

تکامل وب و مقایسه ابزارهای جستجو در اینترنت

فصلنامه اطلاع رسانی. دوره ۱۶، شماره ۴ و ۳

تالیف: جیران خوانساری

کارشناس ارشد کتابداری و اطلاع رسانی - مرکز اطلاع رسانی کتابخانه توانیر

چکیده

این مقاله تکامل وب را در رابطه با وب و فناوری‌های بازیابی اطلاعات بررسی می‌کند. در این بررسی بین عناصر مختلف وب در زمینه نمایه‌سازی و صفحات جستجوی وب وجه تمایز وجود دارد. ده ابزار اصلی و مهم وب براساس معیارهای انتخاب و تعداد اتصال‌ها با یکدیگر مقایسه و ارزیابی شده‌اند. ابزارهای جستجو مانند راهنماها و موتورهای جستجوگر نیز مورد بررسی و مقایسه قرار گرفته‌اند.

کلیدواژه: وب، راهنماهای جستجوگر، موتورهای جستجوگر

مقدمه

کلمه اینترنت مخفف "Interconnected Networks" و به معنی شبکه‌های به هم مرتبط است. اینترنت سیستمی برای توزیع و تبادل اطلاعات در سطح جهان است که براساس آخرین آمار موجود در حال حاضر ۲۲۶ کشور به آن متصل هستند. این مجموعه به همراه فناوری‌های ساده کاربردی، سبب کاهش هزینه‌های نشر الکترونیکی شده است (۱). در هر صورت، وب یک عنصر منسجم و واضح مستقل نیست، بلکه دو عنصر مجزای وب وجود دارند: وب قابل رویت و دیدنی، دیدنی، و وب نامرئی و غیرقابل رویت (۲).

برای فهم کاربرد این دو عنصر و برای ارزیابی اطلاعات ضرورت دارد که از موضوع خارج شویم و ببینیم چگونه صفحات وب تولید می‌شوند.

دو نوع صفحه وب وجود دارند: استاتیک و دینامیک.

صفحات وب استاتیک توسط طراح وب، در خدمات‌دهنده وب نصب می‌شوند و برای هر کس و هر چیزی که وب سایت را مشاهده کند قابل دسترس‌اند و هر تغییری هم باید به صورت دستی انجام شود.

صفحات وب دینامیک توسط رایانه با به کارگیری دستورهای یک زبان دستنوشته (۱) و اغلب «سی‌جی‌آی‌جاوا» (۲) که به عنوان پلی بین سرویسگرهای پروتکل انتقال ابرمتن (۳) و کاربران وب عمل می‌کند، ایجاد می‌شود.

دستورهای یک زبان دستنوشته، نتایج را در یک صفحه خالی وب جا می‌دهند و به کاربر، صفحه وب دینامیک را ارائه می‌دهند. صفحات وب استاتیک اطلاعات عام و کلی را برای کاربر تهیه می‌کند، در حالی که صفحات وب دینامیک،

اطلاعات منحصر به فرد را برای کاربر تهیه می‌نمایند. صفحات وب استاتیک برای همه کاربران قابل رویت‌اند و در دسترس می‌باشند.

صفحات وب استاتیک	صفحات وب دینامیک
دستی تولید می‌شوند.	توسط رایانه تولید می‌شوند
اطلاعات عام و کلی است	اطلاعات منحصر به فرد است
اکثر آن‌ها قابل نمایه‌سازی هستند	قابل نمایه‌سازی نمی‌باشند

اولین مطالعه در موسسه تحقیقاتی «ان‌ئی‌سی» در پرینستون آمریکا در دسامبر سال ۱۹۹۷ نشان داد که صفحات وب قابل رویت، حداقل ۳۲۰ میلیون صفحه و وب را شامل می‌شوند، در حالی که دومین مطالعه تخمین زد که تعداد صفحات وب رشد کرده و به ۸۰۰ میلیون صفحه در فوریه سال ۱۹۹۹ رسیده‌اند. در هر صورت، اکثر ناشرین اطلاعاتشان را بر روی وب به وسیله پایگاه‌های اطلاعاتی بزرگ توزیع می‌کنند. این اطلاعات اساساً جنبه تجاری دارد (۲).

موتورهای جستجوگر (۴) و راهنماهای وب

۱.۲. راهنماهای وب

چه تفاوتی بین راهنمای وب و موتور جستجو وجود دارد؟

راهنمای وب:

الف. فهرستی از سایت‌های وب که از قبل تعیین شده‌اند، می‌باشد؛

ب. توسط ویراستاران جمع‌آوری شده؛

ج. براساس موضوع و عنوان طبقه بندی شده؛

د. انتخابی است.

با توجه به این که راهنماهای وب توسط افراد جمع‌آوری می‌شوند، تصمیم‌گیری درباره کیفیت محتوای هر یک از سایت‌های وب قبلاً انجام شده است. در نتیجه راهنماهای وب در اینترنت بسیار مورد مراجعه قرار می‌گیرند، زیرا کاربران به دنبال موضوعات خاصی در اینترنت هستند که این راهنماها، نقطه شروع خوبی در تشخیص بهترین‌ها در وب می‌باشند.

در کاربرد راهنماهای وب، کاربر می‌تواند در بین فهرست‌های راهنما حرکت نماید. راهنماهای اصلی وب همچنین اجازه عبور از دیگر موتورهای جستجو را هم می‌دهند. مثلاً بزرگ‌ترین راهنمای وب، یعنی «یاهو»، اجازه عبور از موتورهای جستجوی دیگر را می‌دهد.

سایت‌های وبی که به وسیله راهنمای وب نمایه‌سازی شده‌اند، به علت روند تکمیل به صورت دستی، در آن راهنماها باقی می‌مانند؛ مگر این که به احتمال بسیار ضعیف به صورت دستی حذف شوند. در حالی که وجود دائمی یک لیست یا یک فهرست در موتور جستجو همیشه ضمانت نمی‌شود، اما ساختن و حضور دائمی یک فهرست در یک راهنمای وب مثل «یاهو» امکانپذیر می‌باشد.

۲.۲. موتورهای جستجوگر

وقتی که یک موتور جستجوگر را به کار می‌برید، کاربر، پایگاه اطلاعاتی سایت‌های وب را جستجو می‌نماید. همه موتورهای جستجوگر دارای سه جزء ترکیبی هستند:

الف. «اسپایدرها» (۵) که سایت‌های وب را بررسی می‌کنند،

ب. فهرست‌ها و نمایه پایگاه اطلاعاتی؛

ج. نرم‌افزار بازیابی اطلاعات.

۱.۲.۲ اسپایدرهای موتورهای جستجوگر

پایگاه اطلاعاتی موتورهای جستجوگر ابتدا توسط «اسپایدرها» ساخته می‌شود و توسط موتورهای جستجوگر به مأخذ اصلی اعزام می‌شود. «اسپایدرها» برنامه‌هایی هستند که وب را به دنبال صفحات جدید وب جستجو می‌نمایند، کلمات را نمایه‌سازی می‌کنند و اتصال را به صفحات برقرار می‌نمایند، سپس کلمات نمایه‌شده را با آدرسی که به یک صفحه وب یا هر منبعی که بر روی وب جهانی و به طور کلی اینترنت قرار دارد اشاره می‌نماید، (۶) مقابله می‌کنند.

۲.۲.۲ نمایه پایگاه اطلاعاتی موتور جستجو

این بخش عنصر اصلی از هر موتور جستجو است. این همان چیزی است که کاربر به دنبال آن است. در طی سال ۱۹۹۸

تعدادی از تولیدکنندگان موتورهای جستجو آغاز به کار کردند. این شرکت‌ها نمایه‌ها را مطابق با معیارهای گوناگون ساختند. بطور مثال می‌توان از نمایه «دایرکت هیل» و «گوگل» (۷) نام برد و نیز «نمایه نام‌های حقیقی» (۸) که شرکت‌ها را قادر می‌سازد واژه‌های کلیدی خود را برای حمایت از محصولات خود ثبت نمایند.

۳.۲.۲. نرم‌افزار بازیابی اطلاعات

تمامی موتورهای جستجوگر، نرم‌افزار مختص به خود برای پاسخگویی به سوالات دارند. موتورهای جستجوگر مطابق با قوانین مشابه به شرح زیر عمل می‌نمایند: هر سایت وب شامل کلمات یا واژه‌هایی است که سوئال کاربر را با آن‌ها مقابله می‌کند. و در صفحه نمایش، به صورت فهرست‌وار نمایش می‌دهد و با توجه به سوال کاربر و ارتباط موضوعی، فهرستی از مطالب را طبقه‌بندی می‌نماید. تفاوت مختصری در چگونگی کار الگوریتم‌ها بین موتورهای جستجو وجود دارد، که دلیلی بر این مدعاست که کاربران، با یک پرسش واحد، از موتورهای مختلف نتایج متفاوتی را کسب می‌نمایند. در هر صورت، دلیل مهم دیگر اختلاف نتایج جستجو، نسبت پایین همپوشانی، در بین موتورهای جستجوگر می‌باشد.

۳. پورتال‌ها

«پورتال» (۹) ها سایت‌های وبی هستند که خود عرضه‌کننده خدمت خاصی نیستند، ولی مراجعین را به صورت ساده و دوستانه به اطلاعات و سایت‌های موردعلاقه‌شان هدایت می‌کنند. «پورتال» ها نقاط اصلی برای راهبردهای تجارت الکترونیکی هستند و به تدریج از سال ۱۹۹۴ به وسیله سه نوع شرکت اینترنتی به وجود آمدند. این شرکت‌ها که نقش اساسی در توزیع زنجیره اطلاعاتی برای کاربر ایفا می‌کنند به شرح زیر می‌باشند:

الف. تولیدکنندگان خدمات اینترنتی (۱۰)

ب. سایت‌های جستجو (۱۱)

ج. تولیدکنندگان مرورگر (۱۲)

سایت «اکسایت» (۱۳) نمونه‌ای که از یک سایت پورتال موفق است (۱).

۴. توسعه و تکامل فناوری جستجو

همه ما تکامل زندگی بشر بر روی کره خاکی را به خاطر داریم. همانطور که همگی ما می‌دانیم، شکل‌های مختلف زندگی سلولی برای مدت‌های بسیار طولانی قبل از تکامل زیستی وجود داشته است. در هر صورت هنگامی که به این نقطه از تکامل دست یافتیم، تغییرات به فرم سازمان‌یافته، آن در مدت زمان کم‌تری به انجام رسید. همین مطلب برای فناوری جستجوی وب نیز صادق است. با تاکید بر روی پورتالیزیشن و تجارت الکترونیکی، اولین نسل سایت‌های جستجو، عملاً و به تدریج فراموش شدند و این در حالی است که آن‌ها تا سال‌های متمادی، حکومت اصلی را به عهده داشتند؛ اما بدلیل این که نتوانستند خودشان را با تغییرات محیطی وفق بدهند، فراموش گردیدند و انواع جدیدی از تولیدکنندگان جستجو به بهره‌برداری رسیدند.

۱.۴. موتورهای جستجوی ماورا

موتورهای جستجوی ماورا (۱۴) به کاربر این امکان را می‌دهند که از میان موتورهای جستجوی متعدد و راهنماهای وب به طور مقارن و همزمان، جستجو نماید (۳و۲). محبوب‌ترین موتورهای جستجوی ماورا به شرح زیر می‌باشند:

- Dogpile (www.dogpile.com)
- Mamma (www.mamma.com)
- 2Q (www.2q.com)

- Infind (www.infind.com)
- Insleuth (www.isleuth.com)
- Surfby (www.surfby.com)
- Webtaxi (www.webtaxi.com)

۲.۴. نسل اول موتورهای جستجو که توسط سایت‌های وب بصورت «اسپایدر»، نمایه‌سازی را ایجاد کردند، انبوهی از واژه‌ها و کلمات را تجزیه و تحلیل نمودند. راهنماهای وب به طور دستی گردآوری شدند و از آوریل ۱۹۹۸ مورد بهره‌برداری قرار گرفتند: یکی از این راهنماها به نام «دایرکت هیل» (۱۵) یک حرکت ریشه‌ای را آغاز کرد و اصلاحاتی انجام داد و روش خود را «راه سوم» (۱۶) نامید.

این راهنمای وب ادعا می‌کرد که در زمان جستجو، کنترل با کاربر است و نتایج به صورت طبقه‌بندی شده بر روی سایت وب نمایان می‌شوند. این موتور جستجو، آنچنان که ادعا می‌کرد نتوانست خدمات منحصر به فردی را ارائه نماید. بنابراین موتور جستجوی دیگری به نام «هاتبات» (۱۷) به جای آن تولید شد (۲).

۳.۴. جستجوی زبان طبیعی

همانطور که قبلاً بحث شد، اولین نسل موتورهای جستجو با مقابله، واژه‌هایی که توسط کاربران ارائه می‌شدند عمل می‌نمودند. آن‌ها به محتوای واژه‌های جستجو توجه نمی‌کردند. برای مثال به ارتباط منطقی بین واژه‌های جستجو و دیگر معناهای لغوی توجه نمی‌شد، به علاوه، لغات را به صورت تحت‌اللفظی جستجو می‌کردند و به واژه‌ها و توصیفگرهای بولی نظیر **and, or, not** توجه نداشتند. دو موتور جستجوی جدید یکی پس از دیگری در طول یک هفته آغاز به کار کردند. این موتورهای جستجو، به شرح زیر زبان طبیعی را برای جستجو پیشنهاد کردند:

الف. www.askjeeves.com Ask Jeeves

این موتور جستجو در اول ژوئن سال ۱۹۹۸ شروع به کار کرد و به عنوان اولین موتور جستجوی زبان طبیعی بر روی اینترنت معرفی شد. این موتور جستجو سوال کاربر را با استفاده از پایگاه اطلاعات هفت میلیونی سوالات، پاسخ می‌داد. اگر هیچ گونه مقایسه‌ای وجود نداشت، سپس نزدیک‌ترین واژه از پایگاه برای کاربر ارائه می‌شد و سوال مناسب‌ترین آن را انتخاب کند. این موتور جستجو همچنین کاربر را به موتورهای جستجوی دیگری نیز راهنمایی می‌کند، مانند «یاهو»، «لیکاس»، «اینفوسیک» و «آلتاویستا».

ب. دومین موتور جستجو به نام **Electric Monk** چند هفته بعد آغاز به کار کرد. این موتور جستجو با استفاده از زبان طبیعی، الگوریتم سوالات کاربران را پاسخ می‌گوید. این الگوریتم با استفاده از اصطلاحنامه، به واژه‌های مرتبط نیز توجه می‌کند. جستجو از طریق زبان طبیعی در این موتور جستجوگر، سپس به صورت سوالات مرکب بولی درآمد و به «آلتاویستا» تسلیم شد.

۴.۴. تجزیه و تحلیل متکی بر اتصال

موتورهای جستجوی نسل اول برای ساخت نمایه‌های عظیم با هدف پاسخگویی به هر نوع سوال ساخته شدند. آن‌ها به محتوای هر صفحه، خاص تأکید داشتند که چگونه این صفحات به هم مربوط می‌شوند و اتصال می‌یابند. همانطور که قبلاً گفته شد، روش نمایه‌سازی آن‌ها به دلیل بی‌توجهی به ترکیب زبان، ساخت جملات و مترادف‌ها با شکست روبرو شد.

تجزیه و تحلیل متکی بر اتصال سعی دارد که با بررسی ارتباط بین صفحات، بر این مشکلات فائق آید. موتورهای جستجوگر متکی بر اتصال به سرعت رشد کرده‌اند و از نظر کاربران اینترنت در حال حاضر از محبوبیت خاصی برخوردارند.

۴.۴.۱. «گوگل» (۱۸)

این موتور جستجوگر نیز مانند «یاهو» توسط دانشجویان دانشگاه استنفورد ایجاد شد این فناوری با استفاده از روشی با

نام «رتبه‌بندی صفحه» (۱۹) وب، به دیگر موتورهای جستجو متصل می‌شود. اگر شما، به عنوان یک نویسنده، سایت وب، به دیگر سایت‌های وب از جهت اهمیت آن سایت‌ها متصل شوید، می‌توانید داوری‌ها و پیشنهادهایی درباره سایت ارائه دهید. همانطور که راهنماهای وب نظیر «ياهو» توسط ویراستاران به صورت دستی جمع‌آوری و تالیف گردیده‌اند. «گوگل» هم به دنبال اعلام قضاوت میلیون‌ها نویسنده، سایت وب به صورت خودکار طراحی شد.

البته این موتور جستجو در مقایسه با «ياهو» با سایت‌های وب بیش‌تری دسترسی دارد. در حقیقت برخلاف دیگر موتورهای جستجوگر، که هر چه نمایه‌های آن‌ها گسترده‌تر می‌شود، کم‌تر مفید واقع می‌شوند، «گوگل» ادعا دارد که با گستردگی نمایه‌ها و عظیم‌تر شدن آن‌ها، به نتایج بهتری دست یافته است. از دیگر موتورهای جستجو که با استفاده از الگوریتم طبقه‌بندی عمل می‌نمایند می‌توان از «هاتبات» و «اکسایت» نام برد. «گوگل» تنها موتور جستجویی است که به طور بسیار گسترده ای جستجوهایش متکی بر اتصال می‌باشند. تخمین زده شده که نمایه این موتور جستجوگر بین ۷۰ تا ۱۰۰ میلیون صفحه می‌باشد، اما از طریق اتصال، قدرت دسترسی به بیش از ۳۰۰ میلیون صفحه، وب نیز به کاربر اینترنت داده می‌شود. دسترسی بیش‌تر و صحت نتایج جستجوی بیش‌تر سبب گردیده که این موتور جستجو در ردیف ده موتور جستجوی محبوب در شبکه اینترنت قرار گیرد.

Clever (۲۰). ۴.۴.۲

با توجه به «نمایه‌نامه» استنادی علوم» (۲۱) (مطالعه، ارجاعی مقالات علمی با یکدیگر) clever جستجوی واژه‌های کلیدی را براساس تلفیق کلمات و ارتباط مطالب بررسی می‌نماید. این موتور جستجوگر نیز مانند «گوگل» با استفاده از اتصال، به سوالات پاسخ می‌دهد. ولی برخلاف «گوگل» ابتدا سوال را به دیگر موتورهای جستجو نظیر «آلتاویستا» تسلیم می‌کند و سپس کاربر را به صفحاتی که از نتایج جستجو کسب شده‌اند راهنمایی می‌نماید. این موتور جستجو توسط شرکت «آی‌بی‌ام» ایجاد شده.

Focused Crawler. ۴.۴.۳

یکی دیگر از موتورهای جستجویی که توسط شرکت، «آی‌بی‌ام» ایجاد شده. این موتور جستجو البته به اندازه clever گسترش پیدا نکرد. برخلاف دیگر موتورهای جستجو نظیر «گوگل» و clever، این موتور جستجو مجموعه‌ای بسیار مناسب و در ارتباط با موضوع را از طریق وب جستجو می‌نماید. به زبانی دیگر، این موتور جستجو فقط سایت‌های وبی را جستجو می‌کند که در ارتباط با پرسش کاربر می‌باشند.

جستجوی گروه‌های خبری

در حالی که در حال حاضر وب به مثابه مخزن اولیه دانش بشری بر روی اینترنت می‌باشد، تنها مخزن هم به حساب نمی‌آید. گروه‌های خبری، محلی است که گروهی از افراد، تجربیات، افکار و دانش خود را در مورد یک موضوع مورد علاقه، برای بازیابی اطلاعات به اشتراک می‌گذارند. تفاوت بین وب و گروه‌های خبری در این است که وب مجموعه‌ای از دانش بشری را به طور صریح و روشن ارائه می‌کند. در حالی که گروه‌های خبری مجموعه‌ای از دانش بشری را به طور ضمنی و غیرصریح ارائه می‌نمایند.

دانش صریح و روشن می‌تواند به افراد و مؤسسات کمک کند تا از اطلاعات و تجارب گذشته برای برنامه‌ریزی آینده کمک بگیرند، اما دانش غیرصریح و ضمنی مجموعه‌ای از تجارب و عقاید و خلاقیت‌هایی است که برای استخراج اطلاعات صریح و روشن مورد نیاز می‌باشد.

آدرس بعضی از گروه‌های خبری تخصصی:

DejaNews (www.dejanews.com)

Reference.com (www.reference.com)

Liszt's New groups directory (<http://liszt.com/news>)

۴،۴،۴. اطلاعات شرکت‌ها

سایت‌های زیادی درباره معرفی شرکت‌ها و تولیدکنندگان اطلاعات تجاری وجود دارند. هر محقق می‌تواند از این سایت‌ها بازدید کند. چنین سایت‌هایی دارای پایگاه‌های تجاری می‌باشند. یک بررسی بر روی نام‌های شرکت‌ها براساس عملکرد موتورهای جستجوگر پیشرو و راهنماهای وب برای انتقال نتایج مناسب برای پژوهشگران نشان داد که «سراج انجین واچ» (۲۲) و «هاتبات» و «گوگل» در رده‌بندی، در رده‌های ابتدایی قرار می‌گیرند.

در آگوست ۱۹۹۹، موتور جستجوگر «جامپ» (۲۳) آغاز به کار نمود. این موتور جستجوگر دارای یک نمایه تخصصی است که بر اطلاعات و خبرهای مربوط به شرکت‌ها تأکید دارد. این موتور جستجوگر فهرستی از شرکت‌ها، شامل اطلاعاتی نظیر عناوین شرکت‌ها، قدمت، زمینه موضوعی و آدرس پست الکترونیکی و ثبت اختراعات، و شرکت‌های وابسته به آن شرکت را نیز معرفی می‌نماید (۳ و ۲).

۴،۴،۵. فایل‌های تصویری و چندرسانه‌ای

تقریباً ۷۰ درصد از صفحات وب غیرمتنی هستند. آنچه مسلم است کاربرد، مفاهیم را به همراه تصاویر بهتر از اطلاعات فقط به صورت متنی درک می‌نماید. سه موتور جستجوی تخصصی اصلی که فایل‌های تصویری و چندرسانه‌ای را ارائه می‌دهند، در این زمینه به شرح زیر می‌باشند.

Ditto (www.ditto.com)

Scour (www.scour.net)

Altavista photo finder(www.altavista.com)

مقایسه ابزارهای جستجو در اینترنت

این روزها اینترنت منبع اصلی تهیه منابع اطلاعاتی شناخته شده است. ابزارهای مهمی برای کمک به یافتن اطلاعات در این فضای اطلاعاتی وجود دارند. راهنماها و موتورهای جستجوگر پایگاه‌های اطلاعاتی هستند که به روند دریافت اطلاعات سرعت می‌بخشند. این ابزارها اسناد را با استفاده از زبان علامتگذاری ابرمتن (۲۴) جمع‌آوری، ذخیره و تنظیم می‌نمایند. قبل از ظهور اینترنت اکثر پایگاه‌های اطلاعاتی کتابشناختی نقش اصلی را در بازیابی اطلاعات ایفا می‌کردند. با توسعه نرم‌افزارهای جستجو، اطلاعات علمی به راحتی قابل دستیابی شدند. تعداد ابزارهای جستجو در اینترنت به طور مداوم در حال رشد است. در این مقاله تعداد ده ابزار جستجو در اینترنت از قبیل راهنماها و موتورهای جستجوگر براساس تعداد صفحات اتصال، انتخاب شده‌اند.

جدول یک: رده‌بندی موتورهای جستجوگر و راهنماها براساس تعداد صفحات اتصال به آن‌ها

نام موتورهای جستجوگر و راهنماها	تعداد صفحات اتصال به آن‌ها
Yahoo	۱۱۱۰۵۹۹
Excite	۴۵۸۲۳۹
Lycos	۴۳۷۶۱۸
Webcrawler	۴۳۶۱۸۴
Infoseek	۳۵۶۹۶۳
Altavista	۳۵۰۰۸۵
Hotbot	۲۳۸۶۶۸
Nerdworld	۳۴۲۲۹

Alonetfind	۱۶۱۹۲
Northern light	۱۵۷۸۰

معیارهای ارزیابی شامل جمع‌آوری اطلاعات و تجزیه و تحلیل اسناد و بازیابی اطلاعات بوده است.

ابزارهای جستجو در اینترنت	واگذاری به صورت خودکار		واگذاری به صورت دستی					
	دارای Metatag	بدون Metatag	طبقه بندی موضوعی	عنوان	نشانی URL	شرح و توصیف	کشور	زبان
Yahoo			×	×	×	×		
Excite		×	×				×	×
Lycos		×						
Webcrawler		×	×				×	×
Infoseek	×						×	×
Altavista	×							
Hotbot	×							
Nerdworld			×	×	×	×		
Aolnet find	؟	؟						
Northern light		×						

جدول ۲. مقایسه ابزارهای جستجو و راهنماها در زمینه جمع‌آوری اطلاعات و تجزیه و تحلیل اسناد

نتایج ارزیابی ابزارهای جستجوی اینترنت با توجه به معیارهای اصلی بازیابی اطلاعات در جدول یک نشان داده شده است. شش معیار برای ارزیابی وجود دارد و حداکثر نمره ۶ می‌باشد.

تقریباً همه ابزارهای جستجو چهار معیار اصلی را دارند. غیر از Nerdworld که از شش عملگر منطقی بولی فقط از OR استفاده می‌نماید، و «ياهو» که عملگر پرانتز را قبول نمی‌کند، اختلافات دیگری نیز وجود دارند. «اینفوسیک» از عملکرد ستاره (x) استفاده نمی‌کند و «لیکاس» و «وب کراولر» سیستم را ملزم می‌کنند که ستاره را به عنوان اولین حروف کلمات در جستجو استفاده نماید.

جستجو براساس طبقه‌بندی موضوعی، فقط دو راهنما یا دایرکتوری «یاهو» و Nedworld همه، صفحاتشان طبقه‌بندی و دسته‌بندی شده است. و Northern light تنها موتور جستجوگر می‌باشد که اجازه جستجو براساس طبقه‌بندی را نمی‌دهد (۲).

در «نورثرن لایت» این امکان وجود دارد که به وسیله نواحی (۲۵)، جستجو انجام شود. در دیگر موتورهای جستجوگر این امکان کم‌تر وجود دارد. با توجه به جدول شماره ۳، «نورثرن لایت» از تعداد ۹ ناحیه شش ناحیه شامل: عنوان، نشانی، «یوآرال»، کشور، زبان، تاریخ، و نوع سازمان را جستجو می‌نماید.

رده‌بندی ابزارهای جستجوگر با توجه به تعداد نواحی قابل جستجو به ترتیب به شرح زیر می‌باشد:

Altavista	(تاریخ، زبان، نشانی «یوآرال»، عنوان)
Yahoo	(تاریخ، نشانی «یوآرال»، عنوان)
Lycos	(زبان، نشانی «یوآرال»، عنوان)
Infoseek	(کشور، نشانی «یوآرال»، عنوان)
Hotbot	(منطقه، جغرافیایی، تاریخ، زبان)
Excite	(نوع سازمان، زبان)

و بالاخره سه ابزار جستجوی باقی‌مانده یعنی

Web Crawler, Nerdworld and Aolnet find

امکان جستجو از طریق نواحی را ارائه نمی‌دهند.

در اینجا باید توجه کرد مهم‌ترین نواحی مورد توجه در ارتباط با جستجو، عنوان، «نشانی» یوآرال، و زبان می‌باشند. هیچ اختلافی در میان موتورهای جستجوگر و راهنماها در مورد استفاده از ناحیه وجود ندارد. از تعداد دو راهنما در این مقاله، «یاهو» پنج ناحیه قابل جستجو را پیشنهاد می‌کند و «نرد وُرد» هیچ ناحیه‌ای را پیشنهاد نمی‌نماید. ضمناً دیگر موتورهای جستجوگر نیز در مورد استفاده از نواحی ناهمگون هستند.

با توجه به قابلیت مرور (۲۶) و نظراندازی، موتورهای جستجوگر به شرح زیر امکان مرور و نظراندازی را ارائه می‌دهند. «اکسایت»، «لیکاس»، «آلتاویستا»، «هاتبات»، «نورثرن لایت». سه موتورهای جستجوگر نظیر «نورترون لایت» و «آلتاویستا» امکان استفاده از ناحیه کشور را ارائه می‌دهند.

سه ابزار جستجو که قادرند دستورها (۲۷) را به کاربرند شامل «آلتاویستا»، «هاتبات» و «اینفوسیک» می‌باشند. این سه موتور جستجوگر پس از جستجوی یک کلمه‌ای، فهرستی از واژه‌های ترکیبی را برای راهنمایی‌های بعدی و گسترش ارائه می‌دهند.

جدول شماره ۴ رده‌بندی ابزارهای جستجو در اینترنت و راهنماهای وب را براساس نمره‌های آن‌ها نشان می‌دهد. «نورثرن لایت» با کسب بالاترین نمره یعنی نمره ۱۳ در بالاترین رده، و «نردورد» با کسب نمره ۲ در پایین‌ترین رده قرار دارند. تعداد زیادی از نمره‌ها نیز بین ۱۱/۵-۱۰/۵ هستند و به شرح زیر می‌باشند:

«آلتاویستا»، «لیکاس»، «اکسایت»، «اینفوسیک» و «هاتبات».

ابزارهای جستجو	نمره
Northern light	۱۳
Altavista	۱۱/۵
Lycos	۱۱
Excite	۱۰/۵
Infoseek	۱۰/۵
Hotbot	۱۰/۵

Yahoo	۹
Webcrawler	۶
Aolnetfind	۵/۵
Nerdworld	۲

درجه‌بندی موتورهای جستجوگر

نتیجه‌گیری

مهم‌ترین نتیجه قابل توجه این است که هیچ ارتباطی بین محبوبیت و قابلیت جستجوی اطلاعات در میان ابزارهای جستجوی مورد مطالعه وجود ندارد. دو مثال واضح از این یافته‌ها «نورثن لایت» و «یاهو» می‌باشند. اگر چه «نورثن لایت» کم‌ترین شهرت و محبوبیت را در رده‌بندی از نظر اتصال دارا می‌باشد، این موتور جستجوگر اولین مقام را در رده‌بندی از نظر امکانات و قابلیت‌های بازیابی اطلاعات کسب نموده است. در حالی که «یاهو» یکی از محبوب‌ترین ابزارهای جستجو می‌باشد، ولی از نظر بازیابی اطلاعات، در رده پایین قرار دارد.

این بررسی تفاوت عمده‌ای را بین موتورهای جستجوگر یا راهنماهای وب نشان نداد. در جمع‌آوری اطلاعات و تجزیه و تحلیل اسناد، راهنماها قابل اعتمادترند، زیرا سیستم‌های دستی استفاده می‌کنند. موتورهای جستجوگر نیز به دلیل استفاده از دستورهای (۲۸) زبان علامتگذاری فوق‌متن یا ابرمتن‌ها (۲۹) اهمیت دارند. هرچند روشن نیست که نتایج قابل توجهی را ارائه دهند.

اگرچه راهنماها و موتورهای جستجوگر پایگاه‌هایی هستند که صفحات زبان علامتگذاری فوق‌متن را فهرست می‌کنند، از نقطه نظر بازیابی اطلاعات، با توجه به دلایل زیر نسبت به پایگاه‌های اطلاعاتی کتابشناسی معمولی دارای امکانات کم‌تری هستند.

- تعدادی از ابزارهای جستجو تمامی امکانات موجود در پایگاه‌های اطلاعاتی کتابشناسی معمولی را ندارند.

- محدود نمودن بازیابی به وسیله نواحی، چندان در ابزارهای جستجو در اینترنت رایج نیست.

- محققین سوءالاتی درباره نواحی عنوان نشانی که به یک صفحه وب، یا هر منبعی بر روی وب جهانی و به طور کلی اینترنت اشاره می‌کند دارند، که تنها در نیمی از ابزارهای جستجو در اینترنت یافت می‌شوند. نه راهنماها که اطلاعات را از طریق پرسشنامه‌های واگذار شده کسب می‌نمایند و نه موتورهای جستجوگر که از صفحات زبان علامتگذاری فوق‌متن استفاده می‌کنند، در جستجوهایشان از نواحی استفاده نمی‌کنند.

- مرور نمایه‌نامه‌ها یک شکل معمول و رایج نیست.

- استفاده از معانی لغات کنترل شده در ابزارهای جستجوی اینترنت کاربرد ندارد (۳ و ۲).

پی‌نوشت‌ها

- 1-Script
- 2-Java common Gateway Interface (Jave Cgi)
- 3-Search Engines
- 4-Script
- 5-Spiders
- 6-URL (Uniform Resource Locator)
- 7-Direct hill
- 8-Google Real Names Index
- 9-Portal

- 10-Internet Service/Access Providers
- 11-Search Sites
- 12-Browser Providers Ultra Search Engines
- 13-www.excite.com
- 14-Ultra Search Engines
- 15-www.directhil.com
- 16-The Third way
- 17-Hotbot
- 18-www.google.com
- 19-Page Rank
- 20-www.almaden.ibm/cs/k53/clever.htm/
- 21-Science Citation Index
- 22-Search Engine Watch
- 23-Jump
- 24-Hrml (Hypertext Mark up language)
- 25-Fields
- 26-Browsing
- 27-Meta tags
- 28-Meta Tags
- 29-hypertext Mark up Language

منابع

- ۱-قاسم‌زاده، فریدون. آشنایی با اینترنت و تجارت الکترونیک. تهران: ابزار ایران و توسعه ارتباطات افرا، ۱۳۷۹
- 2-Angeles Maldonado Martinez & Elena Fernandez Sanchez. "Comparing Internet Search Tools," Online Information 99 Proceedings. London:1-9 Dec. 1999. 263-266
- 3-Green, David "The evolution of web searching." Online information 99 proceedings, 23rd International Online information Meetings, London: 7-9 December 1999. pp-251-258
- 4-Alta Vista: www.AltaVista.Digital.com
- 5-Aolnetfind: www.Aol.com/netfind
- 6-Excite: www.Excite.com/
- 7-Hotbot: www.Hotbot.com/
- 8-Infoseek: www.guide.Infoseek.com/
- 9-Lycos: www.Lycos.com/
- 10-Nerdworld: www.nerdworld.com/
- 11-Northern light: www.northernligh.com/
- 12-Web Crawler: www.webcrawler.com/



پروپوزیشن گاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی