج الفكري قد أفرزت مجموعة من الدراسات والمقالات القيمة والتطبيقية المكتوبة باللغة الصينية فقط ولكنها تتيح المستخلص والخلاصة وقائمة المصادر باللغة الانجليزية، وبعد الاطلاع على المستخلص والخلاصة؛ وجدت الباحثة إنه من الضروري الاستعانة بهذه البحوث لهذا لجأت إلى ترجمة متن النص عن طريق موقع محرك بحث جوجل Google لترجمة المقالات والبحوث باللغة الصينية إلى اللغة الإنجليزية؛ وذلك بهدف إثراء المحتوى الفكري والمعرفي للباحثة في استخدام خوارزميات التنقيب عن البيانات في مجال المكتبات الرقمية على وجه الخصوص والمكتبات بصفة عامة .كما يجب الإشارة هنا إلى إنه لم يتم الاعتماد على هذه البحوث كدراسات سابقة أو مراجع تم الاستشهاد بها في الدراسة، ولكن تم الاعتماد عليها فقط لزيادة الخبرة الشخصية في هذا الموضوع مًا جاءت بها هذه البحوث من تطبيقات وتحليلات ذات قيمة كبرى.

ثانياً: الجانب التطبيقي:

ويختص هذا الشق بالأدوات العملية التي تم الاستعانة بها لتطبيق ما تم دراسته من أساليب وخوارزميات التنقيب عن البيانات، وقد استخدمت هذه الأدوات في الدراسة التجريبية للخروج بالأنماط المعرفية التي ذ غذيت بها المكتبة الرقمية العربية التجريبية، هذا علاوة على تم الاعتماد علي مختلف برامج التنقيب عن البيانات المشار إليها أدناه ؛ بغرض التنوع حتى يضمن صحة ودقة واتساق البيانات الناتجة مع الواقع الفعلي.

(أ) **برنامج "Microsoft SQL Server 2005"**: ويشتمل على حزمة برمجيات لتسعة من أساليب وخوارزميات التنقيب عن البيانات ويستطيع التعامل مع البيانات باللغة العربية، وقد استخدمته الباحثة في التطبيق العملي للتنقيب عن البيانات أثناء الدراسة النظرية لهذه التقنية، ومن ثم التحقق من مدى كفاءة هذه التقنية وملائمتها مع المحتوى العربي الرقمي من خلال الدراسة التجريبية.

(ب) **برنامج "Xlminer"** في إصدارته الثالثة: أحد البرامج المجانية المتاحة على شبكة الإنترنت ويشتمل على أربعة أساليب للتنقيب عن البيانات، ويعمل من خلال برنامج مايكروسوفت إكسيل Microsoft Excel، وتم استخدامه لتحقيق نفس الغرض السابق.

(ج) **برنامج "Ins-Microsoft Excel 2010 DM Add"**: عبارة عن حزمة برمجية متاحة مجانًا من شركة مايكروسوفت وتتوافق في العمل مع برنامج مايكروسوفت إكسيل Microsoft Excel ولكنه يشترط تحميل برنامجMicrosoft SQL Server 2005، وتم استخدامه لتحقيق نفس الغرض السابق.

(د) **قاعدة بيانات الرسائل العلمية المجازة وقيد الدراسة المتاحة بالمكتبة المركزية جامعة حلوان:** للاعتماد عليها في استخراج بيانات الرسائل الجامعية لاستخدامها داخل برامج التنقيب عن البيانات سواء للدراسة التجريبية أو للدراسة التطبيقية التي أعدتها الباحثة أيضًا لتحليل التسلسل الزمني للرسائل العلمية المتاحة بالمكتبة المركزية جامعة حلوان.

(هـ) **برنامج مكتبة رقمية بلغة Java** : الجزء التجريبي في هذه الدراسة من اجل تجريب استخدام بعض خوارزميات التنقيب لبيانات الرسائل الجامعية المجازة بجامعة حلوان في مجال الفنون الجميلة والتشكيلية باللغة العربية فقط؛ وذلك بغرض اختبار مدى كفاءة وفاعلية وإتقان هذه الخوارزميات عند التعامل مع المحتوى الرقمي العربي من خلال ما أفرزته من أنماط معرفية جديدة عن طريق استخدام كافة البرامج السابق ذكرها على بيانات الرسائل العلمية؛ وقد عمدت الباحثة على استخدام أكثر من برنامج للتنقيب عن البيانات وذلك للآتي:

- التأكد من مدى صحة ودقة واتساق البيانات الناتجة من هذه البرامج عن طريق مقارنة ما تم التوصل إليه من أنماط معرفية مع كل برنامج على حده ومدى درجات الاتفاق والاختلاف بين هذه الأنماط؛ ضمانًا لتغذية المكتبة بالبيانات الصحيحة المتكاملة.
- قياس درجة فاعلية هذه البرامج مع المحتوى الرقمي العربي؛ وتحديد أهم الصعوبات والمشكلات أثناء التعامل مع اللغة العربية داخل هذه البرامج وكيفية التغلب عليها.

وقد تم ترجمة ذلك في صورة برنامج تجريبي لمكتبة رقمية يعرض هذه الأنماط بطريقة مطورة للبحث والاسترجاع لدراسة مدى كفاءة استخدام هذه النتائج داخل المكتبة المركزية جامعة حلوان وكليات التربية الفنية والفنون التطبيقية والفنون الجميلة من خلال اختبارها من قبل الأساتذة والباحثين والطلاب؛ للحصول على رد الفعل بشكل مباشر والتغذية المرتدة حول وجهات نظرهم المختلفة عن هذه المكتبة وما تشتمل عليه من نتائج.

(و) **المقابلات الشخصية المقننة:** أداة اختبارية وقياسية تغطي مجموعة من التساؤلات التي تتداعي تداعيًا منطقيًا من موضوعات عريضة تندرج تحتها موضوعات أكثر تخصصًا ودقة حول البرنامج التجريبي للمكتبة الرقمية؛ وتم إجراء هذه المقابلات مع الباحثين والمتخصصين في مجال الفنون لمعرفة مدى جودة هذا البرنامج وما صحة ما يتيحه من بيانات ودرجات الرضا والثقة الحقيقة لطريقة عرض البيانات وأساليب البحث والاسترجاع والتصفح للموضوعات وعلاقتها مع بعضها، علاوة على اختبار عناوين الرسائل ذات الصلة ببعضها البعض التي نتجت عن خوارزميات التنقيب عن البيانات.

٥ - مراحل الدراسة وخطواتها:

يمكن توضيح المراحل والخطوات التي مرت بها هذه الدراسة على هذا النحو:

١/٥ المرحلة المبدئية:

ينصب موضوع الدراسة في المقام الأول على كيفية الاستفادة من تقنية التنقيب عن البيانات في تطوير المكتبات الرقمية؛ ومن ثم كان لابد من الدراسة النظرية والتجريبية لتقنية التنقيب عن البيانات والتعرف على أساليبها وخوارزمياتها والأسلوب الذي بنيت عليه هذه الأساليب حتى يمكن تطبيق ذلك فعليًا في مجال المكتبات، ومن هذا المنطلق قامت الباحثة بما يأتي:

- ❖ الحصول على دورات تدريبية خاصة في موضوع التنقيب عن البيانات؛ لدراسة لأساليب التنقيب عن البيانات وما اشتملت عليه من خوارزميات مع التطبيق العملي من خلال برامج التنقيب المختلفة السابق ذكرها للتأكد من إتقان استخدام هذه التقنية الحديثة وكيفية التعامل مع البيانات ومعالجتها لاستخراج الأنماط المعرفية الصحيحة منها.
- ❖ إجراء بعض المراسلات البريدية والمقابلات الشخصية مع المتخصصين في التنقيب عن البيانات، ومناقشة ما تم التوصل له من نتائج أثناء التطبيق العملي وعرض الأفكار المبدئية حول كيفية تطبيق هذه التقنية في مجال المكتبات وأكثر الأساليب والخوارزميات ملائمة للتطبيق في هذا المجال على وجه التحديد.

٢/٥ مرحلة جمع المادة العلمية وتحليل البيانات:

تم تجميع المصادر المتعلقة بالموضوع سواء كانت (مقالات – بحوث - أطروحات علمية- دراسات – تقارير - أوراق مؤتمرات – الكتب العلمية)، كما قامت الباحثة بالقراءة المتأنية في أدب الموضوع بشقيه (التنقيب عن البيانات، والمكتبات الرقمية) لتكوين الخلفية النظرية للدراسة التي تمثلت في الجانب التاريخي للتنقيب عن البيانات وشرح لأساليبها وخوارزمياتها وكيفية تطبيقها في مجال المكتبات ومدى جودتها في التعامل مع المحتوى الرقمي العربي، والتجارب الأجنبية لاستخدام هذه التقنية في تطوير المكتبات الرقمية.

وبعد جمع المادة العلمية وتحليل ما تم التوصل له من بيانات من خلال القراءات الدقيقة؛ تم غربلة المادة العلمية التي ذ جمعت وتحديد أهم وأشمل البيانات التي تثري الدراسة، وقد تم الانتقاء والفرز والتصنيف وفقًا لفصول الدراسة، وتعد هذه الجزئية من أهم الصعوبات التي واجهت الباحثة.

٣/٥ مرحلة حصر التطبيقات الفعلية للتنقيب عن البيانات بالمكتبات والمؤسسات:

- حصر التطبيقات والتجارب الأجنبية التي استخدمت بالفعل هذه التقنية داخل المكتبات، وذلك عن طريق استرجاع المقالات والمواقع التي تناول المكتبات والجامعات والهيئات التي اهتمت بتطبيق هذه التقنية كما موضح في فصول الدراسة.
- حصر التطبيقات المصرية في محاولة لرصد مدى الاستخدام والتوجه الفعلي لمثل هذا

الموضوع داخل المكتبات أو الهيئات بمصر، وتم ذلك من خلال إجراء المقابلات الشخصية مع المسئولين والبحث في شبكة الإنترنت والمواقع الإلكترونية.

٤/٥ تجهيز أدوات جمع البيانات:

وهي من أهم مراحل الدراسة؛ حيث فرضت طبيعة هذه الدراسة مجموعة من الأدوات لجمع البيانات التي جاءت تباعًا وتدريجيًا حسب مراحل إجراء الدراسة من أداة لأخرى؛ فنجد تحديد برمجيات التنقيب عن البيانات المناسبة، ثم دراسة سمات الإنتاج الفكري التي ساعدت بشكل جلي في إعداد البرنامج التجريبي للمكتبة الرقمية العربية وما تلاه من أدوات كالمقابلات الشخصية لاختبار هذا البرنامج.

٥/٥ مرحلة إجراء التجربة:

وهي الشق التجريبي والعملي في هذه الدراسة وتعتبر من أهم الأهداف الرئيسية التي سعت الدراسة لتحقيقها للتأكد من إتقان وصحة وكفاءة تقنية التنقيب عن البيانات مع المكتبات الرقمية وخاصة العربية منها.

٦/٥ مرحلة الكتابة:

بعد إتمام كل ما سبق من خطوات بدأت مرحلة الكتابة الفعلية للدراسة العلمية في ضوء الأهداف والتساؤلات المحددة سلفًا حتى خرجت الرسالة في شكلها النهائي.

٦ - فصول الدراسة:

تحتوى الدراسة على جزأين؛ اشتمل الجزء الأول على سبعة فصول بالإضافة إلى المقدمة والخاتمة وقائمة المصادر العربية والأجنبية، أما اجزء الثاني فأختص بالملاحق. وتناولت المقدمة التعريف بموضوع الدراسة وأهميته وأهداف وتساؤلات الدراسة، ومجال الدراسة (الزمني، واللغوي، والجغرافي)، إلى جانب المنهج المتبع وأدوات جمع البيانات، وبالإضافة إلى ذلك تم تقديم وإبراز الدراسات السابقة ذات الصلة الوثيقة بالموضوع محل الدراسة.

الفصل الأول: سمات الإنتاج الفكري لاستخدام تقنية التنقيب عن البيانات في المكتبات:

تناول مسح الإنتاج الفكري المنشور عن موضوع استخدام المكتبات لتقنية التنقيب عن البيانات؛ عن طريق إعداد دراسة ببليومترية لحجم الإنتاج الفكري وتوزيعه زمنياَ وجغرافياَ، وأشكال وأنواع المصادر التي اهتمت بهذا الموضوع، وأهم الموضوعات التي تناولها، علاوة على توضيح إسهامات المؤلفين وتقييم الوثائق.

الفصل الثاني: أساسيات التنقيب عن البيانات:

قدم التعريف بالتنقيب عن البيانات من الناحية اللغوية والاصطلاحية، والجذور التاريخية لهذه التقنية، وأهم العلوم والمصطلحات ذات الصلة بها والتي أثرت بشكل أساسي في تكوينها وتطويرها، وأخيرًا التحديات والصعوبات التي تواجه التنقيب عن البيانات.

الفصل الثالث: أساليب التنقيب عن البيانات:

تم دراسة أساليب ومهام التنقيب عن البيانات بشقيها الوصفية كأسلوب العنقدة وأسلوب تحليل الارتباط، والتنبؤية كأسلوب التصنيف وأسلوب السلاسل الزمنية وأسلوب التنقيب التسلسلي، مع شرح ماهيتهم وأهم خوارزمياتهم خاصة المستخدمة في مجال المكتبات والمعلومات وكيفية عملها، كما تم توضيح عملية تنقيب النصوص Text Mining وأهم الأساليب والخوارزميات المستخدمة فيها.

الفصل الرابع: استخدام التنقيب عن البيانات في المكتبات إلى:

شرح كيفية الاستفادة من هذه التقنية في مجال المكتبات والمعلومات عن طريق توضيح كيفية استخدامها في إدارة وتنمية المقتنيات واتخاذ القرار، وخدمات المعلومات، وتنمية الموارد البشرية، والدعم والعلاقات الخارجية، وإعداد التقارير بالنظم الآلية المتكاملة مع عرض لأهم النظم الآلية المستخدمة لهذه التقنية الحديثة نسبيًا، وأخيرًا تقديم نموذج مقترح لخطوات التنقيب عن البيانات في المكتبات بشكل عام.

الفصل الخامس :تجارب التنقيب عن البيانات في المكتبات بالخارج:

يهدف هذا الفصل إلى دراسة التجارب العملية لاستخدام المكتبات لتقنية التنقيب عن البيانات وذلك عن طريق شرح للمشروعات بالمكتبات المحسبة وهما تجربة مكتبات جامعة "جيا ينج: Jia Ying"، و تجربة مكتبة كلية "بوستيد Boustead" بالصين، وأيضا المكتبات الرقمية حيث تم تحديد عناصر تنقيب بيانات المكتبات الرقمية وتم اختيار تجربتين للمكتبات الرقمية المطبقة للتنقيب عن البيانات وهما تجربة اتحاد المكتبات المصرية (KOBSON)، وتجربة انطولوجيا المقترح الشخصي (PORE) بتايوان، وفي كل هذه التجارب المستعرضة تم شرحها من حيث التعريف بالمشروع والهدف منه وما تم إتباعه من منهجية وتصميم.

الفصل السادس: تجارب التنقيب عن البيانات في المكتبات بمصر:

تناول بالدراسة والتحليل تطبيقين لاستخدام التنقيب عن البيانات بالمكتبات والمؤسسات العلمية داخل مصر وهما مشروع تنقيب الرسائل الجامعية بجامعة القاهرة The Cairo University Theses Mining System (CUTMS)، وتجربة المكتبة المركزية لجامعة حلوان للتطور الزمني المستقبلي للرسائل العلمية، وتم شرح هذين التطبيقين بالتفصيل وكيفية الاستفادة منهم من خلال توضيح ذلك بالرسوم البيانية والتوضيحية.

وأخيرًا الفصل السابع: التجربة العملية لاستخدام التنقيب عن البيانات في تطوير مكتبة رقمية عربية:

عرض الفصل إعداد مكتبة رقمية عربية متطورة في مجال الفنون الجميلة والتشكيلية باستخدام أساليب التنقيب عن البيانات، مع توضيح أهمية وخطوات أهداف هذه التجربة وقدم المراحل التي مر بها البرنامج التجريبي وكيفية

استخراج الأنماط المعرفية من برامج التنقيب عن البيانات وتقييمها، ثم تقديمها للمستفيدين في صورة برنامج تطبيقي على الويب، وتم تقييم هذا البرنامج وما توصل له من نتائج ومعرفة جديدة من قبل المستفيدين، وأخيرًا استعراض مجموعة من شاشات التجربة للتوضيح.

٧- نتائج الدراسة:

من خلال ما تم عرضه توصلت الدراسة لمجموعة من النتائج ، نوجزها في الآتي:

استطاعت الدراسة الخروج بمجموعة من النتائج تم توزيعها على ثلاثة محاور رئيسة جاءت متوافقة ومتناسقة مع أهداف وتساؤلات الدراسة نوضحها في الآتي:

المحو الأول: نتائج تتعمق بعملية التنقيب عن البيانات بشكل عام:

١- أدى الطلب المتزايد على البيانات إلى وجود أدوات متقدمة لتحليل البيانات، نتيجة للنمو الهائل في البيانات التي تم جمعها من مختلف التطبيقات فعلى سبيل المثال وليس الحصر، الأعمال التجارية، والعلوم والهندسة، والكيمياء والفلك، والعلوم الاجتماعية والإنسانية...الخ.

٢- يعد مصطلح (التنقيب عن البيانات) هو أنسب المقابلات باللغة العربية لمصطلح (Data Mining) داخل هذه الدراسة؛ لأنه يعبر عن اكتشاف واستخراج البيانات غير الواضحة وخاصة لاشتماله على حرف الجر (عن) ومن مواضع استعماله الكشف والإظهار عن ما هو خفيًا وغير جلي.

٣- إن هيكلة البيانات ووضعها داخل فئات نوعية وخصائص محددة صحيحة؛ من أهم الخطوات لنجاح عملية التنقيب عن البيانات ككل لأن من الممكن أن يتم اختيار فئات أو خصائص بصورة خاطئة ومضللة؛ مما يسفر عن عدم إتمام عملية التنقيب، أو الحصول على أنماط لا معنى لها.

٤- تتألف مهام ووظائف التنقيب عن البيانات من تصنيف البيانات Classification، والعنقدة Clustering أي جعل البيانات في صورة مجموعات، وتحليل الارتباط Association Analysis والعلاقات بين عناصر البيانات، وتحليل الاتجاهات والتشابه، وتحليل التطرف والانحراف في البيانات، والتوصيف والتمييز للبيانات والتي تعتبر شكل من أشكال التلخيص، ويمكن استخلاص أهم هذه المهام (ما تم دراسته بهذه الأطروحة) في الآتي:

- **العنقدة**: يعتبر من أبرز أساليب التنقيب عن البيانات الوصفية لما يقدمه من وصف للبيانات ووضعها داخل فئات وفق أنواع وخصائص البيانات، ويجب على محللي ومنقبي البيانات أن يحددوا بشكل دقيق أي نوع من الخوارزميات وأشكال العناقيد المراد استخراجها والحصول عليها؛ نظرًا لاشتمال هذا الأسلوب على

عدد كبير من الخوارزميات، أشهرها وأكثرها استخدامًا الخوارزميات الهرمية والتقسيم.

- **التصنيف:** وهي من الأساليب التنبؤية للتنقيب عن البيانات؛ وتعتبر من أقدم أساليب ومهام التنقيب عمومًا وخاصًا خوارزمية شجرة اتخاذ القرار، ويعتمد على تقسيم البيانات داخل فئات محددة سلفًا ناتج عن بيانات متشابهة سابقًا، وهو بذلك يقدم صورة بينة وواضحة عن هوية مجموعة من الكيانات التي تتفق في خصائص متشابهة.
- **تحليل الارتباط:** تعد من الأساليب الوصفية للتنقيب عن البيانات وينتج عن أسلوب تحليل الارتباط مجموعة من القواعد والأنماط المعرفية التي تقدم العلاقات والارتباطات بين عناصر البيانات؛ وكثير من هذه العلاقات لم نستطيع استنتاجها بالخبرة العامة لأنها قد لا ترجع للسببية ولكن لمجرد تكرار حدوث هذه العناصر من البيانات مع بعضها البعض.
- **تحليل السلاسل الزمنية:** 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 أيضًا أن يتم القياس باليوم والشهر حتى يتم قياس المتغير الزمني بشكل منتظم للتنبؤ بصورة منتظمة أيضًا.

٥- يعد حقل التنقيب عن البيانات أحد الحقول متعددة التخصصات، حيث إن له علاقة وثيقة بعدة علوم (علم الإحصاء، والذكاء الاصطناعي وتعلم الآلة والرياضيات والمنطق) وقد استمد من كل علم وآلية ما يخصه ويميزه داخل تقنياته، ويعتبر كثير من المحللين أن حقل التنقيب عن البيانات هو أحد الفروع المتطورة لكل من علمي الحاسب الآلي والإحصاء؛ وساعد كل ذلك على خروجه في صورة حقل متميز جدًا ذو طابع مختلف ومنفرد في تحليل البيانات.

٦- التنقيب في النصوص من أهم الموضوعات ذات الصلة بالتنقيب عن البيانات؛ ذلك لأنها تقدم آلية جديدة للبحث والاسترجاع والتصفح مما يساعد على تفاعل النص مع كل من النظام والمستخدم على حد سواء واستخراج المعلومات بشكل أدق.

٧- لا توجد معايير موحدة لخطوات التنقيب عن البيانات، بل يجتهد المحللون كل وفق تطبيقاته في توحيد الخطوات وتسلسلها لضمان سلامة النتائج النهائية، ويشكل ذلك أهم التحديات والصعوبات التي يواجهها العلماء في بناء وتطوير النظريات العلمية لهذا الحقل حتى يتم وضعه في قائمة العلوم ذات الخطوات المنهجة والقياسية في كل مراحله.

٨- توجد كثير من البرامج التطبيقية والمفتوحة المصدر للتنقيب عن البيانات المتاحة مجانًا على شبكة الإنترنت علاوة على حزم البرامج المتاحة من شركات البرمجيات مثل برنامج SPSS, Weka, Microsoft SQL Server 2005, Xlminer وغيرهم من البرامج.

٩- إن عملية بناء النماذج والتقييم مرتبطتان ببعضها ومتلازمتان فلا يجوز أن تتم عملية استخراج الأنماط المعرفية دون أن تمر مرحلة التقييم على مستوى متخذي القرار والمستخدمين لهذه الأنماط والمعلومات، لأن هذه الأنماط من الممكن أن تكون صحيحة من الناحية النظرية ولكن عند التطبيق وعرضها على الأشخاص المعنيون تكون بعيدة عن الواقع فنجد الإجابات على غير المتوقع مما قد يعيدنا لنقطة البداية أحيانًا لمراجعة أهداف المشروع واختيار الأساليب والخوارزميات السليمة؛ أي إنهما عمليتان متلازمتان.

المحو الثاني: نتائج تتعمق باستخدام المكتبات المصرية والأجنبية للتنقيب عن البيانات:

١- تسير عملية تنقيب بيانات المكتبة بنفس الخطوات المنهجة لعملية اكتشاف المعرفة من قواعد البيانات؛ وذلك بغرض فهم أفضل للمجتمع المستفيد من المكتبة.

٢- يحتاج استخدام التنقيب عن البيانات في المكتبات أدوات استخراج وتحويل وتحميل البيانات من مصادرها المتعددة وهو ما يطلق عليها Tools ETL؛ من أجل الحصول على مستودع للبيانات أو نظام متكامل سهل الاستخدام لجمع كل من مصادر البيانات وأدوات التنقيب عن البيانات معًا داخل نظام واحد؛ مما يشجع أمناء المكتبات لتطبيق آليات التنقيب عن البيانات لتقديم المعلومات المفيدة والضرورية للمتطلبات الإدارية مع التركيز على القضايا المهنية دون التدخل في تقنيات قواعد البيانات والتي قد تكون ضعيفة عند البعض.

٣- تعد أساليب وخوارزميات العنقدة والتصنيف وتحليل الارتباط وتحليل السلاسل الزمنية على وجه الخصوص من أنسب وأكثر مهام وتقنيات التنقيب عن البيانات توافقًا وتماشيًا لطبيعة بيانات المكتبة، وتعتمد هذه الأساليب على وجه الخصوص على توضيح العلاقات بين عناصر البيانات بصورة عامة وهو ما يهمنا بشكل خاص في مجال المكتبات والمعلومات خاصة في عمليات البحث والتصفح والاسترجاع.

٤- إن استخدام التنقيب عن البيانات بالمكتبات يساعد في معرفة اتجاهات الموضوعات الأكثر شيوعًا واستخدامًا بالمكتبة لتمكين المدراء من الاستخدام الأمثل للميزانية وتنمية المقتنيات، ويعتبر تحليل بيانات الاستخدام والاستعارة والإعارة المتبادلة بين المكتبات يساعد في

إعادة تخطيط وتنظيم المجموعات والوقت المستغرق لحركة سير العمل بالمكتبة من أجل التخطيط الأمثل لساعات العمل وتوظيف العاملين، وما إلى ذلك.

٥- تركز عملية التنقيب عن البيانات بالمكتبات على خدمات المستفيدين من أجل وجود خدمات وأنشطة ومصادر معلومات مصممة خصيصًا لتلبية احتياجات مختلف مجموعات المستفيدين مما ينتج عنه تسويق أفضل لخدمات المعلومات.

٦- قدمت مكتبات الجامعات الصينية أهم مثال يحتذي به في تطبيق تقنية التنقيب عن البيانات، فهي تعتبر من أهم وأشمل التطبيقات الأجنبية في هذا الموضوع ؛ حيث عرضت كيفية الاستفادة من أساليب وخوارزميات العنقدة والتصنيف وتحليل الارتباط وتحليل السلاسل الزمنية والتحليل التسلسلي لبيانات المكتبات، وقد كانت لبيانات الاستعارة الركيزة الأساسية التي اعتمدت عليها مكتبات المعاهد والجامعات الصينية لتطبيق التنقيب عن البيانات.

٧- عدم استخدام المكتبات المصرية لعملية التنقيب عن البيانات فعليا؛ بخلاف المكتبة المركزية جامعة حلوان؛ على الرغم من استخدام المكتبة الوطنية وبعض المكتبات لأنظمة آلية تشتمل على تطبيقات للتنقيب عن البيانات.

٨- يقدم مشروع تنقيب الرسائل الجامعية بجامعة القاهرة مثالاً جيدًا لاستخدام خوارزميات التنقيب عن البيانات (خاصة العنقدة والتصنيف) في استعراض الرسائل بموضوعاتها ومؤلفيها ومشرفيها والأقسام التابعة لها، والخروج بمجموعة من الرسوم الجرافيكية التوضيحية عن الأقسام والكليات والموضوعات ومساهمات المشرفين على مستوى الجامعة، وإعطاء صورة واضحة لمتخذي القرار بالجامعة عن أكثر اتجاهات البحث العلمي داخل الجامعة وأكثر الموضوعات صدارة في رسائل الماجستير والدكتوراه وأكثرها عزوفًا لاتخاذ اللازم حيال ذلك، علاوة على أن هذا المشروع يساعد الباحثين للتعرف على الصورة الشاملة عن الموضوعات البحثية داخل الجامعة والاتجاه العام في البحث وقياس درجات الاهتمام سواء على قطاع الإنسانيات والعلوم الاجتماعية والعلوم التطبيقية، مما يبين للباحث أين يقع بحثه بالضبط داخل المنظومة البحثية بالجامعة.

وعلى الرغم من إطلاق هذا المشروع إلا إنه ما زال قيد التنفيذ في بعض الأحيان؛ حيث لا يمكن استخراج الرسوم الجرافيكية التوضيحية وإتمام إجراءات البحث بصورة جيدة خاصة عند البحث عن الموضوعات وما يرتبط بها من موضوعات أخرى ذات الصلة ؛ وترى الباحثة أيضًا من خلال الاستخدام الفعلي إن هذا المشروع لا يعمل بالكفاءة المطلوبة على كل متصفحات الانترنت ونظم التشغيل.

٩- أوضحت الدراسة الببليومترية أنه من المتوقع الزيادة المستقبلية في حجم الإنتاج الفكري للاستخدام التنقيب عن البيانات في المكتبات؛ حيث سجل عامي ٢٠١٠، ٢٠١١ أعلى معدلات في إنتاج البحوث والدراسات العلمية وبلغ عدد البحوث لكل منها (٢٧) بنسبة مئوية بلغت ١٢,٧% (٣٢) بنسبة مئوية بلغت (١٥%) على التوالي من إجمالي ٢١٢ بحث ودراسة علمية .

١٠- كشف التوزيع الموضوعي للإنتاج الفكري المنشور عن استخدام التنقيب عن البيانات في المكتبات عن مدى ارتباط موضوع (التنقيب عن البيانات) بموضوع (المكتبات الرقمية) على وجه الخصوص حيث سجلت الأخيرة عدد (٣٧) بما يعادل من ١٧,٥% إجمالي ٢١٢ بحث ودراسة علمية وأكبر عدد من الدراسات والبحوث حصل عليه أحد الموضوعات المرتبطة باستخدام التنقيب عن البيانات في المكتبات؛ مما يدل على اهتمام الباحثين باستخدام هذه التقنية في المكتبات الرقمية لما يمكن أن تقدمه من فوائد لتطوير أداء وفاعلية وخدمات المكتبات الرقمية، وقد تلى موضوع (المكتبات الرقمية) موضوع (المكتبات) بفارق عشرة بحوث ودراسات بنسبة بلغت ١٢.٧%.

المحو الثالث: نتائج تتعمق باستخدام المكتبات الرقمية للتنقيب عن البيانات:

١- توجد كثير من التطبيقات الأجنبية لاستخدام التنقيب عن البيانات داخل المكتبات الرقمية، بشكل عام ويلاحظ التنوع في تنفيذ هذه التطبيقات.

٢- يمكن الاستعانة بتقنية التنقيب عن البيانات في تطوير بناء مكونات نظام المكتبة الرقمية؛ حيث يتم تضمين داخل مكونات النظام آليات وخوارزميات التنقيب مما يعيد هيكلة وميكنة الإدخال والتعديل والحذف والإضافة والبحث والاسترجاع والتصفح وما غير ذلك من عمليات وخدمات المكتبات الرقمية.

٣- استطاعت هذه التقنية بشكل واضح جدًا تسهيل آليات البحث والاسترجاع سواء على مستوى قواعد البيانات الببليوجرافية أو قواعد بيانات النص الكامل ؛ لأنها اعتمدت على أسلوب (البحث الموجه: Oriented Search) إما لمستفيدين بعينهم أو موضوعات وفئات تم تجميعها بالاستعانة بأساليب التنقيب.

٤- أصبح بالإمكان من خلال استخدام تقنية التنقيب عن البيانات بالمكتبات الرقمية تطوير خدمات هذه المكتبات بشكل ملحوظ؛ حيث أصبحت أكثر تفاعلية / تخصصًا/ توفيرًا لوقت وجهد القراء والباحثين؛ لأن فلسفتها أصبحت قائمة على إعادة توزيع مصادر المكتبة بطريقة مختلفة وفق احتياجات واهتمامات وميول مستخدميها علاوة على اختلاف أسلوب عرض مصادر المعلومات والخدمات لأنها تم بنائها نتيجة استخدام باحثين آخرين مهتمين بنفس الموضوعات، أو

الحصول على روابط وعلاقات بين عناصر البيانات؛ التي لا يمكن الحصول عليها ومعرفتها ذي قبل بالطرق التقليدية المعتادة.

٥- إمكانية تعامل المحتوى الرقمي العربي مع خوارزميات التنقيب ؛ ولكن لابد من إعداده بصورة جيدة ومنقحة وخاصة تجنب النقص في البيانات وإزالة البيانات المشوشة والخاطئة وتوحيد استخدام العلامات؛ ذلك لأنه مازالت اللغة العربية يشوبها بعض القصور عند استرجاع البيانات التي أفرزتها خوارزميات التنقيب؛ وعلى وجه الخصوص عند تطبيق (أسلوب العنقدة، والتنقيب في النصوص بشكل عام)، ولكن هذا لا يمنع من وضع النص العربي داخل آلية التنقيب عن البيانات والخروج بنتائج حقيقية كما أوضحته الدراسة التجريبية.

٦- أوضحت نتائج الدراسة التجريبية للمكتبة الرقمية للفنون الجميلة والتشكيلية؛ أن نسبة رضا الباحثين عن العناوين ذات الصلة بالرسالة العلمية بلغت (٦٤,٨٪) وهو ما يعادل (٦١) متخصص من إجمالي (٩٤)، أما فيما يخص انطولوجيا الفنون فجاءت نسبة الرضا على اتساقها موضوعيًا (٧٥,٥٪) وهو ما يعادل (٧١) متخصص، ورأى بعض المتخصصين بما يعادل (٥٠٪) تقريبًا من إجمالي المستخدمين للمشروع أنه يجب فصل بعض الموضوعات وما ترتبط بها من موضوعات فنية، ورأي كثير من المستفيدين فيما يتعلق بالبحث الموضوعي المتقدم بنسبة مئوية بلغت (٧١,٢٪) بما يساوي (٦٧) متخصص بعدم اتساق بعض الرسائل داخل كل فئة إلى حد ما.

وبشكل عام، خرجت نتائج هذه الدراسة بأن هذا المشروع ساعد في إعادة توزيع الرسائل العلمية بطريقة جديدة وموضوعيه؛ مما يسهل للمستفيدين البحث والتصفح واسترجاع المعلومات، وبالإضافة إلى ذلك، جاءت كثير من التعليقات بلغت (٣٤٪) بما يعادل (٣٢) متخصص بضرورة تطوير المكتبة الرقمية العربية للفنون الجميلة والتشكيلية من خلال وجود علاقات الاستخدام للرسائل والملفت الشخصية للباحثين حتى تعطي المكتبة صورة أكثر وضوحًا حول الروابط بين الرسائل وبعضها البعض.

وخلاصة هذه النتائج جميعًا يمكن إيجازها في الآتي:

- أصبح استخدام التنقيب عن البيانات ضرورة ملحة فرضتها علينا الزيادة غير المعقولة في كم البيانات الآني والماضي، فلابد لأي مؤسسة تهتم بزيادة فاعلية خدماتها أن تستخدم هذه العملية لتحليل بياناتها للخروج بالأنماط المعرفية المساعدة لتطوير الأداء.
- إن تنقيب بيانات الاستخدام بالمكتبات أصبحت مرجعًا ودليلاً أساسيًا لتوجيه وإرشاد متخذي القرار لتحديد اتجاهات البحث والباحثين داخل مكتباتهم ومن ثم المؤسسة الأكبر.
- يجب على المكتبات المصرية أن تحذو حذو المكتبات والهيئات الأجنبية لاستخدامها التنقيب عن البيانات في تحليل بياناتها.