

مقایسه نتایج بازیابی شده از ابرموتورهای کاوش وبی در حوزه علم اطلاعات و دانش‌شناسی

ناهید خوشیان*

کارشناسی ارشد علم اطلاعات و دانش‌شناسی، دانشگاه شیراز

تاریخ پذیرش: ۱۳۹۴/۸/۲۸

تاریخ دریافت: ۱۳۹۴/۶/۲۵

چکیده

هدف: هدف از پژوهش حاضر بررسی ۸ ابرموتور کاوش مطرح به نام‌های Webcrawler, Depperweb, Metagopher, Ask Jeeves, Clusty, Goto, Metalib, Polymeta از نظر پاسخگویی به سؤالات اختصاصی مرجع در رشته علم اطلاعات و دانش‌شناسی است.

روش: تعداد ۱۶ سؤال اختصاصی این رشته در هریک از ابرموتورها جستجو و ده نتیجه اول در هر ابرموتور از نظر ربط، دقت و ریزش کاذب ارزشیابی و مقایسه شد. جمعا ۱۲۸۰ جستجو انجام شد. در تجزیه و تحلیل داده‌ها از فرمول دقت استفاده گردید. توانمندی‌های هر کدام از ابرموتورها، میزان ریزش کاذب و انواع آن، تفاوت ابرموتورهای کاوش و انتخاب بهترین ابرموتور در پاسخگویی به این سؤالات ارزشیابی شد.

یافته‌ها: یافته‌های پژوهش نشان داد که ابرموتورهای کاوش قادر به بازیابی اطلاعات رشته علم اطلاعات و دانش‌شناسی در حد مطلوب نیستند.

نتیجه‌گیری: ابرموتور جستجوی Webcrawler با بازیابی ۸۷/۳۱ درصد مربوط دارای بهترین عملکرد است و نیز هر ابرموتور در بازیابی نتایج، متفاوت با ابرموتور دیگر عمل می‌کند.

کلیدواژه‌ها: ابرموتورهای کاوش، دقت، ربط، ریزش کاذب، بازیابی اطلاعات

مقدمه

اینترنت منبع عظیمی از اطلاعات در موضوع‌های مختلف است. برای دسترسی به اطلاعات هیچگونه نظارتی بر محتوای آن وجود ندارد. لذا با در نظر گرفتن تراکم اطلاعات موجود در اینترنت، بازیابی بهینه اطلاعات مسأله‌ای مهم و حیاتی است (Isfandyari-Moghadam & Parirokh, 2006) و موجب دقت در امر پژوهش و صرفه‌جویی در وقت پژوهشگران می‌شود. برای دستیابی سریع و مؤثر به اطلاعات، ابزارهای متعددی چون موتورهای جستجو، راهنماهای موضوعی، پایگاه‌های اطلاعات وب نامرئی و ابرموتورهای جستجو در اینترنت ایجاد شده‌اند که از این میان ابرموتورهای جستجو بیشتر مورد توجه هستند (مهرداد و بینش، ۱۳۸۶).

ابزارهای جستجو انواع مختلفی دارند و هر دسته از آنها در مواردی چون عمق نمایه‌سازی، صفحات وب، امکانات جستجو، رتبه‌بندی نتایج، و میزان اطلاعات توصیفی برای هرمدخل بازیابی شده، باهم متفاوت هستند (رضوانی، ۱۳۸۳). این تفاوت‌ها نشان‌دهنده قابلیت‌ها و توانایی‌های متفاوت ابزارهای جستجو می‌باشد. لذا بکارگیری ابزارهای جستجو بدون اطلاع و آگاهی از توانمندی‌های هر کدام از آنها چندان سودمند نیست، لذا در همین راستا تعدادی از پژوهشگران با انجام پژوهش‌های مختلف اقدام به شناسایی و معرفی برترین ابرموتورها و موتورهای جستجو کرده‌اند. پژوهش حاضر نیز با همین هدف و نیز بدلیل کاربرد زیاد ابرموتورهای و موتورهای جستجو برای کاربران رشته علم اطلاعات و دانش‌شناسی به‌طور خاص در این حیطه موضوعی انجام شده است.

حجم اطلاعات موجود در وب همچنان افزایش یافته و رو به ازدیاد است به نحوی که امروزه با پدیده‌ای به نام انفجار اطلاعات مواجه هستیم. بر این اساس فرآیند اطلاع‌یابی در محیط وب به امری حیاتی در جوامع اطلاعاتی مبدل شده و دسترسی به منابع مرتبط از میان انبوه منابع موجود در اینترنت یکی از دغدغه‌های جدی کاربران به شمار می‌رود (محمداسماعیل و فیروزی، ۱۳۸۸)، اطلاعات از طریق صفحات وب منتشر می‌شوند و قابل دسترس هستند. برای دسترسی به این اطلاعات باید از محل استقرار آنها آگاه بود. برای حل این مشکل، ابزارهای جستجو در اینترنت پدید آمده‌اند که مجموعه‌ای از اطلاعات منظم شده صفحات وب را در بر دارند (وزیرپور کشمیری و فزونی، ۱۳۸۹). ابزارهای کاوش تنها وسیله برای دسترسی به اطلاعات وب است.

با انجام این پژوهش مشخص خواهد شد که کدامیک از ابرموتورهایی جستجو در حیطه علم اطلاعات و دانش‌شناسی دقت بیشتر و ریزش کاذب کم‌تری داشته و منابع مرتبط‌تری را بازیابی می‌کنند. بر این اساس ابرموتورهای مورد پژوهش رتبه‌بندی خواهند شد و متخصصان این رشته به بهترین ابرموتور جستجو مراجعه می‌کنند و دقیق‌ترین و مرتبط‌ترین نتایج را دریافت خواهند کرد. این امر به صرفه‌جویی در وقت در امر جستجو و بازیابی اطلاعات کمک قابل ملاحظه‌ای می‌نماید. باید توجه داشت که از یک سو هیچگونه سازماندهی اساسی روی اطلاعات موجود در وب صورت نمی‌گیرد، و از دیگر سو روز به روز بر تعداد ابرموتورها و موتورهای جستجو افزوده می‌شود. لذا در چنین شرایطی انتخاب مناسب‌ترین و پرکاربردترین ابزار جستجو و بازیابی اطلاعات مطرح می‌گردد. امروزه تعداد زیادی از کتابداران و پژوهشگران بدون در نظر گرفتن و آگاه‌بودن از نقاط قوت و ضعف ابرموتورهای جستجو به سراغ آنها می‌روند، بنابراین شناخت و آگاهی از کاستی‌ها و توانمندی‌های این ابرموتورها می‌تواند کمک شایان توجهی به امر پژوهش و تحقیق نموده و منجر به نتایج سودمند و کارا برای محققان گردد.

پیشینه پژوهش

محمداسماعیل و کیایی (۱۳۸۸)، با هدف بررسی موتورها و ابرموتورهای کاوش عمومی در بازیابی اطلاعات فیزیکی و میزان هم‌پوشانی آنها، شش موتور کاوش و شش ابرموتور کاوش عمومی را که جزء پراستفاده‌ترین‌ها بودند را برگزیده و میزان هم‌پوشانی آنها در بازیابی اطلاعات را مورد بررسی قرار دادند. نتایج پژوهش حاکی از آن بود که از نظر هم‌پوشانی استفاده از ابرموتورهای کاوش عمومی نسبت به موتورهای کاوش عمومی ارجح است.

میرحسینی و بابایی (۱۳۹۱)، در پژوهشی با هدف مقایسه جامعیت و مانعیت موتورهای کاوش تخصصی پزشکی در بازیابی اطلاعات مربوط به بیماری‌های زنان و مامایی، تعداد ۵ کلیدواژه مربوط به بیماریهای زنان و مامایی را با مشورت پزشکان متخصص زنان انتخاب و در پنج موتور کاوش تخصصی پزشکی مدهانت^۱، امنی^۲، پوگوفراگ^۳، سرچ مدیکا^۴ و تریپ دیتابیس^۵ جستجو شدند. سپس ۱۰ نتیجه نخست در هر موتور بازیابی و جامعیت و مانعیت آنها محاسبه شد. یافته‌ها نشان داد که بین مانعیت موتورهای پزشکی در بازیابی منابع تخصصی مربوط به بیماریهای زنان و مامایی اختلاف معنی‌داری وجود دارد، ولی از نظر جامعیت، بین موتورهای پزشکی با میانگین جامعیت ۲۰/۸ تفاوت معنی‌داری مشاهده نشد.

اسدی قادیکلایی، اسدی، نوروبی و احسانی (۱۳۹۳)، در پژوهشی با هدف مقایسه موتورهای جستجوی عمومی و پایگاه‌های تخصصی رادیولوژی در بازیابی تصاویر رادیولوژی و رتبه‌بندی آنها از نظر میزان دقت ۵ گروه از بیماری‌های قلبی و عروقی، سرطان‌ها، بیماری‌های حول تولد و دستگاه تنفس را انتخاب و با استفاده از نرم‌افزار محقق ساخته ابزار گردآوری داده‌ها و نیز ۸ کلیدواژه با استاندارد سرعنوان موضوعی مش^۶ و ام تری^۷ به جستجو در موتورهای جستجوی عمومی یاهو^۸، گوگل^۹ و بینگ^{۱۰} و موتورهای تخصصی رادیولوژی گولدماینر^{۱۱}، لرنینگ رادیولوژی^{۱۲}، سرچینگ رادیولوژی^{۱۳} و یوتالوک^{۱۴} پرداختند. ۱۰ تصویر اول بازیابی شده از هر پایگاه، بوسیله ابزار محقق ساخته به ۳۰ پزشک متخصص نشان داده شد. ارتباط یا عدم ارتباط تصاویر به کلیدواژه‌ها با استفاده از نرم‌افزار اسپاس^{۱۵} و آزمون فرض‌ها با استفاده از آزمون‌های آماری کای‌دو^{۱۶} و کرامروی انجام و مورد تجزیه و تحلیل قرار گرفت. یافته‌های پژوهش آنها پیشنهاد داد که با توجه به عملکرد بهتر موتورهای کاوش عمومی در بازیابی تصاویر رادیولوژی جهت کسب نتایج بهتر در پژوهش‌ها به موتورهای کاوش عمومی مراجعه شود.

وزیرپور کشمیری و فزونی (۱۳۸۹)، در مقاله‌ای با هدف مقایسه میزان دقت راهنماهای موضوعی وب در بازیابی اطلاعات فنی - مهندسی با روش اسنادی و وب سنجی کلیدواژه‌های تخصصی حوزه فنی - مهندسی از سایت تخصصی انستیتو آف الکترونیکال اند الکترونیکس انجینیرز (آی‌ای‌ای) ^{۱۷} و مجلات واقع در سایت ساینس دایرکت^{۱۸} را برگزیده و در راهنماهای موضوعی یاهو، گوگل، اینفوماین^{۱۹}، اینتیوت^{۲۰} و دیموز^{۲۱} که پر استفاده‌ترین راهنماهای موضوعی بودند، جستجو کردند. ملاک بررسی ۱۰ نتیجه نخست بود. نتایج نشان داد که اختلاف معنی‌داری در میزان دقت راهنماهای موضوعی وب وجود دارد و فرضیه اول تأیید گردید.

1. Medhunt
2. Omni
3. Pogofrog
4. Search Medica
5. Trip Database
6. Mesh
7. M-Tree
8. Yahoo
9. Google
10. Bing
11. Goldminer
12. Learning Radiology
13. Searching Radiology
14. Yottalook
15. SPSS
16. X^2 (Chi-squared test)
17. (Institute of Electrical and Electronics Engineers) IEEE
18. Science Direct
19. Infomine
20. Intute
21. Dmoz

راهنماهای موضوعی وب از نظر دقت به این ترتیب رتبه‌بندی شدند: گوگل، یاهو، اینتیوت، دیموز، و اینفوماین. از نظر میزان خطای مشاهده‌شده، راهنماهای موضوعی یاهو، دیموز، اینتیوت، گوگل، و اینفوماین به ترتیب کمترین میزان خطا را داشتند. در این پژوهش، مقایسه‌ای بین نسبت اقلام بازیابی شده در رده فنی - مهندسی با اقلام بازیابی شده از کلیدواژه‌ها در همه رده‌های موضوعی انجام شد و فرضیه دوم یعنی وجود اختلاف معنی‌دار بین این راهنماها تأیید گردید. همچنین، راهنماهای اینتیوت، اینفوماین، یاهو، دیموز و گوگل به ترتیب بیشترین نسبت اقلام فنی به کل را دارا بودند.

محمداسماعیل و فیروزی (۱۳۸۸)، در پژوهشی با هدف سنجش میزان هم‌پوشانی مدارک بازیابی شده در موتورهای کاوش و ابرموتورهای کاوش در حوزه نانو تکنولوژی^۱، ۵ موتور کاوش و ۵ ابرموتور کاوش را انتخاب و از نظر میزان هم‌پوشانی در حوزه نانو تکنولوژی مقایسه کردند. نتایج نشان داد که موتور کاوش یاهو، با میزان (۳۶/۶۶ درصد) هم‌پوشانی با سایر موتورهای کاوش، رتبه نخست را به خود اختصاص می‌دهد، در حالی که همزمان، موتور کاوش اسک کم‌ترین میزان هم‌پوشانی (۲۲/۰۸) با سایر موتورهای کاوش دارد. در نتیجه، موتور کاوش یاهو بیشترین میزان مدارک منحصر به فرد را در زمینه نانو تکنولوژی بازیابی می‌کند. در ابرموتورهای کاوش نیز، ابرموتور کاوش سرچ^۲، با میزان (۳۲/۰۸ درصد) هم‌پوشانی با سایر ابرموتورهای کاوش، رتبه نخست را به خود اختصاص می‌دهد. به علاوه، ابرموتور کاوش ماما^۳ کمترین میزان هم‌پوشانی (۲۳/۷۵ درصد) با سایر ابرموتورهای کاوش دارد، در نتیجه بیشترین میزان مدارک منحصر به فرد را ابرموتور کاوش سرچ در زمینه نانو تکنولوژی بازیابی می‌کند. در مجموع، میزان میانگین هم‌پوشانی در ابرموتورهای کاوش (۲۸/۸۳ درصد) می‌باشد. در نهایت در بحث تعیین میزان هم‌پوشانی میان موتورها و ابرموتورهای کاوش، ابرموتور کاوش داگ پایل^۴ با میزان (۳۳ درصد) هم‌پوشانی با موتورهای کاوش، رتبه نخست را به خود اختصاص می‌دهد، به علاوه، ابرموتور کاوش متاکراولر^۵ کمترین میزان هم‌پوشانی (۲۲/۳۳ درصد) با سایر موتورهای کاوش را دارد. در نتیجه، بیشترین میزان مدارک منحصر به فرد را ابرموتور کاوش داگ پایل در مقایسه با موتورهای کاوش در زمینه نانو تکنولوژی بازیابی می‌کند.

حریری، امامی و ملک (۱۳۹۴)، در پژوهشی با هدف مقایسه دقت موتورهای جستجوی عمومی در بازیابی تصاویر بیماری‌های مهم غدد درون ریز، پنجاه کلیدواژه در زمینه پنج بیماری مهم غدد درون ریز را در موتورهای جستجوی گوگل، یاهو و بینگ جستجو کردند. ۳۰ نتیجه اول در هر موتور بطور جاگانه ذخیره شدند، سپس نتایج با استفاده از فرمول‌های دقت و جامعیت در نرم‌افزار اکسل^۶ مورد تحلیل قرار گرفتند. نتایج پژوهش نشان داد که موتور جستجوی گوگل و بینگ در رقابت با یکدیگر قرار دارند و متخصصانی که در زمینه بیماری‌های غدد فعالیت می‌کنند، در صورت نیاز به بازیابی تصاویر می‌توانند با جستجو در گوگل و بینگ به دقت بالایی دست یابند. در ضمن کتابداران، در زمینه بازیابی تصاویر غدد می‌توانند این دو موتور جستجو را استفاده کنند و به کاربران پیشنهاد دهند.

Massmann & Rahm (2008)، در مقاله خود با عنوان "ارزیابی تطبیقی راهنماهای موردی وب" به ارزیابی میزان فناوری تطبیقی که می‌تواند برای تعیین خودکار راهنماها مورد استفاده قرار گیرد، پرداختند. هدف آنها یافتن روش‌های بهتر فنون خاص تطبیق براساس مورد، با استفاده از نشانی سایت و نام و توصیف وب‌سایت‌های دسته‌بندی شده است. روش‌هایی بر مبنای موارد خاص برای مقیاس‌های مشابه مختلف ارزیابی کرده و ترکیب آنها را از طریق فراداده مورد مطالعه قرار داده‌اند. مقایسه‌ای را

1. Nanotechnology

2. Search

3. Mamma

4. Dogpile

5. Metacrawler

6. Excel

در چهار راهنمای وب در فروش اینترنتی که عبارت بودند از دیموز، گوگل، وب، و یاهو انجام داده‌اند. با وجود شباهت در حد متوسط و هم‌پوشانی وبسایت راهنماها، مشکلات تطبیق می‌تواند در مقیاس بالایی حل شود.

Kaptien & Kamps (2009)، در مقاله خود با عنوان "راهنماهای وب به عنوان محتوای موضوعی" به بررسی اینکه آیا میتوان از راهنمای دی‌موز برای طبقه‌بندی پرسش‌ها از رده‌بندی‌های موضوعی در سطوح مختلف استفاده کرد و آیا می‌توان از این چهارچوب موضوعی برای بهبود عملکرد بازیابی استفاده کرد، پرداختند. در این مقاله، به کاربرهای مورد آزمون این امکان داده شد تا به وضوح پرسش‌های رده‌های موضوعی را رده‌بندی کنند. رده‌ها از دی‌موز یا از فهرست پیشنهادهایی که توسط فنون رده‌بندی موضوعی به‌طور خودکار ایجاد شده‌اند، انتخاب می‌شدند. نتایج این مطالعه نشان داد که رده‌بندی‌های دی‌موز برای طبقه‌بندی‌های موضوعی مناسب هستند. جستجوهای آزاد و همچنین ارزیابی فهرست پیشنهادها برای ایجاد چارچوب موضوعی به کار گرفته شد. زمانی که رده‌های موضوعی انتخاب شده از طریق جستجوی آزاد به عنوان چارچوب موضوعی به کار گرفته شد، بهبود نتایج بازیابی شده مشهود گردید.

Bar-Ilan (2005)، در پژوهشی با عنوان "هم‌پوشانی، دقت، و جامعیت در موتورهای کاوش، مطالعه موردی از سؤال اردوز^۱ توانایی‌های بازیابی ۶ موتور کاوش را با یک سؤال ساده مورد بررسی قرارداد.

Bharat & Broder (1998)، در پژوهشی با عنوان "تکنیکی برای ارزیابی اندازه نسبی و هم‌پوشانی موتورهای کاوش عمومی وب" نشان دادند که یک راه استاندارد شده برای ارزیابی پوشش و هم‌پوشانی موتورهای کاوش از طریق سؤال‌های تصادفی وجود دارد که می‌تواند توسط اشخاص ارزیابی شود.

Shafli & Rather (2005)، در پژوهشی با عنوان جامعیت و مانعیت پنج موتور کاوش از نظر بازیابی اطلاعات علمی در زمینه بیوتکنولوژی، موتورهای کاوش عمومی آلتاویستا^۲، هاتبات^۳، گوگل و دو موتور کاوش تخصصی بیوتکنولوژی، سایروس^۴ و بایوب^۵ را بررسی کردند. نتایج به دست آمده، حاکی از آن بود که موتور سایروس در بازیابی مدارک علمی در زمینه بیوتکنولوژی بهترین عملکرد را دارد. گوگل بهترین جانشین برای بازیابی مدارک علمی وب بنیاد است. همچنین نتایج نشان داد که مانعیت و جامعیت رابطه معکوسی با یکدیگر دارند به این معنی که اگر مانعیت افزایش یابد، جامعیت کاهش پیدا می‌کند و بر عکس.

Spink و دیگران (2006) در پژوهشی با عنوان "هم‌پوشانی میان موتورهای کاوش اصلی وب" به بررسی میزان هم‌پوشانی میان نتایج بازیابی شده در ۳ موتور کاوش اصلی وب اسک، جیویز، یاهو و گوگل پرداخت هدف از این پژوهش، اندازه‌گیری هم‌پوشانی میان ۳ موتور کاوش اصلی وب از طریق هم‌پوشانی صفحه اول نتایج جستجو، بررسی تفاوت‌های موجود در شمارگسترده‌ای از اصطلاحات جستجو شده توسط کاربر، تعیین تفاوت‌های موجود در صفحه اول نتایج جستجو و رتبه‌بندی آنها در موتورهای کاوش وب است. یافته‌ها نشان داد که نتایج کلی بازیابی شده تنها در یکی از ۳ موتور کاوش وب، ۸۵ درصد بود و نتایج در دو موتور کاوش ۱۲ درصد و در ۳ موتور کاوش وب ۳ درصد بود. میزان اندک هم‌پوشانی، تفاوت‌های اصلی موجود در نتایج رتبه‌بندی را منعکس می‌کند.

Spoerri (2007)، در پژوهشی با عنوان استفاده از ساختار هم‌پوشانی در نتایج جستجو برای رتبه‌دادن به سیستم‌های بازیابی بدون قضاوت ربط^۶ با تعیین ساختار هم‌پوشانی میان نتایج جستجو در سیستم‌های گوناگون، به تفاوت‌های اجرایی آنها پی برد. نتایج نشان داد که درصد مدارک منحصر به فرد هر سیستم، با میزان اثربخشی اجرایی آن سیستم رابطه مستقیم دارد. با استفاده از روش تعیین ساختار هم‌پوشانی می‌توان انواع سیستم‌ها را بدون نیاز به متخصصان برای تعیین ربط، مشخص کرد.

¹ Orduze

² Altavista

³ Hot Bot

⁴ Scirus

⁵ Bioweb

⁶ Relevancy

روش‌شناسی پژوهش

روش مورد استفاده در این پژوهش وب‌سنجی است. ۸ ابرموتور مطرح انتخاب شده و با استفاده از ۱۶ سؤال اختصاصی در حوزه علم اطلاعات و دانش‌شناسی و بررسی معیارهای ربط، دقت (با استفاده از فرمول دقت) و نیز انواع ریزش کاذب شامل (موارد نامربوط محتوایی، پیوندهای مرده، موارد تکراری مربوط و نامربوط، موارد غیرانگلیسی و نیز موارد بازیابی نشده) در رابطه با ۱۰ نتیجه نخست توسط محقق مورد بررسی، ارزیابی و مقایسه قرار گرفتند. دلیل انتخاب این ۸ ابرموتور کارایی بیشترشان در مقایسه با سایر ابرموتورها در پاسخ‌گویی به سوالات کتابداری و اطلاع‌رسانی بر اساس تجارب پژوهشگر و نیز نظرسنجی از اساتید این رشته بود. کلیدواژه‌های استخراجی در تک تک ابرموتورها مورد جستجو قرار گرفته و از بین نتایج بدست آمده تنها ده نتیجه نخست به دلیل مرتبط‌تر بودن مورد بررسی قرار گرفت. پس از بازکردن هر کدام از پیوندها و بررسی آنها موارد مربوط، نامربوط و انواع ریزش کاذب بازیابی شده توسط هر ابرموتور برای هر سؤال محاسبه شد. جهت بررسی صفحات تا حداکثر ۳ پیوند مورد بررسی قرار گرفت. جهت انجام جستجوی عبارتی از گیومه " " و نیز انواع عملگرهای بولی استفاده گردید. برای محاسبه میزان دقت هر یک از ابرموتورها از فرمول دقت استفاده گردید. برای محاسبه سایر موارد از محاسبات آماری و برای ترسیم نمودارها از نرم‌افزار اکسل استفاده گردید.

هدف از پژوهش حاضر آن است که با سنجش و بررسی تفاوت بین ابرموتورهای مورد پژوهش در زمینه بازیابی اطلاعات رشته علم اطلاعات و دانش‌شناسی، ابرموتورهای برتر در این رشته را شناسایی و رتبه‌بندی کند. این امر به کتابداران و نیز پژوهشگران کمک می‌کند تا برای پاسخگویی به سوالات خویش در امر تحقیق ابرموتور برتر را برگزینند که این امر به نوبه خود منجر به کاهش سرگردانی آنها در کار جستجو و بازیابی اطلاعات شده و این امکان را می‌دهد که بتوانند با صرف هزینه و وقت کم‌تر به نتایج مطلوب‌تری دست یابند.

سوالات پژوهش

۱. تعداد پاسخ‌های مربوط هر یک از ابرموتورهای کاوش به سوالات مرجع انتخاب چقدر است؟
 ۲. وضعیت دقت هر کدام از ابرموتورها برای پاسخگویی به سوالات به چه میزان است؟
 ۳. وضعیت ریزش کاذب در نتایج بازیابی شده توسط هر ابرموتور کاوش انتخابی چه وضعیتی دارد؟
- جامعه آماری پژوهش عبارت است از: هشت ابرموتور کاوش مطرح شامل دیروپ^۱، گوتو^۲، وب‌کراولر^۳، متاگوفر^۴، پالی‌متا^۵، متالیب^۶، کلاستی^۷ و اسک‌جیوز^۸ می‌باشد. تعداد ۱۶ سؤال اختصاصی مرجع رشته علم اطلاعات و دانش‌شناسی با پرس و جو از چند تن از اساتید این رشته در مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری انتخاب شد. این سوالات عبارتند از: ملاک بررسی ۱۰ نتیجه نخست بود کلیه جستجوها در تاریخ ۱۳۹۴/۶/۱۲ تا ۱۳۹۴/۸/۲۱ انجام گرفت.

یافته‌های پژوهش

برای پاسخ به سوال اول پس از استخراج کلیدواژه، در هریک از ابرموتورها جستجو انجام شد. قضاوت ربط مدارک توسط پژوهشگر انجام شد، بدین ترتیب که با بازکردن هر پیوند در صورتی که حاوی اطلاعات مختصر یا مفصل مربوط به سوال بود مربوط محسوب گردید. اگر صفحات بازیابی شده حاوی پیوندهایی بودند که به نظر رسید به صفحه یا صفحات مربوط ختم می‌شود حداکثر ۳ پیوند مورد بررسی قرار گرفت.

¹. Deeperweb

². Goto

³. Webcrawler

⁴. Metagopher

⁵. Polymeta

⁶. MetaLib

⁷. Clusty

⁸. Ask Jeeves

عبارت ورچوال رفرنس در مجموع بیشترین رقم را به خود اختصاص داده است و پس از آن عبارت وب معنایی و فاکسونومی بیشترین رقم را کسب کرد. برای پاسخ به سوال دوم فرمول دقت مورد استفاده قرار گرفت، زیرا در پژوهش حاضر ۱۰ نتیجه نخست مورد بررسی قرار گرفته است.

به عنوان مثال برای محاسبه دقت در ابرموتور وب کراولر تعداد موارد بازیابی شده مربوط توسط این ابرموتور برای هر سوال به ترتیب عبارتند از: ۱، ۰، ۷، ۲، ۶، ۹، ۵، ۳، ۲، ۰، ۰، ۴، ۱، ۴، ۵، ۲ که مجموع آن ۵۱ می شود. این رقم را بر عدد ثابت ۱۶۰ یعنی کل اقلام بازیابی شده توسط ابرموتور تقسیم می کنیم، آنگاه حاصل را در عدد ۱۰۰ ضرب می کنیم. عددی که بدست می آید ضریب دقت برای ابرموتور وب کراولر است.

بنابراین دقت در ابرموتور وب کراولر ۳۱/۸۷ درصد است، که در بین سایر ابرموتورها بیشترین رقم را کسب کرده است. بنابراین فرمول دقت برای سایر ابرموتورها در این پژوهش بدین ترتیب است: متاگوفر ۲۹/۳۷ درصد، ۲۸/۷۵ درصد، ۲۱/۸۷ درصد، ۲۰ درصد، ۱۴/۳۷ درصد، ۱۳/۱۲ درصد، ۱۲/۵ درصد. در نمودار ۱ ابرموتورها بر اساس دقت مرتب شده اند.

جدول ۱: کلیدواژه‌های برگزیده هر یک از سؤالات پژوهش

۱	“Open access publishing”
۲	“Virtual Reference”
۳	“Public libraries” and “Reading culture”
۴	“Harnad”
۵	“Global week of open access”
۶	“Iranian females” and “institute of scientific information” and “scientific information”
۷	“Hybrid Library” and “Digital Library”
۸	“Scientometrics” and “Thesis”
۹	“Ilisa fars” and “development of libraries”
۱۰	“Lotka law”
۱۱	“Semantic Web”
۱۲	“Rousseau’s classification system” and “Iran libraries”
۱۳	“Faxonomy”
۱۴	“RFID technology in libraries”
۱۵	“Impact Factor formula”
۱۶	“Website addresses” and “Open access papers”

جدول ۱ کلیدواژه‌های برگزیده برای هر سوال را نشان می دهد که همان گونه که مشاهده می شود تعداد ۱۶ کلیدواژه می باشد.

جدول ۲: توزیع فراوانی موارد بازیابی شده مربوط برای هر سؤال انتخابی در هر یک از ابرموتورهای کاوش

No	Query	WebCrawler	Clusty	Metalib	Ask Jeeves	Goto	Mrtagopher	PolyMeta	Deeperweb
۱	“Open access publishing”	۱	۷	۱	۳	۱	۸	۲	۵
۲	“Virtual Reference”	۰	۲	۳	۲	۲	۲	۲	۱
۳	“Public libraries” and “Reading culture”	۷	۵	۲	۴	۳	۵	۳	۲
۴	“Harnad”	۲	۴	۴	۳	۵	۴	۱	۳
۵	“Global week of open access”	۶	۳	۰	۴	۱	۶	۱	۱
۶	“Iranian females” and “institute of scientific information” and “scientific information”	۹	۸	۱	۱	۰	۳	۲	۲
۷	“Hybrid Library” and “Digital Library”	۵	۹	۱	۵	۰	۱	۳	۱
۸	“Scientometrics” and “Thesis”	۳	۱	۱	۳	۲	۱	۱	۱
۹	“Library association of fars” and “development of libraries”	۲	۱	۱	۲	۴	۱	۱	۲
۱۰	“Lotka law”	۰	۰	۱	۰	۱	۲	۰	۴
۱۱	“Semantic Web”	۰	۰	۰	۴	۱	۵	۱	۱
۱۲	“Rousseau's classification system” and “Iran libraries”	۴	۲	۰	۱	۱	۲	۱	۳
۱۳	“Faxonomy”	۱	۱	۱	۱	۲	۱	۰	۳
۱۴	“RFID technology in libraries”	۴	۱	۲	۰	۰	۳	۱	۳
۱۵	“Impact Factor formula”	۵	۱	۱	۰	۰	۱	۰	۰
۱۶	“Website addresses” and “Open access papers”	۲	۱	۱	۲	۰	۲	۲	۰
مجموع		۵۱	۴۶	۲۰	۳۵	۲۳	۴۷	۲۱	۳۲

در جدول ۲ توزیع فراوانی کلیدواژه‌های مربوط برای هر سؤال آمده است، بر این اساس همان‌گونه که در جدول بالا آمده است، بیشترین تعداد نتایج بازیابی شده مربوط ۵۱ مورد است که توسط ابرموتور وب‌کراولر بازیابی شده است. ۴۷ مورد بازیابی را ابرموتور متاگوفر داشته است. ۴۶ پاسخ مربوط را ابرموتور کلاستی، ۳۵ پاسخ مربوط را ابرموتور اسک‌جیوز، ۳۲ پاسخ مربوط را ابرموتور دیپروپ، ۲۳ پاسخ مربوط را ابرموتور گوتو، ۲۱ پاسخ مربوط را ابرموتور پالی‌متا و ۲۰ پاسخ مربوط را ابرموتور متالایب

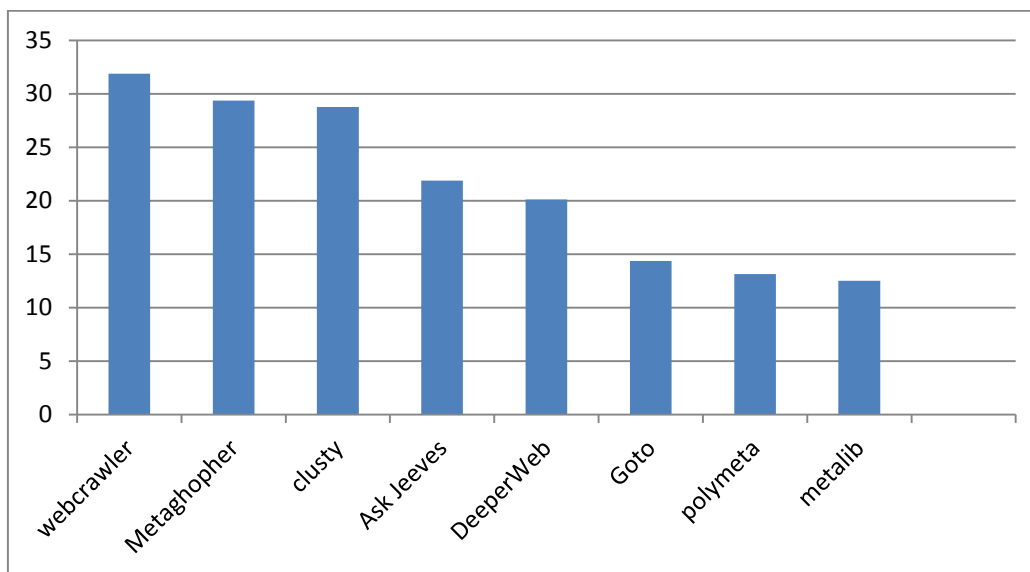
داشته است. بیشترین مقدار بازیابی مربوط با توجه به جدول برای هر سوال ۹ مورد بوده است که این تعداد را ابرموتور وب کراولر برای سوال با کلید واژه فاکسونومی^۱ و اسک جیوز و متاگوفر برای سوال عبارتی وب معنایی^۲ بازیابی کرده‌اند.

جدول ۳: توزیع فراوانی ریزش کاذب برای هر سؤال انتخابی در هر یک از ابرموتورهای کاوش

No	Query	WebCrawler	Clusty	Metlib	Ask Jeeves	Goto	Mirtagopher	PolyMeta	Deeperweb
۱	“Open access publishing”	۷	۱۰	۱۰	۹	۱۰	۱۰	۱۰	۱۰
۲	“Virtual Reference”	۶	۸	۱۰	۱۰	۹	۹	۱۰	۹
۳	“Public libraries” and “Reading culture”	۷	۷	۱۰	۱۰	۱۰	۸	۹	۸
۴	“Harnad”	۵	۱۰	۶	۵	۱۰	۷	۸	۷
۵	“Global week of open access”	۹	۹	۸	۵	۱۰	۵	۱۰	۱۰
۶	“Iranian females” and “institute of scientific information” and “scientific information”	۸	۷	۱۰	۱۰	۸	۹	۱۰	۹
۷	“Hybrid Library” and “Digital Library”	۱۰	۵	۱۰	۸	۹	۶	۱۰	۶
۸	“Scientometrics” and “Thesis”	۱۰	۹	۱۰	۹	۱۰	۱۰	۸	۸
۹	“Library association of fars” and “development of libraries”	۸	۸	۱۰	۱۰	۹	۹	۹	۱۰
۱۰	“Lotka law”	۷	۱۰	۱۰	۹	۱۰	۴	۹	۸
۱۱	“Semantic Web”	۶	۶	۹	۸	۸	۱۰	۸	۶
۱۲	“Rousseau's classification system” and “Iran libraries”	۹	۷	۹	۱۰	۹	۱۰	۱۰	۱۰
۱۳	“Faxonomy”	۹	۹	۹	۱۰	۱۰	۷	۸	۴
۱۴	“RFID technology in libraries”	۷	۵	۱۰	۴	۹	۳	۹	۹
۱۵	“Impact Factor formula”	۶	۹	۹	۷	۷	۱۰	۱۰	۸
۱۶	“Website addresses” and “Open access papers”	۱۰	۱۰	۹	۶	۶	۱۰	۹	۱۰
مجموع		۱۲۴	۱۲۹	۱۴۹	۱۳۰	۱۴۴	۱۲۷	۱۴۷	۱۳۲

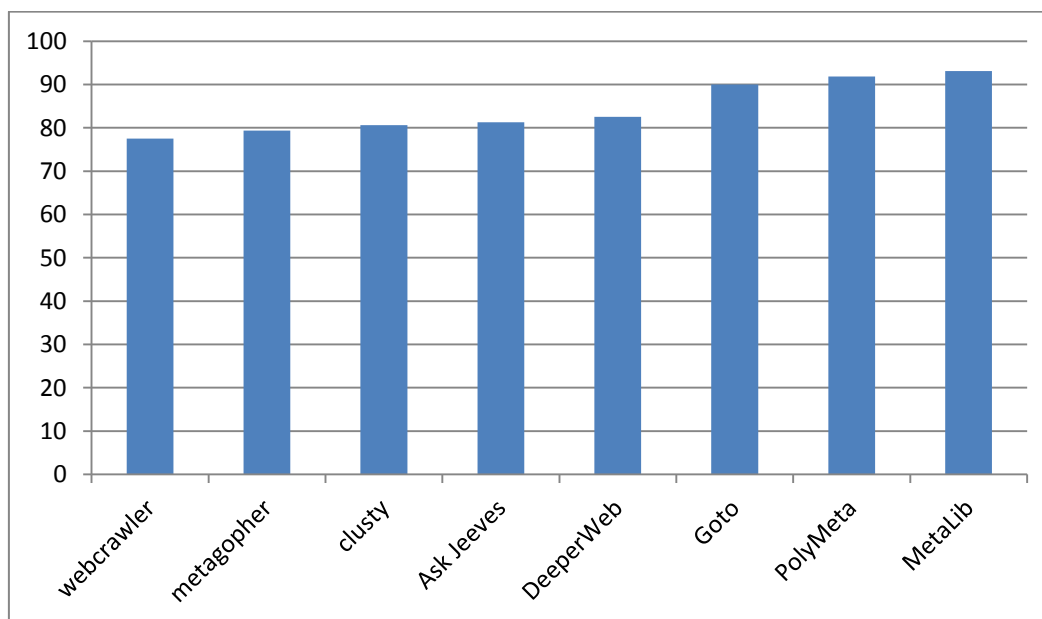
^۱. Faxonomy

^۲. Semantic Web



نمودار ۱: دقت بازیافت در ابرموتورهای جستجوی انتخابی

ریزش کاذب نیز به معنای بازیافت ناخواسته است. بنابراین هر پاسخی که به شکلی مرتبط با نیاز کاربر نباشد ریزش کاذب محسوب می‌شود. برای پاسخ به سوال سوم ۵ نوع ریزش کاذب مشخص شد، که عبارتند از: پاسخ‌های نامربوط به لحاظ محتوا، پیوندهای مرده، پاسخ‌های مربوط و نامربوط تکراری، پاسخ‌های بازیابی نشده و پاسخ‌هایی که به زبان غیرانگلیسی هستند. همان‌گونه که در جدول زیر مشاهده می‌کنیم، ابرموتور کاوش متالایب با ۱۴۹ مورد ریزش کاذب که معادل ۹۳/۱۲ درصد است، بیشترین میزان ریزش کاذب را داشته است. ابرموتور پالی‌متا با ۱۴۷ مورد (۹۱/۸۷ درصد)، ابرموتور متاگوفر با ۱۲۷ مورد (۷۹/۳۷ درصد)، ابرموتور دیپروب با ۱۳۲ مورد (۸۲/۵ درصد)، ابرموتور وب‌کراولر با ۱۲۴ مورد (۷۷/۵ درصد)، ابرموتور اسک‌جیوز با ۱۳۰ مورد (۸۱/۲۵ درصد)، ابرموتور گوتو با ۱۴۴ مورد (۹۰ درصد)، و کلاستی با ۱۲۹ مورد (۸۰/۶۲ درصد) ریزش کاذب داشته است. در نمودار زیر ابرموتورها بر اساس کم‌ترین ریزش کاذب مرتب شده‌اند.

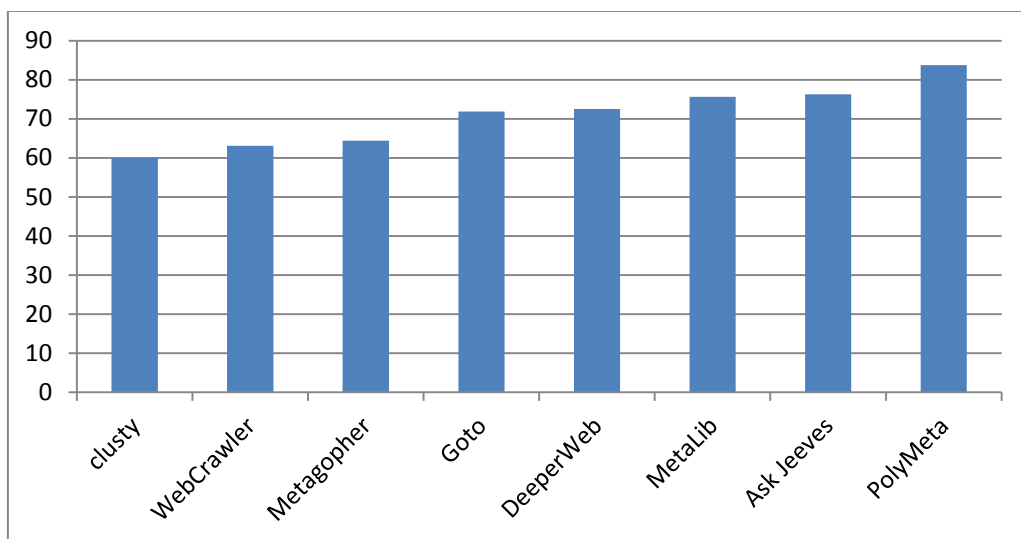


نمودار ۲: ریزش کاذب در هر یک از ابرموتورهای کاوش

میزان انواع ریزش کاذب برای هر ابرموتور:

۱. بازیابی پاسخ نامربوط از نظر محتوا

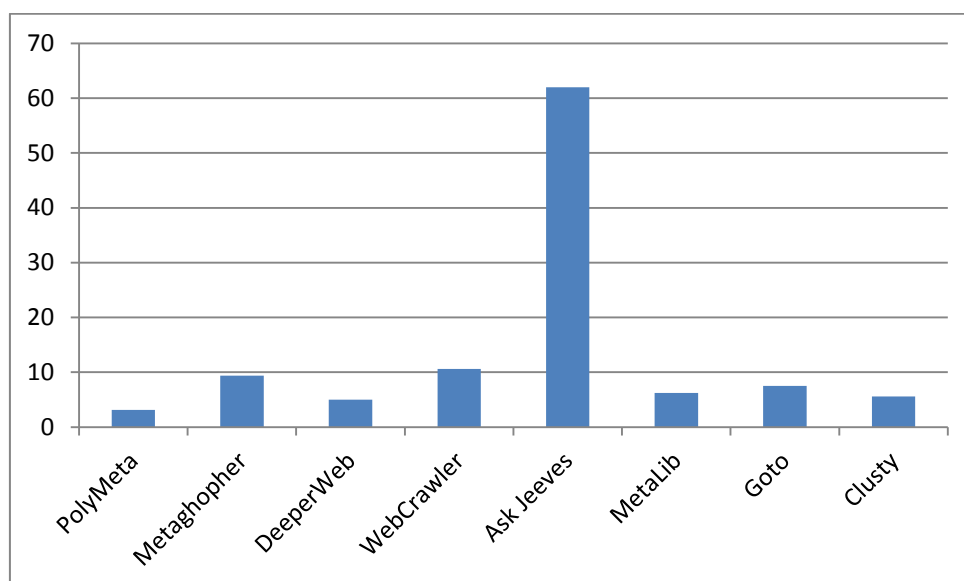
از ۱۰۸۲ مورد ریزش کاذب که در این پژوهش بدست آمده، ۹۰۸ مورد آن به ریزش کاذب محتوایی مربوط می‌شود که در بین انواع دیگر ریزش کاذب دارای بیشترین مقدار بوده است. مقدار محاسبه شده برای هر ابرموتور کاوش عبارت است از: پالی‌متا با ۱۳۴ مورد (۸۳/۷۵ درصد)، متاگوفر با ۱۰۳ مورد (۶۴/۳۷ درصد)، دیپروب با ۱۱۶ مورد (۷۲/۵ درصد)، وب‌کراولر با ۱۰۱ مورد (۶۳/۱۲ درصد)، اسک‌جیوز با ۱۲۲ مورد (۷۶/۲۵ درصد)، متالایب با ۱۲۱ مورد (۷۵/۶۲ درصد)، گوتو با ۱۱۵ مورد (۷۱/۸۷ درصد)، کلاستی با ۹۶ مورد (۶۰ درصد). نمودار زیر، ابرموتورها را به ترتیب نزولی از لحاظ بازیابی موارد نامربوط محتوایی نشان می‌دهد.



نمودار ۳: درصد موارد نامربوط از نظر محتوایی

۲. بازیابی پیوندهای مرده

از ۱۰۸۲ مورد ریزش کاذب، پیوندهای مرده ۶۵ مورد را شامل می‌شود. در زمان جستجو اگر بازکردن پیوند بازیابی شده با پیغام خطای ۴۰۴ مواجه شده بود، پیوند مرده محسوب گردید. میزان پیوندهای مرده در ابرموتورها بدین نحو گزارش می‌شود: پالی‌متا ۵ مورد (۳/۱۲ درصد)، متاگوفر ۱۵ مورد (۹/۳۷ درصد)، دیپروب ۸ مورد (۵ درصد)، وب‌کراولر ۱۷ مورد (۱۰/۶۲ درصد)، اسک-جیوز ۱ مورد (۶۲ درصد)، متالایب ۱۰ مورد (۶/۲۵ درصد)، گوتو ۱۲ مورد (۷/۵ درصد) و کلاستی ۹ مورد (۵/۶۲ درصد). نمودار زیر میزان پیوندهای مرده توسط ابرموتورها را نشان می‌دهد.



نمودار ۴: پیوندهای مرده در ابرموتورهای کاوش

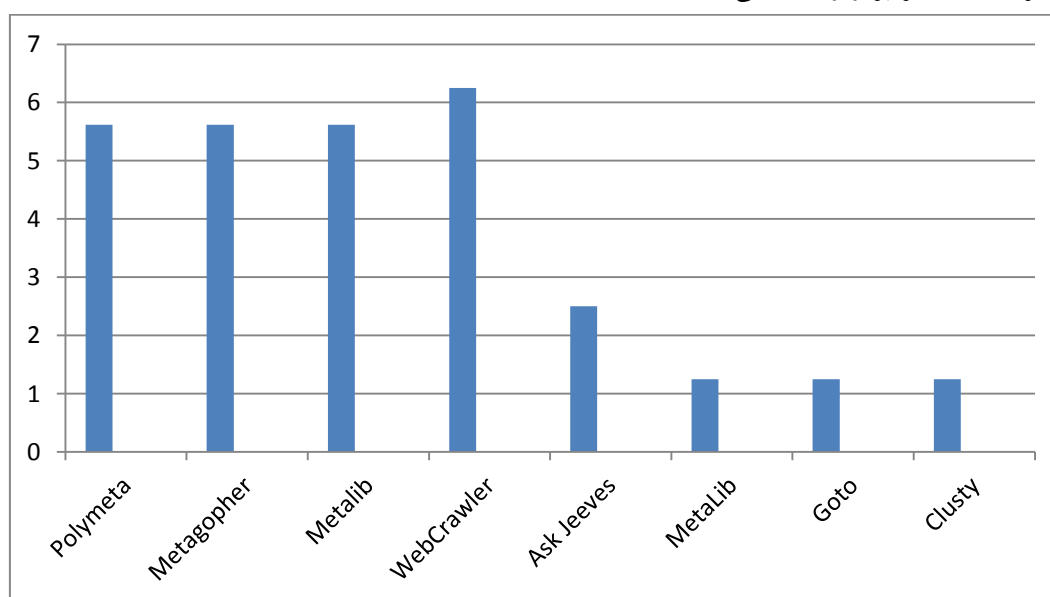
۲. بازیابی صفحه تکراری

موارد تکراری بازیابی شده به دو گروه تقسیم می‌شوند: الف) موارد تکراری مربوط ب) موارد تکراری نامربوط

۱-۳ بازیابی موارد تکراری مربوط

از ۱۰۸۲ مورد ریزش کاذب، موارد تکراری و مربوط ۴۷ مورد است. مواردی که در این پژوهش تکراری و مربوط هستند دو دسته هستند. گروه اول آنهایی که در همان مرحله نخست تکراری بودند و حالت دوم آنهایی که پس از حداکثر ۳ پیوند مربوط و تکراری ارزیابی شدند. بازیابی موارد تکراری و مربوط در ابرموتورها به قرار زیر است:

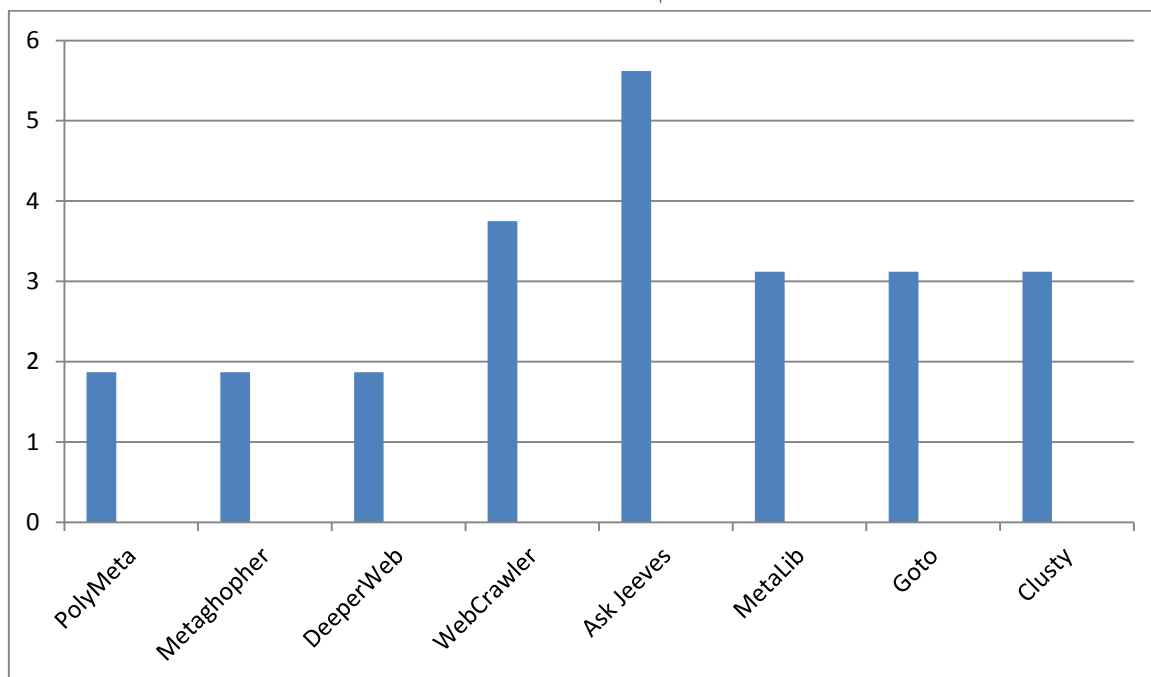
ابرموتورهای پالی‌متا، متاگوفر و دیپروپ هر کدام ۹ مورد (۵/۶۲ درصد)، ابرموتور وب‌کراولر ۱۰ مورد (۶/۲۵ درصد)، اسک-جیوز ۴ مورد (۲/۵ درصد) و ابرموتورهای متالایب، گوتو و کلاستی هر کدام ۲ مورد (۱/۲۵ درصد). نمودار زیر میزان موارد مربوط و تکراری را در هر ابرموتور نشان می‌دهد.



نمودار ۵: موارد تکراری و مربوط در هر ابرموتور کاوش

۲-۳ بازیابی موارد تکراری و نامربوط

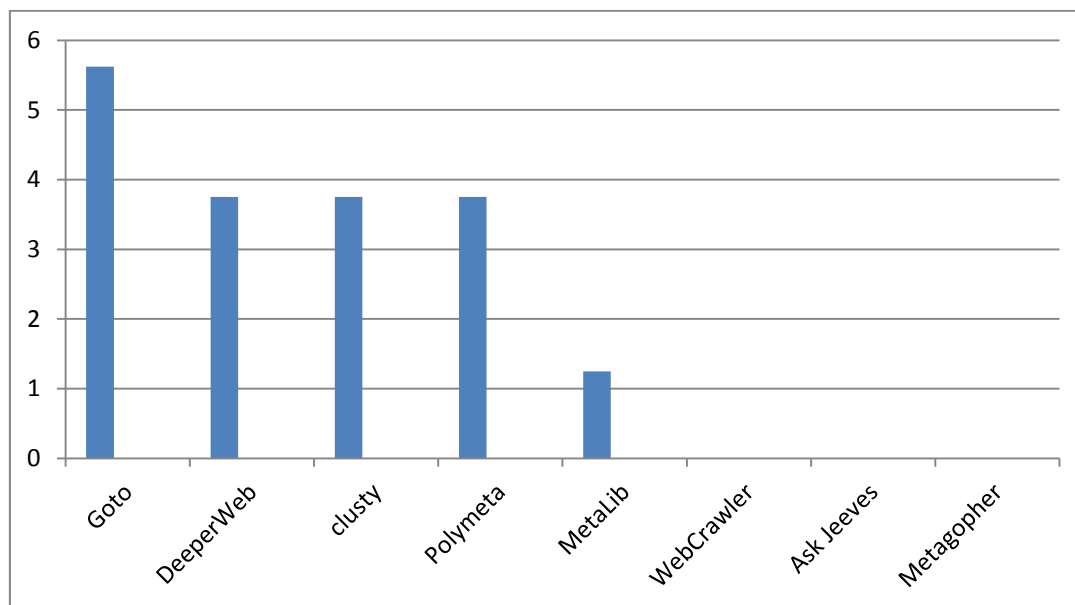
از ۱۰۸۲ مورد ریزش کاذب، ۳۹ مورد، موارد نامربوط و تکراری هستند. تعداد این موارد در هر یک از ابرموتورها به قرار زیر است: پالی‌متا، متاگوفر و دیپروب هر کدام ۳ مورد (۱/۸۷ درصد)، وب‌کراولر ۶ مورد (۳/۷۵ درصد)، اسک‌جیوز ۹ مورد (۵/۶۲ درصد) و ابرموتورهای متالایب، گوتو و کلاستی هر کدام ۵ مورد (۳/۱۲ درصد). نمودار زیر میزان این موارد را نشان می‌دهد.



نمودار ۶: موارد تکراری و نامربوط در هر ابرموتور کاوش

۱. ارقام بازیابی نشده

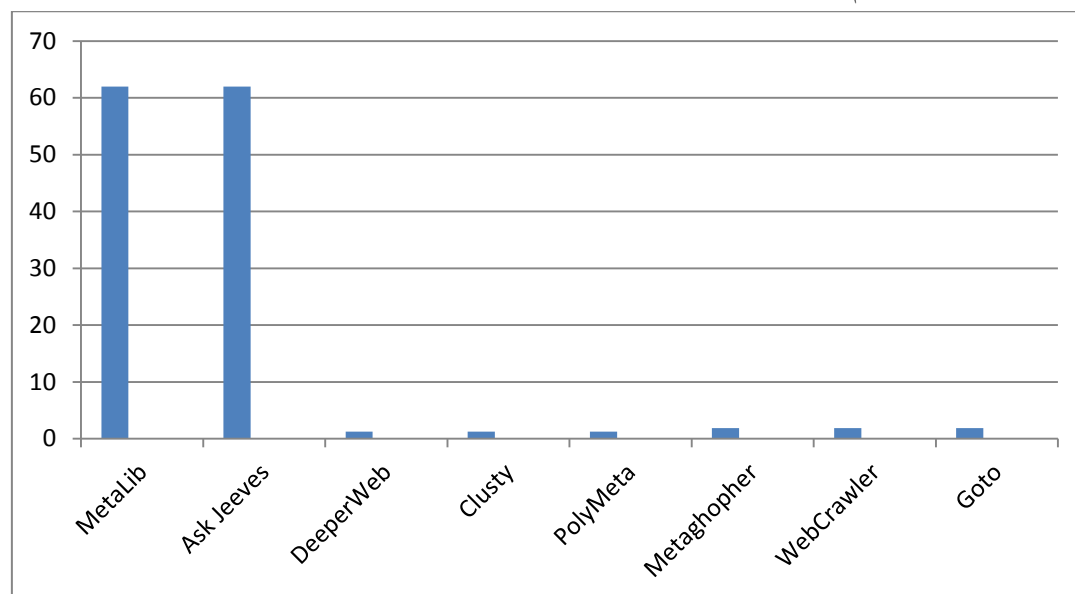
از ۱۰۸۲ مورد ریزش کاذب در این پژوهش، ۲۹ مورد مربوط به مواردی است که بازیابی نشده‌اند، بدین معنا که هر گاه ابرموتوری ۱۰ مورد کامل بازیابی نداشته است، تعداد موارد بازیابی نشده برای آن منظور شده است. تعداد و درصد این موارد در هر ابرموتور نیز عبارتند از: گوتو ۹ مورد (۵/۶۲ درصد)، دیپروب، کلاستی و پالی‌متا هر کدام ۶ مورد (۳/۷۵ درصد) و متالایب ۲ مورد (۱/۲۵ درصد) و ابرموتورهای اسک‌جیوز، وب‌کراولر و متاگوفر نیز هیچ‌کدام دارای موارد بازیابی نشده نبوده‌اند. نمودار زیر میزان این موارد را در ابرموتورها نشان می‌دهد.



نمودار ۷: موارد بازیابی نشده در هر ابرموتور کاوش

۲. بازیابی موارد غیرانگلیسی

از ۱۰۸۲ مورد ریزش کاذب در این پژوهش، ۱۷ مورد شامل موارد غیرانگلیسی زبان بودند که آنها نیز به دلیل اینکه برای اکثریت کاربران غیر قابل استفاده‌اند، جزء ریزش کاذب محسوب می‌شوند. تعداد و درصد این موارد نیز به قرار زیر است: متالایب و اسک‌جیوز هر کدام ۱ مورد (۶۲ درصد)، دیپروبو، کلاستی و پالی‌متا هر کدام ۲ مورد (۱/۲۵ درصد)، و ابرموتورهای وب-کراولر، متاگوفر و گوتو هر کدام ۳ مورد (۱/۸۷ درصد). نمودار زیر میزان موارد غیرانگلیسی را در ابرموتورها نشانی می‌دهد.



نمودار ۸: موارد غیرانگلیسی بازیابی شده در هر کدام از ابرموتورهای کاوش

آزمون سؤالات پژوهش

سوال اول: تعداد پاسخ‌های مربوط هر یک از ابرموتورهای کاوش به سؤالات مرجع انتخاب چقدر است؟ برای آزمون فراوانی تعداد پاسخ‌های مربوط از آزمون کای اسکوئر (خی دو) استفاده گردید.

جدول ۴: آزمون کاسکوئر (خی دو) برای فروانی تعداد پاسخ‌های مربوط

ابرموتورهای کاوش	مقدار خی دو	درجه آزادی	سطح معنی‌داری
Webcrawler	۳,۱۲	۸	.۹۲۶
Metalib	۱۲,۱۲	۸	.۱۴۶
Clusty	۱۴	۴	.۰۰۷
Askjeeves	۱,۲۵	۵	.۹۴۰
Goto	۷,۲۵	۵	.۲۰۳
Metagopher	۶,۷۵	۶	.۳۴۵
Polymeta	۳,۵۰	۳	.۳۲۱
Deeperweb	۵	۵	.۴۱۶

بر اساس نتایج آزمون کای اسکوئر (خی دو) و بر اساس مقدار آماره خی دو برای هر ابرموتور و نیز سطح معنی‌داری تنها ابرموتور Clusty است که با مقدار خی دو (۱۴) و سطح معنی‌داری ۰,۰۷ که از ۰,۰۵ کم‌تر است اختلاف معنی‌داری به لحاظ تعداد پاسخ‌های مربوط با سایر ابرموتورها دارد. سایر ابرموتورها چون دارای سطح معنی‌داری بیشتر از ۰,۰۵ هستند اختلاف معنی‌داری با یکدیگر ندارند.

برای سوال دوم نیز که در رابطه با وضعیت دقت هر کدام از ابرموتورها برای پاسخگویی به سؤالات بود، تعداد موارد بازیابی شده مربوط را تقسیم بر تعداد کل موارد بازیابی شده کرده و در ۱۰۰ ضرب کردیم و بر این اساس میزان دقت در هر ابرموتور بدست آمد.

سوال سوم: ریزش کاذب در نتایج بازیابی شده توسط هر ابرموتور کاوش انتخابی چه وضعیتی دارد؟
برای آزمون ریزش کاذب از آزمون کای اسکوئر (خی دو) استفاده گردید.

جدول ۵: آزمون کاسکوئر (خی دو) برای فروانی میزان ریزش کاذب

ابرموتورهای کاوش	مقدار خی دو	درجه آزادی	سطح معنی‌داری
Webcrawler	۲	۵	.۸۴
Metalib	۲,۷۵	۵	.۷۳
Clusty	۱۱	۳	.۰,۰۱
Askjeeves	۸,۵۰	۶	.۲۰
Goto	۹	۴	.۰,۰۶
Metagopher	۱۱	۷	.۱۳
Polymeta	۸۷	۲	.۶۴
Deeperweb	۵	۵	.۴۱

بر اساس نتایج آزمون کای اسکوئر (خی دو) و بر اساس مقدار آماره خی دو برای هر ابرموتور و نیز سطح معنی‌داری تنها ابرموتور Clusty است که با مقدار خی دو (۱۱) و سطح معنی‌داری ۰,۰۱ که از ۰,۰۵ کم‌تر است اختلاف معنی‌داری به لحاظ

میزان ریزش کاذب با سایر ابرموتورها دارد. سایر ابرموتورها چون دارای سطح معنی‌داری بیشتر از ۰,۰۵ هستند اختلاف معنی‌داری به این لحاظ با یکدیگر ندارند.

بحث و نتیجه‌گیری

جمعا ۱۲۸۰ نتیجه بازیابی شده توسط ابرموتورها مورد بررسی قرار گرفت که از بین این تعداد، ۲۷۶ مورد آن مربوط بوده است. ابرموتور جستجوی وب کراولر بیشترین میزان پاسخگویی را با ۵۱ مورد مربوط داشته است. پس از آن ابرموتورهای متاگوفر با ۴۷ مورد و کلاستی با ۴۶ مورد به ترتیب در رتبه‌های دوم و سوم قرار گرفتند و سپس ابرموتورهای اسک‌جیوز، دیپروب، گوتو، پالی‌متا و متالایب در جایگاه‌های بعدی قرار می‌گیرند. با استفاده از فرمول ضریب دقت، میزان دقت در تمامی ابرموتورها تعیین گردید. ابرموتور جستجوی وب کراولر دارای بیشترین میزان دقت (۳۱/۸۷ درصد) و ابرموتور جستجوی متالایب دارای کم‌ترین میزان دقت (۱۲/۵ درصد) می‌باشند. براساس نتایج بدست‌آمده، بیشترین میزان ریزش کاذب را موارد نامربوط محتوایی به خود اختصاص داده‌اند، و پس از آن پیوندهای مرده دارای بیشترین میزان ریزش کاذب بوده است. سایر موارد ریزش کاذب به ترتیب رتبه عبارتند از: موارد مربوط تکراری، موارد نامربوط تکراری، موارد بازیابی نشده و دست‌آورد غیرانگلیسی. همچنین، به‌طور کلی ۱۰۸۲ مورد از نتایج بازیابی شده توسط ابرموتورها ریزش کاذب بوده است، که از این میان بیشترین مقدار ریزش کاذب مربوط به ابرموتور متالایب (۱۴۹ مورد) که برابر با (۹۳/۱۲ درصد) است، می‌باشد و بنابراین ابرموتور مناسبی برای بازیابی سوالات مرجع رشته علم اطلاعات و دانش‌شناسی نیست و در اولویت آخر قرار می‌گیرد. کم‌ترین میزان ریزش کاذب مربوط به ابرموتور وب کراولر (۱۲۴ مورد) است که برابر با ۷۷/۵ درصد می‌باشد و بنابراین ابرموتور مناسبی برای پاسخ‌گویی به سوالات علم اطلاعات و دانش‌شناسی است. ترتیب سایر ابرموتورها از نظر بازیابی کم‌ترین میزان ریزش کاذب (از راست به چپ) عبارتند از: وب کراولر، متاگوفر، کلاستی، اسک‌جیوز، دیپروب، گوتو، پالی‌متا و متالایب.

۱. از سؤال اول این نتیجه بدست می‌آید که ابرموتورهای جستجو قادر به پاسخگویی به سوالات رشته علم اطلاعات و دانش‌شناسی در حد مطلوب نیستند. این یافته با یافته‌های پژوهش‌های پیشین که آنها نیز بیانگر این مطلب هستند که ابرموتورهای جستجو در هر رشته قادر به بازیابی اطلاعات رشته تخصصی در حد مطلوب نیستند همسو می‌باشد. (Thelwall, ۲۰۰۰; Bharat & Broder, ۱۹۹۸; Tongchim, Sornlertlamvanich, Isahara, King, ۲۰۰۷)

۲. از سؤال دوم این نتیجه استنباط می‌شود که ابرموتورهای جستجو به لحاظ دقت در پاسخگویی به سوالات علم اطلاعات و دانش‌شناسی وضعیت یکسان و ثابتی ندارند. و نتایج نشان می‌دهد که ابرموتورهای پژوهش به ترتیب اولویت از نظر میزان دقت (از راست به چپ) عبارتند از: وب کراولر، متاگوفر، کلاستی، اسک‌جیوز، دیپروب، گوتو، پالی‌متا و متالایب. این یافته با یافته‌های پژوهش‌های پیشین که عدم ثبات ابرموتورها را در بازیابی اطلاعات به لحاظ دقت نشان می‌دهند همسو می‌باشد. (صراطی شیرازی, Veronis, ۲۰۰۶; محمداسماعیل, لفظ قاضی, گیلوری, ۱۳۸۷; نبوی, ۱۳۸۰)

۳. از سؤال سوم چنین نتیجه‌گیری می‌شود که ابرموتورهای جستجوی انتخابی به‌طور کلی به گونه‌ای متفاوت عمل می‌کنند، هر چند در برخی موارد عملکرد یکسانی داشته‌اند و لیکن این برابری در همه جا دیده نمی‌شود. ابرموتور وب کراولر در بین ابرموتورهای مورد بررسی بهترین عملکرد را در پاسخگویی به سوالات اختصاصی علم اطلاعات و دانش‌شناسی دارد. پژوهش‌های زیر همگی بر این مطلب صحنه می‌گذارند (Ilic, Bessel, Silagy, Green, ۲۰۰۳; King, ۱۳۸۲; Veronis, ۲۰۰۶; صراطی شیرازی, ۱۳۸۸; Tongchim, Sornlertlamvanich, Isahara, ۲۰۰۷)

به‌طور کلی از پژوهش حاضر چنین بر می‌آید که ابرموتورهای کاوش بدلیل آنکه کار جستجو را در چندین موتور جستجو انجام می‌دهند همواره متغیر بوده و دارای وضعیت ثابت و یکسانی در کار بازیابی اطلاعات نیستند. روز به روز بر محتوای علمی

اینترنت افزوده می‌شود و میزان مطالب علمی آن که نیاز به ارزشیابی دارد نیز رو به افزایش است، این امر آشنایی هر چه بیشتر کتابداران و پژوهشگران با ابزارهای جستجو و در کنار آن تسلط بر زبان انگلیسی را می‌طلبد.

پیشنادهای کاربردی پژوهش

۱. بدلیل کاربرد زیاد ابرموتورهای کاوش در بازیابی اطلاعات پیشنهاد می‌شود، در حوزه‌های موضوعی دیگر به غیر از علم اطلاعات و دانش‌شناسی نیز این پژوهش انجام شود.
۲. پیشنهاد می‌شود پژوهشی در خصوص هم‌پوشانی نتایج بازیابی‌شده از ابرموتورهای کاوش در حوزه‌های موضوعی گوناگون انجام شود.
۳. پیشنهاد می‌شود پژوهش‌های آتی به بررسی و ارزیابی نتایج بازیابی‌شده از ابرموتورهای کاوشی که به‌طور خاص تنها متعلق به یک حوزه موضوعی هستند پردازند.

منابع

- صراطی شیرازی، منصوره (۱۳۸۸). بررسی مقایسه‌های میزان دقت موتورهای کاوش عمومی و تخصصی پزشکی در بازیابی مدارک مربوط به بیماری‌های کودکان. پایان‌نامه کارشناسی ارشد، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه شهید چمران، اهواز.
- کمیحانی، احمد (۱۳۸۲) مقایسه کارایی موتورهای کاوش عمومی و تخصصی وب در بازیابی اطلاعات کشاورزی. اطلاع‌شناسی، ۱(۳)، ۱۲۷-۱۴۴.
- محمداسماعیل، صدیقه؛ لفظی قاضی، الهام؛ گیلوری، عباس (۱۳۸۷). مقایسه موتورهای کاوش عمومی و ابرموتورهای کاوش عمومی در بازیابی اطلاعات داروشناسی و تعیین میزان هم‌پوشانی آنها. مدیریت اطلاعات سلامت ۵(۲)، ۱۲۱-۱۲۹.
- میرحسینی، زهره و بابایی، الهام (۱۳۹۱). بررسی مقایسه‌ای جامعیت و مانعیت موتورهای کاوش تخصصی پزشکی در بازیابی اطلاعات مربوط به بیماری‌های زنان و مامایی. نظام‌ها و خدمات اطلاعاتی ۱(۲)، ۴۵-۵۷.
- حریری، نجلا؛ امامی، زهرا؛ ملک، مجتبی (۱۳۹۴). دقت و جامعیت موتورهای جستجوی عمومی در بازیابی تصاویر بیماری‌های مهم غدد درون‌ریز. مجله‌ی غدد درون‌ریز و متابولیسم ایران، ۱۷(۲)، ۹۷-۱۰۴.
- اسدی قادیکلایی، ام‌البنین؛ اسدی، سعید؛ نوروزی، عبدالرضا؛ احسانی، روح‌الله (۱۳۹۳). مقایسه دقت موتورهای جستجوی عمومی و پایگاه‌های تخصصی در بازیابی تصاویر رادیولوژی. مدیریت بهداشت و درمان، ۵(۲)، ۷۷-۸۷.
- محمداسماعیل، صدیقه و منصورکیایی، ربابه (۱۳۸۸). مقایسه موتورها و ابرموتورهای کاوش عمومی در بازیابی اطلاعات علم فیزیک و میزان هم‌پوشانی آنها. مجله مطالعات ملی کتابداری و سازماندهی اطلاعات، ۲۲(۳)، ۱۳۱-۱۴۰.
- محمداسماعیل، صدیقه و فیروزی، صغری (۱۳۸۸). مقایسه میزان هم‌پوشانی نتایج بازیابی‌شده در موتورهای کاوش و ابرموتورهای کاوش در بازیابی اطلاعات نانو تکنولوژی. ماهنامه اطلاع‌یابی و اطلاع‌رسانی، ۲۳، ۱۷-۲۴.
- نبوی، فاطمه (۱۳۸۰). مطالعه مقایسه‌ای ابرموتورهای جستجو در بازیابی اطلاعات کتابداری و اطلاع‌رسانی از شبکه جهانی وب. پایان‌نامه کارشناسی ارشد، دانشکده مدیریت و اطلاع‌رسانی پزشکی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی ایران.
- وزیرپورکشمیری، مهرداد و فزونی، بهاره (۱۳۸۹). مقایسه میزان دقت راهنماهای موضوعی وب در بازیابی اطلاعات فنی-مهندسی. فصلنامه پردازش و مدیریت اطلاعات، ۲۷(۳)، ۵۹۷-۶۱۶.

- مهرداد، جعفر و بیشش، مژگان (۱۳۸۶). ارزیابی سایت‌های وب بازیابی‌شده توسط دو راهنمای شبکه وب (Yahoo و ODP) در حوزه علوم کتابداری و اطلاع‌رسانی بر اساس معیارهای محتوا و کارایی. کتابداری و اطلاع‌رسانی، ۱(۱۰)، ۴۱-۶۱.
- رضوانی گیل کلایی، شهلا (۱۳۸۳). مقایسه ۸ ابرموتور جستجوی واقعی اینترنت در پاسخگویی به سؤالات اختصاصی رشته کتابداری و اطلاع‌رسانی. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشگاه شیراز.
- کوشا، کیوان (۱۳۸۱). ابزارهای کاوش اینترنت: اصول، مهارت‌ها و امکانات جستجو در وب. تهران: نشر کتابدار
- Bar-Ilan, Judit (2005). Comparing ranking of search result on the web search. *Information Processing & Management*, 41(6):1511-1519.
- Bharat, Krishna ; Broder, Andrei (1998). A technique for measuring the relative size and overlap of public web search engines. *Computer Networks & Systems*, (30), 1-7.
- Ilic, D., Bessel, T. L., Silagy, C. A., Green, S. (2003). Specialized medical search- engines are not better than General search engine in sourcing consumer information about androgen deficiency. *Human Reproduction*, 18 (3), 557-567. Retrieved from <http://humrep.oxfordjournals.org/cgi/content/full/18/3/557>.
- Isfandyari-Moghadam, Alireza; Pariookh, Mehri (2006). A comparative study on overlapping of search results in metasearch engines and their common underlying search engines. *Libraryreview*, 55(5), 301-306.
- Kaptien, R., and J. Kamps. (2009). Web directories as topical context. *Archives and Information studies*. <http://remote.science.uva.nl/~kamps/publications/2009/kaptweb09.pdf>
- King, David (2007). "Specialized search engine: Alternative to the bigguys". Retrieved May, 2007, from: <http://www.infotoday.com/online/0L2000>
- Massmann, S., and E. Rahm. (2008). Evaluating instance-based matching of web directories. Canada: International Workshop on web and database. <http://citeseerx.ist.psu.edu>.
- Shafi, S. M, Rather, Rafiq A. (2005). Precision and recall of five search engines for retrieval of scholarly Information in the field of Biotechnology. *Webology*, 2(2), 42-47. Retrieved from <http://www.webology.ir/2005/v2n2/toc.html>.
- Spink, Amanda; et al. (2006). Overlap among major web search engines. *Internet Research*, 16(9), 419-426.
- Spoerri, Anselm (2007). Using the structure of overlap between search results to rank retrieval systems without relevance judgment. *Information Processing & Management*, 43(4), 1059-1070. Retrieved from <http://www.scindirect.com>
- Thelwall, Mike (2000). Web impact factors and search engine coverage. *Journal documentation*, 56(2):185-189.
- Tongchim, Shisanu; Sornlertlamvanich, Virach; Isahara, Hitoshi (2007). Improving search performance: A lesson learned from evaluating search engines using Thai queries. *IEICE TRANS. INF. & SYST*, vol. E90-D. no. 10
- Veronis J. A (2006) comparative study of six search engines. Retrieved from <http://www.sites.univ-provence.fr/veronis/pdf/2006-comparative-study.pdf>