محركات بحث الشبكة العنكبوتية

د. خالد عبد الفتاح
قسم المكتبات والمعلومات
جامعة المنيا - مصر
khaledma1@hotmail.com

نظرة عامة على نشأتها وتطورها ومستقبلها

تعمل محركات البحث بصفة أساسية على بناء كشافات لمصادر المعلومات المتشابكة من خلال اشتقاق كلمات أو عبارات من النصوص نفسها لبناء ملفات تسمح ببحث هذه المشتقات بالاعتماد على أساليب البحث والاسترجاع المعروفة مثل المنطق البولينى، وتجاور المصطلحات، والبتر، والجذع وغيرها. والحقيقة أن هذه الملفات لا تتميز عن الأساليب التقليدية التي استخدمت في الاسترجاع منذ أن حل الاسترجاع العشوائي محل الاسترجاع التسلسلي، والتي تشتمل بصفة أساسية على ثلاث ملفات حيوية هي: الملف التسلسلي Serial File، والملف الكشفي Index File، والملف المقلوب Inverted File ومع ذلك فإن التقنيات الحديثة من أجهزة وبرمجيات ساعدت على تحديث وبحث تلك الملفات المقلوبة بسرعة كبيرة، هذا إلى جانب أنها أضافت إلى تلك الملفات مجموعة جديدة من الملفات لتيسير عمليات البحث والاسترجاع مثل ملف الروابط الفائقة، ملف وصف الوثائق ... إلخ [1].

■ نشأة محركات البحث وتطورها:

لقد قام عالم الفيزياء تيم برنر لى بوضع أسس الشبكة العنكبوتية في بداية التسعينات من القرن العشرين لتكون وسيلة أساسية للباحثين في تبادل مسودات البحوث والرسائل الإلكترونية. ومنذ ذلك التاريخ بدأت العديد من الجامعات في استخدام هذه الأداة في بث وتيسير سبل الإفادة من المعلومات. ومع بداية عام 1993 كان هناك بضع مئات من المواقع المتاحة على الشبكة العنكبوتية معظمها مواقع تتعلق بكليات ومعاهد بحثية. وكانت الطريقة الأساسية لتبادل المعلومات بين مستخدمي الشبكة العنكبوتية في ذلك الوقت تتم من خلال بروتوكول تبادل الملفات المعروف بـ File Transfer Protocol (FTP) وهو عبارة عن برنامج يمكن من خلاله نقل الملفات من حاسب إلى حاسب أخر من خلال واجهة تعامل تعمل بالأوامر. فى تلك المرحلة إذا أراد شخص أن يسترجع معلومات من الشبكة العنكبوتية عليه أن يتعامل معها من خلال هذا البروتوكول. وكانت هذه الطريقة فعالة في ظل مجموعات الويب الصغيرة، ولكن مع تزايد المجموعات ونموها لم تصبح هذه الوسيلة فعالة بالدرجة الكافية مما دفع الباحثون للتنقيب عن وسائل أخرى. وتمثل أول تلك الحلول في الاعتماد على أحد محركات البحث التي تم تطويرها قبل نشأة الشبكة العنكبوتية وكان يعرف بالأرشيف Archie إلا أن استخدامه من خلال نظام التشغيل يونكس UNIX فرض ضرورة اختصار الاسم إلى Archie. وقد قام بتطوير هذا المحرك أحد طلاب جامعة ماكجيل McGill بمدينة مونتريال الكندية أسمه ألن إمتاج Alan Emtage. وقد اعتمد هذا المحرك أساساً على قاعدة بيانات بأسماء الملفات المتاحة على الشبكة العنكبوتية، فكانت عملية المضاهاة تعتمد بشكل كبير على البحث في قاعدة البيانات عن أسم الملف الذي يرغب المستفيد في استرجاعه[2]. وقد مرت عملية بناء وتطوير محركات البحث بأجيال متعددة نذكر منه ما يلي:

■ الجيل الأول:

وفى عام 1993طورت جماعة الاهتمام بالحاسبات بجامعة نفادا بالولايات المتحدة محرك بحث جديد أعتمد على نفس البنية المستخدمة في المحرك Archie وعُرف هذا المحرك الجديد بـ Veronica. والاختلاف الوحيد بين Archie و Veronica هو أن الثاني كان يعمل مع ملفات النصوص Plain Text Files بينما كان الأول يعمل فقط على الاسترجاع من قاعدة بيانات تشتمل على أسماء الملفات. ثم ظهر تقريبا في نفس التاريخ محرك ثالث عُرف بـ Jughead وقد أعتمد أيضا على نفس البنية المستخدمة في المحرك Veronica، وقد تم استخدام كلا من Jughead and Veronica لتبادل الملفات من خلال أداة التصفح جوفر Gopher والتي قام بتطويرها مارك ماكهيل Mark McCahill في جامعة ميناسوتا لكي تحل محل المحرك Archie [3].

وفى عام 1993ظهر أول روبوت * على يد ماتثوى جارى Matthew Gary والذى عُرف بمتجول الشبكة العنكبوتية WWW Wanderer. وقد كان الهدف الأساسي من هذا الروبوت هو إحصاء معدلات الزيادة في الشبكة العنكبوتية من خلال تتبع وإحصاء خوادم الويب النشطة Active Web Server. ثم قام ماتثوى بعد ذلك بتعديل الربوت حتى يتمكن من تجميع محددات المصادر الموحدة URL`s. وقد عرفت قاعدة البيانات التي تم تجميعها من خلال هذا الروبوت بـ Wandex. وفى أكتوبر 1993 قام أرتيجن كوستر Artijn Koster بتطوير محرك جديد يشبه في بنيته المحرك Archie وعُرف هذا المحرك بـ Aliweb. وقد أتاح هذا المحرك لأول مرة إمكانية تسجيل الصفحات في محركات البحث، حيث أتاح الفرصة لمعدي صفحات الويب أن يقوموا بتسجيل الصفحات وتكشيفها ووصفها بأنفسهم،

* كمبيوتر روبوت : هو ببساطة برنامج حاسب آلى يستطيع أداء العديد من المهام التكرارية بسرعة كبيرة جداً تفوق إمكانيات مئات بل آلاف الأشخاص إذا حاولوا القيام بنفس الوظيفة يدوياً.

ولكنه واجه مشكلة كبيرة هي أن معدي صفحات ومواقع الويب لم يكن لديهم الخبرة الكافية لتكشيف وتسجيل صفحاتهم بأنفسهم[4].

وبحلول ديسمبر عام 1993 ظهر ثلاث محركات بحث جديدة في نفس الوقت هي على التوالي: JumpStation، و(The World Wide Web Worm (WWWW، و The Re-positoryBased Software Engineering (RBSE). وقد أعتمد المحرك JumpStation على تكشيف عناوين ورؤوس الصفحات Title and Header كما أعتمد في الاسترجاع على البحث الخطي * Linear Search. ومع نمو الشبكة العنكبوتية لم يعد هذا المحرك قادراً على متابعة هذا النمو السريع مما جعله يتوقف سريعاً. أما المحرك WWW Worm فقد أعتمد على تكشيف العناوين ومحددات المصادر الموحدة Page Title and URL`s. ومن العيوب الأساسية في كلا من JumpStation and WWWW أنهما كانا يسترجعا النتائج دون أي ترتيب، حيث كان يتم استرجاع النتائج وفقاً للترتيب الذى وجدت عليه في قاعدة البيانات. أما المحرك RBSE فقد كان أول محرك بحث على الشبكة العنكبوتية يستخدم فكرة نظم الترتيب والفرز Ranking Systems والتى يمكن من خلالها استرجاع النتائج مرتبة وفقاً لمعايير الصلاحية *. ومع نهاية عام 1993ظهر المحرك Excite والذي كان ناتج أحد مشروعات تطوير المحرك Architext والذي بدأه 6 طلاب في جامعة ستانفورد فى فبراير عام 1993. حيث قاموا باستخدام فكرة التحليل الإحصائي Statistical Analysis لعلاقات الكلمات والمصطلحات Word Relationships من أجل جعل البحث أكثر فعالية وكفاءة [5].

■ الجيل الثاني:

ولم تكن كل المحاولات السابقة، في الحقيقة، تمثل مقومات محركات البحث ولم تكن صالحة في الأصل كمحركات؛ نظراً لأن الزاحف Spider أو الروبوت Robot الذي يتولى تجميع الصفحات من الشبكة العنكبوتية لم يكن بالذكاء الكافي الذي يتمكن من خلاله فهم العلاقات القائمة بين الروابط الفائقة Hyperlinks، ومن ثم فإن المستفيد إذا لم يكن يعلم على وجه الدقة عنوان الصفحة التي يرغب في الوصول إليها فإنه كان من الصعب وربما كان من المستحيل عليه الوصول إلى تلك الصفحة.

وفى يناير عام 1994 ظهر أول دليل بحث على الشبكة العنكبوتية الذى عُرف EINet Galaxy. وقد ساعد على نجاح هذا الدليل اشتماله على ملامح البحث التي وفرتها كلا من جوفر Gopher وتلنت Telnet (وهما معا كانا يمثلان أهم أدوات الإنترنت في ذلك الوقت)، هذا إلى جانب ملامح البحث في الشبكة العنكبوتية. وقد شهد أبريل عام 1994 مولد دليل البحث Yahoo على يد كلا من ديفيد فيلو David Filo وجيرى يانج Jerry Yang، والذي لم يكن في بدايته سوى مجموعة من الصفحات والمواقع المخزنة على الحاسبات الشخصية لدى كلا منهما.

■ الجيل الثالث:

وقد شهدت الفترة من عام 1994 حتى نهاية العقد الأخير من القرن العشرين ظهور عدد كبير من محركات وأدلة البحث التي تميزت بقدرتها الفائقة على بحث واسترجاع الصفحات والمواقع على الشبكة العنكبوتية كان أبرزها محرك البحث Google, AltaVista, Alltheweb، وغيرهم. وقد شهدت الفترة من عام 1994 إلى عام 2000 منافسة شرسة بين مجموعة من محركات البحث العالمية على تغطية أكبر قدر ممكن من صفحات ومواقع الويب حيث شهدت تلك الفترة العديد من دراسات المقارنة بين مدى تغطية محركات البحث لصفحات ومواقع الويب [6].

وقد شهدت الفترة من عام 2001 إلى 2005 طفرة جديدة في محركات البحث تمثلت في محاولة معظم المحركات الشهيرة في التحول من مجرد محركات بحث إلى بوابات للويب Web Portals. ويشير مصطلح البوابات إلى مجموعة الأدوات التي تسعى إلى تنظيم مصادر المعلومات المتاحة من خلال تقسيمات موضوعية شاملة بحيث تشتمل البوابة على كافة أنواع المصادر والخدمات التي يحتاج إليها المستفيدين من خدمات الشبكة العنكبوتية من خدمات بريد إلكتروني، دردشة، قوائم خدمات وقوائم بريدية، المواد الإخبارية، أسعار العملات، أحوال الطقس، إلى جانب قوائم موضوعية بمصادر المعلومات المتاحة من خلال البوابة إلى جانب محرك يتيح إمكانية البحث في البوابة. وإلى جانب التنوع في الخدمات التي تقدمها البوابات للمستفيدين منها نجد أن هذه المواقع عادة ما تتضمن برامج تساعد على تحليل استخدامات المستفيدين Web Usage Analyzer بغرض بناء ملفات سمات المستفيدين User Profiles ويمكن من خلال هذه الملفات التعرف على احتياجات المستفيدين

---

* البحث الخطي : هو مضاهاة حروف كلمات الاستفسار حرف بحرف بمعنى أنه إذا كان أحد الحروف غير متشابهة لا يسترجع أي نتائج وهو يشبه في ذلك البحث بإستخدام CLT + F في الويندوز.

* معايير الصلاحية : هى معادلات وخوارزميات رياضية تستخدمها محركات البحث لترتيب النتائج وفقاً لعلاقتها بمصطلحات الاستفسار الذي يدخله المستفيد للبحث فى الشبكة العنكبوتية.

والتنبؤ بها بالتالى اختيار المصادر المناسبة لكل مستفيد من المستفيدين من الموقع. ويمكن أن تقوم تلك المواقع باستخدام تكنولوجيا الدفع Pushing Technology إلى المستفيدين من الموقع. كما يمكن أن تتم عملية الدفع عبر خدمات البريد الإلكترونى التي توفرها تلك المواقع أو إلى الصفحات الأمامية للمستفيدين من هذه المواقع كما يمكن أن يتم الدفع إلى دوسيهات خاصة للمستفيدين من هذه المواقع. من ثم فالبوابات عادة ما تيسر لمستخدمي تلك المواقع كل أنواع الخدمات التي يحتاج إليها بصورة تفاعلية مما يوفر كل احتياجات المستفيد من خدمات ومصادر الشبكة العنكبوتية. وفى مقابل ذلك تسعى البوابات إلى جذب الشركات التي تسعى إلى الإعلان عن منتجاتها وخدماتها لتحقيق الأرباح من خلال تلك المواقع حيث أنه من المعروف أنه كلما زاد عدد مستخدمي الموقع كلما تهافت الشركات على الإعلان عن خدماتها ومنتجاتها من خلال هذه المواقع.

■ الجيل الرابع:

شهدت الفترة من عام 2000 بداية تطوير جيل جديد من أدوات البحث على الشبكة العنكبوتية يعرف بالأعوان الذكية للبحث Intelligent Agent التي تسعى إلى الاستفادة من إمكانيات الذكاء الاصطناعي والنظم الخبيرة لتحقيق متطلبات تشغيل الويب الدلالي Semantic Web في تيسير عمليات البحث والاسترجاع ومازال العمل في هذه الأدوات في طور التجارب المبدئية. وتمثل الأعوان الذكية تطور طبيعي لبوابات الويب بحيث تسعى إلى التعرف على احتياجات المستفيدين من الشبكة العنكبوتية من خلال مجموعة البرامج التي تحلل تلك الاستخدامات وبناء واجهات تعامل ذكية تساعد على تقديم خدمات موجه لمستفيد بعينه وفقا لاحتياجاته المخزنة في النظام. كما تحاول تلك الأدوات الاستفادة من مقومات تكنولوجيا الدفع في توجيه النتائج إلى المستفيدين مباشرة بالتالى لا يحتاج المستفيد إلى إجراء البحث بنفسه حيث يتولى وكيل البحث إجراء البحث بشكل دوري نيابة عن المستفيد. كما تحاول تلك الأعوان أن تقوم بتطوير وسائل تساعد المستفيد في إجراء العمليات والوظائف العامة التي يؤديها من خلال الشبكة العنكبوتية مثل فحص الرسائل الإلكترونية وتنظيمها بحيث يمثل هذا البرنامج السكرتيرة الخاصة للمستفيد مما يمكنها من التصرف في بعض المواقف مثل القيام بعمليات التنبيه وإعداد جداول المواعيد.

فعلى سبيل المثال نجد أن موقع Tracerlock يقوم بالمهام التالية نيابة عنه [7].

- flexible search terms - single words, phrases or advanced search strings
- complete control - at any time you can change your search terms, suspend or cancel your account, change your account details or open new accounts
- personalized searches - in addition to TracerLock's standard, extensive database of sites, you can enter up to 20 URLs that you want TracerLock to monitor for you
- timely alerts - TracerLock can alert you by e-mail within as little as 15 minutes of new occurrences of your search terms [sample alert]
- archives - your past search results are retained online for later access
- webmaster tools - TracerLock can generate newsfeeds (based on your search results) for use on your own web site
- TracerLock Toolbar - a downloadable toolbar for Microsoft Internet Explorer which helps you search as you surf
- TracerLock Desktop - a free downloadable application to manage your TracerLock account and to retrieve your results directly to your PC [More...]

الخلاصة:

تناولت هذه الورقة عرض لتطور محركات البحث على الشبكة العنكبوتية لتوضيح الوضع الحالي لتلك الأدوات State of the Art والتطورات المستقبلية المحتملة في هذا الإطار. وقد أوضحت الورقة أن هناك تطور كبير في هذه الأدوات يسعى إلى توظيف إمكانيات الذكاء الاصطناعي لكى تتناسب في طبيعتها مع الجيل القادم للشبكة العنكبوتية والذي يعرف الويب الدلالي والذي يمثل الهدف الإستراتيجي للجهة المشرفة على تطوير الشبكة العنكبوتية ألا وهي WWW3C.

المصادر:

1- Lancaster, F.W. (1998) Indexing and Abstracting in Theory and Practice. Champaign, Illinois: University of Illinois, Graduate School of Library and Information Science, 412 p.

2- Gromov, Gregory R. History of the Internet and WWW-Part 8: Statistics. The Road and Corssroads. Feburary, 2000. Retrieved from the WWW at May,25, 2005. http://www.netvalley.com/intval/07262/main.htm?sdf=1

3- Lenssen, Phillip. Search Engines History. April, 2004. Retrieved from the WWW at May 14,2005 http://blog-outer-court.com/history/

4- Seo Consultants. Directory and Search Engines History. June 2003. Retrieved from the WWW at May 15, 2005. http://www.seoconsultants.com/search-engines/history/

5- Wall, Aaron. Search Marketing. History of Search Engines & Web History. Retrieved from the WWW at May, 16, 2005. http://www.search-marketing.info/search-engine-history/

6- Hawking, D., Craswell, N., Bailey, P., Griffith, K. (2001). Measuring Search Engines Quality. Information Retrieval. 4(1), pp33-39.

7- Tracerlock. (2005). Welcome to Tracerlock. http://www.tracerlock.com/