

معیارهای ارزیابی موتورهای کاوش¹

علی حسین قاسمی²

چکیده

موتورهای کاوش در اینترنت، عمده‌ترین ابزار برای شناسایی و مکان‌یابی منابع اینترنتی هستند. شناخت ویژگی‌ها و کارکردهای این موتورها موجب خواهد شد که از منابع اینترنتی به شیوه‌ای کارآمدتر استفاده شود. در این مقاله پس از بیان علت به کارگیری و نحوه کار موتور کاوش، معیارهای ارزیابی موتورهای کاوش ارائه شده‌اند. سپس مختصری درباره موتورهای وراکاوش گفته شده و در پایان، با یک مثال، چهار موتور کاوش عمده در جدولی با هم مقایسه و ارزیابی شده‌اند. کلیدواژه‌ها: موتور کاوش، ارزیابی موتورهای کاوش، اینترنت، منابع اینترنتی، موتورهای وراکاوش، «گوگل»، «آل. د. وب. آدونسد»، «توما»، «آلتاویستا. آدونسد».

سخن گفتن از انفجار اطلاعات و حجم عظیم منابع داده‌ای و اطلاعاتی امروزه دیگر یک سخن تکراری است، اما این امر از اهمیت و تأثیر این پدیده در ظهور بسیاری از پدیده‌های دیگر نمی‌کاهد. ظهور اینترنت و وب را می‌توان یک انقلاب ارتباطی دانست که همراه با دیگر پدیده‌های اجتماعی، پیامدهای مثبت (و منفی) بسیاری به همراه داشت که تأمین امکان دسترسی به حجم انبوه داده‌ها و اطلاعات عمومی، تخصصی و شخصی، از جمله پیامدهای مثبت آنها است. اما همین وجه مثبت، در بسیاری اوقات موجب سردرگمی و آشفتگی کاربران می‌گردد. زمانی که منبع اطلاعاتی محدود و دارای اندازه کوچکی است. حتی با تورق می‌توان از محتوای آن آگاه شد؛ زمانی که تعداد منابع بیشتر می‌شود، فهرست مندرجات به یاری استفاده‌کننده می‌آید و با افزایش بازهم بیشتر تعداد منابع در کتابخانه، سازماندهی اطلاعات و دستیابی به محتوای منابع، از طریق فهرستبرگه حاصل می‌شود.

یکی از مشکلات فراروی فهرست‌نویسان کتابخانه‌ها، افزایش صعودی تعداد منابع بود که موجب تدوین راهکارها و راهبردهای گوناگون در گزینش و سازماندهی منابع گردید، اما ظهور وب نشان داد که آن مسائل و مشکلات، چندان دشوار نبوده‌اند. با وجود تعداد میلیاردی صفحات وب، نمایه کردن و مهم‌تر از آن، بازیابی این صفحات، از عهده انسان خارج می‌شود. هنر موتورهای کاوش³ در وب، کمک به انسان در نمایه کردن و سپس، بازیابی این صفحات است.

با این حجم عظیم اطلاعات چه باید کرد؟ چگونه می‌توان به اطلاعات کارآمد، روزآمد، مرتبط، دسترس‌پذیر و مناسب دست یافت؟ آیا می‌توان به بررسی یکایک صفحاتی که با ارائه پرسش به موتور کاوش،

¹ نویسنده لازم می‌داند که بدین وسیله از راهنمایی‌های استاد ارجمند جناب آقای دکتر رحمت‌الله فتاحی که تهیه این مقاله با هدایت ایشان ممکن گردید سپاسگزاری و قدردانی کند.

² دانشجوی دوره دکتری اطلاع‌رسانی و کتابداری دانشگاه فردوسی مشهد.

بازیابی می‌شوند پرداخت و از آن میان، منابع و صفحات سودمند را برگزید؟ هر کس که کمترین تجربه‌ای در کاوش بانک‌های اطلاعاتی داشته باشد می‌داند که پاسخ، منفی است و در مصاف با حجم انبوه حافظه‌ای که رایانه در اختیار ما می‌گذارد، باز هم باید از رایانه بهره گرفت. ارزیابی موتورهای کاوش، قابلیت و توانمندی‌های تعبیه‌شده در هر یک را آشکار می‌سازد و استفاده‌کننده، متناسب با نیازها و انتظارات خود و نیز متناسب با توانایی‌های هر موتور کاوش، به بررسی یافته‌های حاصل از هر کاوش و بهره‌برداری بهینه از آن توانایی‌ها می‌پردازد.

موتور کاوش را نیز همچون هر ابزار دیگر ارزیابی اطلاعات، از ابعاد گوناگون می‌توان بررسی کرد. برای حصول نتیجه بهتر، باید با تدوین معیارهای مشخص، به ارزیابی آن دسته از ویژگی‌هایی پرداخت که در کارکرد مورد انتظار از موتور کاوش، اثرگذارتر هستند.

درباره ارزیابی موتورهای کاوش در اینترنت، در زبان فارسی متأسفانه منابع بسیار اندکی موجود است که همه ترجمه مقاله‌های خارجی هستند. اما در منابع غیرفارسی، مطالب بسیار می‌توان یافت. در این منابع از جنبه‌های گوناگون به موتورهای کاوش، ساختار، نقش و کارکرد، عملکرد، سودمندی، ویژگیها، عیبهات و نکات مثبت همه یا تک‌تک آن پرداخته شده که اشاره به آنها از حد این مقاله بیرون است.

در این مقاله ابتدا به ساختار اینترنت به عنوان یک داده‌پایگاه⁴ بزرگ اطلاعاتی اشاره می‌شود، سپس رویکرد این مقاله به معیارهای ارزیابی موتورهای کاوش تعیین می‌شود. آنگاه به ذکر معیارهای ارزیابی می‌پردازیم و در پایان، ارزیابی چند پایگاه مهم اینترنتی ارائه می‌شود.

اینترنت چیست

اینترنت شبکه‌ای متشکل از شبکه‌هایی است که با استفاده از پروتکل «تی‌سی‌پی/آی‌پی»⁵ رایانه‌ها را به یکدیگر پیوند⁶ می‌دهد. هر رایانه برای ارائه اطلاعات و یا دسترسی به اطلاعات و مشاهده آن از نرم‌افزار استفاده می‌کند. اینترنت محل انتقال اطلاعاتی است که در فایل‌ها یا اسناد موجود در یک رایانه دیگر ذخیره شده‌اند. اینترنت را می‌توان به یک تأسیسات بین‌المللی ارتباطی تشبیه کرد که به رایانه‌ها خدمات می‌دهد. گاهی آن‌را به یک سامانه بین‌المللی غول‌پیکر که از لوله‌های ارتباطی تشکیل شده همانند می‌دانند. خود اینترنت، اطلاعاتی در خود ندارد؛ چندان درست نیست که بگوییم «سندی را در اینترنت پیدا کردم». صحیح‌تر این است که بگوییم از طریق اینترنت، یا با استفاده از اینترنت آن‌را پیدا کردم؛ آنچه یافته‌ایم نه در اینترنت، بلکه در یکی از رایانه‌هایی بوده که به اینترنت متصل‌اند.

موتورهای کاوش در وب، داده‌پایگاههای ایجاد شده بوسیله رایانه هستند که شامل پیوندهایی به هزاران منبع اینترنتی می‌باشند. کاربران موتور کاوش قادرند با این داده‌پایگاه تعامل کنند و پرس‌وجوهایی را به داده‌پایگاه ارائه نمایند و از او «پرسند» آیا به منابعی که با آن پرس‌وجو انطباق دارند، دسترسی دارد یا خیر.

چرا موتورهای کاوش به وجود آمدند

دشواری یا آسانی در یافتن اسناد وب (که به «پایگاه»⁷ها یا «صفحه»های وب معروف‌اند) تا اندازه‌ای به دلیل حجم گسترده‌گی وب است، حجمی که در حال حاضر بیش از دو میلیارد سند برآورد می‌شود. دلیل دیگر

4. Database

5. TCP/IP (Transmission Control Protocol/ Internet Protocol).

6. Link

7. Site

این است که وب با هیچ واژه‌نامه استاندارد نمایه نشده است (برخلاف فهرست کتابخانه‌ها که سرعنوانهای موضوعی پذیرفته شده و استاندارد مثل سرعنوانهای موضوعی کتابخانه کنگره را به کتابها اختصاص می‌دهند، یا نمایه مقاله‌های مجلات که از توصیفگرهای موضوعی استاندارد شده بهره می‌گیرند).

کاوش مستقیم در وب ناممکن است. وب در واقع مجموعه بسیاری از صفحات وب است که بر روی رایانه‌های (موسوم به «خدمتگر»⁸) سراسر دنیا قرار دارند. رایانه شما نمی‌تواند این صفحات را بیابد یا مستقیماً به سراغ آنها برود. آنچه با رایانه خود انجام می‌دهید، دسترسی به یکی از چندین صفحه وب یا داده پایگاههایی است که حاوی گزیده دیگر صفحات وب هستند و سازمان‌یافتگی این اطلاعات گزیده، این امکان را می‌دهد که دیگر صفحات وب، و گاهی دیگر داده پایگاهها را پیدا کنید. شما در این «ابزارهای کاوش» واسطه، کاوش می‌کنید و آنها پیوندهای فرامتنی⁹ («میم»¹⁰) به صفحات دیگر را در اختیار شما می‌گذارد. با کلید زدن بر روی این پیوندها است که اسناد، تصاویر، صدا و ... را از خدمتگرهای مختلف سراسر دنیا بازیابی می‌کنید.

موتور کاوش چگونه کار می‌کند

هر موتور کاوش دارای داده پایگاهی است که به وسیله برنامه‌های آدمواره‌ای¹¹ به نام «عنکبوت»¹² ایجاد می‌شود. این برنامه آدمواره‌ای رایانه‌ای در جستجوی محتوای لازم برای موتور کاوش، در شبکه وب به این سو و آن سو می‌خزد. این خزیدن¹³ یا راهنوردی¹⁴ در وب از طریق دنبال کردن پیوندهایی انجام می‌شود که او را به صفحات وب هدایت می‌کنند و قبلاً در داده پایگاه موتور کاوش اصلی وجود داشته‌اند. بنابراین اگر صفحه‌ای دارای هیچ پیوندی نباشد، این برنامه رایانه‌ای قادر به پیدا کردن و «دیدن» آن صفحه نخواهد بود.

از حاصل کار این برنامه‌های رایانه‌ای، یک داده پایگاه عظیم تهیه می‌شود که در واقع، نمایه‌ای از صفحات اینترنتی می‌باشد. وقتی کاربران، پرسشی را به موتور کاوش ارائه می‌دهند، موتور کاوش با مقابله و تطبیق پرسش با محتوای این داده پایگاه، یافته‌های¹⁵ به دست آمده را در قالب صفحات «نتیجه»¹⁶ به نمایش می‌گذارد. این نتایج و صفحات بر دو نوع‌اند: صفحاتی که بر مبنای پرسش‌های ارائه شده در یک کاوش خاص (و از یافته‌های خاص همان کاوش) پدید می‌آیند صفحات «پویاساخت»¹⁷ نامیده می‌شوند. در این حالت، پاسخ‌های یافت شده برای پرسش شما، در قالب یک صفحه وب که برای همین منظور طراحی شده ریخته و به رایانه شما ارسال می‌شوند. غالباً این صفحه نوساخته در هیچ جایی ذخیره و نگهداری نمی‌شود، زیرا فرض بر این است که محتوای منحصر به فرد آن (که در واقع، پاسخ سؤال شما بوده) احتمالاً برای بسیاری از دیگر افراد و کاربران، استفاده چندانی ندارد. بعلاوه، برای داده پایگاه موتور کاوش، راحت تر این است که به جای نگهداری آن صفحه، در هنگام نیاز دوباره آن را بسازد.

شکل دیگر صفحات، صفحه‌های «ایستاساخت»¹⁸ هستند. صفحات ایستاساخت بر روی خدمتگرها مقیم

8. server
9. Hyper-text Links

۴. میم = مکان‌نمای یکسان منبع (Unique Resource Locator).

11. Robot programs
12. spider
13. crawl
14. navigate
15. hits
16. result
17. Dynamically-generated
18. static

می‌شوند و هریک، با «میم» منحصر به فردی مشخص می‌گردند. زمانی که این «میم» بازخوانی شود، این صفحه‌ایستاساخت نیز بازیابی می‌گردد. اگر صفحه‌های ایستاساخت به هر صفحه‌ای که عنکبوتها آنها را «می‌شناسند» پیوند داشته باشند، عنکبوتها می‌توانند آنها را پیدا کنند؛ آنان این پیوندها را دنبال و صفحه را بازیابی می‌کنند. البته گاهی موتورهای کاوش، به دلایل مربوط به خط‌مشی عملکردی خود، تصمیم به حذف این صفحات ایستاساخت از داده‌پایگاه خود می‌گیرند.

دسته‌بندی ابزارهای کنونی کاوش

موتورهای کاوش بسیاری در وب موجودند که ترکیب کاوش¹⁹ در همه آنها یکسان نیست. هریک از این موتورهای کاوش، اهداف متفاوتی را دنبال می‌کنند و ویژگی‌های موجود در آنها نیز با توجه به همین اهداف در آنها تعبیه شده‌اند. به قولی بیش از 800000 موتور کاوش در اینترنت فعال‌اند که ما تنها با تعداد انگشت‌شماری از آنها آشنا می‌شویم. این موتورها با اهداف تجاری، شغلی، تخصصی، علمی، پژوهشی، نژادی، سیاسی، فرهنگی و حتی شخصی پدید آمده‌اند و بدیهی است که در تعیین و تعبیه ویژگی‌های این موتورها، اهداف گردانندگان نیز مدنظر قرار گرفته است. جدای از این، سطح تخصص و نوع نگاه پدیدآورندگان این موتورها نیز مهم است. مثلاً در موتورهای کاوش دانشگاهی، آنچه اهمیت می‌یابد محتوای منابع اطلاعاتی موردنظر می‌باشد؛ در موتورهای کاوش کتابداران و اطلاع‌رسانان، دسته‌بندی محتوا و افزایش میزان ربط، نمود و جلوه بیشتری می‌یابد، و در موتورهای کاوش تجاری، از هر فرصتی برای بازشدن یک صفحه تبلیغاتی و آگهی‌های گوناگون استفاده می‌شود.

به هر حال، این تنوع و گوناگونی، امکان‌پذیر است. این موتورها را با توجه به اهداف کاربران - و به ویژه کتابداران - فراهم می‌آورد. در دنباله این مطلب به معیارهایی که می‌توان با استفاده از آنها به ارزیابی موتورهای کاوش پرداخت اشاره می‌کنیم. بدیهی است موتوری را نمی‌توان یافت که همه این معیارها را در خود داشته باشد. بنابراین، ارزش هر موتور کاوش، نسبی است و باید در ارتباط مستقیم با اهداف کاربران، تعریف و تبیین شود.

معیارهای ارزیابی و بررسی موتورهای کاوش

با توجه به فراگیری اینترنت و استقبال فزاینده کاربران و کاربران نهایی از آن به عمل می‌آورند، رویکرد ما در ارائه معیارهای ارزیابی موتورهای کاوش، از منظر متخصصان کتابداری و اطلاع‌رسانی است. اما تعیین معیارهایی که به تمام و کمال، اهداف موردنظر کتابداران و اطلاع‌رسانان را در حوزه ذخیره و بازیابی اطلاعات محقق کند، با توجه به وضعیت کنونی اینترنت و موتورهای کاوش - دست کم در کوتاه‌مدت - چندان فایده‌ای ندارد. بنابراین، هدف از ارائه این معیارها آن است که ابزاری به دست کاربران واسطه (مانند کتابداران) و کاربران نهایی داده شود تا با استفاده از آن، موتورهای بهتر و سودمندتر را شناسایی کنند. این معیارها عبارتند از:

1. آیا راهنمای²⁰ کاوش (که در آن، امکانات و شیوه‌های کاوش معرفی شده باشند) وجود دارد؟

1.1. آیا دسترسی به این راهنما آسان و سریع است؟

2.1. آیا این راهنما در سایت موتور کاوش، به صورت درونخطی در اختیار قرار دارد؟

19. Search syntax
20. help

- 3.1. آیا این راهنما هوشمند است (یعنی در هنگام خطای کاربر، به طور خودکار راهنمایی می کند - مثلاً غلطهای املائی مربوط به عبارت کاوش را مشخص می کند)؟
2. آیا موتور کاوش، آسان تر و کاربرپسند²¹ است؟
3. آیا مطلب و محتوای ارائه شده در صفحه، از ساختار و آرایش مناسب برخوردارند؟
4. کاوش در چه مقدار از متن کامل صفحات انجام می شود (یا به عبارت دیگر، آیا در همه یا هریک از فیلدهایی که نمایه شده اند، کاوش انجام می شود)؟ مثلاً در کدامیک از موارد زیر:
- فیلد موضوع - فیلد تاریخ (دوره زمانی) - فیلد عنوان - فیلد نام ها - فیلد نشانی فیلدهای دیگر (مثلاً کشوری که سایت در آن قرار دارد).
5. چند نوع کاوش، امکانپذیر است:
- ساده - پیشرفته - کلیدواژه ای - انواع دیگر
6. میزان دقت در انطباق کلمات مورد کاوش با محتوای صفحات تا چه حد است؟
7. آیا در نمایه/ کاوش، از زبان کنترل شده/ اصطلاحنامه استفاده شده/ می شود؟
8. آیا امکان کاوش با اصطلاحات عام و کلی وجود دارد؟
9. آیا امکان کاوش با اصطلاحات خاص و جزئی وجود دارد؟
10. از چه شیوه های کاوش استفاده می شود:
- کاوش معناشناختی²² - کاوش وزنی²³ / وزن دادن به واژه های کاوش به صورت پیش گزیده - کاوش عبارتی²⁴ - کاوش بولی²⁵
11. در کاوش بولی، چه امکاناتی در دسترس قرار می گیرند:
- AND - OR - AND NOT - NEAR - ANY
12. آیا امکان کاوش فرعی²⁶ / کاوش در یافته ها²⁷ / محدود کردن کاوش²⁸ / پالایش یافته ها²⁹ وجود دارد؟
13. در کاوش، چه گزینه هایی در اختیار کاربر قرار می گیرد:
- امکان کاوش واژه های بازدارنده³⁰ - امکان کاوش اسم خاص - حساسیت به حالت (بزرگی و کوچکی) حروف (لاتین) - دم برش³¹ - ریشه گیری³² - استفاده از جانویسه³³ - تعداد جانویسه های مورد پذیرش - یافتن مجاورت معنایی - امکان دسته بندی واژه های کاوش (مثلاً با قرار دادن آن ها در داخل پرانتز)
14. آیا محدود کردن کاوش به فیلدهای زبان «اچ تی ام ال»، یعنی فیلدهای زیر، امکانپذیر است:
- (رایانه میزبان) Host - (پیوند) Link - (نشانی اینترنتی) Url - (عنوان) Title
 - (تصویر) Image - (متن سند) Text - (متن پیوند) Anchor - (قلمرو) Domain
 - (کلیدواژه) KW - (توضیحات تصاویر) Alt - (بایگه) Site - (برنامک) Applet

-
21. User-friendly
 22. Semantic search
 23. Weighted search
 24. Phrase search
 25. Boolean search
 26. Sub-search
 27. Search within results
 28. Limiting the search
 29. Refining/ filtering the results
 30. Stop words
 31. truncation
 32. Stemming
 33. Wildcard

(شرح 34) AN - (پدیدآورنده) AU - (موضوع) SU -

15. آیا می‌توان کاوش را به شکلهای زیر محدود کرد:
محدود به ترکیبی از چند فیلد - محدود به فیلد موضوع - محدود به دامنه زمانی مشخص - محدود به زبانهای غیر لاتین
16. آیا امکان کاوش همزمان به چند زبان / خط وجود دارد؟
- 16-1. آیا تعداد زبان / خطهای انتخابی محدود است؟
- 16-2. آیا انتخاب نوع زبان / خط اختیاری است یا گزینه‌های مشخصی برای این کار وجود ندارد؟
17. میزان دقت در دستچین کردن پایگاهها به وسیله ویراستاران مجموعه چقدر است؟
- 17-1. میزان ربط³⁵ موضوعی و ارتباط³⁶ محتوایی اسناد با یکدیگر چقدر است؟
18. صفحات نمایه شده دارای ارزش محتوایی (تجاری، علمی، تبلیغاتی، ...) هستند؟
19. آیا موتور کاوش دارای فهرستهای فرعی گوناگونی بر مبنای منطقه جغرافیایی و تقسیمات کشوری است؟
20. آیا دسته‌بندی محتوایی بر روی اطلاعات انجام شده است؟ نحوه این دسته‌بندی چگونه است:
موضوعی - سلسله‌مراتبی - موضوعی / سلسله‌مراتبی
21. آیا از صفحات نمایه شده، شرح و توصیفی نیز ارائه شده است یا نه؟
- 21-1. آیا این شرح، مختصر است یا مفصل؟
- 21-2. آیا این توصیف، تخصصی است یا غیر تخصصی؟
22. آیا حوزه موضوعی صفحات (سطحی / عمومی، تجاری، خاص یا تخصصی) مشخص شده است؟
23. آیا امکان مرور / تورق³⁷ مقوله³⁸ های موضوعی موجود در خود موتور کاوش وجود دارد؟
24. آیا می‌توان از طریق یافته‌های کاوش، به متحویات داده پایگاهی که در جای دیگری بر روی یک رایانه دیگر واقع شده، دسترسی یافت.
25. آیا امکان محدود کردن کاوش از طریق بازیابی صفحات مشابه³⁹ وجود دارد؟
26. آیا تورق و مرور یافته‌ها امکان‌پذیر است؟
27. آیا امکان دسته‌بندی موضوعی یافته‌ها وجود دارد؟
28. آیا امکان ردیف کردن یافته‌ها به شیوه‌های زیر وجود دارد؟
به ترتیب تاریخ - به ترتیب الفبا - بر مبنای فیلدهای مختلف - به طریق دیگر
29. نحوه گردآوری صفحات چگونه بوده است:
- 29-1. کلاً بوسیله «عنکبوت⁴⁰ ها» (برنامه‌های رایانه - روایات) با کم‌ترین نظارت انسانی
- 29-2. بوسیله «عنکبوت‌ها» (برنامه‌های رایانه روایات)، همراه با نظارت انسانی
- 29-3. بوسیله نیروی انسانی
30. حجم (و تعداد صفحات نمایه شده) در داده پایگاه مربوط به موتور کاوش چقدر است؟

34 annotation
35. relevance
36. relationship
37. browse
38. category
39. Similar pages
40. spider

- 1.30. فواصل زمانی تغییر/ میزان تغییر در تعداد این صفحات چقدر است؟
31. سرعت موتور کاوش در پاسخگویی به پرسش/ بازیابی و اخذ نتایج/ بازنمایی نتایج چقدر است؟
32. از نظر محتوا:
- 1.32. میزان سطحی/ عمقی بودن کاوش انجام شده و نتایج بازیابی شده چقدر است؟
- 2.32. رویکرد موتور کاوش به محتوای موردنظر شما، عام است یا تخصصی؟
- 3.32. گستره موتور کاوش در جنبه‌های زیر چه وضعیتی دارد:
- گستره جغرافیایی - گستره موضوعی - گستره زبانی - ابعاد دیگر
33. میزان جامعیت⁴¹ داده پایگاه تا چه حد است؟
34. در بازنمایی یافته‌های کاوش:
- 1.34. آیا امکان تعیین تعداد یافته‌هایی که در هر صفحه نمایش داده می‌شوند وجود دارد؟
- 2.34. آیا امکان تعیین ترکیب محتوایی رکوردها (فیلدها) در هنگام نمایش وجود دارد؟
- 3.34. آیا امکان تعیین ویژگیهای بصری صفحه نمایش وجود دارد؟
35. قالب بازنمایی یافته‌هایی که از موتورهای مختلف پدید آمده‌اند چگونه است؟
36. نحوه بازنمایی یافته‌های فراهم آمده از موتورها/ منابع مختلف چگونه است:
- مجزا/ درهمکرد - با حذف موارد تکراری
37. چه مقدار (چند درصد) از یافته‌های فراهم آمده از هر یک از موتورهای مختلف بازنمایی می‌شوند؟
38. آیا یافته‌ها رتبه‌بندی/ دسته‌بندی می‌شوند؟ بر چه مبنایی:
- 1.38. بر مبنای وزنی که به موجب توالی اصطلاحات کاوش (مثلاً از چپ به راست) به اصطلاحات داده می‌شود.
- 2.38. بر مبنای تعداد دفعات رؤیت هر صفحه/ تعداد مراجعان به هر صفحه
- 3.38. بر مبنای تعداد پیوندهای ارجاع داده شده به هر صفحه
- 4.38. بر مبنای تاریخ نشر یافته‌ها
- 5.38. بر مبنای موضوع یافته‌ها
39. آیا گزینه‌هایی برای تعیین مبنای رتبه‌بندی پیش‌بینی شده است؟
40. آیا دسترسی به متن کامل صفحاتی که در نتیجه کاوش به آنها اشاره شده (یعنی تحویل مدرک⁴²) امکانپذیر است؟
- 1.40. آیا این دسترسی به متن کامل، رایگان است؟
- 2.40. اگر نه، آیا قیمت تحویل مدرک و گزینه‌های مختلف در این رابطه، اراعه شده‌اند؟
41. آیا ترجمه متن اسناد/ یافته‌ها به زبانهای مختلف امکانپذیر است؟

مختصری درباره موتورهای «وراکاوش»

یکی از دلایل استفاده از موتورهای وراکاوش⁴³، حجم گسترده اسناد موجود در وب و ناتوانی موتورهای کاوش در نمایه کردن همه منابع موجود در وب است که موجب نامرئی ماندن بخشی از وب

41. Comprehensiveness
42. Document delivery
43. Meta-search engines

می‌شود. استفاده از موتورهای وراکاوش یکی از راه‌های غلبه بر این نامرئی بودن است. در موتور وراکاوش، شما کلیدواژه‌ها را وارد خانه یا محل کاوش می‌کنید و موتور وراکاوش، کاوش شما را به‌طور همزمان به چندین موتور کاوش جداگانه و داده‌پایگاه صفحات وب ارسال می‌کند. در چند ثانیه، یافته‌های حاصل از تمام موتورهای کاوش مورد نظر برگشت داده می‌شوند. موتورهای وراکاوش، خودشان دارای داده‌پایگاهی از صفحات وب نیستند، بلکه اصطلاحات کاوش را برای داده‌پایگاههایی که در اختیار دیگر موتورهای کاوش است می‌فرستند.

در موتورهای کاوش معمولی، کلیدواژه‌ها را به داده‌پایگاه صفحات وب که در اختیار موتور کاوش است عرضه می‌کنیم و از هر موتور کاوش، نمایش متفاوتی از اسناد دریافت می‌کنیم. یافته‌های حاصل از کاوش‌های مشابه ممکن است تا حد زیادی (نزدیک به 40 درصد) با هم متفاوت باشند، اما درعین حال شامل پایگاههای همسان (در حدود 60 درصد) نیز هستند.

برخی پایگاههای مربوط به موتورهای وراکاوش، خدمات سودمند ردیف دوم، و مجموعه‌های تخصصی از پایگاهها و یا منابع پیشه‌گانی⁴⁴، طراحان وب، مشتریان سینما، و... را نیز عرضه می‌کنند. گروهی دیگر نیز خدمتی را ارائه می‌دهند که می‌توان آن را «شبه وراکاوش» نامید و آن عبارتست از: مجموعه‌ای از خانه‌های کاوش برای موتورهای مختلف کاوش، یا منو⁴⁵یی که امکان می‌دهد از میان فهرست موتورهای کاوش، یکی را برای کاوش خود انتخاب کنید. ما دربارهٔ هیچ‌یک از این خدمات توضیحی نمی‌دهیم، زیرا شبه‌وراکاوشگرها به‌جای این که وراکاوشگر باشند، درواقع مجموعه‌ای از داده‌پایگاههای قابل کاوش هستند.

برخی از ابزارهای وراکاوش برپایهٔ جاهایی که کاوش می‌کنند، سهولت کاربرد، امکان تمرکز بر یک وجه خاص، توانایی انجام هوشمندانهٔ کاوش‌های پیشرفته‌تر (مثلاً ترجمه یا کاربرد دقیق عملگرهای بولی) در جایگاهی برجسته‌تر قرار می‌گیرند.

محدودیت موتورهای وراکاوش

سه عامل عمده در تعیین سودمندی هر موتور وراکاوش وجود دارد که عبارتست از:

1- موتورهای کاوشی که موتور وراکاوش، اصطلاحات کاوش شما را برای آنها می‌فرستد: اندازهٔ داده‌پایگاه مربوطه، محتوای آن، تعداد موتورهایی که کاوش برای آنها ارسال می‌شود، امکان انتخاب موتورهای کاوشی که آنها را ترجیح می‌دهید. همهٔ موتورهای وراکاوش علاوه بر موتورهای کاوش، در فهرستهای موضوعی نیز کاوش می‌کنند و یافته‌های به دست آمده از همهٔ این منابع را درهم می‌آمیزند.

2- شیوهٔ کار موتور وراکاوش بر روی اصطلاحات کاوش و ترکیب کاوش: مثل عملگرهای بولی، یا عبارات و پیش‌گزیده⁴⁶هایی که دارند.

3- شیوهٔ نمایش یافته‌ها: مثلاً امکان رتبه‌بندی یافته‌ها، ادغام آنها در یک سیاههٔ واحد یا گزارش یافته‌های حاصل از هر موتور کاوش به صورت جداگانه. توجه به این نکته ضروری است که در استفاده از موتورهای وراکاوش باید احتیاط کرد؛ زیرا:

1- بیشتر موتورهای وراکاوش تنها زمان بسیار کوتاهی را در هر داده‌پایگاه صرف می‌کنند و غالباً فقط 10 درصد یافته‌های هر یک از داده‌پایگاههای مورد پرسش را بازیابی می‌کنند. این امر معمولاً باعث

44. business

45. menu

46. default

«سرعت» و نیز «به هم ریختگی» در ارائه نتایج کاوش های آنها می شود، اما در عین حال غالباً می توانید از حاصل کار، آنچه را در پی آن هستید بیابید.

2- بیشتر وراکاوشگرها فقط به شکل خطی به اصطلاحات کاوش می پردازند و اگر کاوش شما حاوی بیش از یکی دو کلمه، یا دارای منطق بسیار پیچیده ای باشد، بخش اعظم آن از دست خواهد رفت. پشتیبانی چنان منطق هایی فقط در معدودی از موتورهای کاوش انجام می گیرد (به جدول «ارزیابی موتورهای کاوش» مراجعه نمایید).

3- بالا بودن کمیّت یافته ها به معنای رضایتمندی نیست. اگر یافته هایی بیش از آنچه می خواستید به دست آورید، از موتورهای وراکاوشی استفاده کنید که امکان کاوش فرعی یا «کاوش در یافته ها» را نیز به عنوان گزینه انتخابی ارائه می دهند.

4- به دنبال موتورهای وراکاوشی بگردید که اصطلاحات کاوش شما را برای داده پایگاههای خاص یا غیرعادی (مثل «تاندراستون»⁴⁷، «دایرکت هیت»⁴⁸، «وب کراولر»⁴⁹ و «وات یوسیک»⁵⁰) نیز بفرستند. یکی از مزایای وراکاوشگر همین است. چه بسا خود شما نیز به داده پایگاههایی از این قبیل، که ممکن است دارای پایگاههایی باشند که از چشم موتورهای کاوش بزرگتری پنهان مانده، پی نبرده باشید.

مثالی از ارزیابی موتورهای کاوش

نام موتور کاوش	گوگل www.google.com	آل.د.وب.آدوانس www.alltheweb.com/advanced	تئوما www.teoma.com	آلتاویستا آدوانس www.av.com
پیوند به راهنمای تفصیلی	بله	بله	بله	بله
اندازه، نوع (اندازه) پایگاهها مکرراً و به شکل گسترده تغییر می کند	داده پایگاه عمومی وب که غالباً دارای رتبه بندی سودمندی از یافته ها بر مبنای میزان محبوبیت صفحه ها است، بسیار عظیم می باشد و ادعا می کند که بیش از 2/5 میلیارد صفحه را نمایه می کند، ولی ممکن است همه صفحه ها به شکل کامل نمایه نشده باشند. نمی توان آن را جامع دانست، ولی غالباً «بهترین» صفحات را پیدا می کند.	داده پایگاه عمومی که بسیار عظیم است و ادعا می کند که تعداد صفحاتش به زودی به یک میلیارد صفحه خواهد رسید. گاهی در آزمون ها در رده اول قرار می گیرد. دارای رتبه بندی بسیار عالی است و کاوش پیشرفته آن بسیار سودمند است.	داده پایگاه بزرگی است که در ردیف «آل دوب» و «آلتاویستا» قرار می گیرد.	داده پایگاه عمومی بزرگی که ادعا می کند بزرگترین است. در قسمت کاوش پیشرفته آن می توان با استفاده از عملگرهای بولی به کاوش پرداخت. گزینه هایی نیز برای رتبه بندی یافته ها دارد.
امکان کاوش	بله. اگر از علامت گیومه «» استفاده کنید،	بله. از علامت گیومه «» استفاده کنید. در کاوش	بله. از علامت گیومه «» استفاده کنید.	بله. از علامت گیومه «» استفاده

47. Thunderstone
48. Direct Hit
49. WebCrawler
50. WhatUseek

کنید.		پیشرفته، اصطلاحاتی که در مستطیل های «فیلتر» پائینی قرار گیرند همیشه به صورت عبارتی کاوش می شوند.	حتی واژه های بازدارنده رایج را نیز کاوش می کند	عبارتی
AND (پیش گزیده)، AND، OR NEAR، NOT (تا محدودی 10 کلمه، که فقط در مستطیل بزرگتر کاوش پیشرفته باید از آن استفاده کرد)	نه. همیشه فرض بر وجود AND در بین کلمات کاوش است.	در مستطیل بالایی، عملگر پیش گزیده، AND است. در مورد OR بدون این که واژه OR را بنویسد، کلمه یا عبارت مورد کاوش را در داخل پرانتز قرار دهید. مستطیل های «فیلتر» در کاوش پیشرفته، برخی معادل ها را برای منطق بولی ارائه می کنند: Must include مساوی با And و Must not include مساوی با NOT از نظر رتبه بندی در رده بالایی قرار می گیرد. از OR استفاده نمی کند.	تاحدی. فرض بر وجود AND در بین کلمات است. OR را باید با حروف بزرگ بنویسد. فاقد امکان استفاده از پرانتز برای دسته بندی عبارات کاوش است. در کاوش پیشرفته، از عملگرهای بولی محدودی می توان استفاده کرد. ALL، مساوی با AND، WITHOUT مساوی با AND NOT و AT، و LEAST ONE مساوی با OR. محدودیت به تعداد 10 کلمه در هر کاوش، به استثنای کاوش با OR.	منطق بولی
فقط در کاوش ساده ممکن است. در کاوش پیشرفته، بهتر است از منطق بولی استفاده شود، که بسیار قدرتمندتر است.	علامت + امکان بازیابی واژه های بازدارنده را می دهد	علامت -	اگر پیش از واژه ای علامت منفی (-) بیاورید، صفحاتی که حاوی آن واژه باشند بازیابی نمی شوند. علامت + امکان بازیابی واژه های بازدارنده را می دهد.	علامت + علامت -
بله. از sorted by box در زیر مستطیل کاوش بولی استفاده کنید. یافته های کاوش را ردیف و پالایش می کند.	نه. سیاهه REFINE موضوعهای فرعی یا communities را ارائه می کند.	نه	بله. در پایین صفحه یافته ها بر روی «کاوش درون یافته ها» کلیک بزنید و اصطلاحات بیشتر را وارد کنید	کاوش فرعی
Fuzzy AND خودکار. برخی از یافته های بالاتر، این مکان بالاتر را	بر مبنای محبوبیت، Subject-specific از (دسته بندی موضوعی)، پیوندهای	Fuzzy AND خودکار، ظاهراً از "importance" و پیوندهای صفحات نیز	بر مبنای تعداد پیوندهایی که از سایت های دیگر به آن داده شده، امکان	رتبه بندی یافته ها

خریداری کرده‌اند (و ارتباطی با واژه‌های مورد کاوش شما از این نظر ندارند).	داده شده به صفحه از صفحات مرتبط.	استفاده می‌کند. در کاوش پیشرفته، SHOULD INCLUDE بالاتر را به کلمات یا عبارات داخل مستطیل می‌دهد. هر مستطیل یک عبارت شمرده می‌شود.	استفاده از Fuzzy 51 AND نیز وجود دارد. انطباق و رتبه‌بندی ممکن است بر مبنای نسخه «ضبط شده» صفحات چندان هم جدید نباشد.	
Title: url: Link: Host: Domain: Anchor: Text: Image: Applet:		در کاوش پیشرفته می‌توان در فیلدهای زیر کاوش کرد: Text Title Link name url Link to the url و می‌توان بر مبنای اصطلاحات قلمرو، یافته‌ها را پالایش کرد.	Link: Site: Allintitle: Allinurl: امکان استفاده آسان از این امکانات در قسمت کاوش پیشرفته. برای صفحات مربوط به ایالت‌های آمریکا گزینه Uncle Sam را ارائه می‌دهد.	محدود کردن به فیلد
بله. از علامت ستاره * استفاده کنید.	نه	نه. شکل‌های مختلف کلمه را در مستطیل بالایی در داخل پرانتز قرار دهید تا با استفاده از OR کاوش شوند.	بدون دُم بُرش و بدون ریشه‌گیری. شکل‌های مختلف انجامه‌ها و مترادفها را جداگانه و با استفاده از OR (با حروف بزرگ) کاوش می‌شوند.	دُم بُرش
بله. حروف بزرگ، فقط کلماتی را که با حروف بزرگ نوشته شده بازیابی می‌کنند. حروف کوچک، کلمات با حروف بزرگ را نیز بازیابی می‌کنند. به «اکسنت» و سایر نویسه‌ها نیز حساس است.	نه	نه	نه	حساسیت به حالت (بزرگی و کوچکی) حروف
بله. سیاهه گسترده آن، عمده زبان‌های رومی و غیررومی را در بر می‌گیرد. مجموعه نویسه‌های دلخواه را نیز می‌دهد.		بله. سیاهه گسترده آن، عمده زبان‌های رومی و غیررومی را در بر می‌گیرد. امکان انتخاب مجموعه نویسه‌های دلخواه را نیز می‌دهد.	بله. در کاوش پیشرفته، عمده زبان‌های رومی و غیررومی را می‌پذیرد.	زبان

۱. با استفاده از **Fuzzy AND** اشتراک بین دو مجموعه (و بنابراین تعیین میزان ربط) برخلاف **AND** بولی نه با بله یا خیر، بلکه با عددی بین صفر و یک نشان داده می‌شود.

محدودیت بر مبنای قدمت اسناد	در کاوش پیشرفته.	در کاوش پیشرفته.	نه	در کاوش پیشرفته.
ترجمه	ترجمه از زبان‌های عمده اروپایی به انگلیسی و بالعکس از طریق پیوند translate this page که در دنباله بعضی از صفحات آمده.	نه	نه	بله. بین انگلیسی و دیگر زبان‌ها. بر روی translate که در دنباله یافته‌ها آمده کلیک کنید.

نتیجه

حجم انبوه و روزافزون وب و نیز روندهای فنی حاکم بر آن، استفاده از موتورهای کاوش را در بازیابی اطلاعات، گریز ناپذیر می‌کنند. اما ماهیت کار موتورهای کاوش موجب می‌شود که در هنگام بازیابی، حجم گسترده‌ای از صفحات بازیابی شوند که در نتیجه، از کارآیی کاوش در اینترنت کاسته می‌شود؛ یا این که در پاسخ به یک پرسش واحد، نتایج بازیابی در بسیاری از موتورهای کاوش، متفاوت است. برای غلبه بر این مشکل، باید در حال حاضر از آن موتورهایی استفاده کرد که مزایای بیشتر و معایب کمتری داشته باشند. هرچه میزان انطباق میان نیاز واقعی استفاده‌کننده و یافته‌های کاوش بیشتر باشد، و به عبارت دیگر هرچه میزان ربط یافته‌ها با موضوع مورد کاوش بیشتر باشد، کاوش با موفقیت بیشتری انجام گردیده است.

معیارهایی که در صفحات قبل ذکر کردیم از راههای گوناگون به افزایش میزان انطباق میان کلمات با نمایه موجود در موتور کاوش کمک می‌رسانند و از این طریق می‌توان به نتایج بهتری دست یافت.

منابع

- "Search Engine Submission!". WSuhmiI.com.
<http://www.wssubmit.com/addyourwebsite.cfm> (8 Jan. 2003)
- Ryan, Tracy "XML promises better search, navigation techniques". UNISYS WORLD. June 2000.
http://www.unisysworld.com/monthly/2000/06/xml_search.shtml
(8 Jan. 2003)
- Horn Fell, Pat and Hansen, David T. "Construction of a Theme Keyword Thesaurus for Indexing Search and Retrieval across Networks". ESRI 1997 User Conference in San Diego CA, USA on July 8 to 11 1997.
<http://www.mp.usbr.gov/geospat/p187.html> (8 Jan.2003)
- Abilock, Debbie "Choose the Best Search for Your Information Need". Information Literacy: Search Strategies. 10/6/2002.

- <http://www.noodletools.com/debbie/literacies/information/5locate/adviceengine.html>
(8 Jan. 2003)
5. Gray, Terry A. "How to Search the Web; A Guide To Search Tools".
<http://daphne.palomar.edu/TGSEARCH/> (8 Jan. 2003)
6. "Individual Search Engines: Features and Help". 03/10/2002.
<http://exlibris.colgate.edu/web/finding/finding2.htm> (8 Jan. 2003)
7. Paul, Kathryn. "Internet Searching". August 17, 2002.
<http://gateway.uvic.ca/is/isrev.html> (8 Jan. 2003)
8. "Questions and Answers about the Internet, Web and Search Engines". http://c-library.um.ac.ir/Questions_Answers_Internet.htm (8 Jan. 2003)
9. "What are Search Engines?" http://c-library.um.ac.ir/what_se.htm (8 Jan. 2003)
10. Leighton, H.Vernon and Srivatava, Jaideep. "Precision among World Wide Web Search Services (Search Engines): Alta Vista, Excite, Hotbot, Infoseek, Lycos" 8/29/97
<http://www.winona.msus.edu/library/webind2/webind2.htm> (8 Jan. 2003)
<http://www.winona.msus.edu/library/webind2/wi2pt2.htm#EVALCRIT> (8 Jan. 2003)
<http://www.winona.msus.edu/library/webind2/wi2pt3.htm#RESULTS> (8 Jan. 2003)
<http://www.winona.msus.edu/library/webind2/wi2appen.htm> (8 Jan. 2003)
11. Barker, Joe "Beyond General World Wide Web Searching". Finding Information on the Internet: A Tutorial. 1 October 2002.
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/BeyondWeb.html> (8 Jan. 2003)
12. Barker, Joe "Evaluating Web Pages: Experience Why It's Important". Finding Information on the internet: A Tutorial. 6 January 2003.
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Evaluate.html> (8 Jan. 2003)
13. Barker, Joe. "Things To Know Before You Begin Searching". Finding Information on the Internet: A Tutorial. 18 Aug. 2001.
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/ThingsToKnow.html> (8 Jan. 2003)
14. "What is the Internet, the World Wide Web, and Netscape?". Finding Information on the Internet: A Tutorial. 27 Sept. 2002.
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/WhatIs.html> (8 Jan. 2003)
15. "The BEST Search Engines". Finding Information on the Internet; A Tutorial. 27 September 2002.
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html> (8 Jan 2003)