

الفصل الخامس

البحث في "الإنترنت" المشكلات والأدوات

الفصل الخامس

البحث في "الإنترنت" المشكلات والأدوات

تحتوي إنترنت، وشبكة (ويب) على وجه الخصوص حالياً على ملايين الصفحات من المعلومات معظمها على شكل نصوص. وتنمو كتلة النصوص التي تحتوي عليها إنترنت بمعدلات متسارعة فضلاً عن خضوعها لعمليات التحديث المستمرة على مدار اليوم بل الساعة. ويطرح كل ذلك تحدياً جديداً أمام أولئك الذين يبحثون عما يحتاجونه من معلومات في هذا البحر الزاخر بالمعلومات. غير أن مشكلة البحث عن المعلومات في النصوص لم تنشأ مع ظهور إنترنت، بل تعود جذورها إلى عهد الكتب المطبوعة، ولتسهيل عملية البحث عن موضوع معين في كتاب مطبوع، تضم الكتب الحديثة في صفحاتها الأخيرة فهرساً للكلمات الدليلية الهامة (الكلمات المفتاحية) وأمام كل منها أرقام الصفحات التي وردت بها الكلمة. وقد إتمدت أساليب البحث عن المعلومات بواسطة الكمبيوتر على منهجية مشابهة لمنهجية إستخدام كلمات دليلية (Keywords) كتلك المستخدمة في الكتب التقليدية. وتمتاز الأدلة المستخدمة في نظم إسترجاع المعلومات بإستخدام الحاسب، بإتساع نطاقها مقارنة بأدلة الكتب.

ومشكلات البحث عن المعلومات خلال الإنترنت ناتجة من سببين رئيسيين، السبب الأول هو كثرة المعلومات الموجودة على الشبكة، فهناك ملايين صفحات الوب المتشبكة وكذلك مواقع الجوفر (Gopher) ومواقع (FTP) بالإضافة إلى الملايين من مجموعات الأخبار (News-Group) وقوائم البريد (Mailing Lists). السبب الثاني هو عدم وجود هيئة أو منظمة موحدة تشرف على تنظيم هذه المعلومات، بمعنى أنه لا يوجد كتالوج مركزي لهذه الموارد المتاحة ولا يوجد مكان واحد موحد يمكنك الذهاب إليه للبحث عما يريد.

ومع اتساع قاعدة النصوص. كما هو الحال مع شبكة إنترنت، تصبح عملية إسترجاع المعلومات بإستخدام دليل الكلمات غير عملية وتؤدي هذه الطريقة إلى إسترجاع أعداد كبيرة من النصوص التي تحتوي على الكلمات الدليلية التي تبحث عنها، بالإضافة إلى أن هذه النصوص قد لا يكون لمعظمها صلة وثيقة بالموضوع قيد البحث على الرغم من إحتوائها على الكلمة المطلوبة على سبيل المثال إذا كنت تبحث عن معلومات عن تاريخ بعض الدول وأدخلت أسم الصين. مثلاً، كمفتاح للبحث ستحصل على جميع الوثائق التي تحتوي على كلمة الصين (China) غير أن إحتواء بعض الوثائق على كلمة الصين لا يعني بالضرورة أنها ستكون وثيقة تتعلق بتاريخ الصين، فعندما ترد الكلمة في وثائق من النوعيات أطباق صينية، عادات صينية،

علاقة الصين باليابان، الحروف الصينية... الخ، وكما هو موضح فهي موضوعات ليست لها علاقة بالتاريخ الصيني القديم. ولمساعدتك على الوصول إلى مصادر المعلومات المطلوبة وفرزها وتنقيتها من المعلومات غير المرتبطة بها وباستخدام أدوات للبحث، تستخدم تقنيات مثل "وزن الكلمة" (Word Weighing).

ووفقاً لهذه التقنية لا تعامل سائر كلمات النص على قدم المساواة، بل يسند إلى كل منها (وزن) يتناسب مع درجة أهمية الكلمة المعنية في النص فكلمة الصين الواردة في تقرير حول الصين كانت رقم (1) من قبل (China was No.1 before) مثلاً، سيكون لها وزن أكبر تاريخياً مما للكلمة ذاتها في سياق وثيقة تتضمن مجموعات الاختيار في الصين (Newsgroup in China). وتأتي نتائج البحث القائم على هذا المنهج مرتبة بحسب درجة أهميتها المتوقعة بالنسبة للباحث، ويؤدي الانتشار والنمو المتسارع للإنترنت إلى إزدهار خدمات البحث عن المعلومات في داخل هذه الشبكة الكونية. وستعرض فيما لبعض المواقع التي توفر مثل هذه الخدمات مع عناوين هذه المواقع على الشبكة.

أهم المواقع التي تقدم خدمات البحث في شركة إنترنت

1. اكسيت (Excite)

تقدم هذه الخدمة شركة (Architext Software) وقد تم تقديمها على الشبكة منذ وقت قصير. وتشمل هذه الخدمة أداتين تنتج البحث عن مواقع معينة على شبكة ويب (Web) الأولى: البحث في الشبكة (Network) والبحث في المجموعات الاخبارية (Newsgroups) الخاصة بشبكة (Usenet). ويمكن استخدام كلمات مفتاحية في عملية البحث أو استخدام مفاهيم كمنطلق للبحث مع إمكانية تحديد نطاق للبحث. أما الأداة الثانية فهي مراجعة الشبكة (Netreviews) وهي توفر مخططا هريا لمحتويات شبكة ويب مبنية حسب الموضوعات وتفرعاتها.

ويتم تحديث قاعدة البيانات الخاصة باكسيت اسبوعيا وهي تضم حالي 105 بليون صفحة ويب و 50000 صفحة أخبار وعنوان هذه الخدمة على شبكة ويب هو:

<http://www.excite.com>

2. انفوسيك (Infoseek)

تتوفر هذه الخدمة بشكل مجاني أو مدفوع الأجر. وتغطي الخدمة المجانية شبكة ويب ويوزنت (Usnet). ونتيجة لذلك يتزاحم معظم المستخدمين على هذه الخدمة مما يصعب معه الوصول إليها. وتغطي هذه الخدمة المجانية مايقرب من مائة عنوان للبحث أما الخدمة المدفوعة الأجر فتغطي مايقرب من مائتين عنوان بالإضافة إلى سرعة استجابتها لطلبات البحث. وتقدم " انفوسيك " نتائج عملية البحث في صورة عناوين الموضوعات المرئية بحسب درجة الصلة المتوقعة لموضوعاتها مع مادة البحث.

وعنوان هذه الخدمة على شبكة ويب هو.

<http://www.Infoseek.com>

3. ليكوسن (Lycos)

وتعتبر هذه الخدمة من أهم وأقوى أدوات البحث في شبكة إنترنت. ولهذه الخدمة دليل للكلمات يضم أكثر من ثمانية ملايين كلمة تغطي أكثر من 90% من محتويات شبكة ويب. ولهذه الخدمة مزايا فريدة تشمل إمكانية البحث عن النصوص والصور والمواد الصوتية والفيديو وقائمة تضم حوالي 250 من أكثر مواقع شبكة ويب شعبية ويفضل استخدام هذه الشبكة في أعمال البحث المكثفة وليس البسيطة. عنوان خدمة ليكوس على الشبكة هو :

<http://www.Lycos.cmu.edu>

4. ألتا فينستا

تضم هذه الخدمة أكبر فهرس لشبكة ويب وتستطيع من خلالها البحث بين بلايين الكلمات تضمها مايقرب من 21 مليون صفحة ويب وتتيح أيضا فهرس نص كامل (full text Index) به أكثر من 13 ألف مجموعة إخبارية (Newsgroups)

5. ماجلان (Magellan)

تقدم خدمة ماجلان دليلا مصنفا لعدد هائل من مواقع شبكة" ويب" وFTP وجوفر Gopher ويوزنت ومجموعات الأخبار ويشمل ذلك تقديم نبذة عن محتويات الموقع ودرجة شعبيته بين مستخدمي شبكة إنترنت ويمكن استعراض محتويات الدليل بمتابعة تفرعات شجرة الموضوعات أو البحث فيها باستخدام كلمات مفتاحية. وتستخدم خدمة ماجلان ضوءا أخضرا للإشارة إلى أن محتويات الموقع تلائم قاعدة عريضة من المستخدمين وخلوه من المواد الاباحية أو النصوص الخارجة.

6. من هنا؟ How Is Here

تعتبر خدمة من هنا؟ من أسرع خدات البحث وأكثرها استخداما بالإضافة إلى أنها خدمة مجانية وتستخدم هذه الخدمة في تحديد مواقع الهيئات والأفراد على الشبكة. كذلك تقدم خدمة من أين؟ (Who Where) خدمة متميزة للبحث عن أسماء الأشخاص أو الهيئات عن طريق استخدام الحروف الأولى منها.

7. فهرس النص المفتوح (Open Text Index)

تعتبر هذه الخدمة من أكثر الخدمات شمولاً فلديها امكانية الوصول إلى نحو مليون صفحة في شبكة ويب وتقوم بالبحث خلال 21 مليون كلمة وجملة وتقوم المفهرسات الآلية (Crawlers) بمتابعة آلاف الصفحات يوميا لفهرسة ماتحتويه من كلمات وتقدير " وزن " كل كلمة مقارنة مع الموضوع الرئيسي للنص وعنوان هذه الخدمة هو:

<http://www.opentext.com>

(Web Crawler) مفهرسات الويب 8 .

تقوم شركة أمريكا أون لاين (America Online) بتوفير هذه الخدمة. لذلك فهي تعتبر خدمة مدفوعة الأجر وليست مجانية. ويستخدم فيها ما يشبه أجهزة الروبوت في جمع الكلمات وفهرستها عبر ملايين الصفحات في شبكة ويب " ويب ". وتمتاز هذه الخدمة بسهولة الإستخدام وسرعة الإستجابة لطلبات البحث القائم على المحتوى. كما تحتوي الصفحات الخاصة بهذه الخدمة على معلومات إحصائية هابة عبر شبكة ويب وعنوان هذه الخدمة هو:

<http://www.woberawler.com>

9. البرمجيات (Shareware)

تقوم هذه الخدمة أسرع وسيلة للحصول على البرمجيات من على الشبكة. ويمكن من خلالها البحث خلال 170,000 ملف وعارضة (Browsing) بالإضافة إلى إمكانية تحميل هذه البرمجيات (Download) على حاسبك الشخصي باستخدام خدمة FTP للإستفادة منها وعنوان هذه الخدمة هو:

<http://www.Shareware.com>

10 خدمات الوب (Web Servers Tourist)

يضم هذا الموقع دليلا شاملا للمواقع العلمية والأكاديمية المسجلة لدى الكونسورتيوم ويقوم الذي يتولى الإشراف على شبكة ويب. يضم هذا الدليل دليلا فرعيا من النوع الفهرسي (rectoryWeb Servers Di) مبوبا على أساس جغرافي ودليلا فرعيا آخر على هيئة خريطة للعالم (Virtual Tourist) يمكن التعامل معها بالضغط على مواقع معينة منها بالفأرة. وهذه الخدمة لا توفر آلية للبحث من النوع الذي توفره الخدمات الأخرى.
عنوان هذه الخدمة هو:

<http://www.w3.org/Servers.html/www/Daa-Sources/hypertext/>

11. المكتبة الإلكترونية (Electric Library)

يمكن عن طريق إستعراض محتويات هذه المكتبة عدم الحاجة إلى البحث خلال الوب. فهي تضم قاعدة بيانات تحوى ألف صحيفة نصية (ewspaperText) (N) والعديد من المجالات والدوريات العلمية (Academic Journals) بالإضافة إلى العديد من الصور والمراجع العلمية (Reference Books) والأبحاث والموضوعات الفنية... الخ.

12. آلة البحث العملاقة (Search)

يضم هذا الموقع 250 آلة بحث يتم إستخدامها كآلة بحث عملاقة للغوص خلال أعماق الإنترنت. وتتيح هذه الآلة الولوج إلى كل الخدمات الرئيسية التي تعرضها الشبكة ومئات من قواعد البيانات المتخصصة وعنوان هذه الخدمة هو:

<http://www.Search.co>

13. الكتالوج العام للإنترنت (Whole Internet Catalog)

يتيح هذا الكتالوج 1600 موقع يمكن الولوج إليها مجاناً مع توصيف هذا الموقع بتقديم نبذة عن كل منها والدليل الذي توفره هذه الخدمة منظم بطريقة جيدة ولكنه لا يوفر آلية البحث من النوع الذي توفره الخدمات الأخرى وعنوان هذا الموقع هو:

<http://gnn.com/Wic/wics/index.html>

14 ياهو (Yahoo)

هذا الموقع من أكثر المواقع شعبية على الإنترنت من حيث المترددين عليه. وهذا أول موقع قدم خدمة البحث على الشبكة. لذلك فهو يحتوي على دليل منظم بطريقة جيدة ويضم كل محتويات شبكة ويب بحسب الموضوعات. ويمكن الولوج إلى هذه المحتويات بمتابعة إرشادات مرجعية تقود إلى الموضوعات ذات الصلة. ومن أهم المزايا التي تقدمها هذه الخدمة ميزة إنشاء مواقع على الويب (Web Lunch) وإدراجها في الدليل كما يتضمن دليلاً بمواقع ويب الأكثر شعبية وقوائم بمواقع مميزة ومثيرة للإهتمام وينصح بزيارتها. وعنوان هذه الخدمة هو:

<http://www.Yahoo.com>

3 - فهارس الإنترنت (Net):

يوجد صفحات من الويب تسمى فهارس (Directories) وهي من أهم الصفحات التي تساعد في عملية البحث عن موضوع عام (General Topics) مثل الفنون والتعليم والإدارة والأنشطة ويحتوي الدليل على مواقع (Sites) مرتبة حسب الموضوع.

فمثلاً يعرض دليل إكسيت (Excite) لمواقع تحوي موضوعات عديدة من بينها: الفنون – الأعمال – الحساب – التعليم – التسلية – الصحة والدواء – الأنشطة – المياه – المال والاستثمار – الأخبار – الصفحات الشخصية – السياسة والقانون – الديانات – العلوم – التسوق – الرياضة.

و كما هو واضح تغطي هذه المواقع العديد من مجالات الاهتمام وتنتج عند الولوج إليها مصادر واسعة من المعلومات، وعند النقر عليها ستظهر لك قائمة بالموضوعات الفرعية المندرجة تحت الموضوع الأساسي، وهكذا تتفرع داخل القوائم حتى تصل إلى ما تبحث عنه، وفيما يلي نقدم بعض أدلة الإنترنت الأكثر شعبية:

A2Z – Amazing Enviromental Organization Web Directory ! – Excite
– Gamelan – Infoseek Guide – Magellan – Point – W3 Server – World Wide
– Arts Resourees – Yahoo !

و الجدول الآتي يعرض لعناوين هذه المواقع:

العنوان

<http://www.yahoo.com>

<http://theyellowpages.com/feedback.htm>

الدليل

Yahoo

Yahoo pages.com

http://www.netcenter.com/netcenter/whats-new/whats-news.html	NetCenter
http://www.Lib.umich.edu/chheme.html	Internet resourceGuid
http://galaxy.einet.net	Galaxy
http://www.gold.net/gold/	The Global On-Line Mckinety Groups Magellan
http://www.Pointcom.com/	Point Communications
http://www.Homecom.com/global/pointers.html	Pointersto Pointers
http://www.stpt.com/until/submit.html	Starting Point
http://webspaces.com.nvoyager	Web Vayager
http://nearnnet.gnn.com/wie/index.htm	The whole Internet cataiog

و من افضل الطرق للوصول الى هذه الادلة هو الرجوع الى احد العناوين الآتية:

<http://home.netseape.com/home/internet-directory.html>

<http://www.lip.umich.edu/ehhome.html>

عند استخدام احد هذه الادلة مثل دليل اكسيت في عملية البحث، ومن خلال برنامج نيت سكيب (Net Scap) تظهر الموضوعات التي تحتوي على الكلمة التي تبحث عنها بجانبها نسبة تواجه الكلمة داخل الوثيقة و درجة الثقة (Confidence) و تعرض خدمة اكسيت امكانية تعميق عملية البحث (Refin Search) او اجراء بحث جديد (New Search) كذلك استعراض الاسئلة العامة المتعلقة بالموضوع بالاضافة الى العديد من الاليات التي تسهل عملية البحث و كل مواقع البحث و ما تقدمه من مصادر معلومات تضعك دائما في موقف المتسائل اياً من هذه المواقع يعطي افضل المصادر (Best Resour) و اياص منها يعطي اكثر المصادر (Most Resourees).

و بالرغم من تأثير حجم قاعدة البيانات و جودة الفهرسة على نتائج البحث عنها، فإن كيفية ادخالها ما زالت تؤثر على جودة النتائج، فليست كل آلات البحث تعامل الكلمة او الجملة التي تدخل بنفس الطريقة.

على سبيل المثال يقوم (Infoseek) بالبحث عن الكلمة كلها او اجزاء منها، فمثلا عند البحث عن كلمة (Compression) يمكنه مطابقة الجزء (Comprees) و بمجرد ان يجده يضمه الى نتائج البحث، بينما ليكوس

(Lycos) بمطابقة الكلمة او ما زاد عليها، فمثلا عند البحث عن كلمة ((Fluid يمكنه مطابقة (Fluidity) و ضمه الى نتائج البحث.
و تسمح بعض خدمات البحث باستعمال الكارت العام (Wild Cards) مثل "*"،
"؟" داخل الجملة مسار البحث. ومن الامثلة هذه الخدمات خدمة (Aliweb) و
عنوانها:

<http://web.nexor.co.uk/public/aliweb.html>

و يمكنك في معظم آلات البحث استخدام العوامل المنطقية (Logical Operators) مثل AND او OR او NOT.
و تفرد انفوسيك (Infoseek) و فهرس النص المفتوح (Open Text Index) بإمكانية البحث عن الكلمات المتقاربة (near) او المتواليه (Followd Next) من بين الكلمات المحددة في عبارة يمكنك استخدام آلية من آليات البحث المتعدد (Meta Search)، و التي تتيح استخدام اكثر من آلة للبحث في وقت واحد، و من اشهر صفحات الويب التي تعرض هذه الخدمة المدمجة صفحة (Savvy Search) و عنوانها:
<http://www.es.colostate.edu/dreiling\smartform.html>
و صفحة Meta Crawler و عنوانها:

<http://www.es.washington.edu/research/projects/melaerawler/www/home.html>

و باستخدام ادوات البحث المدمجة لا يسمح لك باستخدام مميزات آلة البحث المنفردة مثل العوامل المنطقية و مثيلاتها مما قد يؤثر بالسلب على دقة البحث.
سبعة نصائح مفيدة أثناء عملية البحث:

آلة البحث او خدمة البحث تستخدم قاعدة بيانات عبارة عن فهرس يضم مجموعة من الكلمات او العبارات يصاحبها عناوين لمواقع على شبكة انترنت. و عند استخدامك لهذه الادوات يجب ان تراعي جيدا اختيار و وضع الكلمات المناسبة لكي يتم مطابقتها و الحصول على فائدة اكبر من عملية البحث، لذلك إليك سبعة نصائح مفيدة لتوفير اقصى استفادة من ادوات البحث:

- 1- اختار كلمات بحث مميزة و ليست دارجة، و كلما كانت الكلمة التي تدخلها مميزة و غير معتادة و تعبر عن حاجتك كلما كانت النتائج افضل و ادق و اسرع.
- 2- لا تستخدم لغة التخاطب العادية: بالرغم من ان بعض ادوات البحث تتيح لك استخدام اللغة العادية و التي يمكنك من ان تسأل عما تبحث و كأنك تحدث إنسان مثلك لكن يفضل دائما استخدام الكلمات الهامة التي تعبر عن موضوع بحثك.
- 3- استخدام اكثر من آلة بحث: كرر تجربة البحث باستخدام اكثر من آلة للتأكد من حصولك على افضل و اقرب النتائج.
- 4- اقرأ تعليمات استخدام آلة البحث: توفر في كل آلة مجموعة من التعليمات التي تساعدك في اجراء البحث بكفاءة مثل المعاملات (Operators) و المحددات

(Delimiters) و قواعد البحث (Rules)، لذلك يجب ان تقرأ جيدا هذه التعليمات قبل استخدامك لآلة البحث.

5- الكلمات ذات المقطعين: عند البحث عن كلمة ذات مقطعين مثل Turbo-Propeller حاول تجربة كل الاشكال الممكنة لكتابتها مثل:

Turbo-Propeller - Turbo-Propeller -Turbo-Propeller

6- استخدام المعاملات المنطقية: تدعم بعض آليات البحث استخدام المعاملات المنطقية، و ينصح باستخدام المعامل (Not) اذا كان متاحا و الذي يمكنك من اسبعاد احد من البحث: فمثلا عند البحث بالصيغة الآتية:

Aircraft NOT Civil Subsonic

فأنك تحصل على معلومات عن كل الطائرات و قد تم استبعاد المعلومات الخاصة بالطائرات المدنية.

7- احرص على ذكر البدائل و المرادفات: و قد يكون ما تبحث عنه موجود بإسم مرادف او اسم بديل لذلك احرص على إدخال معظم الكلمات التي تدل على ما تبحث عنه لكي يزيد فرصتك في الحصول عليه.