

توسعه روشی برای ارزیابی دقت موتورهای جستجو در اینترنت

سعید منصور، استادیار دانشکده مهندسی صنایع دانشگاه صنعتی امیرکبیر
علی عباس خسروی، کارشناس ارشد سازمان پژوهشهای علمی و صنعتی ایران

امروزه اطلاعات وسیعی به شکل صفحات وب توسط مراکز کامپیوتری جمع‌آوری و شاخص‌گذاری می‌شود. تعداد سرویس‌گرها هر سال و حجم اطلاعات روی اینترنت هر ۶ ماه دو برابر می‌شود. به همین منظور ابزارهایی برای جستجوی اطلاعات بر روی اینترنت به نام موتورهای جستجو ایجاد شده است. در این مقاله به توسعه روش‌ها و ابزارهایی برای سنجش و ارزیابی کیفیت موتورهای جستجوی وب پرداخته شده است. معیارهای ارزیابی موتورهای جستجو عبارتند از: پوشش؛ دقت؛ فراخوانی؛ سرعت پاسخ‌دهی؛ سهولت استفاده و نمایش نتایج. با توجه به اهمیت تعدادی از این عوامل؛ چند موتور جستجو با توجه به عوامل پوشش صفحات؛ دقت در پاسخ و تعداد مراجعات مقایسه می‌شوند. عوامل مختلف بر حسب کمیت عددی سنجیده شده و سپس با احتساب مجموع وزنی آنها با یکدیگر مقایسه می‌شوند.

با توجه به اهمیت معیار دقت روشی برای ارزیابی آن ارائه می‌شود. ارزیابی با انتخاب تعدادی پرسش نمونه و ارائه آنها به شش موتور جستجو و ارزیابی بیست نتیجه اول بازگردانده شده از موتورهای جستجو انجام می‌شود. به همین منظور رابطه‌ای برای سنجش دقت توسعه داده می‌شود و توسط آن امتیاز هر موتور جستجو برای هر پرسش و سپس میانگین امتیازات برای موتور جستجو محاسبه می‌شود. از آنجا که برای مفهوم ارتباط تعاریف مختلفی وجود دارد؛ آزمایش‌های مختلفی انجام می‌شود. برای بررسی معنی‌دار بودن اختلاف بین میانگین‌های بدست آمده، از آزمون ناپارامتری بلوک کامل تصادفی شده <کوئند> استفاده شده است.



۱- مقدمه

با عمومی شدن استفاده از صفحات وب^۱، نیاز به پیدا کردن این صفحات به یک مسأله جدی برای کاربران اینترنت تبدیل شده است. از این رو ابزارهایی به نام موتورهای جستجو^۲ و فهرست‌های راهنمای موضوعی^۳ ایجاد شده است که صفحات وب موجود در شبکه اینترنت را شناسایی کرده، روی آنها شاخص می‌سازند. در نتیجه کاربران شبکه اینترنت به کمک این ابزار می‌توانند منابع مورد نظر خود را جستجو کنند.

دهها و شاید صدها ابزار جستجو در اینترنت وجود دارد که غالباً کاربران را دچار سردرگمی ساخته و آنان را با این پرسش ساده روبرو می‌سازد: تفاوت این ابزار در چیست و کدامیک کارآمدتر است؟^۴ از این رو در این مقاله سعی شده است تا روشی برای ارزیابی و مقایسه ابزارهای جستجو ارائه گردد.

از مهم‌ترین شاخص‌هایی که برای سنجش کیفیت اطلاعاتی که ابزار جستجو در اختیار کاربران می‌گذارند، دو عامل <دقت>^۴ و <فراخوانی>^۵ است. هدف یک موتور جستجو ارائه پاسخ‌هایی مناسب (نتایجی که به صورت صفحات وب است) درمقابل پرسش^۶ کاربر است. <دقت> احتمال آن است که همه نتایج ارائه شده توسط موتور جستجو به درخواست کاربر مرتبط باشد. <فراخوانی> احتمال آن است که تمام نتایجی که به پرسش کاربر مربوط است بازیابی شود. اندازه‌گیری <فراخوانی> به دلیل ماهیت در حال تغییر وب قابل اندازه‌گیری نیست. بنابراین سعی شده است که روشی کارآمد برای محاسبه و تخمین <دقت> موتور جستجو و ارزیابی تعدادی از موتورهای جستجو با استفاده از نمونه‌گیری و آزمون‌های مختلف آماری ارائه شود.

-
- 1- Web Pages
 - 2- Search Engines
 - 3- Subject Directories
 - 4- Precision
 - 5- Recall
 - 6- Query



۲- مطالعه کتابخانه‌ای

موتورهای جستجو وب از سال ۱۹۹۴ با ظهور WWWorm پا به عرصه وجود نهادند [1]. کمی پس از پیدایش اولین موتور جستجو، نوشتارها و مقالات مربوط به آنها نیز منتشر شد. بررسی اجمالی این مقالات مشخص می‌سازد که تعداد مطالعات ارزیابی انجام شده روی موتورهای جستجو ناچیز است و اکثر آنچه منتشر شده است طبیعتی توصیفی دارند.

با بررسی روند مطالعات و ارزیابی‌ها مشخص می‌شود که به تدریج علاوه بر توصیف موتورهای جستجو گام‌هایی در جهت ارزیابی آنها برداشته شده است. در مرجع [2] قابلیت‌های موتورهای جستجوی CUI, CUSI, HarvestBroker, Worm, World-Wide Web و WebCrawler بر اساس مستندات ارائه شده توسط این موتورهای جستجو مورد ارزیابی قرار گرفته است.

در مرجع [3] پانزده موتور جستجو و فهرست راهنما از جمله WebCrawler, Yahoo, AltaVista, Lycos, Infoseek, Inktomi و AliWeb مقایسه و ارزیابی شده‌اند. از معیارهای مورد مقایسه می‌توان به سرعت رابط کاربر^۷، تعداد اسناد شاخص‌گذاری شده، قابلیت‌های جستجو و سهولت کاربر را نام برد. طبق نتایج به دست آمده در این مرجع موتورهای جستجوی Yahoo و AltaVista بالاترین امتیاز را کسب نموده‌اند.

بر اساس مطالعه دیگری که بر روی موتورهای جستجو انجام شده است، عملاً معیار ارزیابی «دقت» به کار گرفته شده است [4]. در این مرجع موتورهای جستجوی WebCrawler, Lycos, Infoseek و World Wide Web-Worm با استفاده از ۸ پرسش مرجع از کتابخانه دانشگاهی به عنوان پرسش‌های جستجو ارزیابی شده است. نتایجی که از این بررسی بدست آمده است آن است که «Lycos و Infoseek تقریباً یک میزان دقت را دارند» در حالیکه دقت WebCrawler بسیار پایین‌تر است. «World Wide Web-Worm معمولاً برای پرسش‌های معینی دست کم یک یا دو نتیجه را با دقت بالا بازیابی می‌کند».



۳- موتورهای جستجو

در این قسمت به بررسی موتورهای جستجو بر حسب اجزای تشکیل دهنده آنها و انواع گوناگون این موتورها پرداخته می‌شود.

۳-۱- اجزای موتورهای جستجو

به طور کلی یک موتور جستجو را می‌توان یک «سیستم جمع‌آوری و ارائه اطلاعات» نامید. هدف اصلی موتور جستجو جمع‌آوری صفحات وب، شاخص‌گذاری کلمات موجود در آن و ارائه آن به کاربران با امکان جستجوی کلمات کلیدی آن می‌باشد. اجزای یک موتور جستجو در ادامه آورده شده است [5].

الف جمع‌آوری اطلاعات

اطلاعات و یا صفحات HTML که بر روی سرویس‌گرهای شبکه جهانی وب قرار دارد به صورت خودکار و به کمک برنامه‌هایی که اصطلاحاً روبات^۸ نامیده می‌شود، جمع‌آوری می‌شود. در برخی از سرویس‌های جستجو و ارائه اطلاعات، جمع‌آوری اطلاعات نیز توسط انسان انجام می‌شود. صفحات جمع‌آوری شده در محلی موقت قرار داده می‌شوند تا برنامه شاخص‌گذار^۹ به آنها مراجعه نموده و اطلاعات آنها را شاخص‌گذاری نماید.

ب- شاخص‌گذاری اطلاعات و ایجاد بانک شاخص‌ها

کلمات موجود در صفحات جمع‌آوری شده توسط برنامه شاخص‌گذار تفکیک شده و در «بانک اطلاعات شاخص» قرار داده می‌شود. پس از شاخص‌گذاری کلمات، صفحه جمع‌آوری شده از محل نگهداری موقت حذف می‌شود.



ج- ارائه اطلاعات به کاربران

معمولاً به صورت فراهم آوردن امکان جستجوی کلمه کلیدی^{۱۱} از طریق رابط کاربر عرضه می‌شود. همچنین امکان مرور صفحات از طریق دسته‌بندی اطلاعات نیز عرضه می‌گردد. هر یک از بخش‌های جمع‌آوری، شاخص‌گذاری و ارائه اطلاعات می‌تواند به صورت خودکار و یا توسط انسان و یا ترکیبی از این دو صورت پذیرد. جدول شکل (۱) مثال‌هایی از ترکیب این حالت‌ها را مشخص می‌سازد:

۲-۳- انواع موتورهای جستجو

با توجه به اجزایی که شرح داده شد، معماری سیستم‌های جمع‌آوری و ارائه اطلاعات یا اصطلاحاً موتورهای جستجو با توجه به نحوه فعالیتشان به انواع زیر دسته‌بندی می‌شوند که توضیحات مبسوط پیرامون آنها در مرجع [5] آورده شده است.

الف- موتورهای جستجو بر پایه روبات

ب- فهرست‌های راهنمای موضوعی

ج- فراموتورهای جستجو

جمع‌آوری خودکار	جمع‌آوری انسانی	شاخص‌گذاری خودکار	شاخص‌گذاری انسانی
✓		✓	
	✓		✓
✓			✓

موتور جستجوی مبتنی

بر روبات: Altavista

فهرست: Yahoo

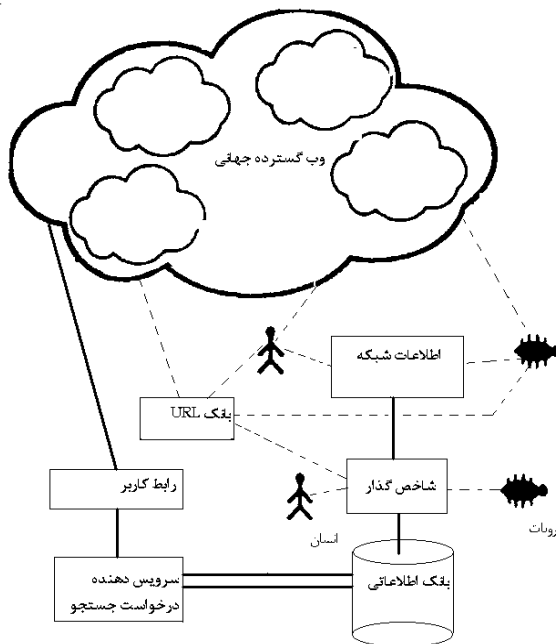
ترکیبی: Aliweb

شکل ۱: مثال‌هایی از سیستم‌های جمع‌آوری و ارائه اطلاعات

د- موتورهای جستجوی هوشمند

موتورهای جستجوی بر پایه روبات با استفاده از برنامه‌های خودکاری که اصطلاحاً روبات نامیده می‌شوند، به منابع اطلاعاتی موجود در شبکه اینترنت سرکشی می‌کنند و شاخص‌هایی از تعدادی یا تمامی واژه‌های موجود در آن منبع می‌سازند. آنگاه بر اساس تقاضای کاربران، با جستجوی کلمه‌ای در شاخص‌های خود منابع مورد نظر کاربران را پیدا کرده، آنها را معرفی می‌کنند. به عنوان نمونه، موتورهای جستجوی Excite، Infoseek و AltaVista از این دسته هستند. در این مقاله ارزیابی موتورهای جستجوی بر پایه روبات مورد توجه قرار گرفته شده است.

در شکل (۲) انواع موتورهای جستجو (سیستم جمع‌آوری و ارائه اطلاعات) و ارتباط میان اجزای آن آورده شده است.



شکل ۲: شمای کلی یک سیستم جمع‌آوری و ارائه اطلاعات



۴- معیارهای ارزیابی موتورهای جستجو

موتورهای جستجو ویژگی‌های متعددی دارند که معمولاً با توجه به این ویژگی‌ها مورد ارزیابی و مقایسه با یکدیگر قرار می‌گیرند. معیارهایی که بر اساس آنها بازدهی سیستم‌های ارزیابی اطلاعات^{۱۱} سنجیده می‌شود عبارتند از [6]:

الف پوشش^{۱۲}

ب- دقت و فراخوانی

ج- زمان پاسخ‌دهی

د- تلاش کاربر^{۱۳} (قابلیت‌ها و سهولت استفاده)

ه- نمایش نتایج^{۱۴}

موتورهای جستجو در واقع سیستم‌های ارزیابی اطلاعات در مقیاس گسترده و حجم اطلاعاتی بسیار بالا می‌باشند. به نظر می‌رسد که معیارهای فوق برای ارزیابی موتورهای جستجو کاملاً مناسب می‌باشند. بازدهی ارزیابی بر اساس سه عامل دقت، فراخوانی و زمان پاسخ‌دهی سنجیده می‌شود.

۴-۱- پوشش

عموماً ویژگی پوشش در موتور جستجو علاوه بر تعداد صفحاتی که موتور جستجو شاخص‌گذاری می‌کند، به جنبه‌های مختلفی از جمله پوشش جغرافیایی، پوشش دامنه و پوشش موضوعی اطلاق می‌گردد.

برخی از موتورهای جستجو فقط اطلاعات مربوط به یک منطقه خاص را جمع‌آوری و شاخص‌گذاری می‌کنند، یعنی فقط سرویس‌گرهایی که در آن منطقه قرار می‌گیرند را پوشش

11- Information Retrieval System

12- Coverage

13- User Effort

14- From of Out Put



می‌دهند. برای مثال موتور جستجوی 15 NWI یک موتور جستجو منطقه‌ای است که فقط سرویس‌گرهای مستقر در اروپای شمالی را شاخص‌گذاری می‌کند [7]. موتورهای جستجوی مهمی که در این مقاله ارزیابی می‌شوند اطلاعات موجود بر روی کلیه سرویس‌گرها در سراسر دنیا را جمع‌آوری و شاخص‌گذاری می‌نمایند.

برخی از موتورهای جستجو نیز شاخص‌گذاری صفحاتی را که URL آنها در یک دامنه مشخص قرار دارد را انجام می‌دهند. برای مثال موتور جستجوی AC/DC¹⁶ فقط سرویس‌گرهایی را که پسوند «.ac.uk» دارند، یعنی دانشگاه‌های انگلستان را، شاخص‌گذاری می‌نماید [8].

برخی دیگر از موتورهای جستجو موضوعات خاصی مانند موضوعات علمی و فنی را مورد توجه قرار داده و فقط اطلاعات مربوط به آن موضوع را شاخص‌گذاری می‌کنند. برای مثال موتور جستجوی AC/DC علاوه بر شاخص‌گذاری سرویس‌گرهایی که پسوند «.ac.uk» دارند، موضوعات علمی و فنی را شناسایی و شاخص‌گذاری می‌نماید [8].

۴-۲- دقت و فراخوانی

فراخوانی و دقت دو عامل بسیار مهم در ارزیابی موتورهای جستجو می‌باشند. طبق تعریف، فراخوانی عبارت است از احتمال اینکه همه اسناد مرتبط با پرسش بازیابی شوند، در حالی که دقت عبارت است از احتمال اینکه همه اسنادی که بازیابی شده‌اند مرتبط با پرسش باشند [9]. یا به طور دقیق‌تر:

$$\text{دقت} = \frac{\text{تعداد اسناد بازیابی شده مرتبط}}{\text{کل تعداد اسناد بازیابی شده}} \quad (1)$$

15- Nordic Web Index
16- Academic Directory



$$(۲) \quad \text{تعداد اسناد بازبازی شده مرتبط} \\ \text{فراخوانی} = \frac{\text{تعداد اسناد مرتبط}}{\text{کل تعداد اسناد مربوط}}$$

در شکل (۳) شمایی از رابطه دقت و فراخوانی با مجموعه کل اسناد و مجموعه اسناد بازبازی شده نشان داده است.

۳-۴- زمان پاسخ‌دهی

میانگین زمان انتظار، بین صدور یک دستور جستجو و نمایش اولین دسته نتایج جستجو بر روی صفحه نمایش، زمان پاسخ‌دهی نامیده می‌شود. زمان پاسخ‌دهی به عوامل زیر بستگی دارد:

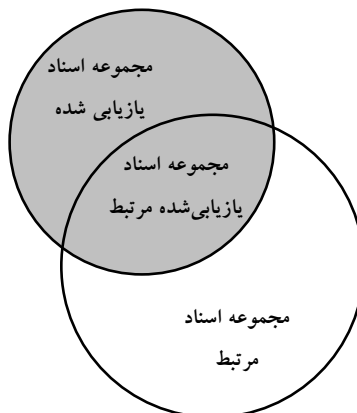
الف - پهنای باند موتور جستجو

ب- سخت افزارها و نرم افزارهای موتور جستجو

ج- پهنای باند مورد استفاده کاربر

امروزه با توجه به سخت‌افزارها و نرم‌افزارهای قدرتمند و پهنای باند بسیار بالایی که

موتورهای جستجو فراهم آورده‌اند، این زمان بین ۱ تا ۵ ثانیه کاهش یافته است [5].



شکل ۳: رابطه دقت و فراخوانی

۴-۴- تلاش کاربر

تلاش کاربر به سهولت استفاده، وجود مستندات و راهنمای کافی برای جستجو و قابلیت‌های جستجو که توسط موتور جستجو عرضه می‌شود اطلاق می‌گردد. موتورهای جستجو قابلیت‌های مختلفی را ارائه می‌دهند، که کاربران به کمک آنها می‌توانند عبارات جستجو را با توجه به پرسش مورد نظرشان ساخته و به موتور جستجو تحویل دهند. امکانات جستجوی متداول که اغلب موتورهای جستجو دارا می‌باشند، بسیار متنوع بوده که از حوصله این بحث خارج می‌باشد [5].

۴-۵- نمایش نتایج

موتورهای جستجو معمولاً نتایج را بر اساس میزان ارتباط با پرسش مرتب کرده و در دسته‌های ده‌تایی به کاربر ارائه می‌دهند. معمولاً به دو یا سه حالت امکان نمایش وجود دارد: حالت ساده و حالت پیشرفته. در حالت ساده عنوان و آدرس سند نشان داده می‌شود و در حالت پیشرفته



سایر اطلاعات از جمله اندازه، تاریخ، خلاصه و نشان داده می‌شود. برخی از موتورهای جستجو تعداد پاسخ‌هایی را که نمایش می‌دهند، محدود می‌سازند. اغلب موتورهای جستجو حداکثر ۵۰۰ نتیجه اول را نشان می‌دهند.

۵- ارزیابی معیار «دقت»

در این بخش شش موتور جستجوی وب گسترده جهانی به نام‌های Infoseek, Lycos, HotBot, Excite, AltaVista و WebCrawler از نظر «کیفیت دقت» مقایسه می‌شوند. مقایسه براساس یک سری آزمایش، با توجه به تعریف و استنباط‌های مختلف از مفهوم ارتباط، صورت پذیرفته است. در هر آزمایش به هر یک از این موتورهای جستجو، چند پرسش ارائه می‌شود. برای مقایسه نتایج بازگردانده شده از موتورهای جستجو، از معیاری به نام «دقت بیست نتیجه نخست» استفاده می‌شود. این معیار موتورهای جستجو را بر مبنای درصد میزان ارتباط اولین بیست نتیجه بازگردانده شده به درخواست جستجو، ارزیابی می‌کند [5]. در این معیار رتبه‌بندی نتایج توسط موتورهای جستجو نیز مورد توجه قرار گرفته است.

۵-۱- متدولوژی ارزیابی

در ادامه، روشها و راهکارهایی که به کمک آنها آزمایش انجام شده است، شامل نحوه انجام آزمایش، روش جستجو و روش ارزیابی ارائه می‌شود.

۵-۱-۱- روش انجام آزمایش

آزمایش در دو مرحله انجام می‌شود: انتخاب موضوع جستجو و سپس تعیین عبارت جستجو مطابق با موضوع مورد نظر. نظرات و سلیقه‌های شخصی ممکن است آگاهانه و یا ناآگاهانه در هر یک از مراحل اعمال شود. برای مثال فرد جستجوگر احتمال دارد موضوعاتی را انتخاب



کند که می‌داند یک موتور جستجوی خاص ممکن است اطلاعات بیشتری را نسبت به سایر موتورهای جستجو ارائه دهد و یا احتمالاً شکل عبارت را به گونه‌ای انتخاب کند که از قابلیت و ویژگی یک موتور جستجوی خاص استفاده کند و سایر موتورهای جستجو فاقد آن قابلیت باشند.

برای انتخاب موضوعات سعی شده است پرسش‌هایی که در حوزه علمی و فنی طرح می‌شوند، استفاده شوند. برای این منظور موضوعاتی که توسط مراجعین به یک مرکز اطلاع‌رسانی، برای جستجو در منبع اطلاعاتی اینترنت درخواست جستجو شده است، طی چند روز متوالی ثبت شده و سپس تعدادی از آنها توسط کارشناس مرکز مربوطه انتخاب و برای تحقیق ارائه شده است. این پرسش‌ها به طور تصادفی انتخاب شده‌اند و فردی که آنها را انتخاب نموده است از قابلیت‌های موتورهای جستجوی مختلف اطلاعی ندارد. همچنین برخی از پرسشها از مراجع مختلفی که همین شیوه را بکار برده‌اند، اخذ شده است. هرچه تعداد پرسش‌ها بیشتر باشد، دقت آزمون‌ها بیشتر خواهد شد ولی نیازمند انجام محاسبات بیشتری خواهد بود.

برای انجام آزمایش پرسش‌ها به دو دسته زیر تقسیم شده‌اند:

- الف - ساخت یافته شامل کلماتی که به کمک عملگرهای منطقی ترکیب شده‌اند.
- ب- ساده شامل یک کلمه یا تعدادی کلمات که با فاصله از یکدیگر ارائه می‌شوند. از هر یک از دسته‌های فوق چند پرسش و مجموعاً ۱۵ پرسش برگزیده شده‌اند [5].

۵-۱-۲- روش انجام جستجو

برای ارائه پرسش به موتور جستجو بهترین حالت آن است که پرسش در یک زمان به همه موتورهای جستجو داده شود. دلیل این امر آن است که اگر پس از جستجو از یک موتور جستجو، صفحه مفیدی بر روی وب قرار گیرد، هنگامی که جستجو از دومین موتور جستجو انجام شود، ممکن است آن صفحه جزء نتایج بازیابی شود. بنابراین جستجوها بلافاصله پس از هم انجام شده است.



دومین مسأله بررسی صفحاتی است که پیوندهای نتایج به آنها اشاره می‌کند. این بررسی باید بلافاصله پس از بازگرداندن نتایج صورت پذیرد، چرا که هر چه فاصله بررسی صفحات طولانی‌تر شود ممکن است صفحه‌ای که هنگام بازگرداندن نتایج واقعاً فعال است هنگام بازیابی برای بررسی صفحه غیرفعال شود و با آن به عنوان یک صفحه غیر فعال برخورد شود. به همین منظور بلافاصله پس از جستجو صفحات نتیجه بازیابی می‌شوند تا بعداً مورد بررسی و ارزیابی قرار گیرند. در این آزمایش صفحات به گونه‌ای ارزشیابی می‌شوند که فرد ارزیابی‌کننده نمی‌داند که این صفحه از کدام موتور جستجو به دست آمده است.

۵-۱-۳- روش انجام ارزیابی

پیش از ارزیابی پیوندها، مجموعه‌ای از ملاک‌های تصمیم‌گیری برای دسته‌بندی پیوندها تعیین می‌شود، که در قسمت معیارهای ارزیابی مورد بررسی قرار گرفته‌اند. به منظور اجتناب از اعمال نظر شخصی در دسته‌بندی صفحات بازگردانده شده توسط موتورهای جستجو، روشی به کار رفته است که برای یک صفحه معین فرد ارزیابی‌کننده نمی‌داند که کدام موتور جستجو آن را به عنوان نتیجه بازگردانده است.

برای تشخیص پیوندهای تکراری به برخی از مشخصه‌های صفحه مانند عنوان صفحه توجه می‌شود که اگر در هنگام ارزیابی به هم شباهت داشتند، می‌توانند به عنوان صفحات مشابه قلمداد شوند.

۵-۲- معیارهای ارزیابی

برای بررسی میزان ارتباط نتایج جستجو با پرسش کاربر، واحدهای سنجش ارتباط بایستی تعریف شوند. در این قسمت ملاک‌های کلی برای نتایجی که «مربوط» و «مفید» شمرده می‌شوند، تعریف می‌گردند؛ و سپس بر اساس آن، نتایج جستجو دسته‌بندی می‌شوند. با به دست آمدن داده‌ها و دسته‌بندی آنها، رابطه‌ای برای مشخص نمودن معیار دقت ارائه می‌شود.



از این رابطه که از آن با عنوان «دقت بیست نتیجه نخست» نام می‌بریم، برای اندازه‌گیری دقت برای هر پرسش و برای هر موتور جستجو استفاده می‌شود. اساساً نتیجه یک جستجو در یکی از دسته‌های زیر قرار می‌گیرد [5]:

- الف پیوندهای تکراری
- ب- پیوندهای غیرفعال یا پوچ
- ج- پیوندهای نامربوط (۰)
- د- پیوندهای از نظر فنی مربوط (۱)
- ه- پیوندهای بالقوه مفید (۲)
- و پیوندهای با احتمال زیاد مفید (۳)

۵-۲-۱- اندازه‌گیری دقت

پس از انجام قضاوت درباره ارتباط نتایج جستجو با درخواست جستجو، لازم است معیاری برای مقایسه بازدهی موتورهای جستجو تعیین گردد. در روشی که در این فصل توسعه داده می‌شود، میزان ارتباط در بیست نتیجه نخست بازگردانده شده از موتورهای جستجو مورد بررسی قرار می‌گیرد. سپس سه آزمایش مختلف بر روی بیست نتیجه نخست انجام می‌شود، به گونه‌ای که می‌توان مشاهده کرد چگونه موتورهای جستجو نتایجی را که «مربوط» یا «مفید» شمرده می‌شوند ارائه می‌دهند.

در عمل یک کاربر پرحوصله پنجاه یا صد پیوند نخست را بررسی می‌کند. کاربران بسیاری، از تعداد صفحات زیادی از نتایج صرف‌نظر می‌کنند تا از دسته‌هایی از پیوندها که گاه از یک سرویس‌گر بازگردانده می‌شود، اجتناب کنند. به نظر می‌رسد که بررسی یک گروه بیست‌تایی از نتایج بازگردانده‌شده امکان‌پذیر و قابل توجیه باشد.



۵-۲-۲- توسعه رابطه ای برای اندازه گیری دقت

طبق تعریف «دقت» برابر است با نسبت تعداد نتایج مربوط بازیابی شده از یک سیستم بازیابی اطلاعات به کل نتایج بازیابی شده یا به عبارت دیگر:

$$p = \frac{m}{n} \quad (۳)$$

که در آن m برابر است با تعداد نتایج مربوط بازیابی شده و n برابر است با کل نتایج بازیابی شده. در این رابطه عبارت «مربوط» مبهم است و به صورت صریح تعریف نشده است و در واقع بسته به نظر افراد متفاوت است. منظور از «نتیجه مربوط» آن است که پرسش را ارضاء نماید. در این رابطه یا پرسش را ارضاء می سازد و یا ارضاء نمی سازد و از نظر ریاضی، یا صفر است و یا یک. بنابراین لازم است تعاریف مختلفی برای مفهوم ارتباط ارائه داد و بر اساس آنها آزمایش‌های مختلفی برای سنجش دقت انجام داد. بنابراین «مربوط بودن» می تواند مقادیر صفر یا یک را کسب کند. میزان ارتباط را با β نشان می دهیم. در صورتی که نتیجه نامربوط باشد β برابر صفر و در غیر این صورت برابر ۱ است.

به این ترتیب با اعمال معیار «میزان ارتباط» در رابطه (۳)، رابطه دقت به صورت زیر نوشته می شود:

$$p = \frac{\sum_{i=1}^n \beta_i}{n} \quad (۴)$$

نکته دیگر، عامل رتبه بندی نتایج است که در این رابطه مورد توجه قرار نگرفته است. موتورهای جستجو معمولاً نتایج بسیار زیادی را بازیابی می کنند، به گونه ای که بررسی همه نتایج امکان پذیر نیست. در صورتی که ده یا بیست نتیجه نخست ربطی به پرسش نداشته باشد، معمولاً کاربر یا پرسش خود را تغییر می دهد و یا به موتور جستجوی دیگری مراجعه می کند.



بنابراین اگر موتور جستجوی «الف» بیست نتیجه بازگرداند و فقط دو نتیجه اول آن به پرسش مربوط باشد باید مقدار دقت آن به مراتب بیش از موتور جستجوی «ب» که فقط از بیست نتیجه بازگردانده شده آن دو نتیجه آخر آن مربوط به پرسش است، باشد. در این صورت باید نتایج نامربوطی که در ابتدای فهرست نتایج آورده می‌شوند، جریمه شوند. برای این منظور ضریبی به عنوان «ضریب ارتباط» تعریف می‌شود. در این صورت رابطه دقت به صورت زیر تعریف می‌شود:

$$p = \frac{\sum_{i=1}^n \beta_i r_i}{\sum_{i=1}^n r_i} \quad (5)$$

ساده‌ترین مقداری که می‌توان برای « β » یا ضریب ارتباط در نظر گرفت، آن است که اگر بیست نتیجه بازگردانده شده است، به اولین نتیجه ضریب ۲۰، به دومین نتیجه ضریب ۱۹ و همینطور به بیستمین نتیجه ضریب ۱ تعلق گیرد. در این حالت رابطه دقت به صورت زیر تعریف می‌شود:

$$p = \frac{\sum_{i=1}^n \beta_i (n-i+1)}{\sum_{i=1}^n i} = \frac{2 \left[\sum_{i=1}^n \beta_i (n-i+1) \right]}{n(n+1)} \quad (6)$$

در این رابطه نسبت ضریب ارتباط آخرین نتیجه n برابر اولین نتیجه است که باعث ایجاد پراکندگی زیاد در داده‌ها می‌شود. برای این منظور از یک پارامتر برای کاهش این نسبت استفاده می‌کنیم. مقدار «ضریب ارتباط» را مقداری بین ۰ تا ۱ در نظر می‌گیریم. مقدار ضریب ارتباط برابر است با:



$$r_i = \frac{n+k-i+1}{n+k} \quad (7)$$

که در آن هر چه مقدار k افزایش یابد، نسبت ضریب ارتباط اولین نتیجه به آخرین نتیجه کاهش می‌یابد. در مثال قبل اگر بخواهیم این نسبت به ۵ کاهش یابد، مقدار k برابر است با ۵، در این صورت رابطه دقت به صورت زیر بیان می‌شود:

$$p = \frac{\sum_{i=1}^n \beta_i (n+k-i+1)}{\sum_{i=1}^n (n+k-i+1)} = \frac{2 \left[\sum_{i=1}^n \beta_i (n+k-i+1) \right]}{n(n+2k+1)} \quad (8)$$

همچنین برای $n=20$ و $k=5$ داریم:

$$p = \frac{2 \sum_{i=1}^{20} \beta_i (26-i)}{310} \quad (9)$$

۵-۳- طرح آزمایش‌ها

از آنجا که روشهایی متعددی برای تعریف یک پیوند به عنوان پیوند مفید یا پیوند غیر مفید وجود دارد، چند آزمایش متفاوت بر روی داده‌ها انجام شده است تا موتورهای جستجو با استفاده از تعاریف مختلف مقایسه شوند.

در آزمایش نخست، به تمامی پیوندهای دسته‌های یک، دو و سه؛ عدد ۱ نسبت داده می‌شود. این آزمایش نشان می‌دهد که هر سرویس چه اندازه پیوندی که پرسش جستجو را در کمترین صورت ممکن ارضاء می‌کند ارائه می‌دهد. به سایر پیوندها نیز مقدار صفر داده می‌شود. در آزمایش دوم مقدار ۱ به پیوندهای دسته دو و سه داده می‌شود. این آزمایش نشان



می‌دهد یک موتور جستجو چه اندازه پیوندهایی را ارائه می‌دهد که برای موضوع مورد پرسش به‌طور بالقوه مفید هستند. آزمایش سوم تنها مقدار یک را به پیوندهای دسته سه نسبت می‌دهد. این آزمایش نشان می‌دهد که هر موتور جستجو چه اندازه پیوندهای مفید را در اولین بیست نتیجه بازمی‌گرداند.

۴-۵- روش تحلیل نتایج

با اندازه‌گیری دقت به کمک رابطه به‌دست آمده و تعیین میانگین دقت نتایج بازگردانده شده ۱۵ پرسش برای هر موتور جستجو، می‌توان دقت موتورهای جستجو را با یکدیگر مقایسه نمود. نکته‌ای که باید به آن توجه داشت آن است که آیا این نتایج معلول تصادف‌اند یا اختلاف‌های واقعی هستند. بنابراین برای بررسی معنی‌دار بودن اختلاف بین میانگین‌های دقت موتورهای جستجو نیاز به استفاده از یک آزمون کارآمد می‌باشد. در این تحقیق از آزمون ناپارامتری کوئد برای بررسی معنی‌دار بودن اختلاف دقت موتورهای جستجو استفاده شده است [5].

روش ناپارامتری «کوئد» بستگی به رتبه‌های مشاهده‌ها در درون هر بلوک (موتورهای جستجو) و رتبه‌های دامنه‌های بلوک‌های نمونه دارد. بنابراین می‌توان آزمون «کوئد» را یک تجزیه واریانس دوطرفه روی رتبه‌ها نامید [10].

۵-۵- تحلیل نتایج

آزمایش‌های مختلف نشان می‌دهد که اختلاف امتیازات بستگی به چگونگی تعریف «ارتباط» دارد. آزمایش اول در صورتی یک پیوند را مفید می‌نامد که حداقل از نظر فنی عبارت جستجو را ارضاء کند. در این آزمایش میانگین کل برابر ۰/۷۱ و بهترین امتیاز دقت متعلق به AltaVista برابر ۰/۹۲ است. اگر تعریف «ارتباط» تنها به صفحاتی که به صورت بالقوه مفیدند



موتور جستجو	میانگین ۱	مجموع ۱	میانگین ۲	مجموع ۲	میانگین ۳	مجموع ۳
AltaVista	۰/۹۲۰۰	۲۱۰	۰/۴۶۷۷	۶۳	۰/۰۹۳۱	۹۱
Excite	۰/۸۶۱۱	۱۸۲	۰/۵۳۹۴	۱۸۹	۰/۱۲۰۰	۱۴۱
Lycos	۰/۴۵۴۲	-۲۱۱	۰/۲۵۲۵	-۱۳۵	۰/۰۴۳۰	-۱۰۱
HotBot	۰/۶۵۴۸	-۶۰	۰/۲۸۲۲	-۹۵	۰/۰۶۵۶	-۶۸
Infoseek	۰/۸۲۰۴	۷۰	۰/۵۱۱۲	۱۰۸	۰/۱۴۶۵	۱۵۷
WebCrawler	۰/۵۶۷۷	-۲۰۱	۰/۲۳۶۳	-۱۵۶	۰/۰۴۸۲	-۱۲۹
میانگین کل	۰/۷۱۳۰	---	۰/۳۸۱۵	---	۰/۰۸۶۱	---
کمترین اختلاف معنی دار	---	۸۲	---	۱۴۲	---	۱۶۳

شکل ۴: میانگین عددی و مجموع رتبه‌های نسبی کوئد برای هر موتور جستجو

اطلاق شود، میانگین کل به مقدار ۰/۳۸ تنزل می‌یابد و بالاترین امتیاز که متعلق به Excite است برابر ۰/۵۳ می‌باشد. اگر معیار صفحه‌ای باشد که احتمال مربوط بودن آن زیاد است، میانگین به مقدار ناچیز ۰/۰۹ کاهش می‌یابد و بالاترین امتیاز مقدار ۰/۱۴ است که متعلق به Infoseek می‌باشد. جدول شکل (۴) میانگین‌های تخمینی و مجموع اندازه نسبی رتبه‌های کوئد را برای هر موتور جستجو، میانگین‌های کل و کمترین اختلاف معنی دار در مجموع رتبه‌ها را برای هر آزمایش که در بخش قبل محاسبه شده‌اند، نشان می‌دهد.

در جدول شکل (۴) میانگین‌ها و میانگین کلی برای هر آزمایش آورده شده است. همچنین مجموع رتبه‌های نسبی کوئد و اندازه‌های کمترین اختلاف معنی دار بین مجموع رتبه‌ها آورده شده است. اگر اختلاف بین مجموع رتبه‌ها برای دو موتور جستجو بیشتر از این عدد

آزمایش	کمترین	بیشترین
۱	<u>HOT</u> <u>Web</u> <u>Lyc</u>	<u>Alt</u> <u>Exc</u> <u>Inf</u>
۲	<u>HOT</u> <u>Lyc</u> <u>Web</u>	<u>Exc</u> <u>Inf</u> <u>Alt</u>
۳	<u>HOT</u> <u>Lyc</u> <u>Web</u>	<u>Inf</u> <u>Exc</u> <u>Alt</u>

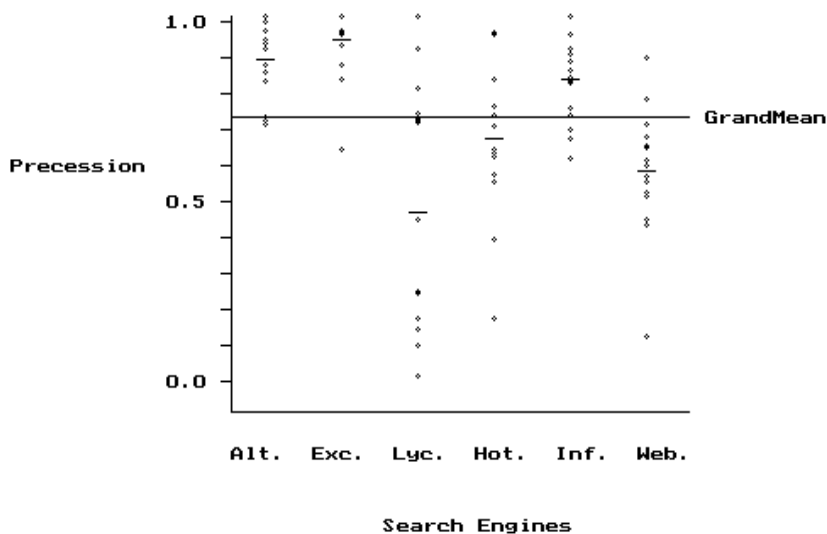
شکل ۵: رتبه‌بندی و اختلاف‌های معنی‌دار بین موتورهای جستجو



باشد، اختلاف معنی دار است. در جدول شکل (۵) هر سرویس بر اساس آزمایش رتبه‌بندی شده است. زیرخط‌ها نشان دهنده آن است که اختلاف معنی‌داری بین موتورهای جستجو وجود ندارد.

با توجه به نتایج به دست آمده در شکل (۶) مشاهده می‌شود که AltaVista، Excite و Infoseek دارای بالاترین امتیازهای میانگین دقت هستند. مقادیر آنها بیشتر از HotBot نیست ولی به طور قابل ملاحظه‌ای بیش از Lycos و WebCrawler می‌باشند. یکی از دلایلی که می‌توان برای این موضوع برشمرد، نرخ به‌روز رسانی بالای صفحات وب توسط موتورهای جستجوی AltaVista، Excite و Infoseek است که تعداد صفحات پوچ کمتری دارند.

جدول شکل (۵) رتبه‌بندی موتورهای جستجو را بر اساس هر آزمایش نشان می‌دهد. اگر زیرخط دو سرویس را به هم متصل کند، میانگین‌های تخمینی‌شان اختلاف معنی‌داری ندارند. برای مثال در آزمایش اول میانگین‌های Excite و AltaVista با میانگین HotBot، Infoseek و Lycos و WebCrawler اختلاف معنی‌داری دارد. بنابراین بر حسب اینکه پیوندهای فعال



شکل ۶: نمودار میانگین و محدوده تغییرات برای موتورهای جستجو در آزمایش اول



غیر تکراری ارائه می دهند که حداقل از نظر فنی پرسش جستجو را ارضاء کند، بهتر هستند. Infoseek در وسط قرار می گیرد که از نظر آماری بدتر از ۲ تایی بالاتر است ولی بهتر از HotBot و Lycos و WebCrawler است.

در این آزمایش AltaVista به هیچ وجه متغیر نیست، که نشان دهنده آن است که در ارائه صفحاتی که حداقل از نظر فنی عبارت پرسش را ارضاء می کند مفید است، علت این موضوع نرخ به روزرسانی بسیار بالای موتور جستجوی AltaVista است. تنها Lycos تغییر پذیری زیادی دارد.

با نگاهی به نتایج، دیده می شود که به طور کلی Excite، AltaVista و Infoseek از نظر کیفیت ارتباط برتری دارند که یکی از دلایل عمده آن نرخ به روزرسانی بالای آنها و در نتیجه تازگی صفحات است که از تعداد پیوندهای غیر فعال و نامربوط می کاهد. AltaVista بهترین نتایج را در آزمایش اول به دست می آورد. در آزمایش سوم بهترین صفحات مربوط، Infoseek بالاتر از سایر موتورهای جستجو قرار گرفت، که نشان دهنده الگوریتم رتبه بندی بهتر آن نسبت به سایر موتورهای جستجو است. تعداد پیوندهای غیر فعال Lycos زیاد است که ناشی از نرخ به روزرسانی پایین آن است. ملاکهای ارزیابی که منجر به انجام سه آزمایش مختلف گردید، تأثیر زیادی در نتایج کلی نداده است. HotBot در همه آزمایش ها در بین موتورهای جستجو قرار گرفت. Lycos و WebCrawler در هر سه آزمایش پایین ترین رتبه را به خود اختصاص داده اند.

۶ - نتیجه گیری و پیشنهاد برای کارهای آتی

به طور خلاصه نتایجی که از ارزیابی معیار «دقت» موتورهای جستجوی Infoseek، HotBot، Excite، AltaVista و WebCrawler به دست آمده است عبارتند از:



- ۱- به طور کلی Excite, AltaVista و Infoseek از نظر «کیفیت ارتباط» برتری دارند که یکی از دلایل عمده آن نرخ به‌روزرسانی بالای آنها و در نتیجه تازگی صفحات است که از تعداد پیوندهای غیرفعال و نامربوط می‌کاهد.
 - ۲- AltaVista بهترین نتایج را در آزمایش اول به دست آورد. یعنی آنکه کمترین تعداد صفحات تکراری و یا غیرفعال را داشته است. همچنین در صفحات بیشتری کلمات پرسش وجود داشته است.
 - ۳- در آزمایش سوم یعنی بهترین صفحات مربوط، Infoseek بالاتر از سایر موتورهای جستجو قرار گرفت، که نشان‌دهنده الگوریتم رتبه‌بندی بهتر آن نسبت به سایر موتورهای جستجو است.
 - ۴- تعداد پیوندهای غیرفعال Lycos و WebCrawler زیاد است که ناشی از نرخ به‌روزرسانی پایین آنهاست.
 - ۵- ملاک‌های ارزیابی که منجر به انجام آزمایش‌های مختلف گردید تأثیر زیادی در نتایج کلی نداده است. HotBot در همه آزمایش‌ها امتیازی متوسط نسبت به سایر موتورهای جستجو کسب کرده است. Lycos و WebCrawler در هر سه آزمایش پایین‌ترین رتبه را به خود اختصاص داده‌اند. همچنین Excite, AltaVista و Infoseek همواره بهترین امتیازات را کسب کرده‌اند.
- در این مقاله سعی بر آن بوده است تا روشی قاعده‌مند و مستدل برای ارزیابی موتورهای جستجو، خصوصاً از نقطه نظر معیار <دقت> ارائه شود. البته با در نظر گرفتن عوامل دیگری می‌توان متدولوژی ارائه شده را بهبود بخشید. از سایر کارهایی که برای بهبود نتایج حاصل می‌توان انجام داد آن است که جهت انجام بررسی‌ها تعداد موتورهای جستجو بیشتری را به مجموعه اضافه نمود. از جمله Google, Fast, OpenText و NorthernLight. طبق آمارهایی که اخیراً منتشر شده است موتور جستجوی Google بیش از یک میلیارد صفحه وب را شاخص‌گذاری کرده است و به جمع بزرگترین موتورهای جستجو پیوسته است. کار دیگری



که می‌توان انجام داد بررسی آن است که آیا ارتباطی میان نوع پرسش و دقت یک موتور جستجوی خاص وجود دارد یا خیر.

همچنین ساخت موتور جستجو در داخل کشور با توجه به رشد روز افزون تعداد سرویس‌گرهای وب داخل کشور پیشنهاد می‌شود. بهتر است این کار در دو مرحله انجام شود ابتدا ساخت یک فهرست موضوعی از سرویس‌گرهای داخل کشور به دلیل سهولت و تعداد کم سرویس‌گرها و سپس آن را می‌توان به صورت یک موتور جستجوی بر پایه ربات توسعه داد.

منابع

- 1- Brin, S., Page, L., *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, <http://www.roedu.net/~cmatei/search/google/google.html>, Computer Science Department, Stanford University.
- 2- Notess, Greg R., *Searching The World Wide Web : Lycos, WebCrawler and More*. Online, 19 (4), 48-53., July /August 1995.
- 3- Slot, M., *Web Matrix: Overview Matrix*.
<http://www.ambrosiasw.com/~fprefect/matrix/overview.html>, 6/8/1996.
- 4- Leighton, H. V., *Performance of four World Wide Web(WWW) index Services: InfoSeek, Lycos, WebCrawler, and WWWWorm.*,
<http://www.winona.msus.edu/services-f/library-f/webind.htm>, 1995
- 5- خسروی، ع.، *موتورهای جستجوی وب: بررسی تحلیل و متدولوژی ارزیابی*. پایان‌نامه کارشناسی ارشد، دانشکده مهندسی صنایع، دانشگاه صنعتی امیرکبیر، شهریور ۷۹.
- 6- Chu, H., Rosenthal, M., *Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology*,
<http://www.asis.org/ElectronicProceedings/chu.html>, ASIS 1996 Annual Conf. Proceedings.
- 7- Ard., A., Lundberg, S., *A Regional Distributed WWW Search and Indexing Service- the DISIRE Way*, <http://nwi.dtv.dk/www7/default.htm>.
- 8- Beckett, D., Smith, N. G., *The Academic Directory*,
<http://www.ariadne.ac.uk/issue3/acdc/default.htm>, University of Kent, 1996.



- 9- Blair, D. C., Maron, M. E., *An Evaluation of Retrival Effectiveness for a Full-Text Document Retrieval System*, Commun. ACM, 28 (3), pp. 289-299, Mar. 1985.

۱۰- والپول، ر.، فروند، ج.، *آمار ریاضی*، ترجمه علی عمیدی، مرکز نشر دانشگاهی، ۱۳۷۶، صفحات ۴۸۷-۵۱۰.