Cybrarians Journal
ISSN 1687-2215
دورية إلكترونية فصلية محكمة ومتخصصة في مجال المكتبات والمعلومات

العدد 2، سبتمبر 2004

محركات البحث على شبكة الانترنت

فاطمة الزهراء محمد عبده

تمهيدي ماجستير - تخصص معلومات - جامعة القاهرة

f.elzahraa@cybrarians.info

مستخلص

دراسة عامة حول محركات البحث على الانترنت، تتناول سبعة نقاط رئيسية هي : الفرق بين محركات البحث والأدلة، مكونات محركات البحث على شبكة الإنترنت وكيفية عملها، نسبة المواقع التي تغطيها محركات البحث من مجموع مواقع الشبكة،فهرسة محركات البحث لمواقع الشبكة، الوقت اللازم لظهور صفحات الويب الجديدة في محركات البحث، آليات البحث في محركات البحث، واخيرا أنواع محركات البحث.

الاستشهاد المرجعي بالبحث

فاطمة الزهراء محمد عبده . محركات البحث على شبكة الانترنت - . cybrarians journal - . ع 2 (سبتمبر 2004) . - تاريخ الاتاحة <اكتب هنا تاريخ اطلاعك على الصفحة - . <متاح في: www.cybrarians.info/journal/no2/searchengines.htm

---

إن محركات البحث search engine هي أدوات شائعة الاستخدام إلى الوصول إلى صفحات النسيج الشبكي العالمي world wide web لكنها تعود بمئات النتائج والوصلات التي تكون في الغالب غير ذات صلة بموضوع البحث الرئيسي مما يجعل الباحث في حيرة ويأس من وجود ما يبحث عنه، فإن استخدام أدوات البحث المتوفرة على الإنترنت بدون استراتيجية بحث محددة تشبه إلى حد كبير من يسير في مكتبة ضخمة بشكل عشوائي يحاول العثور على كتاب معين.

وسوف يناقش هذا البحث النقاط التالية:

-1الفرق بين محركات البحث والأدلة.
-2مكونات محركات البحث على شبكة الإنترنت وكيفية عملها.
-3نسبة المواقع التي تغطيها محركات البحث من مجموع مواقع الشبكة.
-4فهرسة محركات البحث لمواقع الشبكة.
-5الوقت اللازم لظهور صفحات الويب الجديدة في محركات البحث.
-6آليات البحث في محركات البحث.
-7أنواع محركات البحث.

أولا : الفرق بين محركات البحث والأدلة: [1]

يسخر مستخدمو الإنترنت أدوات البحث المختلفة على الوب للوصول إلى المعلومات المفيدة الموزعة على ملايين الأجهزة الخادمة على الإنترنت، وهذه الأدوات هي محركات البحث search engineوالأدلة directories التي تساعد في تسريع عملية البحث بدرجة كبيرة، وفي هذا السياق لابد تقديم تعريف مبسط لهاتين الأداتين قبل أن دخل في تفاصيل محركات البحث.

محركات البحثSearch Engines

وهي يمكن تعريفها بأبسط صورة هي عبارة عن قواعد بيانات ضخمة بعناوين ومواقع، ومع وصف مصغر لصفحات الإنترنت المختلفة، والتي بواسطتها أي محركات البحث يمكن البحث ع موضوع معين في حقل من الحقول المختلفة في الشبكة بشكل دائم بغرض إيجاد دليل معين لمثل هذه الصفحات، ولأنها تعمل بشكل آلي وتقوم بفرز وفهرسة كم هائل من الصفحات ..فسيلاحظ كل شخص يستخدم هذه المحركات على أنها تحتوي على كثير من المعلومات غير المتوفرة في الأدلةdirectories وقد يمثل ذلك جانب إيجابي يجعل من محركات البحث أداة فعالة أكثر من الأدلة.[2]

الأدلةDirectories

الأدلة لا تعمل بشكل آلي بل تتم إدارتها من قبل أشخاص متخصصين، وما يحدث هو أن العديد من المواقع يتم تسليمها إلى دليل ما، ومن ثم يتم فرزها وتبويبها تحت تصنيف معين، ولأن هذه الآله يتم إدارتها بشكل بشري فإنها قادرة دوما على توفير معلومات أكثر دقة وما موقع yahoo المشهور إلا المشهور إلا مثال واحد لهذه الأدلة.[3]

وتقدم الأدلة للمستخدم طريقة سريعة للبدء بعمليات البحث عن المعلومات بواسطة تفحص المواضيع المصنفة التي يعرضها ، إذ يندرج تحت كل موضوع لائحة من المواضيع الفرعية فيمكن للمستخدم أن يتفحصها تباعا إلى أن يصل إلى المعلومات المطلوبة، وفي حال عدم وجود المعلومات تحت الموضوع الذي اختاره المستخدم، يتراجع ويختار موضوعا رئيسيا آخر ليقوم بالبحث في تفرعاته من جديد، وهكذا...

وفي هذا السياق لابد من توضيح من أن هناك العديد من أدلة البحث تعمل أيضا كمحركات منها..... yahoo – altavista – excite :

ثانيا : مكونات محركات البحث وكيفية عملها: [4][5]

يوجد على شبكة الإنترنت كميات ضخمة من المعلومات تتوزع على شكل صفحات وتنمو بشكل سريع (في عام 1996 بلغ معدل النمو 400%)، و في أوائل 1998 كان هناك ما يتعدد بـــ 300 مليون صفحة متوفرة على النسيج، وفي أواخر اعم 2000 قدر عدد الصفحات بأكثر من مليار صفحة، وهو الآن يقدر بأن يكون حوالي مليار ونصف في بداية عام 2002، ويقدر عدد الصفحات التي تضاف أو تحدث يوميا بمليون ومائتين صفحة، وهذا يدل على معدل نمو الإنترنت الذي يتضاعف بكل سنوي.

هناك أكثر من 2500 أداة بحث متوفرة ويقدر تغطية تلك الأدوات للإنترنت بـــ 5 % للأدوات الصغيرة و 15 % لتلك العملاقة، ومن هنا نعلم أنه مهما كان سمعة أداة البحث فإنها لا يمكن أن تغطي كافة المعلومات المتوفرة على الشبكة، وتظهر محركات البحث نتائج مختلفة، لأن كل منها يبحث في قاعدة بياناته الخاصة به، التي تتضمن جزء من مواقع الوب ، ويتألف محرك البحث من ثلاثة أجزاء رئيسية هي:

1. برنامج العنكبوت(Spider program) :
2. برنامج المفهرس(Indexer program) :
3. برنامج محرك البحث.

1.برنامج العنكبوت:

تستخدم محركات البحث برنامج العنكبوت لإيجاد صفحات جديدة على الوب إضافتها، ويسمى هذا البرنامج أيضا " الزائر" لأنه يبحر في الإنترنت بهدوء، لزيارة صفحات الويب والإطلاع على محتويتها، ويأخذ هذا البرنامج مؤشرات المواقع من عنوان الصفحة " page title " والكلمات المفتاحية "key words " التي تحتويها ، إضافة إلى محتويات محددات الميتا "meta tags " فيها.
ولا تقتصر زيارة برنامج العنكبوت على صفحات الموقع، بل يتابع البرنامج تعقب الروابط (links)الموجودة فيها لزيارة صفحات أخرى. أما الغاية من هذه الزيارات فهي وضع النصوص المتاحة على تلك المواقع على فهارس محرك البحث ليتمكن المحرك من العودة إليها
فيما بعد، ولم تغب فكرة تغيير المحتوى في الموقع عن البال مصممي محرك البحث، إذ يقوم محرك البحث بزيارات دورية للمواقع الموجودة في الفهارس للتأكد من التعديلات التي تصيب المواقع المفهرسة.

2.برنامج المفهرس:

يمثل برنامج المفهرس(index program) ، الكتالوج (catalogue) أحيانا، قاعدة بيانات ضخمة توصيف صفحات الويب ويركز التوصيف على المعلومات التي حصلت عليها من برنامج العنكبوت (spider)كما تعتمد على بعض المعايير مثل الكلمات الأكثر تكرارا من غيرها، وتختلف محركات البحث عن بعضها في هذه المعايير، إضافة إلى اختلافها في خوارزميات البحث searching algorithms .

3.برنامج محرك البحث:

يبدأ دور برنامج محرك البحث (search engine program) عند كتابة كلمة مفتاحية في مربح البحث (search box) إذ يأخذ هذا البرنامج الكلمة المفتاحية ويبحث عن صفحات الويب التي تحقق الاستعلام، الذي كونه برنامج المفهرس في قاعدة بيانات الفهرس، ثم يعرض نتيجة البحث المتمثلة

بصفحات الويب التي طلبها المستخدم في نافذة المستعرض "browser window " ويقوم أيضا بعملية الترتيب لهذه الصفحات.

وتختلف محركات البحث عن بعضها في أسلوب العمل، فمثلا: تحتفظ قاعدة بيانات altavista بكل تفاصيل صفحة الويب المخزنة عليها، أما محركات البحث الأخرى فقد تحتفظ بالعاوين الرئيسية للصفحة فقط، مما يؤدي إلى اختلاف شكل ودقة نتائج البحث الظاهر للمستخدم.

ثالثا :نسبة المواقع التي تغطيها محركات البحث من مجموع مواقع الشبكة[6]

أكدت دراسة أجراها معهد NEC للبحوث الواقع في مدينة برينيتون في ولاية نيوجرسي الأمريكية عام 1999، على 11 محرك بحث مستخدمة في الإنترنت، إن هذه المحركات لم تتمكن من التأقلم مع النمو الكبير لمواقع الويب، وبينت الدراسة أن محركات البحث لا تغطي سوى قسم ضئيل من صفحات الويب، وان إدراج الصفحات الجديدة بينها يستغرق وقتا طويلا.

ووجدت الدراسة أن محرك البحث "northern light" هو أكثر المحركات شمولا لكنه على الرغم من ذلك لا يغطي سوى 6/1 من صفحات الإنترنت التي يمكن أن تصل الإنترنت إليها، ويذكر أن هذه النسبة قبل 18 شهر كانت تصل إلى 3/1 من صفحات الإنترنت.

ويقول ستيف لورانس وسي لي جاليز، المتخصصان في علوم الكمبيوتر في معهد NEC والقائمين بهذه الدراسة، أنه من المحتمل أن يكون تأخر إدراج الصفحات الجديدة ضمن محركات البحث أمرا مقصودا، إذا تمتلئ قواعد البيانات التابعة لمحركات البحث هذه وأصبحت عملية إضافة موقع جديد إليها عملية مكلفة، خاصة إن هذه الإضافة لا تزيد في معظم الأحوال من الأرباح الإعلانية.

أسم المحرك	نسبة التغطية المئوية	دعم اللغة العربية	عنوان موقع الويب للمحرك
NORTHERN LIGHT	%16.0	لا	WWW.NORTHERNLIGHT.COM
SNAP	%15.5	لا	WWW.SNAP.COM
ALTAVISTA	%15.5	نعم	WWW.ALTAVISTA.COM

HOT BOT			
WWW.HOTBOT.COM	لا	%11.3	
MICROSOFT			
WWW.MSN.COM	لا	%8.5	
INFOSEEK			
WWW.INFOSEEK.COM	نعم	%8.5	
GOOGLE			
WWW.GOOGLE.COM	نعم	%7.8	
WWW.YAHOO.COM	لا	%7.4	[15]YAHOO
LYCOS			
WWW.LYCOS.COM	لا	%2.5	
EUROSEEK			
WWW.EUROSEEK.COM	نعم	%2.2	
EXICIT			
WWW.EXCIT.COM	نعم	%5.6	
WWW.ALLTHEWEB.COM	نعم	%15.5	ALL THE WEB

رابعا : فهرسة محركات البحث للمواقع:[7]

ومن خلال العرض السابق لمكونات محركات البحث، تبين أن برنامج العنكبوت أو عنكبوت الشبكة " web spider "وبرنامج المفهرس "cataloger" هما المسئولان عن إيجاد صفحات الويب الجديدة وفهرستها وإدخالها داخل قاعدة بيانات المحرك ... ولكن كيف ؟

تستخدم واصفات <META> الخاصة بكلاً من "وصف الموقع " و " كلمات المفتاح " في صفحة HTML ، حتى تضمن إدراج المواقع بشكل سليم في محركات البحث، ووضع وصف واضح له فيها، ولا يؤدي استخدام واصفة <META> بعكس كثير من واصفات HTML إلى ظهور أي عنصر مرئي على صفحتك، لكن عمل هذه الواصفة يتم في الكواليس، حيث تقدم وصفاً ن المعلومات التي تتضمنها الصفحة إلى المتصفحات وعناكب الشبكة ومحررات. HTML

إذا الوظيفة الأساسية هي وصف الصفحة وكلمات المفتاح فيها وتقديمها لمحركات البحث.
عندما يزور عنكبوت الشبكة صفحة لا تحتوي على وصف لها باستخدام واصفة <META> فإنه يسجل الأسطر القليلة الأولى من النص الموجود في بداية الصفحة ويعتبر ذلك وصفا لها ، ولا يعطي هذا الأسلوب غالبا وصفاً دقيقاً للصفحة. وأحد الأسباب الرئيسية لذلك "وجود ملف صورة في بدايتها"، إذ لا يتعرف عنكبوت الشبكة إلا على النصوص، ولا يستطيع قراءة الملفات الرسومية، التي تتضمن نصوص على هيئة صور، وهذا يوضح استخدام واصفات <META> وخاصتاً عندما تحتوي الصفحة على كثير من الصور.

خامسا : الوقت اللازم لظهور صفحات الويب الجديدة في محركات البحث.

من خلال بحث أجرته مجلة الطبيعة الأمريكية في عدد JOLY من عام 1999 والتي أوضحت أن المتوسط العمري لظهور الموقع في محركات البحث هو 57 يوم تقريبا [8]وحاز محرك البحث ALTAVISTA على المركز الأول فيما بينها والإحصائيات كانت على النحو التالي:

ATAVISTA	33يوم		
EXCITE	47يوم		
NORTHERNLIGHT	48يوم		
BOT	51يوم	HOT	
MSN	57يوم		
INFOSEEK	60يوم		
YAHOO	76يوم		
SNAP	91يوم		

LYCOS	147يوم
UROSEEK	غير معروف
GOOGLE	غير معروف

كما أوضحت الدراسة أن الوقت اللازم لتحديث قاعدة بيانات كل محرك بحث هي نفس المعدلات السابقة، حيث لا بد أن يتم تحديث قواعد بيانات محركات البحث لملاحقة التغيرات أو التعديلات على المواقع الموجودة داخل قاعدة البيانات وكذلك الحصول على مزيد من المواقع الجديدة على الويب.

سادساً : آليات البحث في محركات البحث

ذكرنا سابقا أن محركات البحث تستخدم في بحثها عن مواقع الويب ما يدعى الكلمات المفتاحية "KEY WORDS"التي يمكن ان تكون عبارة، وتستخدم آليات البحث عادة بعض المعاملات "OPERATORS"مع هذه الكلمات المفتاحية، لتوفير خيارات إضافية لعملية البحث، وهناك طريقتين للبحث في محركات البحث هما:

البحث البسيطSIMPLE SEARCH

وهذا النوع من أنواع البحث يقوم به معظم الناس وخاصتا المبتدئين مع استخدام الإنترنت وهم من يجهلون تقنيات البحث المتقدم. وهو يكون عن طريق وضع كلمة بدون أي علامات أو شارات ثم البحث عنها ولكننا سنجد النتيجة العديد من النتائج العلائقية والروابط المتينة.

البحث المتقدمADVANCED SEARCH

تزيد خصائص البحث المتقدمة التي يوفرها الكثير من محركات البحث إمكان العثور على المعلومات، كما إنها طريقة فعالة للبحث عن معلومات محددة "SPECIFIC INFORMATION" إذ تتيح للمستخدمين إمكان البحث عن عدة كلمات مفتاحية معاً.

وفيما يلي قائمة لمحركات البحث ومصطلحات البحث المتقدم التي تستخدمها:[9]

	yahoo	altavista	hotbot	excite	lycos	infoseek
and	x	x	x	x	x	
Or	x	x	x	x	x	
Not		x	x	x	x	
Near		x			x	
And not				x		
" "	x	x	x	x	x	x
+	x	x	x	x	x	x
-	x	x	x	x	x	x
*	x	x				

سابعا : أنواع محركات البحث:[10]

يمكن القول أنه قد ولى الزمان الذي كان يمكن فيه الاكتفاء بموقع بحث واحد ليلبي مختلف احتياجاتك في البحث بالشبكة، فمع التنامي المطرد للإنترنت جرى نوع من التخصص في وظائف البحث، بحيث أصبح كل موقع لمحركات البحث يحتوي على مجموعة من الخدمات التي تحتاجها للتنقيب عن المعلومات المطلوبة بسرعة ويسر، وسيتم عرض تلك الأنواع وهي كالآتي:

محرك بحث يبحث عن الصور:[11]

يمكن القول بأن موقع WWW.DITTO.COM وهو محرك بحث يرتكز على الصور، من إدخال كلمة، أو قطع للبحث عنه، وبدلا من إعادة نتائج البحث بشكل نصي يعيد الموقع النتائج بشكل عدد من الصور المصغرة Thumbnails ، وعند النقر على أي صورة مصغرة تعرض صفحة الويب التي تتضمنها، ويضم الفهرس في هذا الموقع أكثر من مليوني صورة، وهذا طبقا لإحصائية تم عملها في عام 1999، كما أن هناك العديد من محركات بحث الصور منهاwww.snap.com

***محرك بحث الملفات:**

قامت شركة c2v بتصميم برنامج جديد أطلق عليه أسم textomattom وهو من محركات البحث متعددة اللغات وليس من بينها العربية، وهو يعمل على إيجاد الملفات الهامة عن طريق محتوياتها أيا

كانت نوعها أو مواقعها على الويب، كما يعمل على تنفيذ عمليات البحث اعتمادا على إدخال كلمات مفاتيح المفردات المتخصصة، وبالتالي يستطيع البرنامج من خلال المحتوى دون اللجوء إلى عناوين الملفات ودون الحاجة إلى فتحها أياُ كان نوع الإصدارة، وكذلك يسمح باستخدام كلمات البحث المتقدم. and , or , not

***محركات بحث متخصصة:**

كما أن هناك محركات بحث عامة ، فهناك محركات بحث متخصصة في موضوعات معينة مثل:

-محركات بحث طبية:

www.nlm.nih.gov و www.healthfinder.org

-محركات بحث في مناطق أو دول معينة:

افريقيا . www.rubani.com : و www.ananzi.com

أسيا www.asianconnect.com : و www.4euroseek.com

- 1أدوات البحث على الإنترنت . – مكتبات نت . – ع 6 ، 7 (يونيو ويوليو 2000.(

- 2محركات البحث العربية : أنترنت العالم العربي . – ع مايو / أغسطس (1998) . تاريخ الاطلاع مارس 2002 .- متاح في: www.ditnet.co.ae/arabic/internet/studyd0110_3.html

- 3أحمد عبد القادر الخالد . نبش أعماق الإنترنت .- 2002 .- تاريخ الاطلاع مايو 2002 .- متاح في:www.computers.arabcomputing.com/search.html

- 4دليلك إلى محركات البحث في الشبكة : إنترنت العالم العربي .- ع مايو / أغسطس (1998) . تاريخ الاطلاع مارس 2002 .- متاح في: www.ditnet.co.ae/arabic/internet/studyd0110_2.html

- 5نبيل بدر سنبل . مقدمة إلى محركات البحث في الشبكة . – الرياض : جامعة الملك فهد ، 2001.- تاريخ البحث مايو 2002 .- متاح في:http://Users.kfupm.edu.sa

- 6أمنية طلعت . الطريق الآمن لدخول عالم الإنترنت . – جريدة البيان . – 27 يناير 2002 .- تاريخ البحث مارس 2002 .- متاح في:www.albayan.co.ae/albayan/2002/01/271mnw/9.html

- 7التحكم بفهرسة محركات البحث لموقعك .- انترنت العالم العربي . – ع مايو (1999). تاريخ الاطلاع مارس 2002 .- متاح في:www.ditnet.co.ae/arabic/internet/tutor0702.html

- 8محركات البحث لا تتأقلم مع نمو الإنترنت . – مجلة إنترنت العالم العربي . – ع أغسطس (1999) .- تاريخ الاطلاع مارس 2002 .- متاح في:www.idwmag.com/issue0211/nal.html

- 9البحث على الإنترنت . – لغة العصر . – (8 أغسطس 2000) . تاريخ الاطلاع أبريل 2002 .- متاح في:www.ahram.org.eg/ict

- 10مواقع البحث .– لغة العصر .– (26 مارس 2002) .- تاريخ الاطلاع أبريل 2002 .- متاح في:www.ahram.org.eg/ict

- 11أريزونا عبد العزيز أبانمي . خطوات بسيطة تضمن الاستفادة القصوى من محركات البحث . – جريدة الجزيرة . – ع 9993 (فبراير 2000) .- تاريخ الاطلاع مايو 2002 .- متاح في:www.suhuf.net.sa