مجموعات قواعد بيانات المكتبات الرقمية على الويب :
الوصول والمعالجة والإفادة باستخدام محركات بحث الويب غير المرئية

د. سيد ربيع سيد إبراهيم
مدرس بقسم المكتبات والوثائق
كلية الآداب – جامعة بني سويف

المستخلص:

عمل ظهور الويب غير المرئية بمختلف أنواعها وأشكالها على طرح مشكلة قواعد بيانات المكتبات الرقمية الببليوجرافية وغير الببليوجرافية التي تحويها مواقع المكتبات الرقمية على الويب، حيث تنتمي قواعد بيانات المكتبات الرقمية إلى فئة مواقع الويب التي لا تراها محركات بحث الويب، ومن ثم فإنه يمكن تسميتها بمجموعات المكتبات الرقمية غير المرئية. وعلى ذلك فإنه ثمة حاجة إلى دراسة آليات التفاعل والتوافق بين قواعد بيانات المكتبات الرقمية غير المرئية وبين محركات بحث الويب حتى يعود ذلك على المكتبات الرقمية بالنفع من نواحٍ كثيرة؛ أهمها: استخدام واجهات محركات بحث الويب كنوافذ تصل من خلالها مجموعات وخدمات المكتبات الرقمية إلى مستخدمي الويب أينما كانوا. وتركز هذه الدراسة على الربط بين الويب غير المرئية ومواقع المكتبات الرقمية، ثم آليات المعالجة والتنظيم ومن ثم البحث عن مجموعات المكتبات الرقمية الببليوجرافية وغير الببليوجرافية بواسطة محركات البحث.

الكلمات المفتاحية : المكتبات الرقمية، قواعد بيانات الويب، الويب غير المرئية، محركات البحث.

منهجية الدراسة :

١/١ التمهيد:

مكنت الويب المكتبات الرقمية من تخطي الحدود المكانية وتقديم خدمات المعلومات إلى المستفيدين أينما وُجدوا، ومن ثم فإن قواعد بيانات المكتبات الرقمية على الويب وأنظمة المعلومات المتكاملة قد دخلت ضمن نظم ومصادر معلومات الويب

التي تتأثر بمختلف ظواهرها المعلوماتية، ومن أهم ظواهر الويب الآن هي: الويب غير المرئية التي طرأت بخصائص جديدة تستدعي إعادة توصيف وتأطير لنظم معالجة وبحث الويب. وقد أضحت الفهارس المتاحة على الخط المباشر جزءاً من قواعد البيانات الببليوجرافية المستهدفة من محركات بحث الويب. وكما اختلفت مصادر المعلومات في الشكل الرقمي بين النصية والمصورة والصوتية، فقد اختلفت أيضًا في حجم وطريقة الوصول إليها، حيث أفرزت الإنترنت صفحات الويب html وملفات الصور والصوت إلى جانب ظهور عدد غير نهائي من قواعد البيانات العلمية والمتخصصة، وتحمل هذه القواعد معلومات ذات قيمة تعلو غيرها من صفحات الويب نظرًا لإخضاعها لتحكيم وإشراف الهيئات العلمية المختلفة.

لقد نمت أعداد قواعد بيانات الويب عامة وقواعد بيانات المكتبات الرقمية خاصة إلى الدرجة التي أصبح معها حجم هذه القواعد يفوق حجم صفحات وملفات الويب الأخرى بما يقدر بخمسمائة مرة. إلى جانب اختلاف هذه القواعد في طبيعة بنيتها وتنظيمها وطرق الوصول إلى معلوماتها عن طبيعة تنظيم وبحث صفحات الويب الأخرى، ونظرًا لهذا التضخم في الحجم والاختلاف في الطبيعة فقد شكلت هذه القواعد جزءاً منفصلاً عن الويب الحالية أطلق عليه مصطلحات كثيرة مثل: (الويب الخفية والويب غير المرئية والويب العميقة) وهي على التوالي Hidden web, Invisible web, Deep web. ويوما بعد يوم شكلت مواقع المكتبات الرقمية وليدًا جديدًا للويب غير المرئية يحتاج إلى رعاية أكثر حرصًا ودأبًا على تحقيق تنظيمه وتسهيل الوصول إلى ما يحويه من معلومات محجوبة عن مستخدمي الويب ومحركات البحث الآنية. حيث بات ما يعرف بالويب الأكاديمية ومصادر المعلومات العلمية هي قِبلة الباحثين في الوصول إلى مصادر الويب، ودون التفاعل الناجح بين المكتبات الرقمية ومحركات البحث، فإن مستخدمي المكتبات الرقمية يظلون بعيدين عن استرجاع مجموعات المكتبات الرقمية من خلال واجهات محركات البحث ذات القابلية العالية للمستفيدين.

وقد شكلت الويب غير المرئية (مع قواعد بيانات المكتبات الرقمية) حاجزًا منيعًا أمام محركات بحث الويب الحالية في التعامل مع ملفات تلك القواعد، وذلك لأن هذه القواعد تعمل على إتاحة مصادرها من خلال واجهات البحث search interfaces الخاصة

بها ، إلى جانب اختلاف احتواء هذه القواعد على كثير من أنواع الملفات الرقمية مثل: pdf, doc, ppt. فضلاً عن خصوصية الملكية الفكرية التي تتميز بها قواعد البيانات عن غيرها من مصادر المعلومات على الويب، كما أن قواعد بيانات الويب غير المرئية تحتاج إلى مستويات أكثر تعمقًا في التحليل من برامج الزاحف التي قلما تتعدى المستوى الثاني أو الثالث في تحليل روابط صفحات الويب URLs. وعلى ذلك فإننا بصدد إما إهمال هذا الكم من الملفات الرقمية عالية القيمة لباحثي الويب أو التعامل مع هذه الظاهرة بتطوير أدوات العمل والتحليل لمصادر المعلومات الرقمية على الويب، بما يمكن معه مسايرة هذا الجزء المتطور في الويب.

تشكل قواعد بيانات الويب غير المرئية كمًّا هائلاً من مصادر المعومات التي لا يمكن لمؤسسات المكتبات أو مراكز المعلومات العمل بدونها؛ حيث يعني كشف وإتاحة هذه القواعد على الويب مصدرًا جديدًا يضاف إلى مقتنيات المكتبات من المصادر الإلكترونية على الخط المباشر. وثمة جانب آخر لا يمكن إغفاله وهو المكتبات الرقمية التي تضع قواعد بياناتها على الويب، ويتزايد عدد هذه القواعد بشكل سريع وفي وقت قصير، وكونها تبقى معزولة عن برامج الزاحف فهي بعيدة عن تكشيف محركات البحث وبعيدة بدورها عن مستخدمي الويب، هذا في الوقت الذي تسعى المكتبات فيه إلى الوصول إلى المستفيدين في أماكنهم. فضلاً عن قواعد البيانات التي تحمل مصادر معلومات متباينة الأشكال مثل: قواعد بيانات المصادر الرقمية المصورة الثابتة والمتحركة العامة أو العلمية، قواعد بيانات مصادر المعلومات الصوتية، قواعد الهيئات المتحفية والطبية والعلمية... وغيرها.

٢/١ مشكلة الدراسة :

تمثل قواعد بيانات المكتبات الرقمية على الويب جزءاً كبيرًا مهمًّا من حجم الويب غير المرئية؛ حيث تتسق مجموعات التسجيلات الببليوجرافية داخل قواعد الفهارس الرقمية مع غيرها من مجموعات قواعد بيانات الويب التي تعجز أدوات بحث الويب عن التفاعل معها معالجةً واسترجاعًا. وتشكل قواعد بيانات الويب غير المرئية كمًّا هائلاً من مصادر المعومات التي لا يمكن لمرافق المكتبات أو مراكز المعلومات العمل بدونها ،

حيث يعني كشف وإتاحة هذه القواعد على الويب مصدرًا جديدًا يضاف إلى مقتنيات المكتبات من المصادر الإلكترونية على الخط المباشر. وثمة جانب آخر لا يمكن إغفاله في المكتبات الرقمية التي تضع قواعد بياناتها على الويب، هو: تزايد عدد هذه القواعد بشكل سريع وفي وقت قصير واتجاه خدمات المكتبات إلى اتخاذ الويب نافذة للوصول إلى فئة كبيرة من المستفيدين تتعدى الحواجز الجغرافية، وكونها تبقى معزولة عن برامج الزاحف فهي بعيدة عن تكشيف محركات البحث وبعيدة بدورها عن مستخدمي الويب، هذا في الوقت الذي تسعى المكتبات فيه إلى الوصول إلى المستفيدين في أماكنهم.

فضلاً عن قواعد البيانات التي تحمل مصادر معلومات متباينة الأشكال مثل: قواعد بيانات المصادر الرقمية المصورة الثابتة والمتحركة العامة أو العلمية، قواعد بيانات مصادر المعلومات الصوتية، قواعد الهيئات المتحفية والطبية والعلمية... وغيرها. ومن ثم فإن قواعد بيانات المكتبات الرقمية على الويب التي تحمل البيانات والتسجيلات الببليوجرافية ستبقى بعيدًا عن متناول مستخدمي الويب، مما يعني ضياع وسيلة مهمة من بين يدي المكتبات الرقمية تستطيع استخدامها للتنويه والإعلام عن مجموعاتها من مصادر المعلومات. وهذا فضلا عن خسارة القيمة التسويقية لخدمات المكتبات الرقمية التي يمكن استغلالها بإدخال تسجيلات الفهارس الرقمية ضمن نتائج محركات بحث الويب غير المرئية. ومما سبق يمكن إيجاز تساؤلات الدراسة فيما يلي :

١- ما خصائص الويب غير المرئية كبيئة جديدة خلاف ما يعرف بالويب المرئية؟

٢- ما ملامح قواعد بيانات المكتبات الرقمية على الويب كجزء من الويب غير المرئية؟

٣- ما جوانب التأثير والتأثر بين الويب غير المرئية وقواعد البيانات الببليوجرافية داخل مواقع المكتبات على الويب؟

٤- ما معايير وآليات التوافق بين معالجة واسترجاع التسجيلات الببليوجرافية ومحركات بحث الويب غير المرئية؟

٥- ما ملامح التفاعل بين فهارس المكتبات الرقمية ومحركات بحث الويب غير المرئية؟

٦-ما مواصفات محركات بحث الويب غير المرئية المتوافقة وبحث قواعد البيانات الببليوجرافية؟

٣/١ أهداف الدراسة :

يسعى الباحث إلى تحقيق الأهداف التالية :

١-التعرف إلى ملامح بيئة جديدة من الويب داخل الإنترنت تندرج تحت اسم " الويب غير المرئية"

٢-فهم الطبيعة الخاصة لقواعد بيانات الويب واختلافها عن غيرها من مصادر الويب الأخرى.

٣-تحديد جوانب الاستفادة من بيئة الويب غير المرئية وانعكاسها على قواعد بيانات مواقع المكتبات على الويب.

٤-التعرف إلى ملامح العلاقة والتعامل بين أدوات بحث الويب غير المرئية وقواعد البيانات الببليوجرافية للمكتبات الرقمية.

٥-وضع ملامح جديدة لمحركات بحث الويب غير المرئية، تستطيع من خلالها إخضاع محتوى قواعد البيانات الببليوجرافية داخل مواقع المكتبات الرقمية للبحث من جانب مستخدمي الويب.

٤/١ أهمية الدراسة :

تنبثق أهمية هذه الدراسة من مجموعة العوامل التالية:

١-أن الويب غير المرئية التي تشكل قواعد البيانات النسبة الأعلى فيها تعبر عن ظاهرة معلوماتية تسترعي الاهتمام من حيث المعالجة والتنظيم وآليات الاسترجاع أمام مستخدمي الويب.

٢-أن قواعد بيانات المكتبات سواء التي تحوي التسجيلات الببليوجرافية أو مصادر المعلومات لا بد وأن تصل المستفيدين أينما كانوا، وذلك تماشيًا مع استخدام متصفحات الويب كنوافذ لإيصال خدمات المكتبات الرقمية إلى مختلف المستفيدين، دون اعتبار للحدود الجغرافية.

٣- يقتضي تواجد المكتبات الرقمية على الويب استخدام أدوات تنظيم وبحث الويب باعتبارها روابط يصل بها المستفيد إلى إمكانية التعامل مع المحتوى الببليوجرافي للفهارس الرقمية من خلال محركات وأدلة بحث الويب.

٤- مازالت محركات بحث الويب تقف عاجزة أمام الاستفسارات الببليوجرافية المقننة، أو تلك التي توافق طبيعة تسجيلات قواعد بيانات الويب، ومن ثم فإنه من الضروري البحث عن آليات عمل تمكن محركات بحث الويب من التفاعل مع محتوى ومجموعات قواعد بيانات الويب عامة، وقواعد بيانات المكتبات الرقمية خاصة.

٥/١ منهج الدراسة وأدواته:

١/٥/١ منهج الدراسة:

تستخدم الدراسة منهجي البحث الوصفي التحليلي والتجريبي، حيث تعتمد الدراسة أولاً: على استقراء خصائص ظاهرة البحث والتعرف إلى جوانب موضوع المكتبات الرقمية على الويب من ناحية ودراسة خصائص الويب غير المرئية ومحركات البحث داخلها من ناحية أخرى، وذلك سعيًا إلى تحليل الأدبيات المختلفة للوصل إلى تفسير وتعليل هذه الظاهرة. ثانيا : تسعى الدراسة إلى كشف واقع عمل بعض محركات ونظم استرجاع متعددة تختص بالبحث في مجموعات قواعد بيانات المكتبات الرقمية على الويب، وقد حدد الباحث نموذجين من أدوات بحث مجموعات المكتبات الرقمية كعينة قصدية لملاحظة ومعايشة واقع عمل هذه الأدوات.

٢/٥/١ عينة الدراسة:

اختار الباحث نموذجين من محركات بحث مجموعات المكتبات الرقمية على الويب، حيث لم يجد أي من المحركات الأخرى التي تتفاعل مع مجموعات التسجيلات الببليوجرافية بشكل مخصص. ويمثل الجدول رقم (١) عينة الدراسة.

محركات بحث قواعد المكتبات الرقمية
http://infomine.ucr.edu
http://www.scirus.com/

٦/١ مصطلحات الدراسة :

- **قواعد البيانات:**

هي ملف أو أكثر من ملفات البيانات التي يتم تنظيمها في شكل تسجيلات يمكن استخدامها والرجوع إليها إذا دعت الحاجة الموضوعية. وتعتمد قاعدة بيانات الويب على العنصر البشري في بناء وتسيير ملفات قاعدة البيانات. وبطبيعتها تعتمد قواعد البيانات على آليات تختلف عن تلك المستخدمة داخل محركات بحث الويب.

- **المكتبات الرقمية:**

تمثل المكتبة الرقمية على الويب مجموعات قواعد البيانات الببليوجرافية أو غير الببليوجرافية التي تدار بواسطة نظام آلي متكامل يسمح للمستفيدين البحث والاسترجاع للتسجيلات. وقد توجد المكتبة الرقمية في شكل محدود يمثله الفهرس المتاح على الويب، أو شكل غير محدود تمثله قواعد المعلومات التي تحوي النصوص الكاملة أو / و ملفات الوسائط المتعددة.

- **الويب:**

تعد الويب أحد مكونات شبكة الإنترنت، وهي مجموعة الصفحات و/ أو المواقع التي تتميز باستخدام النص الفائق، والوسائط المتعددة داخل صفحاتها ذات بنية html.

- **الويب المرئية:**

هي مجموعة صفحات و / أو مواقع الويب ذات القابلية للتنظيم والبحث والاستدعاء بواسطة محركات البحث، حيث لا تعتمد هذه الصفحات على قواعد بيانات لتخزين ملفاتها أو معلوماتها.

- **الويب غير المرئية:**

هي مجموعة الصفحات و/ أو مواقع الويب التي تعتمد على قواعد بيانات مختلفة الحجم، ولذلك لا تستطيع محركات البحث التعامل معها نظرًا لعدم قدرة برنامج الزاحف على التعامل مع آليات التنظيم والبحث في قواعد بياناتها.

محركات البحث:

هي برامج آلية لتنظيم واسترجاع صفحات و/ أو مواقع الويب، معتمدة على برامج فرعية مثل الزاحف والمكشف لأداء مهام محددة. وتؤدي محركات البحث عملها بشكل تام مع مواقع الويب غير أنها لا تستطيع أداء مهامها بشكل دقيق داخل قواعد البيانات التي تحكمها آليات عمل مختلفة.

٧/١ الدراسات السابقة :

استعرض الباحث أدبيات الموضوع ووجد أن أدبيات الإنتاج الفكري العربي والأجنبي لم تتناول قواعد بيانات المكتبات الرقمية على الويب وعلاقتها بكل من الويب غير المرئية ومحركات البحث، سواء كان ذلك في الأطروحات الأكاديمية أو الدراسات البحثية. غير أن الإنتاج الفكري الأجنبي قد حاول علاج قضايا الويب غير المرئية عامة ومحركات بحث الويب غير المرئية، دون التركيز على قواعد المكتبات الرقمية. وكان أهم ما أفرزه الإنتاج الفكري الأجنبي في هذا الموضوع عن طريق بحث مستخلصات الرسائل الجامعية Proquest Digital Dissertation Abstracts International (UMI)[١]، ومع استخدام مصطلحات مثل: digital libraries search engines, deep web ... digital libraries وأفرز ذلك الدراسة التالية:

Ipeirotis, Panagiotis G. Classifying and searching hidden-web text databases , advisor: Prof. Luis Gravano , Columbia University, 2004. [٢]

غير أن هذه الدراسة عملت على استقراء محتوى قواعد البيانات النصية فقط، محاولة تحديد التقسيمات الموضوعية المهمة لمحتوى تلك القواعد، دون الإشارة القريبة أو البعيدة إلى قواعد بيانات المكتبات الرقمية. كما اعتمدت في الإجراءات على برنامج QProber لبحث واستكشاف محتوى قواعد البيانات النصية. وباستطلاع نتائج الاستفسارات البحثية الموجهة من هذا البرنامج إلى محتوى قاعدة البيانات، وتحليلها إحصائيا، يستطيع البرنامج تحديد أهم الكلمات المفتاحية الدالة المستخدمة في محتوى هذه القاعدة.

١ / ٨ خطوات الدراسة :

تسير هذه الدراسة وفقاً لمجموعة من الخطوات هي :

١-دراسة واستقراء واقع الويب غير المرئية والتعرف إلى ملامحها المعلوماتية من حيث البنية والتكوين والتنظيم.

٢-دراسة واستقراء خصائص قواعد بيانات المكتبات الرقمية على الويب، سواء تسجيلات الفهارس الرقمية أو قواعد مصادر المعلومات للنصوص الكاملة أو الوسائط المتعددة.

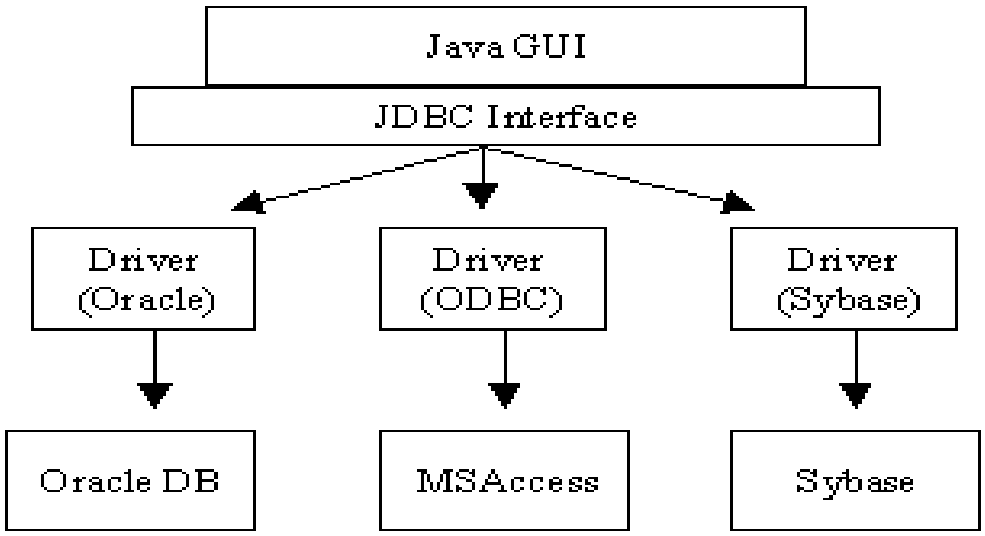
٣-دراسة تطبيقية لنموذجين من مواقع بحث الويب غير المرئية المتخصص في التعامل مع حقول تسجيلات المكتبات الرقمية من حيث البنية والتكوين وإمكانات البحث المتاحة أمام المستفيدين.

٤- طرح بعض الرؤى حول العلاقة والإطار والتنظيم الجامع بين قواعد بيانات المكتبات الرقمية وبين الويب غير المرئية من ناحية، وبين قواعد بيانات المكتبات الرقمية ومحركات البحث من ناحية أخرى.

٢- قواعد البيانات في بيئة الويب:

تختلف قواعد البيانات في طبيعتها عن بنية الويب؛ حيث تعتمد قواعد البيانات على نظم تشغيل مثل: MS SQL, Oracle, MS Access, My SQL . وهي بدورها لا تتعامل مع بنية النص الفائق في تنظيم ملفاتها. ويترتب على ذلك عدم استطاعة نظم استرجاع الويب الحالية التعامل مع محتوى هذه القواعد نظرًا لاختلاف طبيعتها عن بنية برنامجي الزاحف والمكشف في محركات البحث. ولدخول هذه المصادر الرقمية (قواعد البيانات) إلى الويب تحتاج إلى بنية وسيطة تجمع بين خصائص واجهات النص الفائق من جانب وآليات عمل قواعد البيانات من جانب آخر. وسميت هذه الواجهات بـ CGI داخل خادم الويب من جانب المستخدم، بالإضافة إلى استخدام the JDBC interface من جانب قواعد البيانات. ويمثل ذلك الشكل رقم (١) (٣).

إن ثمة مستويين للاسترجاع تمر بهما قواعد البيانات على الويب؛ **الأول** : يتمثل في أساليب استرجاع الملفات الداخلية المكونة لقاعدة البيانات ويقوم بذلك النظام الداخلي لإدارة قاعدة البيانات DBMS. **والثاني** : يتمثل في استرجاع قاعدة البيانات كوحدة واحدة ثم التعامل مع ما بداخلها من ملفات وتقوم بذلك محركات وأدلة بحث الويب غير المرئية. ويمثل الجانب الثاني المشكلة التي تواجهها نظم الاسترجاع الآن على الويب، ويطلق على هذه المشكلة ظاهرة الويب غير المرئية نظرًا لعدم تمكن هذه النظم من رؤية محتوى قواعد البيانات والقدرة على أداء عمليتي المعالجة والتنظيم اللازمتين لاسترجاع ملفات قواعد البيانات أمام المستفيدين.



الشكل رقم (١) الواجهات الوسيطة بين الويب وقواعد البيانات

٣- أشكال محتوى الويب:

تقسم محركات البحث محتوى الويب إلى قسمين؛ الأول : الويب المرئية وهو كل ما تستطيع محركات البحث الوصول إليه ومعالجته وتنظيمه على اختلاف أنواع ملفات المعلومات أو وسائط حملها. الثاني: الويب غير المرئية وهو كل ما لا يمكن لمحركات البحث التعامل معه أو إضافته إلى مستودعاتها. فالقدرة على الوصول إلى المعلومات أو عدم الوصول إليها عند محركات البحث، تمثل قدرة المحركات على رؤية أو عدم

رؤية المعلومات الرقمية. ومن ثم فإن معيار الحكم على المعلومات الرقمية بكونها مرئية أو غير مرئية داخل المحركات، هو وصول أو عدم وصول محركات البحث إليها. ويمكن توضيح شكلي المعلومات المرئية وغير المرئية كما يلي :

١/٣ الويب المرئية:

يطلق على الويب المرئية مسميات أخرى مثل الويب السطحية surface web، وهو المعنى المقابل لأحد مصطلحات الويب غير المرئية المسمى بالويب العميقة the deep web. وقد أوردت موسوعة Wikipedia encyclopedia تعريفًا للويب المرئية هو: "إن الويب السطحية – التي تعرف أيضًا بالويب المرئية أو المكشفة – هي ذلك الجزء من الويب القابل للوصول والمكشف بواسطة محركات البحث. وتكون محركات البحث مستودعاتها من مصادر الويب باستخدام برامج الزاحف أو العنكبوت التي تبدأ عملها بالاعتماد على قائمة مواقع معلومة مسبقًا، ويقوم الزاحف بالوصول إلى صفحات الويب ونسخها إلى مستودع محرك البحث. وتستخدم برامج الزاحف الروابط الحديثة في الوصول إلى صفحات أخرى، ومن ثم تكون كل صفحات الويب قابلة للوصول والتكشيف بواسطة محركات البحث. وفي حالة وجود عوائق تمنع برامج الزاحف من أداء عملها مثل كلمات المرور أو مستويات التحليل العميقة أو تقنيات متطورة... فإن محرك البحث لا يرى هذه الصفحات وتدخل بذلك ضمن ما تعرف بالويب غير المرئية[٤]." ويتضمن هذا التعريف بعض الجوانب مثل:

١- الربط بين مصطلحي الويب المرئية والويب المكشفة، بما يؤكد أن الرؤية إنما هي رؤية محرك البحث لصفحات الويب حتى يتمكن من التعامل معها.

٢- اشتمل التعريف السابق على كل جوانب عمل محرك البحث، حيث تكون كل صفحات الويب قابلة للوصول والتكشيف بواسطة برنامج الزاحف.

وتشمل الويب المرئية مختلف أشكال صفحات ومواقع وملفات الويب والقليل من قواعد البيانات على الويب؛ حيث تكون هذه المعلومات متاحة أمام برامج الزاحف للتعامل معها ونسخها داخل مستودعات محركات البحث. وتستخدم برامج الزاحف تقنيات الروابط الفائقة بين صفحات الويب؛ لتشكل نسيجًا من مصادر المعلومات يظهر بعضه بعضًا.

٢/٣ بيئة الويب غير المرئية:

يعرف قاموس مصطلحات المكتبات والمعلومات ODLIS الويب غير المرئية تحت مصطلح الويب العميقة على أنها: "المعلومات المتاحة للوصول من خلال شبكة الويب العالمية، غير أنه لا يمكن استرجاعها بواسطة محركات البحث التي تعتمد على برامج الزاحف أو العنكبوت، مثل أشكال المعلومات التي توجد في ملفات pdf داخل قواعد البيانات التي تعتمد فقط على أسئلة البحث بشكل محدد. وتحوي الويب العميقة كمًا من المعلومات يفوق الويب السطحية بـ٤٠٠-٥٠٠ مرة، يوجد أكثر من نصفها داخل قواعد البيانات، وتستخدم الويب بعض الخدمات لاسترجاع هذه المعلومات مثل .CompletePlanet and ProFusion" (٥) ويظهر التعريف السابق بعض الجوانب مثل :

١- إن الهدف من المعلومات الرقمية هو الوصول إليها من خلال محركات البحث التي تعد نافذة مستخدمي الويب لرؤية مصادر المعلومات الرقمية.

٢- إن السبب الرئيس لظهور الويب العميقة هو وجود المعلومات الرقمية في أشكال متنوعة مثل ملفات ...pdf , mdb, ppt.

٣- إن المصدر الرئيس لتكوين الويب غير المرئية هو قواعد البيانات، حيث تتصف بخصائص مخالفة لطبيعة العمل داخل محركات البحث بما يصعب على المحركات الوصول إلى محتواها.

ويرى الباحث أن الويب المرئية تتكون من مجموعة من المصادر تعتمد في بنيتها الأساسية على بنية html التي تستخدمها محركات البحث للتعرف إلى المصادر حديثة الإضافة، إلى جانب ما بداخل هذه الصفحات من ملفات ذات روابط محددة يمكن لمحركات البحث المتخصصة الوصول إليها بحسب نوع الملفات، وذلك خلاف ما تكون عليه قواعد البيانات.

ويرى الباحث أن مصطلح قواعد البيانات بمثابة المرادف الموضوعي لمصطلحات الويب غير المرئية والويب العميقة والويب الخفية. ويوضح هذه الفكرة واقع العمل بين محركات البحث وقواعد البيانات، حيث تستطيع محركات البحث الوصول إلى أي مصدر معلومات دون قواعد البيانات لما لها من طبيعة بناء وتصميم لا تتوافق مع محركات البحث في العمل ومن ثم يشكل محتوى هذه القواعد بيئة جديدة غير مرئية أمام محركات البحث.

وقد أورد الباحث في دراسته للماجستير تفصيلاً لمستويات وأنواع قواعد بيانات الويب، حيث يمكن على أثرها تقسيم مستويات الوصول إلى معلومات الويب ابتداءً بالمعلومات ذات الرؤية غير الشفافة ثم انتهاءً بالشكل غير المرئي تماما لقواعد البيانات. وكانت مستويات رؤية المعلومات كالتالي[٦]:

١/٢/٣ الويب غير الشفافة:

وهي تشمل قواعد البيانات كبيرة الحجم على الويب، التي تتكون بدورها من مجموعات كبيرة من مواد المعلومات، وعلى الرغم من ملاءمة هذه المواد لعمل برنامج الزاحف وإمكانية اقتنائها داخل قواعد محركات البحث، إلا أن برامج الزاحف لا تستطيع الوصول إليها ويرجع Paul Pedley ذلك إلى عوامل مختلفة منها:

- عمق عملية الزحف: تعمل محركات البحث على زيادة عمق التحليل والبحث لبرامج الزاحف لتحقيق أفضل أشكال الاقتناء، خاصة في الوقت الذي تتزايد فيه أعداد صفحات مواقع الويب بشكل مستمر. وتعمق برنامج الزاحف داخل موقع الويب وتحليل كامل صفحاته لأجل جمعها من شأنه رفع تكاليف العمل داخل محركات البحث. وتتجنب محركات البحث إظهار مدى العمق الذي تعمل به برامج الزحف داخل صفحات الويب.
- تأخر التحديث الدوري: فإضافة صفحة جديدة إلى محرك البحث يحتاج إلى وقت طويل قد يصل إلى شهر أو أكثر حتى يتم تمييزه ببرنامج الزاحف، وعلى الجانب الآخر فإن الرجوع إلى مواقع سابقة للبحث عن التحديثات الأخيرة بها من صفحات مضافة يحتاج إلى وقت أطول من الوقت الذي يستغرقه التحديث. وعلى ذلك فإن جزءاً من مواد المعلومات المحدثة داخل مواقع الويب التي تم إضافتها مسبقًا سيظل بعيدًا عن أيدي الباحثين مدة من الوقت، وهو ما يحدث في قواعد البيانات كبيرة الحجم.
- ازدياد النتائج المستدعاة Maximum Number of Viewable Results : تتسم محركات البحث بكبر حجم قواعد بياناتها، ويتضح ذلك في أعداد النتائج المسترجعة داخل قاعدة البيانات التي قد يصل عددها إلى ملايين الصفحات، مما

يعني معه أن محركات البحث تعمل على الإهمال المتعمد واستبعاد النتائج الأقل صلة بموضوع البحث، ومثل هذه النتائج تظل دائما الجزء غير المرئي أمام الباحثين.

- زيادة الروابط غير النشطة: وهذا بدوره لا يرجع إلى عمل محرك البحث وإنما خطأ في تفعيل روابط بعض الصفحات التي لا يستطيع برنامج الزاحف الوصول إليها إلا باستخدام مسار الصفحة URL أو الإضافة اليدوية عن طريق مصممي صفحات الويب.

٣ / ٢ / ٢ الويب الخاصة:

وهي تتكون من مجموعة من صفحات الويب ذات الطابع الشخصي سواء التي تتبع أفرادًا أو هيئات رسمية، ومن ثم لا توافر إمكانات التفاعل مع محركات البحث. وعلى الرغم من قابلية هذه الصفحات للتكشيف بواسطة برنامج الزاحف، إلا أنها تستخدم حالة من الحالات الثلاث التالية لمنع دخول برنامج الزاحف إليها. وهذه الحالات هي:

- استخدام كلمات المرور للتعرف إلى محتوى الصفحة.
- تهيئة ملف robots.txt المحدد لعمل برنامج الزاحف لمنع التعامل مع صفحة الويب.
- استخدام حقل noindex meta tag للتوقف عن إضافة هذه الصفحة.

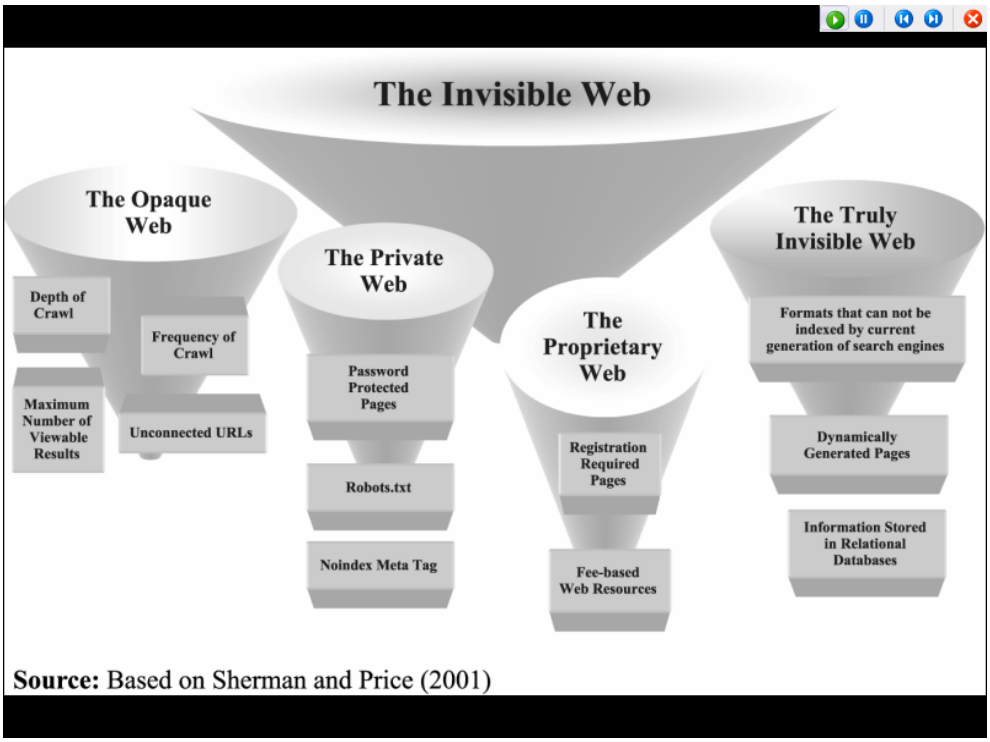
٣ / ٢ / ٣ الويب ذات الملكية:

ويقصد بهذا النوع مجموعة قواعد البيانات التي تعتمد على وجود تسجيل أو ربط Registration مع موقع القاعدة. وعلى الرغم من أن هذا النوع يقدم خدمات مجانية، فضلا عن سهولة إتمام عملية الربط بالقاعدة، إلا أن أفضل برامج الزاحف حتى الآن لا تستطيع إقامة رابط لقواعد البيانات. مما يجعل من الصعب عبور الزاحف إلى محتوى هذه القواعد.

٣ / ٢ / ٤ الويب غير المرئية الحقيقية: Thr Truly Invisible Web

يعد هذا الجزء من مواقع الويب مثل مواقع المكتبات الرقمية هو الواقع الفعلي لمفهوم الويب غير المرئية. ويمكن لأي جزء من صفحات الويب أن ينجلي أمام الباحثين

إذا ما امتدت إليه برامج محركات البحث؛ لتكشيفه وإخضاعه للبحث أمام المستفيدين، ويختلف الجزء الحالي من الويب في البناء الفني عن الشكل المستخدم لدى برامج الزاحف فلا يستطيع برنامج الزاحف استخدام محتوى تلك الصفحات. ويشير Paul Pedley إلى أن ذلك يتوقف على احتواء صفحات الويب على أنواع من الملفات لا تتعامل معها محركات البحث مثل الملفات المضغوطة وملفات pdf وملفات الفلاش Flash، ولا يكفي الجزء البسيط من حقول الميتاداتا المصاحب لهذه الملفات في إعانة برامج الزاحف على تحليلها وتكشيفها التي تحتاج إلى التعامل مع كامل المتن لصفحة الويب في تكشيفها وبحثها. وقد أطلقت بعض الدراسات مسمى الويب الأكاديمية على مواقع المكتبات الرقمية ومواقع مصادر المعلومات العلمية. ويوضح الشكل رقم (٢) مكان الويب غير المرئية الحقيقية بين مختلف مستويات الويب غير المرئية الأخرى(٧).



الشكل رقم (٢) الويب غير المرئية الحقيقية بين مختلف مستويات الويب غير المرئية الأخرى(٨).

وقد حددت[٩] Laura Cohen فئات عامة لمصادر الويب غير المرئية، وهي التي تشكل الخصائص العامة لتكوين هذه الويب. وجاءت الفئات كما يلي:

الفئة الأولى : محتوى قواعد البيانات : يتكون محتوى قاعدة البيانات على الويب من معلومات تم اختزانها داخل جداول صممت باستخدام برامج مثل: Access, Oracle, SQL Server and DB2، ويتم الوصول إلى هذا المعلومات باستخدام استفسارات البحث الملائمة لبنية قاعدة البيانات. وعلى ذلك فمحتوى قواعد البيانات يختلف عن محتوى صفحات الويب الذي يتم الوصول إليه مباشرة.

الفئة الثانية : ملفات الويب غير النصية : مثل ملفات الوسائط المتعددة أو الرسومية أو البرامج أو تلك الملفات في شكل pdf , mdb, ppt and doc.

الفئة الثالثة : المحتوى محدد الوصول : وهي مجموعة المواقع التي تم حمايتها باستخدام كلمات مرور أو برامج حماية نظرًا لطبيعة المحتوى الحساسة. أو تلك الصفحات التي تتيح محتواها فقط من خلال استمارات الاشتراك لمستخدمي الويب.

الفئة الرابعة : المحتوى التفاعلي (الديناميكي) : وهي صفحات الويب التي تضع استمارات إدخال للمعلومات وتتغير بياناتها بواسطة مستخدمي الموقع.

الفئة الخامسة : الصفحات فاقدة الروابط Unlinked content – pages **:** حيث يسبب افتقاد هذه الصفحات إلى روابط فائقة مع الصفحات الأخرى قطع طريق الوصول أمام برامج الزاحف.

الفئة السادسة : صفحات ذات تقنيات محددة Scripted content - pages **:** حيث يتم الوصول إلى هذه الصفحات بواسطة روابط رسومية قد لا يتمكن برنامج الزاحف من التعامل معها.

١/٤/٢/٣ خصائص الويب غير المرئية:

إن مصطلح الويب غير المرئية قد نال اهتمامًا بالغًا في أدبيات الإنتاج الفكري حول بيئة الويب، وعلى الرغم من ذلك فإن بعض الباحثين مثل[١٠] Cohen, Laura يرون أن

مصطلح "invisible Web" يعاني من فقر في الدلالة والتعبير عن واقع الويب. وذلك يرجع في نظرهم إلى الأسباب التالية:

١- يعبر مصطلح "invisible Web" فقط عن آلية عمل محركات البحث؛ حيث يفترض المصطلح أن كل ما تصل إليه محركات البحث من مصادر الويب هو الويب المرئية، وان غير المرئي من المصادر هو ما لا تستطيع الوصول إليه وتكشيفه داخل مستودعاتها. ومن ثم افتراض أن مستخدمي الويب يستطيعون فقط استرجاع معلومات الويب من خلال محرك البحث، ولا أمل لهم في استرجاع معلومات أخرى خارج محرك البحث، إلا أن يؤدي الحظ دوره في ذلك.

٢- لا يمكن القول أن هناك معلومات رقمية مسجلة يطلق عليها غير مرئية، وإنما يمكن القول أن هناك معلومات أصعب بدرجة ما في الحصول عليها من المعلومات الأخرى.

٣- إن قواعد البيانات على الإنترنت موجودة منذ وقت طويل مما جعل مستخدمي الويب على قدر من التمرس بالتعامل وطرح الاستفسارات البحثية المباشرة لاسترجاع محتوى هذه القواعد.

٤- إن محتوى محركات البحث ذاته يتم اختزانه في قواعد بيانات، وهذا ألا يشكل بطبيعته جزءاً من الويب غير المرئية، حيث لا يستطيع مستخدم الويب استرجاع ما بداخل محرك البحث سوى باللجوء إلى محرك البحث وطرح الاستفسارات البحثية داخل واجهة البحث.

غير أن الباحث يتبنى رؤية مخالفة تمامًا ترد على الجوانب السابقة؛ حيث تجاهلت الأسباب السابقة حقيقة أن محركات البحث هي واسطة العقد في عمل الويب بين منتج لصفحة الويب ومستخدم لها : فمن غير الممكن أبدًا أن يحفظ مستخدمو الويب جميع مسارات صفحات ومواقع وقواعد بيانات الويب. هذا إلى جانب أن عددًا قليلاً من محركات البحث يمكنه نظريًا تغطية كل المعلومات الرقمية الواقعة في فضاء الويب آخذين في الاعتبار التقدم والتطوير الذي تشهده محركات البحث.

إن ثمة حقيقة يجب التعرض لها وهي: الربط بين قواعد البيانات عامة وبين الويب غير المرئية؛ فمصطلح الويب غير المرئية يطلق فقط على قواعد البيانات التي تعمل في بيئة الويب html, xml، وذلك لأن هناك كثيرًا من قواعد البيانات التي تعمل على الخط المباشر والفارق جلي بين النوعين، حيث تحمل قاعدة بيانات الويب خصائص خلاف ما توجد عليه قواعد الخط المباشر، ويتضح ذلك من أساليب الاسترجاع والعمل لكل منهما. تختلف بيئة الويب عن الخط المباشر فيما يلي(١١):

- يعني مصطلح الخط المباشر تلك القواعد المتاحة على أقراص ضوئية CDs أو قاعدة محلية الاستخدام أو فهرسًا متاحًا على شبكة داخلية للمكتبة OPAC.
- ترتبط قواعد بيانات الخط المباشر بالتعامل مع الشكل العلائقي لقواعد البيانات RDBMS.
- يعني مصطلح الويب تلك الملفات المعالجة في بيئة html, xml، التي يتم الوصول إليها عن بعد أو المختزنة في قواعد بيانات الويب.
- يعمل الاسترجاع على الخط المباشر بخصائص البحث والمضاهاة لمجموعات البيانات الببليوجرافية اعتمادًا على آليات البحث البولياني. أما الاسترجاع على الويب فغالبًا مع يعمد إلى التكشيف والاسترجاع للنص الكامل، وهو ما تقوم به محركات بحث الويب.
- تعمل بيئة الويب باعتبارها بيئة تفاعلية دائمة التغيير والإضافة والتحديث، خلاف ما يوجد عليه الخط المباشر من ثبات أو بطء في التغيير.

٤- العلاقة بين المكتبات الرقمية والويب غير المرئية:

لقد أتت الويب غير المرئية كسحابة تُظِل أرجاءً كثيرة من أرجاء الويب، وكان أهم هذه الأرجاء هي قواعد بيانات المكتبات الرقمية، حيث عجزت محركات بحث الويب العامة عن التفاعل مع قواعد بيانات النصوص الكاملة والببليوجرافية من حيث الوصول والمعالجة واسترجاع المجموعات الداخلية لقواعد البيانات. وقد نقلت المكتبات الرقمية قواعد بيانات النظم المتكاملة من البيئة الشبكية المحدودة إلى بيئة الويب التي

تضطلع محركات البحث داخلها بمهام التنظيم والبحث، وعلى ذلك فإن قواعد بيانات فهارس المكتبات الرقمية أضحت في مواجهة التفاعل مع محركات البحث معالجة وبحثًا واسترجاعًا للتسجيلات الببليوجرافية. وقد تعامل الباحث مع ظاهرة الويب غير المرئية واحتياجها لجيل متطور من محركات البحث بشكل مفصل في أطروحته للدكتوراة، وانتهى العمل بتقديم تصور مفصل لبنية وآليات عمل محركات بحث الويب غير المرئية.

وبات من المسلم به لمحركات البحث أن مختلف قواعد بيانات الويب سواء تلك الببليوجرافية أو قواعد النصوص الكاملة أو متعددة الوسائط، إنما هي الصورة الحقيقية للويب غير المرئية. ونظرًا لأن المكتبات الرقمية تعتمد على أنظمة المعلومات المتكاملة التي تقف على قواعد البيانات الببليوجرافية، فإنه يمكن القول أن قواعد بيانات المكتبات الرقمية هي ويب غير مرئية لمحركات البحث. ومن ثم فإننا بصدد الحديث عن وافد جديد من الموضوعات للمكتبات الرقمية على الويب يتمثل في معالجة وبحث قواعد بيانات المكتبات الرقمية وفهارسها باستخدام محركات بحث الويب غير المرئية.

١/٤ الفهارس الرقمية OPACs كجزء من الويب غير المرئية:

تتجه الويب في هذه الآونة إلى معالجة المحتوى العلمي من مصادر الويب، وذلك من خلال تفعيل تنظيم واسترجاع محتوى مواقع الويب الأكاديمية؛ حيث عملت بعض محركات بحث الويب على الربط بين قواعد بياناتها وبين مواقع الدوريات العلمية والصحف الإلكترونية e_gournals ومجموعات النقاش والمواقع الشخصية لكثير من الباحثين إلى جانب مخاطبة مواقع الهيئات والمؤسسات العلمية. ويعزى ذلك كله إلى توجه محركات البحث لإظهار عمق الويب غير المرئي من مصادر المعلومات ذات البنية المهيكلة كقواعد البيانات، وذلك ليس فقط في المعلومات العلمية الأكاديمية وإنما في مختلف مجالات مصادر المعلومات غير المرئية. ولقد حرص الاتحاد الأمريكي للمكتبات الرقمية American Digital Library Federation (ADLF) في لقائه الدوري في ربيع عام ٢٠٠٤م على تبني فكرة الموافقة والربط بين قواعد معلومات المكتبات

الرقمية على الويب وبين أدوات بحث وتنظيم الويب، وذلك بهدف إدخال المكتبات الرقمية كواحدة من حلقات نشر وتنظيم المعلومات الرقمية العلمية على الويب في دائرة تنظيم محتوى الويب الأكاديمية غير المرئية(١٢).

إن نجاح محركات بحث الويب غير المرئية في التوافق وبنية قواعد البيانات على الويب، قد أثمر عن إمكان الربط بين محركات بحث الويب وبين قواعد بيانات المكتبات الرقمية. ولقد ظلت المكتبات الرقمية على الويب تعمل على إتاحة فهارسها على الخط المباشر حتى تقدم لمستخدمي الويب ما تملكه من مصادر المعلومات، غير أن حركة تداول المعلومات الرقمية قد جعلت المكتبات تتجه إلى ما هو أبعد من ذلك وأفضل، وهو تزويد قواعد بيانات المكتبات الرقمية بمجموعات من النصوص الكاملة لمصادر المعلومات سواء النصية أو غير النصية. ومن ثم أصبحت نظم المكتبات المتكاملة على الويب على استعداد تام للتفاعل مع استفسارات محركات البحث والرد عليها بمجموعات التسجيلات أو النصوص الكاملة التي تلبي الحاجة المعلوماتية للاستفسار.

تقف نظم المكتبات المتكاملة خلف إدارة قواعد بيانات المكتبات الببليوجرافية، ومن ثم فإن تركيب وإدارة الفهارس الرقمية على الويب WEB OPACs يتوقف على مدى نجاح النظام المتكامل في التوافق مع محركات بحث الويب غير المرئية، ويرى الباحث أن هذه المسألة لا بد أن تأخذ قدرًا كبيرًا من الشركات المنتجة لنظم المكتبات المتكاملة. وهو أيضًا ما يجب أن تضعه مختلف المكتبات في شروطها ومواصفاتها في النظم المتكاملة المطلوبة، حتى تدعم إدارة مواقعها الرقمية على الويب.

٢/٤ قواعد بيانات مصادر المعلومات:

لم تعد المكتبات التقليدية أماكن لحفظ وتنظيم مصادر المعلومات المطبوعة أو غيرها، كذلك فإن المكتبات الرقمية ينبغي أن تتخلى عن كونها واجهة لبحث واسترجاع مجموعات التسجيلات الببليوجرافية فقط أو إعلام المستفيدين عن مواعيد عمل المكتبة التقليدية أو تقديم خدمات الإحاطة الجارية وغيرها من الخدمات

الإرشادية، وإنما أضحت المكتبة الآن جزءاً حيويًّا من كيان معلوماتي هائل يسمى الويب. هذه الويب الآن لا تحتاج إلى مواقع للمكتبات الرقمية تعمل بنمطها الحالي، وإنما تحتاج إلى مواقع للمكتبات أكثر تفاعلية وديناميكية مع أدوات بحث وتنظيم الويب من محركات وأدلة وبوابات، ومن ثم فإن قاعدة بيانات المكتبات الرقمية لن تحتوي فقط على مجموعات من التسجيلات الببليوجرافية، وإنما ستصل إلى حدود قواعد النصوص الكاملة وملفات المعلومات من الوسائط المتعددة. وهذه القواعد سوف تعتمد على آليات متعددة متقدمة في خصائص الإضافة والمعالجة والاسترجاع تتوافق من خلالها مع مواقع المعلومات العلمية والدوريات الرقمية على الويب من ناحية ومحركات البحث من ناحية أخرى.

إن قواعد المكتبات الرقمية من مصادر معلومات النصوص الكاملة تأخذ اتجاهين من التفاعلية مع الويب غير المرئية؛ يتمثل أولهما: في تنمية مجموعات النصوص الكاملة وملفات المعلومات من الوسائط المتعددة بشكل آلي بشكل منتظم، ويمكن عمل ذلك بفتح آفاق التعامل بين موقع المكتبة الرقمية وبين مصادر الويب من الدوريات والمواقع العلمية والمواقع الشخصية للباحثين بما يمكن معه النظر إلى المكتبة الرقمية على أنها بوابة كبيرة لتنظيم ومعالجة المعلومات العلمية على الويب. أما الاتجاه الثاني فهو: توافق وتفاعل قاعدة مصادر المعلومات داخل موقع المكتبة مع آليات البحث المتعدد التي تتمثل في محركات البحث المتعدد ومحركات بحث الويب غير المرئية. وعندئذ يمكن لقاعدة مصادر المعلومات الإضافة والاسترجاع باعتماد نوافذ متعددة تسمح للمكتبة الرقمية باتساع حدود العمل والتخلي عن مبدأ الحصر والتضييق في التعامل مع التسجيلات الببليوجرافية. وثمة دعائم محددة لإنجاح هذه الفكرة، ويشكل غياب هذه الدعائم معوقات ينبغي أن تتعامل معها المكتبات الرقمية بأدق قدر وأسرع وقت، وهذه الدعائم كالتالي[١٣] :

١- **قدرات البحث للأشكال المتعددة والنصوص الكاملة :** حيث تفتقر مواقع المكتبات الرقمية ونظمها المتكاملة إلى البحث في أشكال المعلومات المتباينة والوسائط المتعددة، إلى جانب تدني قدرات البحث في النصوص الكاملة.

٢- **التغطية الشاملة لأنواع محتوى الويب :** تحتاج المكتبات إلى تفاعل أكثر مع مواقع المصادر العلمية مثل: الدوريات الرقمية ومجموعات النقاش ومواقع الهيئات والاتحادات العلمية.

٣- **قدرات البحث والاسترجاع المتعدد وآلياته :** حيث لا تعمل نظم المكتبات المتكاملة بشكل فعال مع نظم البحث المتعدد، وهو ما يمثل ركنًا أساسيًا في إنجاح التفاعل ومحركات بحث الويب غير المرئية.

٤- **تطوير آليات البحث :** حيث تتجه محركات بحث الويب غير المرئية إلى استخدام آليات بحث أكثر تقدمًا من معاملات المنطق البولياني التي تتوافق أكثر مع حقول البحث الببليوجرافية، مثل استخدام معاملات التحليل اللغوي والبحث بالتقارب بين الكلمات وتوسيع سؤال البحث... وغيرها.

٥- **تحسين قدرات الترتيب والتنظيم والعرض لنتائج البحث :** حيث تؤثر عمليات ترتيب النتائج ونمط استرجاعها على كفاية آليات الدمج والترتيب في عملية البحث المتعدد التي تعد محور أداء محركات بحث الويب غير المرئية.

٥- واقع محركات بحث قواعد بيانات المكتبات الرقمية:

يسعى الباحث في هذا الجزء إلى دراسة آليات العمل داخل محركات بحث الويب غير المرئية المنوطة بالتفاعل مع قواعد بيانات المكتبات الرقمية، مستخدمًا في ذلك وصف البرامج التي تعمل بها محركات الويب. وسوف ينصب التركيز على الجوانب المؤثرة في تفاعل محرك بحث الويب غير المرئية مع قواعد بيانات المكتبات الرقمية وصولاً ومعالجة وبحثًا، ومن هذه الجوانب التعرف إلى خصائص واجهات البحث التي ستكون الواجهة العامة لبناء إستراتيجية البحث المزمع إرسالها بعد ذلك إلى مختلف قواعد بيانات المكتبات الرقمية لبحث تسجيلاتها، أيضًا التركيز على برنامج الزاحف ومحاولة رصد الإشارات الدالة على نجاحه أو إخفاقه في التعامل مع محتوى قواعد بيانات المكتبات الرقمية وواجهات عملها. بعد ذلك يأتي تناول واجهات عرض النتائج والخدمات التي تقدمها محركات البحث لاستكمال خصائص المعالجة

والتنظيم لمجموعات قواعد بيانات المكتبات الرقمية على الويب. وتتمثل جوانب الدراسة التطبيقية لمحركي البحث عينة الدراسة في العناصر التالي:

- **واجهات بحث محركات بحث الويب غير المرئية:**

١- نموذج البحث.

٢- أساليب البحث.

٣- آليات البحث المستخدمة.

٤- حقول البحث الببليوجرافية.

- **برامج الزاحف وقواعد المكتبات الرقمية:**

١- مسح روابط الويب.

٢- التعامل مع بحث قواعد البيانات الببليوجرافية.

٣- تكشيف قواعد البيانات الببليوجرافية.

- **واجهات عرض نتائج بحث قواعد البيانات الببليوجرافية:**

١- ترتيب نتائج البحث.

٢- خصائص الفرز والاستبعاد.

٣- أنماط عرض الوحدات المسترجعة.

- **تنظيم الوصول إلى المكتبات الرقمية:**

١- التقسيم الموضوعي العام لقواعد المكتبات الرقمية.

٢- التكشيف وإتاحة التصفح الموضوعي.

١/٥ واجهات بحث محركات بحث الويب غير المرئية:

تتمتع واجهة البحث داخل أدوات البحث عامة بقدر كبير من الأهمية قد تفوق به بعض المكونات الرئيسية لأداة بحث الويب. وتتمثل هذه الأهمية في كونها أول ما يجده المستفيد من تفاعل بينه وبين أداة البحث، فضلاً عن كونها منفذًا وسبيلاً للتعبير عن

الحاجة الموضوعية والمعلومات التي يبتغيها المستفيد. وإذا ما نجح المستفيد في إدراك الجوانب الوظيفية لواجهة البحث، كان بحثه عن المعلومات مثمرً بما يتصل بموضوع البحث. وإذا كانت درجة الأهمية لواجهة البحث في أدلة ومحركات بحث الويب المرئية كبيرة، فإن أهمية واجهة البحث ستكون أكبر عند الحديث عن واجهات محركات بحث قواعد بيانات المكتبات الرقمية؛ حيث ستكون واجهة البحث هنا بديلاً عن واجهات كثيرة تقدر بعدد واجهات المكتبات الرقمية التي يصل إليها محرك البحث، حيث يجب أن تبلغ من التوافق مع واجهات قواعد بيانات المكتبات الرقمية الدرجة التي تجعل المستفيد قادرًا على إرسال كلمات وحقول البحث المفتاحية إلى مختلف قواعد البيانات الببليوجرافية.

الجدول رقم (٢) شكل واجهة محرك بحث قواعد بيانات المكتبات الرقمية

أساليب البحث	محركات بحث المكتبات الرقمية	
	infomine	scirus
واجهة بسيطة مباشرة	- - -	- - -
واجهة معقدة مركبة	√	√

يتضح من الجدول رقم (٢) أن محركي البحث يعتمدان على استخدام واجهات البحث المركبة أو المعقدة، ويعزو الباحث ذلك إلى أن آليات ومعاملات البحث التي تميل إلى الشكل المقنن أو الأكاديمي دائمًا ما تستخدم واجهات البحث المعقدة حتى تتيح للمستفيد استخدام عناصر البحث بالحقول واستخدام معاملات البحث البولياني، إلى جانب رسم إستراتيجية البحث من حيث الحدود الزمنية أو حصر البحث على واحدة أو أكثر من المكتبات المتوافقة ومحرك البحث. هذا إلى جانب تواجد واجهات البحث البسيطة التي تمثل الشكل العام للبحث دون التعرض لتفاصيل أو محددات بحث كثيرة. وقد اتفق محركا البحث في هذا الجانب.

١/١/٥ أساليب البحث:

تتنوع آليات وأساليب البحث على الويب عامة بين أسلوبين أساسيين هما: البحث الحر بالكلمات المفتاحية والبحث بالتصفح. ويندرج تحت كل منهما مختلف الآليات

التي تعتمد عليها محركات بحث المكتبات الرقمية؛ فعلى سبيل المثال سواء اعتمدت أداة البحث على التقسيم الموضوعي بالمصطلحات أو الأشكال المصورة مثالا أو التصفح من خلال الوحدات والتسجيلات ذاتها ، فإن كل ذلك يسمى في النهاية بأسلوب البحث بتصفح المعلومات. كما أن استخدام الحقول الببليوجرافية أو آليات البحث البولياني مباشرة أو اعتماد آليات أخرى تتكون من كلمات مفتاحية ، فإن كل ذلك يسمى نهاية بأسلوب البحث الحر بالكلمات الدالة.

الجدول رقم (٣) أساليب البحث داخل محركات بحث قواعد بيانات المكتبات الرقمية

أساليب البحث		محركات بحث المكتبات الرقمية	
		infomine	scirus
البحث الحر		√	√
البحث بالتصفح	بالمصطلحات	√	√
	عنوان الموقع	- - -	- - -

يتضح من الجدول رقم (٣) أن محركي البحث عملا على توفير أسلوبي البحث الحر والتصفح المعتمد على التقسيم الموضوعي العام لمستوى أو مستويين من تقسيمات المعرفة. ويرجع ذلك إلى الحرص على توفير أسلوبي البحث الحر الملائم للمستفيدين ذوي المهارات المتقدمة في البحث والمحددين بدقة لاتجاهاتهم واحتياجاتهم الموضوعية. أما البحث بالتصفح فيلائم أكثر المستفيدين ذوي قدرات البحث المحدودة أو غير القادرين على الصياغة والبناء المباشر لإستراتيجية البحث. غير أن محرك البحث **infomine** قد عمل على توفير تقسيم موضوعي مقنن يعتمد على الترتيب الهجائي لرؤوس الموضوعات، وقد لاحظ الباحث شدة ضبط المصطلحات التي تقترب إلى توفير عامل التفريع أو التجزئة لرؤوس الموضوعات كما في قائمة رؤوس الموضوعات العربية الكبرى. ويقترح الباحث اعتماد محرك بحث المكتبات الرقمية على تقسيم منطقي للمعرفة يتبع في نسقه أحد أنظمة التصنيف كتصنيف ديوي أو العشري العالمي. حتى يمكن اتساق عمليات البحث مع تنظيم وبناء مجموعات المكتبات الرقمية ، وتوفير درجة أكبر من معاملات التحقيق والدقة في البحث.

٢/١/٥ نموذج البحث:

يبنى نموذج البحث في أدوات بحث الويب عامة داخل واجهات تلك الأدوات. ويتكون نموذج البحث من مجموعة من عناصر وحقول بحث تتكامل معًا لاسترجاع المعلومات؛ حيث تستقبل هذه العناصر الكلمات المفتاحية بحسب توزيعها بين حقول المؤلف والموضوع والشكل... إلخ، ثم توجه معًا كإستراتيجية بحث تهدف إلى استرجاع معلومات ذات سمات محددة تتفق واستفسار البحث. ويتكون نموذج البحث من مكونين رئيسيين هما؛ أولا : مسمى عنصر البحث، وهو يدل على طبيعة المكان الذي ستصل إليه الكلمة الدالة مثل مسمى حقل المؤلف، حقل العنوان، حقل الموضوع. أما المكون الثاني فهو عنصر البحث ذاته، وقد يكون عبارة عن فراغ نصي text box لتلقي الكلمات الدالة، أو عنصر تحديد خصائص البحث chek box . وتتدرج نماذج بحث أدوات الاسترجاع على الويب بداية من النماذج البسيطة التي تقتصر على فراغ البحث لاستقبال الكلمات المفتاحية، ثم الانتهاء بنماذج البحث المتقدمة المعقدة التي تتكون من عناصر بحث مختلفة الوظيفة والأداء مثل؛ فراغات الكلمات المفتاحية، وأزرار الخيارات chek box وأزرار التحديد radio buttons.

الجدول رقم (٤) نموذج البحث في محركات بحث قواعد بيانات المكتبات الرقمية

نموذج البحث		محركات بحث المكتبات الرقمية	
		infomine	scirus
نموذج عام		- - -	- - -
نموذج متعدد	فراغ النص	√	√
	زر التحديد	- - -	- - -
	زر الخصائص	√	√
	قائمة الاختيارات	√	√
	عنصر الترتيب والعرض	√	√

يتضح من الجدول رقم (٤) أن محركي البحث قد عملا على استخدام معظم عناصر البحث التي تتواجد في نماذج البحث المركبة أو المعقدة، حيث تساهم مختلف عناصر بحث النموذج في ضبط وتدقيق إستراتيجية البحث النهائية، ومن ثم فإن

معاملات التحقيق والاستدعاء تكون قابلة للضبط اليسير من قبل المستفيد، وهو ما يعني تحكم أكثر في طبيعة المواد المسترجعة من المكتبات الرقمية الخاضعة لعمليات البحث من قبل المحرك.

٣/١/٥ آليات البحث المستخدمة:

تمثل آليات البحث المكون الثالث من مكونات محركات البحث إلى جانب برنامجي الزاحف والمكشف. وترتبط آليات البحث باعتماد محرك البحث على أسلوب البحث الحر بالكلمات الدالة، ومن ثم تأتي آليات البحث لتساعد على تكوين الشكل الملائم للبحث بالكلمات البحثية. وتتعدد أنواع آليات البحث بين تلك التي تستخدم مع الكلمات المفردة مثل؛ البحث بحساسية الحروف، وآلية البحث الحر wild card، وتلك الآليات التي تستخدم مع الكلمات المتعددة مثل البحث بالتطابق، والبحث بالعبارات واستخدام المنطق البوليني boloean logic. وإذا كانت آليات البحث قد أخذت أهمية في عمل أدوات بحث الويب المرئية، فإنها ستأخذ أهمية أكبر في أدوات بحث قواعد بيانات المكتبات الرقمية على الويب؛ ذلك لأن دور آليات البحث هنا يكمن في التعامل مع نماذج بحث قواعد البيانات الببليوجرافية التي تملك أشكالاً مختلفة من آليات البحث، ومن ثم وجب التوافق بين الآليات التي تعتمد عليها محركات بحث الويب غير المرئية، وبين تلك المستخدمة في قواعد بيانات المكتبات الرقمية.

الجدول رقم (٥) آليات البحث المستخدمة في قواعد بيانات المكتبات الرقمية.

آليات البحث المستخدمة	محركات بحث المكتبات الرقمية	
	infomine	scirus
البحث بالتطابق	√	√
البحث الحر ? , *	❖ في آخر الكلمة	؟ في آخر الكلمة
حساسية الحروف	- - -	- - -
بحث بالعبارات والتوقف	- - -	- - -
معاملات المنطق البوليني	√	√
البحث برابط التسجيلة	√	- - -
التصحيح الإملائي	√	- - -

يشير الجدول رقم (٥) إلى أن محرك بحث scirus لا يهتم بتوفير آليات بحث المجموعات والمقالات بالقدر الذي يعمل به محرك infomine، حيث قدم الأخير إمكان استدعاء التسجيلات برابط مخصص لتسجيلة محددة يمكن للمستفيدين عن طريقه استدعاء وحدات من المقالات أو مصادر المعلومات دون غيرها، وهو ما لم يتوافر في محرك scirus الذي يتجه بالبحث إلى مقالات الدوريات العلمية أكثر منه إلى المكتبات الرقمية.

٤/١/٥ حقول البحث الببليوجرافية:

يقصد بالبحث بالحقول هنا الاعتماد على رموز حقول الوصف الببليوجرافي في استرجاع وحدات من المعلومات تشترك معًا في مجموعة من الحقول مثل تاريخ نشر محدد أو مؤلف واحد أو موضوع مخصص. تختلف محركات بحث الويب غير المرئية عن محركات بحث الويب المرئية في أن الثانية تمتلك قاعدة بيانات تحتوي بداخلها على مجموعات المعلومات التي يملكها محرك البحث، وهذه المعلومات جاءت إليه نتيجة لعمل برنامج الزاحف في نسخ وإضافة مواقع وصفحات الويب إلى قاعدة بيانات المحرك. غير أن محرك بحث الويب لقواعد بيانات المكتبات الرقمية سيتوقف هنا عند حد الوسيط في البحث بين المستفيد وبين تسجيلات قاعدة البيانات، فضلاً عن أن هذا المحرك لن يملك القدرة على احتواء مختلف تسجيلات قواعد بيانات الببليوجرافية والنصوص الكاملة التي سيصل إليها، إنما الدور المنوط به فقط هو إجراء البحث في أكثر من قاعدة بيانات في آن واحد، تمامًا مثل ما يقوم به محرك البحث المتعدد عند إجراء البحث في محركات بحث الويب المرئية. ومن ثم فإن احتواء نموذج البحث في محركات المكتبات الرقمية على خيارات البحث بالحقول الببليوجرافية، إنما هو للتوافق مع قواعد البيانات الببليوجرافية، وليس لبحث مجموعاتها الداخلية.

الجدول رقم (٦) البحث بحقول البحث الببليوجرافية في قواعد بيانات المكتبات الرقمية.

حقول البحث الببليوجرافية	محركات بحث المكتبات الرقمية	
	infomine	scirus
المؤلف أو المنشئ	√	√

العنوان	√	√
رؤوس الموضوعات	√	√
الجهة أو الناشر	√	√
النوع أو الشكل	√	√
رابط ومحدد التسجيلة ID	√	- - -
تواريخ مرتبطة	√	√
الترقيم الدولي	---	√

يتضح من الجدول رقم (٦) أن صفة البحث والاسترجاع العلمية المقننة في محركي البحث قد انعكست على مجموعات حقول البحث الببليوجرافية المخصصة لإجراء عمليات الاستفسار والاسترجاع لمصادر المعلومات داخل المحركين. ويمكن القول أنه كلما اتجه محرك البحث إلى اعتماد عناصر وحقول الوصف الببليوجرافي في جزء من صياغة إستراتيجية البحث، فإن ذلك يدل قطعًا على تفاعل المحرك مع تسجيلات ببليوجرافية مبنية وفقًا لمعايير وصف ببليوجرافي محددة. كما أن ذلك يعطي إمكان طرح استفسارات البحث من محرك البحث إلى قاعدة بيانات المكتبة الرقمية بتوافق يوفر أعلى درجات التحقيق والدقة في استرجاع نتائج البحث.

٢/٥ برامج الزاحف وقواعد المكتبات الرقمية

إن برنامج الزاحف في الويب غير المرئية عامة ومحركات بحث المكتبات الرقمية خاصة سوف يعمل على اكتشاف وتحليل نماذج البحث search forms بدلاً من اكتشاف وتعقب الروابط الفائقة URLs. وللقيام بمهام تكشيف وتحليل واجهات وقواعد المكتبات الرقمية، ينبغي على برنامج الزاحف التحلي بمجموعة من آليات العمل التي تختلف كلية عما كان عليه داخل محركات الويب المرئية. ومن هذه الآليات؛ إمكان التعرف إلى الواجهات التي تحوي نماذج البحث داخل مواقع وقواعد المكتبات الرقمية، وتمييزها عن غيرها من مواقع الويب الأخرى، أيضًا التحلي بآلية القدرة على تحليل وتكشيف هذه النماذج لأجل تضمينها ضمن مجموعات محرك المكتبات الرقمية. إن محركات بحث الويب المرئية أو غير المرئية عامة ترتبط ارتباطًا وثيقًا بواقع برنامج الزاحف بها؛ حيث يمثل تواجد برنامج الزاحف وجه الاختلاف

الأساس بين المحرك وبين الدليل كأداتين لبحث الويب. ومتى كان محرك البحث متوقفًا عن العمل أو لا يوجد ضمن مكونات محرك البحث، فإن محرك البحث هذا لا يعد محرك بحث للويب من الناحية الدلالية والتطبيقية. وهذا الوصف ينسحب أيضًا على محركات بحث الويب غير المرئية، وليس فقط محركات الويب المرئية(١٤).

الجدول رقم (٧) برامج الزحف داخل محركات بحث قواعد بيانات المكتبات الرقمية.

برامج زحف محركات البحث		محركات بحث المكتبات الرقمية	
		infomine	scirus
فاعلية الزاحف	نشط	- - -	- - -
	غير نشط	√	√
اتجاه العمل بروابط الويب	للصفحات	- - -	- - -
	للنماذج	- - -	- - -
تعامل المحرك مع قواعد البيانات	تكشيف عام	√	√
	تكشيف المحتوى	√	- - -
تكشيف نماذج البحث		- - -	- - -

يكشف الجدول رقم (٧) عن حقيقة أن محركي بحث العينة لا يملكان برنامج الزاحف ضمن مكوناتهما الداخلية، وهو ما تعكسه طبيعة وحدات النتائج المسترجعة؛ حيث يقوم محركا البحث بأداء دور محرك البحث المتعدد الذي يرتبط مسبقًا بمجموعة من المحركات أو قواعد بيانات الويب، ثم يعمل على توجيه استفسارات البحث إليها ولتلقي النتائج ودمجها وترتيبها. كما عكس ارتباط محركي بحث العينة بمجموعة محددة من مصادر المعلومات العلمية على الويب، ممثلة في الدوريات الرقمية أو غيرها، أن محركي البحث مقيدان بالبحث في مصادر محددة لا يتم تحديثها على مدد قريبة، وهو ما كان سيوفره برنامج الزاحف من قدرات التحديث والإضافة السريعة. وقد عكست قوائم المصطلحات الموضوعية المرتبة هجائيًا في محرك البحث infomine أن عمليات التكشيف وتحليل محتوى مصادر المعلومات تتم بشكل أقرب لليدوي، وهو ما يصعب في حضور برنامج الزاحف.

٣/٥ واجهات عرض نتائج بحث قواعد البيانات الببليوجرافية:

تولي أدوات بحث الويب جميعها أهمية خاصة لواجهات عرض نتائج البحث؛ حيث تمثل هذه الواجهات ثمرة الرد على استفسار البحث الخاص بالمستفيد. ولا تقتصر عناية أدوات بحث الويب بواجهات عرض النتائج على الشكل العام لواجهة العرض، وإنما يكون الاهتمام الأكبر بطرائق ترتيب وتنظيم الوحدات المسترجعة بحسب درجة صلتها بموضوع البحث. ويزيد هذا الأمر في أهميته عندما تسترجع أداة البحث على الويب مجموعات مختلفة من نتائج البحث يتم تلقيها من قواعد بيانات مختلفة، مما يعني أن تقوم أداة البحث هنا بعمليات الفرز والاستبعاد والتنظيم وإعادة الترتيب مرة أخرى، وهو ما يحدث تمامًا في محركات البحث المتعددة، وما سيحدث في محركات بحث مجموعات المكتبات الرقمية. ومن ثم فإن الاهتمام بآليات عرض النتائج لا يقل أهمية عن باقي العمليات التي يقوم بها محرك بحث قواعد بيانات المكتبات الرقمية.

الجدول رقم (٨) واجهات عرض النتائج المستدعاة لمحركات بحث قواعد بيانات المكتبات الرقمية

واجهات عرض النتائج المستدعاة		محركات بحث المكتبات الرقمية	
		infomine	scirus
خيارات عرض وترتيب النتائج	أثناء الاستفسار	√	√
	أثناء العرض	- - -	- - -
خصائص الفرز والاستبعاد	ذاتية	√	- - -
	اختيارية	- - -	- - -
أنماط ترتيب النتائج	بالموضوع	√	√
	الحداثة	- - -	- - -
	أخرى	عنوان الوحدة	- - -

تميز محرك بحث infomine بتوفير خيارات التحكم في عرض النتائج بدرجة أكبر من محرك بحث scirus الذي وضع بعض خيارات عرض النتائج المحدودة تحت رابط الخيارات Preferences. وإذا كانت آليات البحث ذات أهمية في استرجاع النتائج، فإن خيارات عرض النتائج ذات أهمية أكبر في توفير جهد ووقت ومتابعة الوحدات

المسترجعة؛ حيث يفضل بعض المستفيدين تصفح نتائج البحث وفقًا لدرجة الصلة بالموضوع، بينما يفضل البعض الآخر تصفحًا وفقًا لحداثة المعلومات... الخ. ويمكن القول أن خيارات عرض النتائج واحدة من أهم الخدمات المساعدة في بحث المعلومات.

٤/٥ تنظيم الوصول إلى المكتبات الرقمية:

تعمل محركات بحث المكتبات الرقمية على بحث مجموعات التسجيلات الببليوجرافية داخل فهارس الويب، وعلى الجانب الآخر؛ فإن محركات بحث المكتبات الرقمية يمكنها عمل أدلة أو أدوات للحصر والتقسيم والوصول إلى قواعد بيانات المكتبات الرقمية على الويب. ومن ثم فإن هذا الجزء يركز على مختلف الآليات التي يوفرها محرك بحث المكتبات الرقمية للمستفيدين حتى يوفر بديلاً آخر لعملية البحث المتعدد متمثلاً في قوائم حصرية وروابط مباشرة لواجهات المكتبات الرقمية والتعامل معها بشكل مباشر.

١/٤/٥ التقسيم الموضوعي العام لقواعد المكتبات الرقمية:

يمثل التصفح برؤوس الموضوعات أو المصطلحات داخل الويب نموذجًا من نماذج الاعتماد على اللغة المقيدة في التكشيف وتحليل محتوى مصادر الويب، ذلك لما يرتبط به من تقييد مستخدم نظام الاسترجاع بقائمة مصطلحات هرمية التنظيم يتم ربطها بمجموعة مصادر المعلومات التي تتوافق مع دلالة كل من هذه المصطلحات. وهناك من أدوات بحث الويب المرئية مثل Yahoo ما يعتمد على قائمة من المصطلحات الخاصة بالموقع تندرج تحتها صفحات الويب التي تلبي الحاجة الموضوعية الخاصة لهذه المصطلحات. ويمكن في هذه الحالة الاعتماد على أي من المكانز الخاصة بالتكشيف في مجال محدد أو مجالات متعددة بصياغة شكل من أشكال التصفح الداعم للتحكم في مجموعة ثابتة من المصطلحات تتكون على أثرها قوائم رئيسية ثم فرعية تدرج تحتها محتويات قاعدة البيانات ككل. وقد اعتمدت بعض أدوات البحث على نظم تصنيف معدة مسبقًا مثل: تصنيف ديوي والتصنيف العشري العالمي، أو أن تعتمد بعض المواقع الأخرى على صياغة أشكال من نظم التصنيف تلائم المحتويات الخاصة بها من مصادر الويب.

الجدول رقم (٩) البحث بالتصفح داخل محركات بحث قواعد بيانات المكتبات الرقمية.

البحث بالتصفح داخل دليل البحث		محركات بحث المكتبات الرقمية	
		infomine	scirus
أنماط التقسيم الموضوعي	بالمصطلحات	√	√
	بالعناوين	√	√
	بالروابط	- - -	- - -
	بالنماذج المصورة	- - -	- - -
	أخرى	عناوين المكتبات	عناوين الدوريات
أنظمة التقسيم الموضوعي	خطط تصنيف	- - -	- - -
	رؤوس موضوعات	√	- - -
	تقسيم عام	- - -	√

يتضح من الجدول رقم (٩) أن محركي بحث العينة حددا قواعد بيانات النصوص دون غيرها من قواعد بيانات الويب باعتبارها مصادر لاسترجاع المعلومات العلمية الأكاديمية، وهو ما يتضح من تخصيص نمط التقسيم الموضوعي بالمصطلحات أو اعتماد الحصر والترتيب لعناوين قواعد البيانات التي يتعامل معها كل محرك. ويلاحظ الباحث أن محركي البحث اهتما بتوفير الحد الأدنى من أسلوب البحث بالتصفح، وذلك تأكيدًا من محركي البحث على أهمية التصفح في استرجاع المعلومات العلمية. غير أن الاهتمام بالتقسيمات الموضوعية لم يكن على القدر المطلوب؛ حيث إنه لم يوفر أي من المحركين تصنيفاً أو ترتيبًا منطقيًا للمصطلحات الموضوعية، واكتفى محرك بحث infomine بالحصر والترتيب الهجائي للمصطلحات، بينما غاب التقسيم الموضوعي تمامًا في محرك البحث scirus الذي اكتفى بذكر بعض المجالات الموضوعية العامة، دون تنظيمها في ترتيب محدد.

٢/٤/٥ التكشيف وإتاحة التصفح الموضوعي:

يوجد مستويان من مستويات تحليل وتكشيف محتوى قاعدة البيانات يمكن اتباعهما عند بناء أدلة بحث قواعد بيانات الويب غير المرئية؛ وهما:

المستوى الأول ويتمثل في التكشيف أو التحليل العام : وهذا يتم باستخدام المصطلحات الموضوعية والكلمات الكشفية للتعبير عن المجال أو التخصص الموضوعي لقاعدة البيانات ككل.

الثاني وهو التكشيف الداخلي لمجموعات قاعدة البيانات : وهنا يتم تكشيف وتحليل محتوى كل مقالة أو وحدة داخلية داخل قاعدة البيانات، والهدف من ذلك هو أن يتمكن المستفيد من الوصول إلى استرجاع المقالات أو الوحدات الداخلية دون الوقوف عند قاعدة البيانات ذاتها.

الجدول رقم (١٠) تكشيف وتصنيف قواعد بيانات المكتبات الرقمية.

تكشيف وتصنيف قواعد البيانات		محركات بحث المكتبات الرقمية	
		infomine	scirus
الحصر والترتيب		√	√
التكشيف	آليا	- - -	- - -
	يدويا	√	- - -

يوضح الجدول رقم (١٠) أن محركي عينة البحث تعاملا مع قواعد بيانات الويب العلمية كنظم استرجاع للمعلومات أكثر من كونها مجموعات من مصادر المعلومات العلمية؛ حيث لم يقدم أي من محركي البحث ما يشير إلى عمليات تكشيف متعمقة لمجموعات قواعد البيانات. وقد اقتصرت عمليات تحليل المحتوى داخل المحركين على الحصر والترتيب بشكل موضوعي. ويقر الباحث بأن عمليات تحليل محتوى مجموعات قواعد البيانات هي من الصعوبة بالدرجة التي لا تستطيع محركات بحث المكتبات الرقمية القيام بها، وإنما يمكن تحقيق ذلك من خلال تحليل محتوى الإشارات الببليوجرافية المصاحبة لمجموعات مصادر المعلومات داخل المكتبات الرقمية.

٦ – مقترحات معالجة قواعد المكتبات الرقمية داخل محركات الويب غير المرئية :

إن الهدف المرجو من هذه الدراسة هو الإشارة إلى أهمية التفاعل بين قواعد بيانات المكتبات الرقمية سواء الفهارس الببليوجرافية أو قواعد بيانات النصوص الكاملة

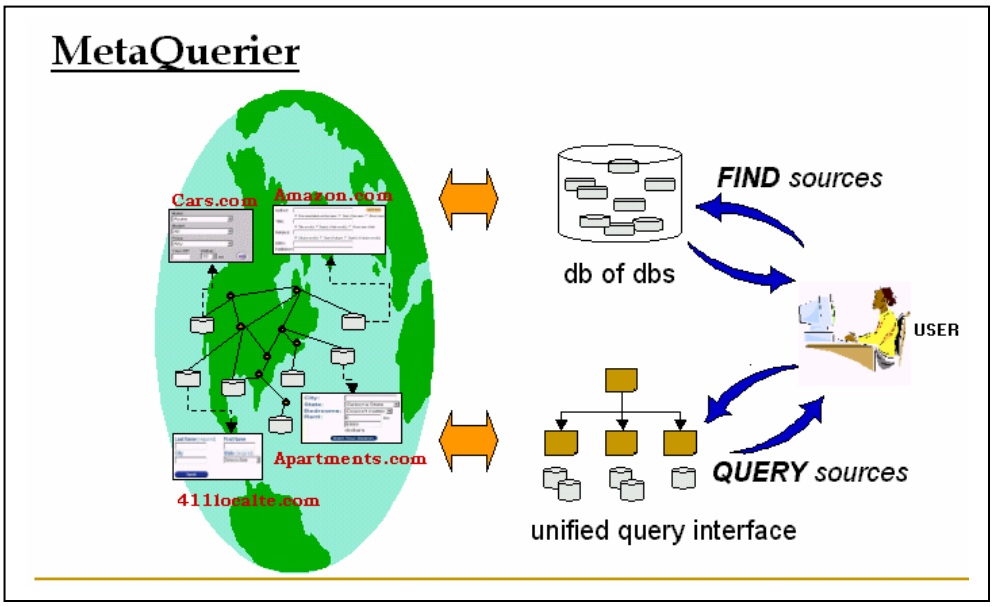
وبين محركات بحث الويب غير المرئية. وسينتج عن هذا التفاعل نجاح مؤكد لفتح آفاق الوصول والإتاحة لمصادر المكتبات الرقمية على الويب، التي قد تقتصر على البحث في الفهارس مباشرة أو استرجاع مجموعة من مصادر المعلومات التي تتيحها المكتبة على موقعها الافتراضي. ويقدم الباحث في هذا الجزء توضيحًا لمنهج المعالجة والتنظيم لقواعد بيانات المكتبات الرقمية من قبل محركات بحث الويب غير المرئية.

١/٦ معالجة وبحث قواعد المكتبات الرقمية.

تمثل قواعد بيانات المكتبات الرقمية هدفا مزدوج التنظيم والاسترجاع لمحركات بحث الويب غير المرئية، حيث يكون التفاعل في التنظيم والاسترجاع على مستويين هما.

أولاً: تنظيم واسترجاع قاعدة البيانات ككل.

ثانيًا: تنظيم واسترجاع محتوى وتسجيلات قاعدة البيانات. وفي الحالة الأولى يقف محرك البحث عند استرجاع قاعدة البيانات ذاتها وتنظيمها كوحدة واحدة داخل مجموعات بوابة أو نظام استرجاع الويب غير المرئية، ويختص دليل البحث بهذا الجزء، حيث سيعمل على حصر وترتيب قواعد البيانات التي يتعامل معها النظام في شكل قوائم أو تقسيمات موضوعية مرتبة في مستويات متعددة من الموضوعات العامة إلى المتخصصة. أما الحالة الثانية فإن قاعدة البيانات سوف تسمح لمحركات البحث بالتعمق إلى محتوى التسجيلات الببليوجرافية والقدرة على الوصول إلى ملفات المعلومات النهائية، ويرى الباحث أن هذا الجزء سوف يختص بالعمل به أسلوب البحث الحر بالكلمات المفتاحية من خلال توجيه استفسارات البحث إلى قاعدة البيانات والرد عليها بالنتائج المسترجعة. ويوضح الشكل رقم (٣) طريقتي تعامل المستفيد مع استرجاع قواعد البيانات. ومن ثم فإن الباحث يقترح تبني مستويي العمل السابقين للتنظيم واسترجاع محتوى قواعد البيانات غير المرئي، وبذلك يكون وصف العمل المقترح في العناصر التالية:



شكل (٣) مستويا التنظيم والبحث لقواعد البيانات

٦ / ١ / ١ أنماط تكشيف وتحليل محتوى قواعد بيانات المكتبات الرقمية

يستخدم مصطلح تصنيف محتوى قواعد بيانات الويب أكثر من مصطلح التكشيف للتعبير عن التحليل الموضوعي للمجموعات الداخلية لقاعدة بيانات الويب. وذلك يرجع في رأي الباحث إلى أن محركات بحث الويب غير المرئية لا تهدف إلى تكشيف التسجيلات أو الوحدات الداخلية لقاعدة البيانات لكي تضمها إلى قاعدة المحرك، وإنما الهدف من التحليل الموضوعي لمجموعات قاعدة البيانات هو التعرف إلى المجالات الموضوعية التي يمكن أن تدرج أسفلها قاعدة بيانات المكتبة الرقمية، ومن ثم يمكن إخضاعها للبحث عنها وتوجيه استفسارات المستفيدين إليها. وعلى ذلك فإن تعامل أدوات بحث الويب مع قواعد البيانات يأتي في سياق التحليل الموضوعي بغرض التصنيف، وليس بغرض التعريف بالمحتوى في مصطلحات وهو ما يعرف بعملية التكشيف.

٦/١/١/١ معالجة واسترجاع قاعدة البيانات:

اتجهت أدوات بحث عينة الدراسة إلى حصر وترتيب قواعد بيانات المكتبات الرقمية التي تتعامل معها، وهي بذلك تعتمد على مبدأ الحصر والإتاحة لمسميات قواعد البيانات

أمام المستفيدين؛ مما يعني أن هذه الأدوات تكتفي بدورها كأدلة أو أدوات حصر وليس بحث مباشر لمحتوى قاعدة البيانات. ومن ثم فإن هذا المستوى من التحليل العام لمجال وطبيعة قاعدة البيانات يقتضي معه أن تُتيح أدلة البحث بعضًا من المعلومات الواصفة لقاعدة بيانات الويب. وهذا الوصف الفني سيكون بمثابة التسجيلات الببليوجرافية الملحقة بكل قاعدة بيانات للتعبير عنها والتعريف بها أمام المستفيد قبل أن يتجه للبحث داخلها. غير أن بناء هذه التسجيلات سوف يعتمد إلى حد كبير على مراجعة وتقييم العنصر البشري لهذا الإجراء؛ وذلك لأنه لا توجد معايير موحدة تتعامل مع قواعد بيانات المكتبات الرقمية على كونها مصدرًا للمعلومات، بل إن التعامل مع قواعد بيانات الويب يسير دائمًا في اتجاه أنها أحد نظم استرجاع المعلومات الرقمية على الويب.

ويمكن أن تتكون تسجيلات قواعد البيانات وعناصر الوصف بها كالتالي[١٥]:

الجدول رقم (١١) حقول التسجيلة الببليوجرافية عن قاعدة بيانات المكتبة الرقمية.

Database record fields	حقول تسجيلة قاعدة البيانات
Title	عنوان قاعدة البيانات.(مسمى المكتبة الرقمية)
Agency/Owner	المؤسسة أو المالك.
Database Abstract	مستخلص عن قاعدة البيانات. (الفهارس)
Database Update Frequency	دورية وفترات التحديث لقاعدة بيانات المكتبة.
Legal Constraints on Database Access	خصائص الإتاحة الشرعية للوصول إلى قاعدة بيانات المكتبة.
Form(s) of Digital Copies of Database	الأشكال الرقمية لقاعدة البيانات.
Database Digital Copy Media and Costs List	تكاليف النسخ أو مقابل التحميل.
Report Reproduction Cost List	تكاليف إعادة الإنتاج لقاعدة البيانات.
Custom Services Costs	تقرير مقابل الحصول على خدمات قاعدة البيانات.
List of Database Data Modules	قائمة نماذج البيانات داخل قاعدة البيانات.
Annotated List of Fields (one list per data module)	قائمة حقول التسجيلات داخل قاعدة البيانات

وعلى ذلك فإن الإجراءات المتبعة في معالجة قواعد بيانات المكتبات الرقمية ككل كالتالي:

١- البحث والتعرف إلى قواعد بيانات المكتبات الرقمية المختلفة في الموضوع، سواء أكان ذلك باستخدام الزاحف أم عن طريق العنصر البشري.

٢- تحليل المجال الموضوعي لقاعدة البيانات والتعرف إلى خصائص العمل وبنية المعلومات.

٣- صياغة مختلف خصائص وعناصر وصف المجال الموضوعي والتقني في شكل عناصر أو حقول وصف ببليوجرافي.(الجدول رقم ١١).

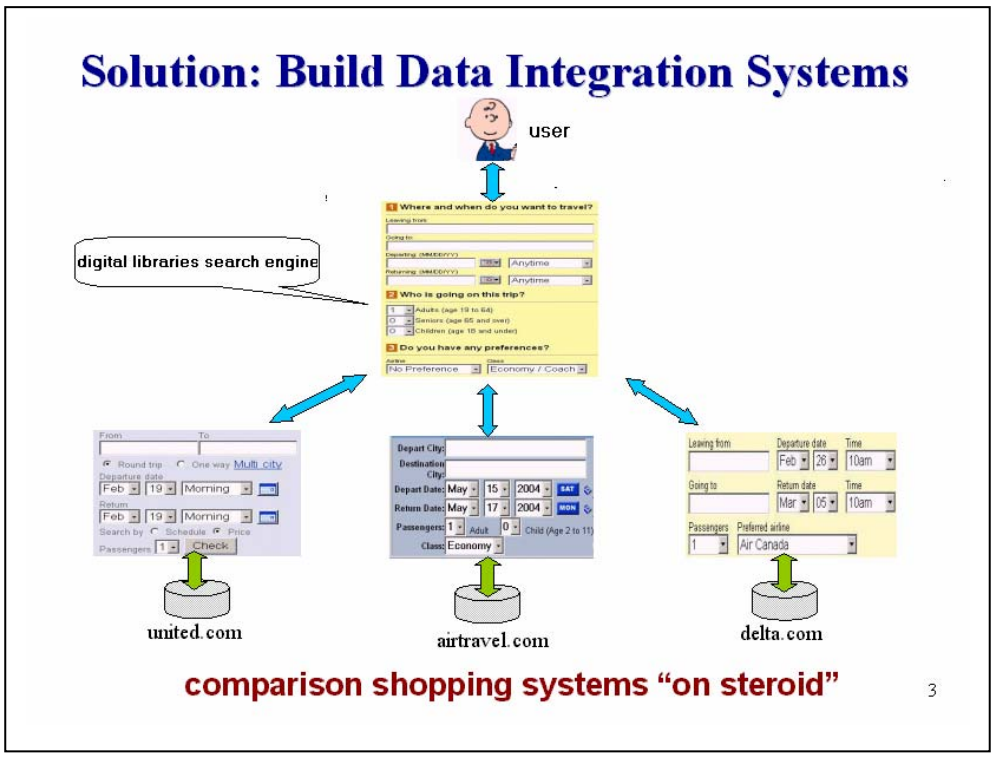
٤- ربط كل قاعدة بيانات على حدة بالمصطلح أو رأس الموضوع المتوافق معها في قوائم رؤوس موضوعات دليل البحث.

٦ / ١ / ١ / ٢ معالجة واسترجاع تسجيلات ومصادر قاعدة البيانات:

يمثل محتوى قواعد بيانات الويب القسم الأكبر من معلومات محتوى الويب غير المرئي، وتأتي مواقع الويب التفاعلية ومتغيرة المحتوى في المرتبة الثانية. ومن ثم فإن محركات بحث الويب غير المرئية تهدف إلى طرق باب هذا المحتوى بمجموعة من الاستفسارات البحثية التي يطرحها برنامج محدد حتى يمكن استقراء الموضوعات المتخصصة لمجموعات قاعدة البيانات وتصنيفها اعتمادًا على تصنيف النتائج المسترجعة.

وقد تبنت جامعة كولومبيا Columbia University نموذج Qprober لتحليل محتوى قواعد البيانات النصية، مستخدمة أسلوب طرح الاستفسار وتحليل النتائج المستدعاة. يطرح هذا النظام كلماته البحثية على قاعدة البيانات للتعرف إلى وحدات النتائج المستدعاة، ويستقبل البرنامج الفرعي classifiers هذه النتائج ثم تحليلها والتعرف إلى موضوعاتها ثم وضعها في تقسيم هرمي للموضوعات التي توجد بداخله. غير أن التحليل لا يكون من خلال مضاهاة الكلمات وإنما من استقراء أعداد النتائج التي تخرج مع كل مصطلح بحث. وتعكس أعداد النتائج المستدعاة ترتيب قاعدة البيانات في الشكل الهرمي للموضوعات. وهو مبدأ استخدام مصطلحات البحث واستقراء الأعداد المسترجعة، ثم مكان قاعدة البيانات في التوزيع الهري للموضوعات[١٦].

إن ثمة تطابقاً بين نموذج بحث قاعدة البيانات داخل المكتبة الرقمية وبين النمط الذي صيغت عليه بنية التسجيلات الببليوجرافية والملفات الداخلية، ومن ثم فإن لكل قاعدة بيانات آليات وعناصر البحث الخاصة بها، التي يعبر عنها في النهاية نموذج البحث، وعلى ذلك فإن محرك بحث المكتبات الرقمية سوف يقابل كثيرًا من الأنماط والمسميات المختلفة داخل نماذج بحث قواعد بيانات الويب. ويقوم محرك البحث بإجراء المقابلة والمطابقة بين عناصر البحث في كل نموذج ثم استنباط النموذج العام الذي يعمل به محرك بحث الويب غير المرئية المتخصص في بحث المكتبات الرقمية. ويوضح الشكل رقم (٤) عملية المطابقة بين نماذج البحث والوقوف على الشكل الموحد الذي يمكنه التعامل مع أكثر من نظام متكامل للمكتبات(١٧).



الشكل رقم (٤) واجهة محرك بحث المكتبات الرقمية مع نماذج بحث لمكتبات رقمية متعددة

وقد ذكرت بعض الدراسات كدراسة Denis Shestakov مدى الاهتمام الذي يتطلبه نموذج بحث محرك المكتبات الرقمية عند بنائه داخل صفحات الويب، حيث يمكن

للمستفيد التعامل معه وملء القيم الملائمة للبحث، ويطلق على هذا النموذج مصطلح html form. وتحوي صفحات الويب هذا النموذج مبتدئًا ومنتهيًا بتيجان form. وتعمل لغات التحويل هنا على متابعة التعامل مع نموذج البحث؛ حيث تتلقى هذه اللغات المعلومات الواردة من نموذج البحث ثم تجري عمليات المعالجة لاستفسار البحث، متابعة لاسترجاع النتائج في نماذج أخرى يتم بناؤها باعتبارها واجهات عرض النتائج. وقد أشارت هذه الدراسة أيضًا إلى طبيعة بناء حقول نموذج البحث form fields وهي مثل: (checkboxes, radio buttons, menus, etc.)، حيث تحتوي هذه الحقول على ملامح محددة ينبغي الاهتمام بها عند بناء نموذج البحث ككل، وتتمثل في[١٨]:

١- **مجال الحقل** field domain وهو كم المعلومات المنتظر إدخاله في هذا الحقل وطبيعته.

٢- **مسمى أو عنوان الحقل** field label وهو عنوان الحقل الموضح لهويته عن بقية حقول النموذج.

٣- **القيم الذاتية** Initial field set وهي المعلومات الذاتية التي تملأ الحقل إذا لم يدخلها المستفيد.

وقد أكد الاتحاد الأمريكي للمكتبات الرقمية (DLF) في لقائه أن هناك مجموعة من المتطلبات الإضافية التي يجب توافرها في برامج محرك بحث المكتبات الرقمية. تتمثل هذا المتطلبات فيما يلي[١٩] :

١-**الاهتمام بمعالجة وتكشيف المصادر العلمية فقط** : حيث ستتعامل محركات بحث المكتبات الرقمية مع الإنتاج العلمي المنشور في شكل كتب رقمية ومقالات للدوريات العلمية، الأمر الذي يتطلب من محركات البحث تدقيق عملياتها الفنية وخاصة التركيز على تكشيف محتوى المصادر العلمية عالية القيمة والمحتوى.

٢-**قدرات التعامل مع محتوى متعدد متغير** : حيث ستتعامل محركات بحث المكتبات الرقمية مع مجموعة من مصادر المكتبات الرقمية تختلف بين الميتاداتا والصور والصوت... وغيرها، مما يتطلب معه كفاية التفاعل مع مختلف مصادر المعلومات الرقمية داخل قواعد بيانات المكتبات الرقمية.

٣-توفير إمكانات التصفح والإبحار المحكمة : تتعامل محركات بحث الويب غير المرئية عامة ومحركات بحث المكتبات الرقمية خاصة مع أنظمة كثيرة وقواعد بيانات متعددة، وهو الأمر الذي يتطلب من محركات البحث تقسيم وترتيب هذه المجموعات من المكتبات الرقمية في قوائم ذات تنظيم موضوعي منطقي ييسر على المستفيدين التعرف إلى مجموعات كل نظام أو مكتبة رقمية بشكل مفصل.

٤-مرونة آليات الدمج وإعادة ترتيب النتائج المسترجعة : تحتاج محركات بحث المكتبات الرقمية إلى التحلي بصفات محركات البحث المتعدد التي تستدعي نتائج متنوعة متباينة من وجهات مختلفة، وهو ما يتطلب قدرات عالية من إجراءات الدمج وإعادة الترتيب ووزن أهمية النتائج المسترجعة والتعامل مع المكررات.

٥- الاقتباس الآلي لمجموعات الميتاداتا : تأخذ ميتاداتا المصادر العلمية والأكاديمية أهمية أكبر من تلك المتواجدة في المصادر غير العلمية أو ملفات الويب الأخرى. ويرجع ذلك إلى معايير الوصف الببليوجرافي الأكثر ضبطًا وتحكمًا؛ مثل مارك MARC 21 وغيره، وهي دائمًا ما تستخدم مع مصادر المعلومات العلمية دون غيرها من مصادر المعلومات الأخرى، الأمر الذي يتطلب حسن التوافق والتفاعل معها من جانب محركات بحث المكتبات الرقمية.

٢/٦ خصائص الزاحف وواجهات المكتبات الرقمية:

يعمل برنامج الزاحف داخل محركات بحث الويب على كشف وإضافة وتحليل مواقع الويب. غير أن الفارق في العمل بين زاحف الويب المرئية وغير المرئية (بيئة المكتبات الرقمية)، يكمن في وجهة العمل في كل من الاثنين؛ فزاحف الويب المرئية يعمل باستخدام الروابط الفائقة والمسارات URLs للوصول إلى صفحات الويب ذات بنية html والعمل على تحليل الكلمات الدالة الواردة بها في حقول الوصف المختلفة metadata. أما زاحف الويب غير المرئية ومواقع المكتبات فيسعى للوصول إلى نماذج بحث مواقع وقواعد البيانات web search forms مستخدمًا لذلك خصائص التعرف إلى حقول البحث وتحليلها والتعرف إلى بنية كل نموذج من نماذج البحث المختارة search

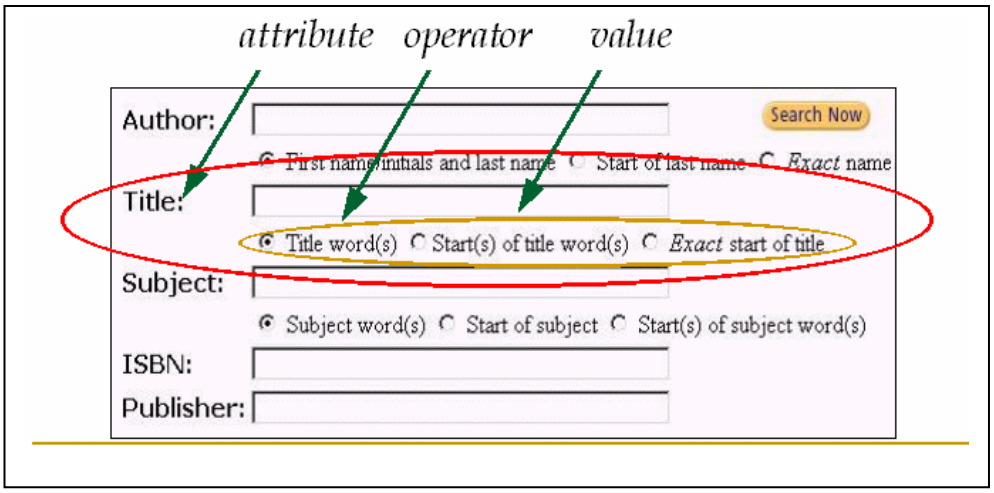
form fields structure. وهذه هي وجهة الزاحف في التعامل مع قواعد بيانات المكتبات الرقمية التي تحمل واجهات للبحث تحوي في معظمها آليات البحث بالحقول الببليوجرافية وأنماط التقسيم الموضوعي المختلفة للتسجيلات الببليوجرافية. ويعرف الزاحف عامة على أنه: " البرنامج الذي يجتاز الويب بشكل آلي العمل automatically لإضافة صفحات الويب إلى محرك البحث، ويستخدم في ذلك الروابط الفائقة للتنقل بين صفحات الويب، غير أنه يفتقر إلى عنصر الروابط الفائقة في التعرف إلى نماذج البحث مع الويب غير المرئية ". وعندما اختلفت بيئة عمل برنامج الزاحف وتحولت من وجهة مرئية ذات روابط فائقة، إلى وجهة خفية ذات نماذج بحث مستترة، وقد واجه برنامج الزاحف كثيرًا من الصعوبات في القيام بالمهام المنوط بها داخل محرك البحث، مثل[٢٠] :

١- القدرة على تحليل ومعالجة نماذج بحث واجهات المكتبات الرقمية.

٢- القدرة على تكوين وطرح استفسارات البحث على نماذج بحث المكتبات الرقمية.

٣- القدرة على التحليل الذاتي للنتائج المسترجعة Automatic response analysis التي غالبًا ما تكون تسجيلات الوصف الببليوجرافي.

إن بداية اللقاء بين محرك بحث الويب غير المرئية وبين المكتبات الرقمية تبدأ عند برنامج الزاحف؛ حيث يتبنى الزاحف مهام الوصول إلى مواقع المكتبات الرقمية والتعرف إليها وتحديد القيمة الأكاديمية والعلمية لمصادر ومجموعات المعلومات بها. وينفذ الزاحف إلى عمق المكتبات الرقمية على الويب من خلال نموذج بحث واجهة النظام الآلي المتكامل للمكتبة الرقمية. ويرى الباحث أنه كلما اتجهت نماذج بحث واجهات قواعد بيانات المكتبات الرقمية إلى المعيارية والتوحيد في آليات وعناصر البحث، كلما استطاع برنامج الزاحف الوقوف وإجراء المفاضلة والترجيح بين نماذج بحث الويب بأفضل أداء. ويوضح الشكل رقم (٥) صورة مبسطة لنموذج البحث داخل إحدى قواعد بيانات الويب[٢١].



الشكل رقم (٥) نموذج بحث لقواعد بيانات المكتبات الرقمية

٣/٦ معايير وآليات بحث قواعد المكتبات الرقمية:

خرج معيار **Z39.50** في أول إصداراته عام ١٩٩٢م من خلال المعهد القومي الأمريكي للتقييس والمنظمة العالمية للتوحيد والتقييس **ANSI/NISO standard**. وهو برتوكول مخصص للعمل مع تطبيقات بحث واسترجاع المعلومات المتنوعة؛ حيث يسمح بتعيين الإجراءات الضرورية وتنفيذ البحث داخل أنظمة المعلومات المختلفة، ومنها على سبيل المثال قواعد البيانات المختلفة سواء أكانت الببليوجرافية أم غير الببليوجرافية مع إمكانية عرض نتائج الاستفسار في شكل الاتصال العميل/ الخادم، حيث إن كل مستخدم يمتلك برنامج عميل **Z39.50** يمكنه إجراء البحث داخل قواعد البيانات المختلفة من خلال خادم **Z39.50**. ومن المكتبات والمراكز البحثية ومراكز حفظ المعلومات التي قامت باستخدام وتطبيق هذا المعيار نذكر على سبيل المثال: مكتبة الكونجرس الأمريكي، والمكتبة الوطنية الكندية(٢٢).

وقد بلغ هذا المعيار من التطور في العمل على الإنترنت مبلغًا يؤهله للدخول ضمن تطبيقات نظم استرجاع قواعد البيانات التي تهدف هذه الدراسة إلى رسم ملامحها، ويتمثل هذا التطور في الملامح التالية:

أولاً: اتخذ كثير من مؤسسات المعلومات قراره بتنصيب معيار **Z39.50** مع برتوكول **TCP/IP** بدلاً من اندماجه مباشرة مع نموذج **OSI**. وهناك كثير من الأسباب التي توضح هذا الاختيار منها على سبيل المثال: أن العمل من خلال برتوكول **TCP/IP** هو الأكثر شيوعاً وانتشاراً في الوقت الراهن مقارنة بنموذج **OSI**، إلى جانب أن التوافق مع **TCP/IP** يجعل استخدام معيار **Z39.50** على شبكة الإنترنت أكثر سهولة.

ثانيًا : لغة الاستفسار (البحث) العام (**Common Query Language (CQL**، حيث يتم من خلال هذه اللغة تنفيذ البحث داخل أنظمة استرجاع المعلومات المختلفة. وقد صدرت الإصدارة الأولى من هذه اللغة **Version 1.1** في فبراير ٢٠٠٤م، وخضعت لبعض التعديلات والإضافات.

ثالثًا: اندماج **Z39.50** إلى الشبكة العنكبوتية العالمية (الويب) وذلك بحسب المنهجين التاليين :

المنهج الأول: يرتبط باستخدام خادم ويب مع معيار **Z39.50**، ويتضمن خادم الويب عميل **Z39.50**، الذي يستخدم في إعادة صياغة الاستعلامات القادمة من عميل الويب والمتجهة إلى خادم **Z39.50** وفي العودة يقوم بتنسيق المعلومات القادمة من خادم **Z39.50** في شكل **HTML** ويقوم بإعادته إلى العميل المسئول على عرض نتائج الاستعلام.

المنهج الثاني: يرتبط بعميل ويب مع دعامة من **Z39.50** وذلك من خلال عنوان **URL** وهنا يقوم عميل الويب بالعمل بالشكل التالي « **z39.50://...** » . وتجدر الإشارة إلى أن نظام **VTLS** يستخدم هذا المنهج.

٧ – رؤى ونتائج ومقترحات الدراسة :

خرجت الدراسة بمجموعة من الرؤى وجوانب الطرح تتمثل في العناصر التالية:

١– إن مواقع المكتبات الرقمية على الويب مازالت تعمل في أضيق الأطر العملية والتطبيقية لها حتى الآن؛ حيث تعمل مواقع المكتبات الرقمية باعتبارها قواعد للبيانات الببليوجرافية ولا تتعدى كونها فهارس متاحة على الخط المباشر، وإذا ما تعدت هذا الإطار قليلاً؛ فإنها تقدم بعض الخدمات الببليوجرافية مثل: عمليات الإحاطة الجارية.

اقتراح الباحث : أن تخرج مواقع المكتبات الرقمية من هذه الأطر المحدودة إلى أوسعها وأجودها في تقديم خدمات المعلومات الرقمية؛ حيث يجب أن تعمل المكتبة الرقمية بوصفها بوابة للمعلومات العلمية، في احتواء مختلف مصادر المعلومات الأكاديمية؛ فضلاً عن توفير البحث في قواعد البيانات ذات النصوص الكاملة والارتباط مع الهيئات العلمية، مما يسمح للمكتبات الرقمية بفتح آفاق تقديم المعلومات والوصول إلى المستفيدين على تعدد فئاتهم.

٢- إن المكتبات الرقمية على الويب تعمل على استرجاع واستدعاء مجموعاتها من المعلومات التي غالبًا ما تقتصر على التسجيلات الببليوجرافية، وهذا الدور إنما تتجه به فقط إلى المستفيدين من خدماتها، غير أن المكتبات مازالت لا تتعامل مطلقًا مع محركات بحث الويب التي تعد نظم المعالجة والتنظيم لمصادر الويب جميعها، وتشكل مجموعات المكتبات الرقمية الآن جزءاً مهمًّا من مصادر الويب العلمية أو الأكاديمية غير المرئية التي تختص بها محركات البحث. وعلى ذلك فالأمر يتطلب من المكتبات ازدواجية الأداء بالعمل في اتجاه محركات البحث إلى جانب اتجاهها إلى المستفيدين.

اقتراح الباحث : أن المكتبات يجب أن تضم مجموعة من التقنيات ومعايير العمل التي تجعلها على قدر من المستوى التقني في التفاعل مع محركات البحث؛ بحيث تتلقى استفسارات البحث الموجهة إليها ثم الرد على هذه الاستفسارات بمجموعات المصادر والنصوص الكاملة أو الملفات أو حتى التسجيلات الببليوجرافية. أيضًا يجب أن تتحلى محركات البحث بصفات الضبط والتقنين في العمل بمعايير وحقول وعناصر الوصف الببليوجرافي المقننة.

٣- تعمل مواقع المكتبات الرقمية على الويب بشكل أقرب إلى الجمود في الأداء وهو ما لا يتناسب مع بيئة ذات طبيعة ديناميكية متغيرة كالويب عامة أو الويب غير المرئية خاصة. وهذا يرجع إلى العمل وتقديم الخدمات بشكل منعزل عن بيئة الويب التي توجد بها المكتبة؛ فالترابط بين المكتبة وبين باقي الخدمات العلمية الأخرى مثل مجموعات النقاش والمدونات ومواقع الهيئات العلمية يجعل المكتبة الرقمية أكثر زخرًا وغنىً بمصادر المعلومات التي تقدمها للمستفيدين.

اقتراح الباحث : أن تمتد عمليات الإضافة وتنمية المجموعات للمكتبات الرقمية إلى التواصل مع مواقع مصادر المعلومات الأخرى بما يسمح بالتشابك والترابط بين المكتبات الرقمية وبين شبكات المعلومات الأكاديمية كالجامعات والهيئات العلمية الأخرى.

٤- إن مواقع المكتبات الرقمية على الويب غالبًا ما تُبنى لتكون مرآة رقمية تعكس واقع مصادر المعلومات وخدماتها التي تقدم في المكتبة التقليدية صاحبة الموقع. وهو الأمر الذي يؤثر سلبًا على تحرر وانطلاق حركة المكتبات الرقمية على الويب، فهذا دائمًا ما يجعل مواقع المكتبات الرقمية على الويب محدودة في مصادر معلوماتها وفي خدماتها وفي تفاعلها مع محركات البحث.

اقتراح الباحث : يجب أن تبنى مواقع المكتبات الرقمية بشكل أكثر تقدمًا عما هو عليه حالها الآن؛ حيث يجب الفصل بين طبيعة وخدمات ومصادر المكتبة التقليدية الأم للموقع، وبين طبيعة وخصائص وقدرات وعمل موقع المكتبة الرقمية في بيئة الويب، حيث تفرض البيئة الرقمية مجموعة من أنماط وتوجهات العمل تختلف تماما عن طبيعة العمل في البيئة التقليدية، وذلك مثل التفاعل مع مجموعات أكثر من المستفيدين عبر فضاء الويب، والتعامل مع أدوات استفسار واستدعاء للمعلومات مثل محركات البحث... الخ.

٥- إن برنامج الزاحف في محرك بحث قواعد بيانات المكتبات الرقمية سيحمل عبء الوصول إلى نماذج البحث في واجهات مواقع وقواعد بيانات المكتبات الرقمية، وذلك لتحليلها وتكشيفها، ومن ثم فإن اتجاهه في العمل سيكون على تتبع نماذج البحث وليس صفحات الويب ذات بنية **html**. فضلاً عن أن وصول برنامج الزاحف إلى نماذج بحث مواقع وقواعد بيانات المكتبات الرقمية ليس هو الغاية في حد ذاته، وإنما تكمن الغاية في إمكانية تحليل وتكشيف هذه النماذج للتعرف إلى خصائص ومكونات وعناصر البحث التي تتركب منها. وهذا كله يهدف بدوره إلى فهم هذه النماذج من حيث آليات البحث بها وطبيعة عملها، حتى تتمكن محركات بحث الويب

غير المرئية من توجيه استفسارات البحث إليها، واسترجاع تسجيلات قواعد البيانات الببليوجرافية.

٦- إن النجاح في الوقوف على أهم آليات وملامح عمل محركات وأدلة بحث الويب غير المرئية، هو بمثابة مفتاح بوابة الدخول إلى قواعد بيانات المكتبات الرقمية، ومن ثم فهو آلية جديدة يمكن لقواعد بيانات المكتبات الرقمية من فهارس ونصوص كاملة أن تنخرط بين مصادر معلومات الويب، وتمتزج معها محققة لمستخدمي المكتبات الرقمية إمكانية التعامل معها واسترجاع محتواها من خلال نظام استرجاع الويب غير المرئية.

٧- مكنت التقنيات الرقمية قواعد بيانات المكتبات الرقمية من القدرة على العمل بنمط تفاعلي ومتغير خلاف حالة الثبات التي تتصف بها مواقع المكتبات الرقمية الحالية؛ حيث يمكن لقواعد بيانات المكتبات الرقمية عرض محتوياتها الداخلية الآن في واجهة واحدة وإن اختلفت المعلومات المعروضة في طبيعتها، كما يمكن لقاعدة بيانات الويب صياغة المحتوى الداخلي وتغييره بشكل ديناميكي يمكن المستفيد من تجميع معلومات مختلفة والاطلاع على موضوعات متعددة في آن واحد. ولقد شكل المحتوى الديناميكي أو التفاعلي لمواقع الويب عامة حجر عثرة أمام محركات بحث الويب التي تجد أنفسها أمام محتوى سريع التغيير، لا يمكن نسخه أو التحكم به، وهذا العامل هو أهم عوامل ظهور الويب غير المرئية بما فيها مواقع المكتبات الرقمية.

اقتراح الباحث : تزويد محركات بحث الويب بمختلف التقنيات الحديثة التي تتعامل مع المحتوى الديناميكي المتغير في الشكل والمحتوى من مجموعات قواعد بيانات الويب. ويقع هذا العبء تحديدًا على شركات بناء محركات بحث الويب، وهيئات رعاية الويب **W3C** في تقديم الملامح الإرشادية للقيام بذلك.

٨- إن محركات بحث مجموعات المكتبات الرقمية على الويب هو جيل من محركات بحث الويب ينتظر التزود بالملامح والسمات المتطورة المتوافقة وقدرات التفاعل مع قواعد بيانات المكتبات الرقمية، ويوصي الباحث مختلف شركات تقنيات

محركات بحث الويب بالتوجه نحو البحث العلمي الأكاديمي الذي يتعامل فيه محرك البحث مع الدوريات العلمية والمكتبات الرقمية وقواعد بيانات الاتحادات العلمية والمؤسسات الأكاديمية. ويرى الباحث أن هذا التوجه قد لاحت بوادره في الأفق متمثلا فيما اتجه إليه محرك البحث الشهير **google** من بناء الباحث العلمي **google scholar** الذي يمكنه البحث في المقالات العلمية وقواعد بيانات الدوريات أو الاتجاه نحو مواقع المكتبات الرقمية.

٩- يتم تحليل وتكشيف محتوى قاعدة البيانات على مستويين يمكن اتباعهما عند بناء أدلة بحث قواعد بيانات الويب غير المرئية؛ أما **المستوى الأول** فيتمثل في التكشيف أو التحليل العام : وهنا يتم استخدام المصطلحات الموضوعية والكلمات الكشفية للتعبير عن المجال أو التخصص الموضوعي لقاعدة البيانات ككل، وليس مجموعاتها أو مقالاتها الداخلية. ويستخدم هذا المستوى في دليل البحث حتى يتم إدراج عناوين قواعد البيانات أسفل المصطلح أو المصطلحات الموضوعية في قوائم التصفح التابعة لدليل البحث، ومن ثم يتمكن المستفيد من الوصول إلى قواعد البيانات الملائمة إذا ما اختار مصطلحًا أو أكثر من هذه المصطلحات. أما **المستوى الثاني** فهو التكشيف الداخلي لمجموعات قاعدة البيانات : وهنا يتم تكشيف وتحليل محتوى كل مقالة أو وحدة داخلية داخل قاعدة البيانات، والهدف من ذلك هو أن يتمكن المستفيد من الوصول إلى استرجاع المقالات أو الوحدات الداخلية دون الوقوف عند قاعدة البيانات ذاتها، ويصلح هذا المستوى أكثر في حالة اعتماد دليل البحث على أسلوب البحث الحر إلى جانب قوائم التصفح بالمصطلحات أو الكلمات الدالة.

هوامش ومصادر البحث

١- http://wwwlib.umi.com/dissertations/, cited a 1/3/2006.

٢- Ipeirotis, Panagiotis G. Classifying and Searching Hidden-Web Text Databases , Advisor: Prof. Luis Gravano , Columbia University, 2004. http://wwwlib.umi.com/dissertations

٣- Sol, Selena. Introduction to databases for the web, 1998, cited at 10/6/2007, cited at http://www.databasejournal.com/sqletc/article.php/26861_1431601_1

٤- Surface Web. Wikipedia : the free encyclopedia, 2007, cited at 20/6/2007, cited at http://en.wikipedia.org/wiki/ Surface_ Web

٥- ODLIS —Online Dictionary for library and information science. OP.cit

٦- سيد ربيع سيد. **محركات بحث الصور الثابتة : دراسة تحليلية** .- ط١.- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧.-ص ١٢٠

٧- Lewandowski, Dirk. Exploring the academic invisible web, 2007, cited at 25/3/2010, cited at http://www.emeraldinsight.com/10.1108/07378830610715392

٨- Ford, Nigel and Mansourian, Yazdan. The invisible web: an empirical study of "cognitive invisibility", Journal of Documentation,2006, cited at 9/12/2008, cited at http://www.emeraldinsight.com/10.1108/00220410610688732

٩- Cohen, Laura. The Deep. internet tutorials, 2006, cited at 20/6/2007, cita at http://www.internettutorials.net/deepweb.html

١٠- Cohen, Laura. The Deep.OP.cit.

١١- online DB and searching.2004, cited at 15/7/2007, cited at http://web.simmons.edu/~benoit/LIS530/602slides/17online.htm

١٢- Lossau, Norbert. Search Engine Technology and Digital Libraries, D-Lib Magazine, 2004, cited at 2/3/2010, cited at http://www.dlib.org/dlib/june04/lossau/06lossau.html

١٣- Lossau, Norbert. Lbid.

١٤- Ntoulas, Alexandros. Crawling and searching the hidden Web, supervising Gregory H. Leazer, UNIVERSITY OF CALIFORNIA, 2006, cited 208/2007, cited at http://www.dia.uniroma3.it/~vldbproc/017_129.pdf

١٥- Google and the deep web. Blogger, 2007, cited 1/7/200٨, cited at http//glinden.blogspot.com/2007/03/google-and-deep-web.html

١٦- Gravano, Luis and Ipeirotis, Panagiotis G. QProber: A System for Automatic Classification of Hidden-Web Databases, Columbia University, 2003, cited 1/8/2007, cited at htt//qprober.cs.columbia.edu/publications/tois2003.pdf

١٧- Wu ,Wensheng WebIQ: Learning from the Web to Match Deep-Web Query Interfaces, University of Illinois,2006, cited at 1252009, cited at http://www.cs.binghamton.edu/~meng/pub.d/icde06-webiq.pdf

١٨- Shestakov, Denis and Bhowmick, Sourav S. DEQUE: querying the deep web, Nanyang Technological University, Singapore, 2004, cited at 2/4/2009, cited at www.sciencedirect.com

١٩- Lossau, Norbert. OP.CT.

٢٠- Ntoulas, Alexandros. Crawling and Searching the HiddenWeb, University of California, 2006, cited at 20/5/2009, cited at

http://portal.acm.org/citation.cfm?id=1061951.1061952

٢١- Barbosa, Luciano. Combining Classifiers to Identify Online Databases, University of Utah, 2005, cited at 20/5/2009, cited at http://www2007.org/htmlpapers/paper603/

٢٢- أحمد فرج أحمد. **معيار Z39.50 : أداة أساسية لاسترجاع المعلومات عبر بوابات المكتبات**. - cybrarians journal، ع ١١ (ديسمبر ٢٠٠٦م). تاريخ الاستشهاد ٢٠٠٧/١٠/٢٠. - مسار : http://www.cybrarians.info/journal/no11/z39.50.htm