

بررسی رابطه ویژگیهای موتور کاوش اطلاعات و انتظارات کاربران*

غلامعلی منتظر** / مهدی ذبیحی***

بستر اصلی انتقال اطلاعات میان شبکه‌های مختلف رایانه‌ای به‌کارگرفته شده است. تا پیش از حضور اینترنت، سه رسانه اصلی: «روزنامه/ کتاب»، «رادیو/ تلفن» و «تلویزیون» وظیفه جابه‌جایی اطلاعات را به سه شکل نوشتاری، شنیداری و دیداری - شنیداری بر عهده داشتند؛ لیکن اینترنت، به عنوان رسانه جدید، علاوه بر داشتن ویژگیهای توأم رسانه‌های یادشده، دو مشخصه ممتاز نیز دارد: نخست اینکه، امکان ارتباط دو طرفه^۱ و متعامل^۲ میان کاربر و رسانه را به وجود می‌آورد؛ دیگر آنکه، جابه‌جایی شکل‌های دیگر داده‌ها (خارج از صورتهای نوشتاری، دیداری و شنیداری) را نیز شدنی می‌کند. در تشریح ویژگی نخست، اشاره به این نکته ضروری

چکیده: شناسایی چگونگی مواجهه کاربران با اینترنت و انتظارات آنان از موتورهای کاوش از مهمترین نکاتی است که هم موتورهای کاوش اصلی در بازیابی اطلاعات را تعیین می‌کند و هم دیدگاه جدیدی را برای طراحی موتورهای کاوش ارائه می‌دهد. این موضوع، به ویژه، با توجه به هدف طراحی موتور کاوش ملی در ایران اهمیت بسیار دارد. در این مقاله، ابتدا، محتوای اطلاعاتی اینترنت بررسی شده و سپس نتایج تحقیقات بر حدود یک هزار نفر از کاربران حرفه‌ای اینترنت در زمینه‌های مختلف کاربرد موتور کاوش، شامل میزان رضایت از موتورهای کاوش برای بازیابی اطلاعات، چگونگی کار با موتورهای کاوش و میزان محبوبیت آنها، بیان شده است. سپس نتایج هریک از این مطالعات مورد تبیین و تدقیق قرار گرفته و از این رهگذر، راهکارهای کلی طراحی موتور کاوش ملی ایران بیان شده است.

کلیدواژه: کاربران حرفه‌ای، اینترنت، موتور کاوش، روبات اطلاعاتی، رفتارشناسی، میزان محبوبیت، بازیابی اطلاعات.

مقدمه

اینترنت، شاخصترین نماد عصر اطلاعات، با سرعتی حیرت‌افزا، در طی دهه اخیر، رشد کرده و به عنوان

* این تحقیق به مدد امکانات پژوهشکده فناوری اطلاعات دانشگاه تربیت مدرس اجرا شده است. از حمایت‌های این پژوهشکده صمیمانه سپاسگزاری می‌کنیم.

** عضو هیئت علمی دانشکده فنی مهندسی دانشگاه تربیت مدرس.

*** دانش‌آموخته کارشناسی ارشد دانشکده فنی مهندسی دانشگاه تربیت مدرس.

1. bidirectional
2. interactive

شبکه متصل هستند و از اشکال مختلف اطلاعاتی، مانند متن، صوت، تصویر، عکس و فیلم استفاده می‌کنند (Micheal, 2002). آخرین آمار استخراج شده (Nov. 2003) توزیع محتوایی اطلاعات را روی شبکه اینترنت به شرح ذیل نشان می‌دهد:

اطلاعات تجاری	۸۲/۱ درصد
اطلاعات علمی	۶ درصد
اطلاعات پزشکی - بهداشتی	۲/۶ درصد
صفحات شخصی	۲/۳ درصد
انجمنهای علمی	۱/۹ درصد
مباحث اجتماعی	۱/۴ درصد
اطلاعات دولتی	۱/۲ درصد
اطلاعات مذهبی	۰/۸ درصد
سایر موارد	۱/۵ درصد

باتوجه به نکات یادشده، و با عنایت به سرعت گسترش اینترنت و ضرورت دسترسی سریع به اطلاعات صحیح، راهکارهای جدیدی در بازیابی اطلاعات پدید آمده است. یکی از این راه حلها استفاده از موتورهای کاوش^۳ است؛ موتورهای کاوش برنامه‌هایی نرم‌افزاری هستند که وظیفه جستجو، بازیابی و دسته‌بندی اطلاعات مرتبط را از پایگاههای مختلف اطلاعاتی موجود در شبکه اینترنت برعهده دارند (Cooky & Mobaser, 1997: 451-455). تاکنون بیش از سه هزار موتور کاوش شناخته شده (Calari & wells, 2000: 10/22-27)، که برخی به صورت خصوصی و صرفاً برای بازیابی اطلاعات از پایگاهی خاص، مانند ایستگاه اطلاعاتی یک شرکت یا مؤسسه، طراحی شده و برخی دیگر نیز موتورهای کاوش قوی هستند که می‌توانند اطلاعات را از روی شبکه جهانی وب^۴ جمع‌آوری کنند و به کمک این موتورها می‌توان هرگونه اطلاعاتی را در شبکه اینترنت بازیابی کرد. به‌رغم فزونی تعداد موتورهای کاوش، تقریباً همه آنها ساختار

است که رسانه‌های موجود تنها امکان بهره‌برداری یکسویه از اطلاعات موجود را فراهم می‌آورند؛ مثلاً، بیننده تلویزیونی صرفاً کاربری منفعل است که در معرض دریافت امواج اطلاعاتی قرار گرفته و نمی‌تواند خود نیز، همزمان، بر رسانه کاربردی مؤثر واقع شود؛ اما محیط اینترنت افزار مناسبی را برای تعامل میان کاربر و رسانه فراهم می‌آورد که از طریق آن هم قدرت گزینش کاربر افزایش می‌یابد و هم خود می‌تواند تولیدکننده (و هم توزیع‌کننده) اطلاعات برای استفاده دیگران باشد. این ویژگی یکی از مخاطرات اصلی پدید آمدن رسانه‌های ارتباطی را نیز برطرف می‌سازد. بدین معنا که گرچه پیدایش رسانه‌های مختلف در بدو امر سبب افزایش ارتباطات جمعی و نزدیکی بیشتر افراد جامعه گردید، درگذر زمان، جریان یکسویه اطلاعات و عدم تأثیرگذاری کاربر بر رسانه، سبب انزوای بیش از پیش انسانها شد و بر تنهایی آنها افزود؛ اما ظهور اینترنت و حضور فعال کاربر در محیط این رسانه، افزایش ارتباطات فعال و حضور هوشمندانه کاربر را به همراه دارد و این خود مانع از عزلت انسانهاست.

از سوی دیگر، یکی از مهمترین امتیازات اینترنت انتقال صورتهای ترکیبی داده، مانند آمیزه‌ای از اطلاعات نوشتاری، دیداری و شنیداری است؛ ضمن اینکه اینترنت محمل مناسبی را برای جابه‌جایی اطلاعات بویایی و لامسه پدید آورده است (Markus, 2001). بدین ترتیب، این رسانه جدید، با در اختیار داشتن ویژگیهای ممتاز خود سعی در پدیدآوردن بستری جامع و فراگیر، برای ایجاد ارتباطی مناسب و هوشمندانه با کاربران دورترین گوشه‌های جهان داشته است.

درحال حاضر، شبکه اینترنت شامل همه رایانه‌ها و شبکه‌های کوچکی است که از پروتکل TCP/IP یا پروتکل‌های مشابه استفاده می‌کنند. تخمین زده می‌شود که اکنون ۲۲۶ کشور، با بیش از ۷۰۰ میلیون کاربر، به این

3. search engines
4. world wide web

الف. روبات خزنده^۹ یا روبات اطلاعاتی. این روباتها برنامه‌هایی هستند که با جستجوی مستمر و مداوم در صفحات اینترنت و نیز پیگیری پیوندهای میان صفحات، یافته‌های خود را به موتور کاوش باز می‌گردانند. این فرایند به صورت منظم در دوره‌های مشخص زمانی تکرار می‌شود. چنین روالی را گردآوری^{۱۰} نیز می‌نامند. در حال حاضر حدود ۲۰۰ نوع روبات مختلف برای جمع‌آوری اطلاعات در موتورهای کاوش به کار می‌رود که هر یک از این روباتها می‌توانند یک یا چند پروتکل مختلف را شامل: http, wais, ftp, gopher حمایت و تا ۱۰ میلیون صفحه را در روز مشاهده کنند (Mar, 2001).

ب. نمایه (پایگاه) اطلاعاتی. روبات اطلاعاتی یافته‌های خود را به بخش نمایه موتور منتقل می‌کند؛ از این رو، نمایه شامل نسخه‌ای از همه صفحاتی است که روبات یافته است. نحوه نمایه‌سازی صفحات موتورهای کاوش مشابه روند نمایه‌سازی در پایگاههای اطلاعاتی متداول است. در این فرایند، دو نکته اصلی وجود دارد: یکی نوع و حجم اطلاعات استخراج شده؛ دیگر نحوه ساماندهی این اطلاعات در پایگاه اطلاعاتی. روشهای متفاوت تحقق این دو مهم موتورهای کاوش را با یکدیگر متفاوت می‌کند.

ج. نرم افزار کاوش. این نرم افزار برنامه‌ای است که مورد کاوش را از میان همه صفحات موجود در نمایه، یافته و براساس میزان تناسب و ارتباط آن با درخواست، واژه یا عبارت مورد نظر رتبه‌بندی می‌کند. چنین فرایندی را بازیابی^{۱۱} می‌نامند. بدیهی است که روند بازیابی کاملاً به نحوه نمایه‌سازی اطلاعات در پایگاه اطلاعاتی موتور کاوش وابسته است. هرچه نمایه دقیقتر و منظمتر تدوین شده باشد

یکسانی دارند و از سه بخش اصلی روبات اطلاعاتی^۵، نمایه^۶ و نرم افزار کاوش^۷ تشکیل شده‌اند. در این میان، توانمندیهای روبات اطلاعاتی و نمایه مبنای قوت محتوایی موتور کاوش است؛ در حالی که امکانات و قابلیت‌های نرم‌افزار کاوش سبب کاربرپسندی^۸ موتور خواهد گردید. البته بدیهی است هرچه توانمندی موتور کاوش در بازیابی و دسته‌بندی اطلاعات بیشتر باشد بر محبوبیت آن نیز نزد کاربران افزوده خواهد شد. با توجه به این نکات، حدود ۱۰۰ موتور کاوش می‌توان نام برد که مقبولیت بیشتری نزد کاربران دارند (برای توضیحات بیشتر، نک: منتظر، ۱۳۸۲: ۲۳-۲۱).

نکته مهم دیگر در دستیابی به اطلاعات دقیق و صحیح از طریق موتورهای کاوش، نحوه عملکرد کاربر و استراتژی اتخاذشده در اجرای کاوش است. انتخاب دقیق موتور کاوش، آشنایی با امکانات کاوش موتور، برگزیدن کلیدواژه، پرسش یا عبارت مناسب در کاوش و نیز دقت در دسترسی به پیوندها و راهنماهای موجود، از مهمترین شاخصه‌های بازیابی دقیق اطلاعات است؛ لذا، در این مقاله، سعی شده است که ساختار کلی موتورهای کاوش و نتایج مطالعه تحقیقی که بر موتورهای مختلف کاوش انجام شده است بیان شود. در این تحقیق به نکات مهم زیر توجه شده و انتظارات کاربران مختلف از موتورهای کاوش اطلاعات تبیین گردیده است.

الف. میزان رضایت کاربران در بازیابی اطلاعات از موتورهای کاوش اصلی؛

ب. رفتار کاربران به هنگام استفاده از یک موتور کاوش؛

ج. رفتار کاربران به هنگام کاوش اطلاعات در اینترنت؛

د. محبوبیت موتورهای کاوش.

ساختار موتورهای کاوش

۵هر موتور کاوش دارای سه بخش اصلی است:

5. Info-bat
6. Index
7. search software
8. user friendliness
9. crawler
10. harvesting
11. retrieve

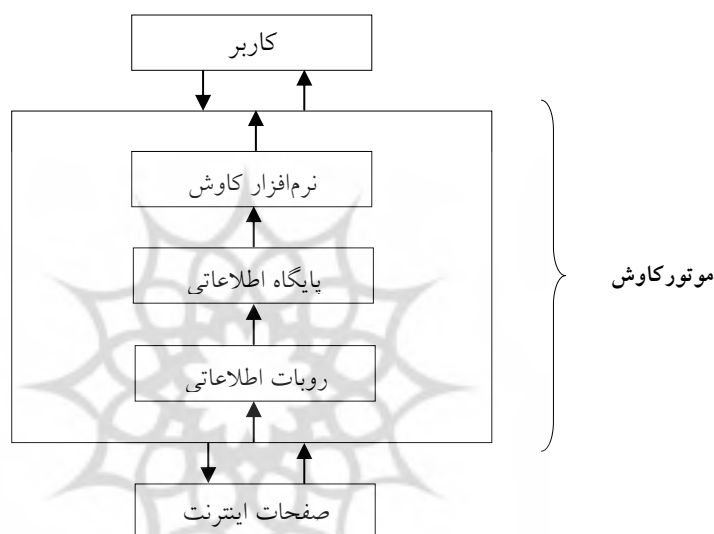
کاوش اطلاعات را نه از اینترنت بلکه از پایگاه اطلاعاتی خود جستجو و بازیابی می‌کند؛ بدین ترتیب، گرچه این پایگاه اطلاعات خود را از صفحات موجود در اینترنت به دست می‌آورد، لیکن، همچنانکه پیشتر اشاره شد، گردآوری اطلاعات از اینترنت دوره‌ای و در فواصل زمانی معین رخ می‌دهد. از این رو می‌توان گفت که جستجو از طریق موتورهای کاوش، کاوشی ناپیوسته (برون‌خط^{۱۲}) است.

بازیابی اطلاعات نیز سریعتر و صحیحتر خواهد بود (Calari & Wells, 2000: 10/22-27).

با توجه به توضیحات، ملاحظه می‌شود که نرم‌افزار کاوش بیشتر به کاربر نزدیک است و روبات اطلاعاتی بیشتر به منابع اطلاعاتی اینترنت؛ بنابراین، می‌توان معماری کلی موتورهای کاوش را مطابق شکل (۱) ترسیم کرد.

نکته مهم اینکه، برخلاف تصور غالب، موتور

شکل ۱. معماری موتورهای کاوش



دهند. نتیجه این نظرخواهی در نمودار (۱) نشان داده شده است. درخور ذکر است که پاسخگویان به این پرسش، کاربران ۱۳ موتور کاوش ذیل بوده‌اند:

Altavista, AOL Search, Ask Jeeves, Excite, Go, Google, Go To, Hot Bot, Lycos, MSN Search, Netscape Search, Webcrawler, Yahoo!

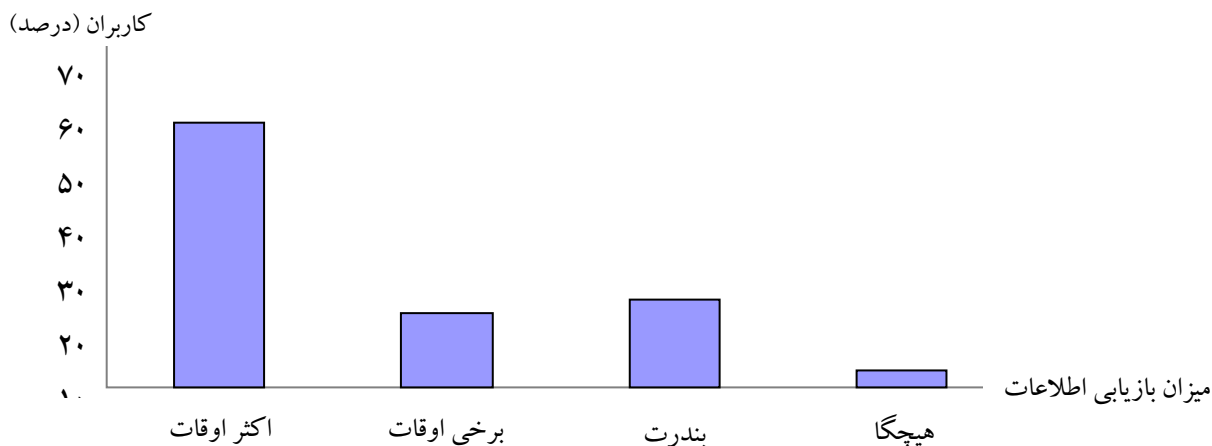
این نمودار نشان می‌دهد که بیش از ۶۰٪ کاربران در اکثر مواقع به پاسخ مطلوب خود در اینترنت دست می‌یابند و فقط نزدیک به ۳٪ کاربران پاسخ جستجوی خود را در اینترنت نمی‌یابند.

امکان بازیابی اطلاعات

یکی از پرسشهای اساسی در مورد منابع اطلاعاتی موجود در اینترنت این است که آیا همواره می‌توان به اطلاعات موردنظر دست یافت؟ بدیهی است که پاسخ این پرسش با درجه‌ای از احتمال همراه خواهد بود. هیچگاه نمی‌توان با قطعیت تمام در مورد پاسخگویی دائمی اینترنت اظهارنظر کرد. مؤسسه NPD (Feb, 2002) برای بررسی احتمال بازیابی اطلاعات در اینترنت، از ۳۳ هزار کاربر اینترنت خواسته بود تا سؤال هنگام کاوش در اینترنت آیا موفق می‌شوید؟ را با یکی از گزینه‌های اکثر مواقع، برخی مواقع، به ندرت و هیچگاه، پاسخ

12. off-line

نمودار ۱. میزان موفقیت کاربران در استفاده از موتورهای کاوش در بازیابی اطلاعات

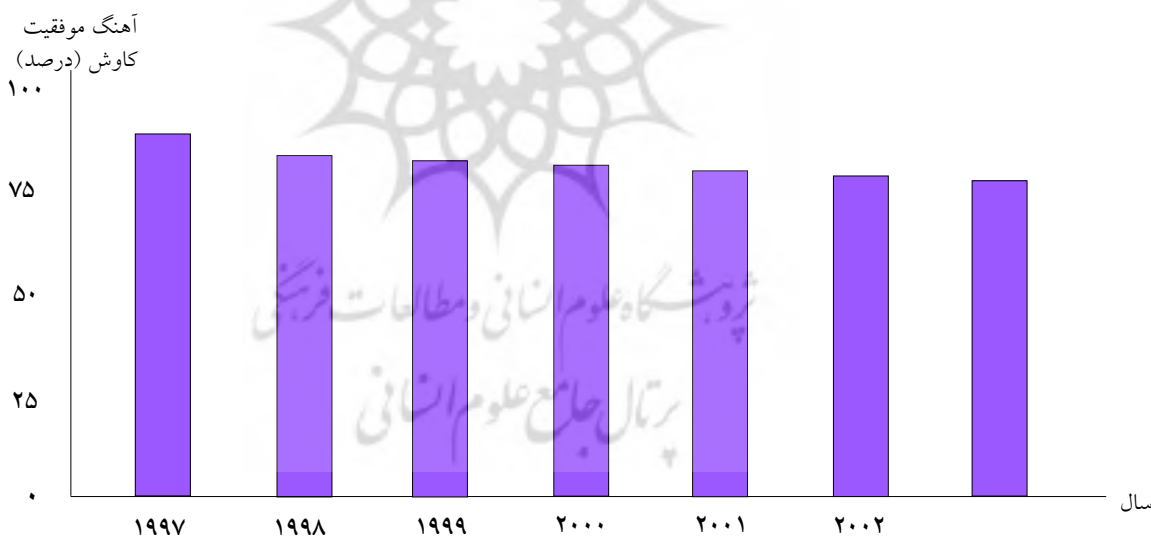


نمودار ۲ نشان‌دهنده آهنگ موفقیت کاوش در فاصله سالهای ۱۹۹۷ تا ۲۰۰۳ م برای ۱۳ موتور کاوش مهم اینترنت است (Apr, 2003).

با توجه به این نکته، شاخصی جدید به نام آهنگ موفقیت کاوش^{۱۳} را به شکل زیر تعریف می‌کنیم:

درصد کاربران با توفیق برخی اوقات + درصد کاربران با توفیق اکثر اوقات = آهنگ موفقیت کاوش

نمودار ۲. تغییرات آهنگ موفقیت کاوش



کاوش است. در این باب باید به حجم پایگاههای اطلاعاتی موجود در اینترنت، در فاصله سالهای ۱۹۹۱ تا ۲۰۰۳ م. اشاره کرد که با بیش از ۲۰ برابر افزایش به ۸۰۰ میلیون پایگاه رسیده است و نیز حجم کل صفحات موجود در اینترنت، که به بیش از ۲۰

چنانکه ملاحظه می‌شود، میزان آهنگ موفقیت کاوش، به تدریج، از سال ۱۹۹۷ تا ۲۰۰۳ م. از ۸۵/۸٪ به ۷۷٪ رسیده است؛ به تعبیر دیگر، میزان بازیابی اطلاعات مطلوب، و بالطبع میزان رضایت کاربران، به تدریج کاهش یافته است. مهمترین دلیل آن افزایش حجم اطلاعات موجود در اینترنت و محدودیت توان بازیابی اطلاعات به توسط موتورهای

13. Search Success Rate (SSR)

کاوش واژه منفرد^۴، کاوش واژه‌های ترکیبی^۵، کاوش جمله‌های از قبل تعیین شده^۶، کاوش زبان طبیعی^۷، استفاده از هر یک از این امکانات نقش مهمی در بازیابی بهینه اطلاعات دارد؛ لیکن معمولاً هریک از کاربران رفتار خاصی برای استفاده از این امکانات دارند که ارزیابی آن می‌تواند در تعیین استراتژی طراحی موتورهای جدید کاوش بسیار مؤثر باشد. نتایج بررسی نحوه بازیابی اطلاعات در موتورهای کاوش در نمودار (۳) آمده است. این نتایج از حدود ۱۰۰۰ مورد نظرسنجی از کاربران حرفه‌ای اینترنت حاصل شده است. در این تحقیق، جامعه‌ای از کاربران حرفه‌ای اینترنت در نظر گرفته شده است، و شامل حدود ۶۵۰ کاربر دانشگاهی، که از امکانات اینترنت در درون مراکز رایانه پنج دانشگاه تهران استفاده می‌کنند، و نیز حدود ۳۵۰ کاربر حرفه‌ای، که از طریق کافی‌نت‌های موجود در شهر تهران به اطلاعات موردنیاز دست می‌یابند. پرسش اصلی در این نظرسنجی این بود که، به هنگام جستجو در یک موتور کاوش، از کدامیک از حالت‌های کلید واژگان ترکیبی، کلید واژه منفرد، جمله‌های از قبل تعیین شده و سؤال استفاده می‌کنید؟

میلیارد صفحه بالغ شده است و این در حالی است که بزرگترین موتور کاوش موجود در اینترنت (Google) حدود هشت میلیارد صفحه را تحت پوشش قرار می‌دهد. به همین دلیل نیز درصد کاربرانی که نمی‌توانند از اینترنت برای بازیابی اطلاعات استفاده کنند در فاصله سالهای ۱۹۹۷ تا ۲۰۰۱ از ۰/۲٪ به ۳/۲٪ رسیده است (Clawer, 2004: 191-196). این امر کاهش توانمندی نسبی موتورهای کاوش را در تأمین رضایت کاربران تأیید می‌کند.

بررسی عملکرد کاربران در کار با موتورهای کاوش

بازیابی اطلاعات از شبکه وب، علاوه بر توانمندی نرم‌افزار جستجوی موتور کاوش به چگونگی کار کاربران با موتور کاوش نیز بستگی دارد. در این بخش به بررسی دو موضوع، در پیرامون چگونگی مواجهه کاربران، می‌پردازیم:

الف. نحوه جستجو در موتور کاوش: یکی از مسائل مهمی که بر چگونگی بازیابی اطلاعات اثر دارد نحوه جستجو از طریق موتور کاوش است؛ هر یک از موتورهای کاوش امکانات مختلفی دارد، چون

نمودار ۳. چگونگی جستجوی کاربر از طریق موتور کاوش



14. mono keyword 15. multiple keyword
16. predefined sentences 17. natural language

کاوش هستند. نکته مهم دیگر این است که موتور کاوش تنها واسطی میان کاربر و دنیای اطلاعات است؛ از این رو، طبیعی است که کاربران بیشتر به استفاده از یک موتور کاوش تمایل داشته باشند. در تحقیق از کاربران سؤال شده که چنانچه با استفاده از روش کاربردی در بند قبل، یعنی استفاده از کلیدواژه‌های ترکیبی / منفرد، به پاسخ مطلوبی در یک موتور کاوش دست نیافتند کدام یک از سه روش ذیل را برمی‌گزینند؟

الف. استفاده از همان موتور کاوش و تغییر کلیدواژه؛

ب. استفاده از همان کلیدواژه و تغییر موتور کاوش؛

ج. تغییر کلیدواژه و تغییر موتور کاوش.

نتایج حاصل از این نظرسنجی در نمودار (۴) آمده است. چنانکه ملاحظه می‌شود، حدود ۸۰٪ از کاربران ترجیح می‌دهند از همان موتور کاوش استفاده کنند و صرفاً کلیدواژه‌های کاربردی را تغییر دهند.

نمودار ۴. چگونگی استفاده از یک یا چند موتور کاوش در بازیابی اطلاعات

کاربران (درصد)



کار با یک موتور کاوش و عدم تغییر آن بسیار مهم است.

پیچیدگی ساختار

یکی از نکات تأثیرگذار بر توانمندیهای بازیابی اطلاعات،

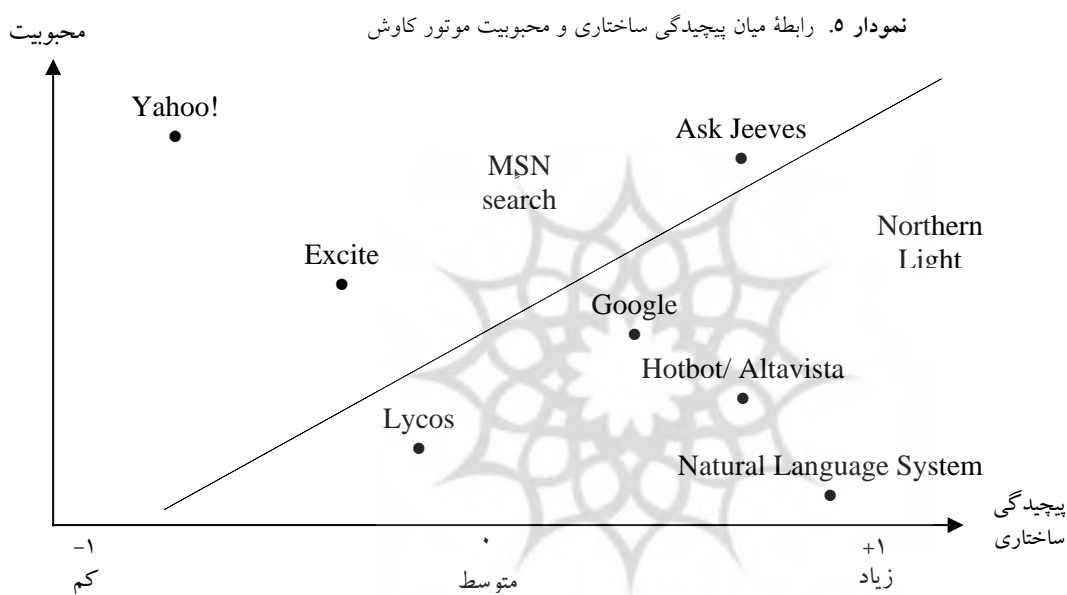
همچنانکه ملاحظه می‌شود، حدود ۸۰٪ کاربران صرفاً با استفاده از کلیدواژه به بازیابی اطلاعات می‌پردازند؛ لیکن باید گفت که امکان بهره‌گیری از سؤالات مختلف یکی از راههای بازیابی دقیقتر اطلاعات است و استفاده از آن، در طی سالهای اخیر، با آهنگ ملایمی در حال رشد است. به طوری که آمار نشان می‌دهد میزان استفاده از این امکان، از حدود ۱٪ در سال ۱۹۹۷ به حدود ۹/۵٪ در سال ۲۰۰۳ رسیده است (Parker, 2004: 83-84). پیش‌بینی می‌شود که به تدریج، به یکی از روالهای اصلی بازیابی اطلاعات تبدیل شود.

ب. نحوه کاوش در اینترنت: یکی از نکات مهم در کاوش اطلاعات، عادت کاربران به استفاده از یک موتور کاوش خاص است. امکانات مختلف موتور کاوش، سادگی و کاربرپسندی آن و جذابیت محیط استفاده، از مهمترین عواملِ خوگرفتن کاربران به استفاده از یک موتور

بدین ترتیب ملاحظه می‌شود خوگرفتن کاربر با یک موتور کاوش، تأثیر بسیار بر نحوه استفاده از آن دارد، از این رو، در تعیین الگوی رفتار کاربران اینترنت توجه به خصلت آنها در

نحوه کاوش انسان و با استفاده از قوانین هوش مصنوعی^{۲۰}، قابلیت جستجوی بهینه و هدفمند را به مقدار بسیار زیادی افزایش داد. بدیهی است این امکانات سبب دستیابی به اطلاعات دقیقتر می‌شود، لیکن استفاده از موتور کاوش را دشوارتر می‌کند. در یک دید کلی می‌توان ارتباط پیچیدگی ساختاری موتورهای کاوش را با میزان کاربر پسند بودن آن در نمودار (۵) نمایش داد (Sep.2003).

ساختار روبات اطلاعاتی و ویژگیهای نرم‌افزار جستجوی موتور کاوش است. امکان کاوش در منابع مختلف اطلاعاتی، امکان نمایه‌سازی بخشهای مختلف سامانه‌های اطلاعاتی، دوره روزآمدسازی و... از مهمترین ویژگیهای موتور کاوش هستند؛ علاوه بر این، امکان کاوش هوشمند^{۱۸} و استفاده از شیوه‌های جستجوی مبتنی بر قواعد فازی^{۱۹}، این توانمندی را به موتور کاوش می‌دهد تا بتوان با مدل‌سازی



درونی، کمتر از ۱۵٪ کاربرد در میان کاربران را دارند؛ اما موتورهای کاوش سمت چپ، به رغم سادگی ساختاری، ۸۵٪ کاربرد دارند. علاوه براین، در تحقیقی از کاربران مختلف، محبوبترین موتورهای کاوش کاربردی آنها سنجش شد که نتایج آن در جدول ۱ ذکر شده است.

همچنانکه ملاحظه می‌شود، میزان محبوبیت موتور کاوش با پیچیدگی ساختاری آن نسبت عکس دارد؛ به طوری که موتور کاوش Yahoo دارای ساده‌ترین شکل ساختاری، ولی بیشترین میزان محبوبیت است. به عکس، موتورهای کاوش مبتنی بر زبان طبیعی، کاملترین شکل ساختاری، اما کمترین میزان محبوبیت را دارند. در این نمودار موتورهای کاوش به کمک خط چین به دو دسته تقسیم شده‌اند: موتورهای کاوش سمت راست، با بهترین و پیچیده‌ترین توابع

18. intelligent search
19. fuzzy rules
20. artificial intelligence

رتبه	موتورهای کاوش میزان محبوبیت	موتورهای کاوش درصد استفاده کاربران
۱	Yahoo!	۵۵/۸
۲	Google	۴۰
۳	Altavista	۱۷
۴	Excite	۹/۷
۵	Hotbot	۹/۳
۶	Infoseek	۸/۹
۷	Lycos	۵/۱

برای مقایسه دقیقتر، میزان پوشش اطلاعات تحت وب موتورهای مختلف کاوش در جدول ۲ ذکر شده است (Tunstel & Seraji, 2003).

جدول ۲. میزان پوشش موضوعی ده موتور کاوش پُر کاربرد

رتبه	موتور کاوش	درصد پوشش
۱	Google	۴۰
۲	Northern Light	۱۶
۳	Snap	۱۵/۵
۴	Altavista	۱۵/۳
۵	Hotbot	۱۱/۳
۶	Microsoft	۸/۵
۷	Infoseek	۸
۸	Yahoo!	۷/۴
۹	Excite	۵/۶
۱۰	Lycos	۲/۵

ب. حدود ۴۵٪ از کاربران اینترنت در بازیابی اطلاعات از کلیدواژه‌گان ترکیبی و حدود ۳۰٪ از واژگان منفرد استفاده می‌کنند و تنها حدود ۲۵٪ از عبارتهای از قبل تعیین شده سؤال برای بازیابی اطلاعات استفاده می‌کنند. این موضوع، اهمیت نمایه‌سازی دقیق منابع اطلاعاتی و تدوین ساختار منظم پایگاه اطلاعاتی را، بیش از پیش نمایان می‌کند. ضمناً، باید گفت که، استفاده از پرسشهای مبتنی بر زبان طبیعی، به تدریج، به‌عنوان یکی از راهکارهای بازیابی اطلاعات، مورد استفاده کاربران قرار می‌گیرد.

ج. حدود ۸۰٪ از کاربران مایل به استفاده از یک موتور کاوش هستند و حتی اگر به‌هنگام کاوش در آن موتور به پاسخ مناسب دست نیابند، بدون تغییر موتور کاوش، صرفاً با تغییر کلیدواژه، به جستجوی مجدد اطلاعات می‌پردازند. از این‌رو، به‌رغم تعداد بسیار زیاد موتورهای کاوش (بیش از ۳ هزار موتور)، ایجاد جذابیت‌های لازم برای کاربرپسند کردن موتور کاوش بسیار مهم است.

د. پوشش موضوعی موتور کاوش، یکی از عوامل تعیین‌کننده کیفیت موتور کاوش است، که خود به معماری طراحی و در نظرگرفتن ویژگیهای خاص در نرم‌افزار کاوش موتور بستگی دارد. با این‌همه، نظرسنجی از کاربران نشان می‌دهد که محبوبیت موتور کاوش، ارتباط کمتری با پیچیدگی ساختاری و توانمندیهای کاوش دارد. از این‌رو، موتور کاوش Yahoo، با اینکه کمتر از ۸٪ اطلاعات شبکه وب را پوشش می‌دهد، صرفاً به دلیل سادگی کاربرد و امکانات عمومی، محبوبترین موتور کاوش اینترنت است.

البته این نکته نیز مهم است که سادگی کاربرد با توانمندی ذاتی موتور کاوش منافاتی ندارد؛ بدین معنا

از مقایسه جدولهای ۱ و ۲، نیز با توجه به نمودار ۵، ملاحظه می‌شود که هیچگونه رابطه مستقیمی میان پیچیدگی ساختاری موتور جستجو، میزان پوشش موضوعی و محبوبیت آن وجود ندارد، بلکه سادگی کاربرد و جذابیت ظاهری آن تأثیر بسزایی در جذب مخاطب داشته است. چنانکه ملاحظه می‌شود، موتور کاوش Yahoo، به‌رغم پوشش بسیار کم آن (حدود ۷/۵٪ اطلاعات وب)، به دلیل سادگی کاربرد و در اختیار گذاشتن بسیاری از امکانات عمومی، مانند امکان استفاده از پیام‌نگار^{۲۱}، پیام‌آور^{۲۲}، بیشترین محبوبیت را در میان کاربران دارد.

بررسی نتایج

در این مقاله، به نتایج تحقیق بر ویژگیهای کاربری

21. e-mail

22. messenger

موتور کاوش برای بازیابی اطلاعات از اینترنت توجه شد. این نتایج در طراحی موتورهای کاوش علمی ایران، که بازیابنده اطلاعات علمی فارسی در گستره شبکه وب خواهد بود، اهمیت بسیار دارد. مهمترین نکات به دست آمده از تحقیق آن به شرح ذیل است:

الف. بیش از ۷۵٪ کاربران اینترنت استفاده از موتورهای کاوش را مهمترین راهکار بازیابی اطلاعات از اینترنت می‌دانند و از استفاده از این افزار رضایت دارند. البته، باید گفت که، به دلیل افزایش سریع حجم اطلاعات در اینترنت و توانایی محدود موتورهای کاوش، آهنگ موفقیت کاوش اندک اندک در حال کاهش است، لیکن پدیدارشدن راهکارهای جدید ناشی از کاربرد روشهای خبره و هوشمند در بازیابی اطلاعات می‌تواند این نقیصه را جبران کند.

مستلزم شناخت انتظارات کاربران از محیط وب و پایگاههای اطلاعاتی موجود در آن و نحوه بازیابی اطلاعات مورد نیاز است. در این مقاله عوامل مختلف مواجهه با موتور کاوش از سوی کاربران شامل: نحوه جستجوی اطلاعات در موتور کاوش، چگونگی استفاده از یک یا چند موتور کاوش در بازیابی اطلاعات و نیز رابطه پوشش موضوعی، سادگی کاربرد و پیچیدگی ساختاری با محبوبیت موتور کاوش در میان کاربران بررسی شد؛ با توجه به آنها می توان نتیجه گرفت که طراحی موتور کاوش به دو عامل اصلی، یعنی توانمندی فنی - شامل قوت و ویژگیهای فنی و نرم افزاری در طراحی الگوریتمهای جستجو، سازماندهی نمایه ها و نحوه بازیابی اطلاعات - و کاربرپسندی - شامل تنوع راهنما و موضوعهای نمایه شده، سادگی استفاده از نرم افزار کاوش، جذابیت ظاهری و وجوه امکانات عمومی - ارتباط مستقیم دارد. به همه وجوه این دو عامل باید در طراحی و پیاده سازی موتور کاوش ملی توجه کرد.

که ممکن است موتور کاوش ساختار درونی قوی، همراه با توابع عملیاتی پیچیده و بهره مند از روشهای پردازشی خیره (و حتی هوشمند) داشته باشد؛ اما دارای انعطاف پذیری و سادگی کاربرد باشد. نمونه بارزی از ترکیب بیان روشهای قدرتمند جستجو همراه با سادگی استفاده کاربر در موتور کاوش google نشان داده می شود. از این رو، بسیار ضروری است که موتور کاوش طراحی شده از دید کاربر ساده و از دید توان جستجو در وب و نمایه سازی داده های بازیافتی قوی باشد.

با توجه به نکات یادشده، بدیهی است که طراحی موتور کاوش جدید نیز باید مبتنی بر اصول فوق باشد؛ از این رو، با توجه به سرعت رشد ایجاد پایگاههای اطلاعاتی تحت وب در کشور و افزایش منابع اطلاعاتی فارسی موجود در اینترنت، لزوم طراحی موتور کاوش فارسی بسیار ضروری به نظر می رسد. طراحی این موتور، هرچند که باید جامع و دربرگیرنده حجم عظیمی از اطلاعات مختلف - مانند اطلاعات علمی، تجاری، ورزشی - باشد، باید مبتنی بر الگوی رفتاری کاربران و در برگیرنده عادات آنها در استفاده از موتور کاوش باشد. از این رو، لازم است موتور طراحی شده ساختاری ساده، امکانات عمومی مناسب - مانند ارائه دسته بندیهای مختلف اطلاعات، امکان ارائه پیام نگار و پیام آور مبتنی بر وب، صندوق پستی شخصی - داشته باشد و درعین حال، روزآمدسازی اطلاعات و نحوه نمایه سازی اطلاعات در آن کیفیت مطلوبی داشته باشد. بدیهی است این ویژگیها، هم در معماری طراحی روبات خزننده و هم در حجم پایگاه اطلاعاتی موتور تأثیر بسزایی خواهد داشت.

نتیجه گیری

طراحی موتور کاوش برای مقاصد علمی یا عمومی

منابع

منتظر، غلامعلی (۱۳۸۲)، *موتورهای کاوش اینترنت: درآمدی بر بازیابی بهینه اطلاعات*، چاپ دوم، انتشارات کویر، تهران؛

Calari, S. & Wells, A.T. (2000), "New Tools for Collection Development: The Internet Scout Project", *Electronic Commerce World*, Vol. 10, No.5;

Clawer, T. (Dec.2004), "Data Fetching in the Web", *Proc. of 27th of Online Information*, London;

Cooky, R. & Mobaser, B. (1997), "Web Mining: Information and Pattern Discovery on World Wide Web", *Proc. of IEEE Conf. on Artificial Intellingence*, U.S.A.;

"**Evaluation of Selected Internet Search Tools**" (Apr. 2001), [online] available at: <http://www.Library.nwv.edu/resources/internet/evaluation.html>;

"**Information Retrieval Technology and Intelligent Concept Extraction Searching**" (Mar.2001), [online] available at: <http://www.excite.com/ice/tech.html>;

"**Internet Information Content**" (Nov.2003), [onLine] available at: <http://www.sp.edu.sg/ict/dit-is-modules.htm>;

Mahjer, S. T. (Sep.2003); "Internet Search and Data Mining", [online] available at: <http://www.rtsinc.org/learning/nw/shapiro.doc>;

Markus, K. (Dec. 2001), "Uses and Users of Internet", [on line] available at: <http://www.rit.com/internet>;

Micheal, R. (Feb.2002), "*Information Development*", [on line] available at: <http://www.smart.com/info/it.htm>;

"**NPD Report**" (Feb.2002), [online] available at: <http://www.searchenginewatch.com/reports/npd.html>;

Parker, S. et al (Dec.2004); "Web Crawler and Web Finding Methods", *Proc of 27th Online Information*; London ;

"**Search Engine Comparison**" (May.2004), [online] available at: <http://www.searchengineguide.com/searchenginecomparisonchart>;

"**Search Engines Harvesting Mechanism**" (Apr.2003), [online] available at: <http://web.nexor.co.uk/mak/dos/robots/robots.htm>;

Tunstel, E. & Seraji, H. (2003), "Soft Computing Approach to Safe Navigator of Autonomous Planetary Rovers", *Intelligent Control System*, CRC Prees. ■