

میزان هم‌پوشانی نتایج بازیابی شده‌ی کلیدواژه‌های تخصصی پزشکی در موتورهای کاوش عمومی وب*

علیرضا اسفندیاری مقدم^۱، زهره بهاری موفق^۲

چکیده

مقدمه: اطلاع از میزان هم‌پوشانی نتایج بازیابی شده در ابزارهای بازیابی اطلاعات در جلوگیری از اتلاف وقت کاربران، امری مهم تلقی می‌گردد و سهم اطلاعات پزشکی - بهداشتی در میان انبوه اطلاعات موجود در وب از جایگاه ویژه‌ای برخوردار است. از این رو، پژوهش حاضر بر آن بود که میزان هم‌پوشانی نتایج بازیابی شده‌ی اطلاعات پزشکی از طریق کلیدواژه‌های تخصصی برگرفته از سرعنوان‌های موضوعی پزشکی مش (MESH یا Medical subject headings) را در موتورهای کاوش عمومی وب مورد بررسی قرار دهد.

روش بررسی: تحقیق حاضر، کاربردی و از نوع پیمایشی بود که با رویکرد توصیفی - تطبیقی در سال ۱۳۸۹ انجام شد. جامعه‌ی مورد بررسی در این پژوهش را چهار موتور کاوش عمومی وب یعنی Yahoo، Live search، Google و Ask تشکیل دادند. به منظور تعیین میزان هم‌پوشانی نتایج بازیابی شده در این چهار موتور کاوش، ابتدا ده کلیدواژه یا عبارت پزشکی از سرعنوان‌های پزشکی به صورت کاملاً تصادفی انتخاب گردید و سپس، ده نتیجه‌ی اول هر موتور کاوش به صورت سیاه‌های، لیست شدند و پس از بررسی، با استفاده از نرم‌افزار Excel، میزان اشتراکات رکوردهای بازیابی شده در موتورهای محاسبه و به عنوان میزان هم‌پوشانی در نظر گرفته شد. برای تعیین هم‌پوشانی نتایج بازیابی شده‌ی کلیدواژه‌های تخصصی پزشکی در چهار موتور کاوش، ده کلیدواژه و عبارت مشترک در آن‌ها مورد کاوش قرار گرفتند. سپس، ده نتیجه‌ی اول بازیابی شده در صفحه‌ی نخست موتورهای کاوش، در یک سیاهه، گردآوری و با تطبیق و بررسی نتایج به دست آمده و استفاده از یک معادله‌ی محقق ساخته، میزان هم‌پوشانی نتایج بازیابی شده محاسبه شد. قابل ذکر است روایی آن بر اساس مشاهدات پژوهشگران، مشورت با برخی استادان صاحب‌نظر در این حوزه و به ویژه استفاده از متون مرتبط مورد تأیید قرار گرفت. همچنین با توجه به رویکرد توصیفی - تطبیقی اتخاذ شده در این پژوهش، آمار توصیفی با محوریت یک معادله‌ی آماری مورد استفاده قرار گرفت.

یافته‌ها: موتورهای کاوش در بازیابی کلیدواژه‌های پزشکی در نخستین صفحه، ۴۸ درصد نتایج بازیابی شده را در یک موتور بازیابی می‌نمودند که نیمی از مرتبط‌ترین نتایج را شامل می‌شد. موتور کاوش Yahoo دارای بیشترین نتیجه‌ی بود که تنها در خود آن بازیابی شده و موتور کاوش Google با کمترین نتیجه‌ی منحصر به خود، از بالاترین میزان هم‌پوشانی با سایر موتورها در نتایج برخوردار بود. میزان هم‌پوشانی نتایج بازیابی شده‌ی کلیدواژه‌های پزشکی برگرفته از مش در صفحه‌ی نخست موتورهای کاوش Google، Ask، Live و Yahoo به میزان ۱۱ درصد بود.

نتیجه‌گیری: موتورهای کاوش دارای هم‌پوشانی کمی نسبت به هم بودند. از این رو به کاربران پیشنهاد می‌گردد در هنگام کاوش اطلاعات پزشکی در دنیای اینترنت، جست‌وجوی خود را در چند موتور کاوش پیگیری کنند تا به مدارک مرتبط از میان انبوه منابع موجود در وب، دسترسی یابند. از طرف دیگر، به کاربران توصیه می‌شود در جست‌وجوی اطلاعات پزشکی از موتورهای جست‌وجوی تخصصی نیز استفاده کنند.

واژه‌های کلیدی: اینترنت؛ بازیابی اطلاعات؛ کلیدواژه؛ موتورهای کاوش.

نوع مقاله: پژوهشی

دریافت مقاله: ۱۸/۱۰/۱۹ اصلاح نهایی: ۳/۵/۹۰

پدیرش مقاله: ۲۲/۶/۹۰

* این مقاله حاصل پایان‌نامه‌ی دانشجویی در مقطع کارشناسی ارشد است.

۱. استادیار، علوم کتابداری و اطلاع‌رسانی، دانشگاه آزاد اسلامی واحد همدان، همدان، ایران. (نویسنده‌ی مسؤل)

Email: ali.isfandyari@gmail.com

۲. دانشجوی کارشناسی ارشد، علوم کتابداری و اطلاع‌رسانی، دانشگاه آزاد اسلامی واحد همدان، همدان، ایران.

ارجاع: اسفندیاری مقدم علیرضا، بهاری موفق زهره. میزان هم‌پوشانی نتایج بازیابی شده‌ی کلیدواژه‌های تخصصی پزشکی در موتورهای کاوش عمومی وب. مدیریت اطلاعات سلامت ۱۳۹۱؛ ۹ (۲): ۲۱۴-۲۰۳.

مقدمه

به وجود آمدن اینترنت و رشد و گسترش روز افزون شبکه‌ی جهانی وب، فرصت‌ها و چالش‌های نوینی را در فرایند اطلاع‌یابی (Information seeking) پدید آورده است و امکان دسترسی به حجم فراوانی از اطلاعات در فعالیت‌های مختلف از جمله زمینه‌های آموزشی، تحقیقاتی، حرفه‌ای و ... را تحت تأثیر خود قرار داده است. اگر در گذشته مشکل عمده، حجم کم مطالب بازیابی شده بود؛ امروزه این مشکل تبدیل به بازیابی بیش از حد اطلاعات (Information overload) شده است. شاید در گذشته نیاز به دانستن مهارت‌های اطلاع‌یابی به اندازه‌ی امروز نبود. اما، امروزه به دلیل افزایش حجم گسترده‌ی اطلاعات در دسترس، داشتن مهارت‌هایی برای بازیابی اطلاعات مورد نیاز و مناسب الزامی است. این فعالیت‌ها در علم اطلاع‌رسانی، اطلاع‌یابی نامیده می‌شود و مستلزم فراگیری مهارت‌ها و داشتن ابزارهای خاصی به منظور جست‌وجو و دسترسی به اطلاعات مورد نظر است (۱). به نقل از Bazac در بیشتر موارد، بهترین و سریع‌ترین راه برای یافتن اطلاعات، جست‌وجو در وب است (۲).

موتورهای کاوش (Search engines) از موتورهای کاوش عمومی گرفته تا موضوعی خاص، کاشفان عمده‌ی منابع روی وب می‌باشند. آشنایی با موتورهای مختلف جست‌وجو به منظور بازیابی سریع و صحیح اطلاعات و جست‌وجوی رکوردهای مرتبط مطلوب و اجتناب از بازیابی رکوردهای نامرتب، یکی از نیازهای اساسی استفاده‌کنندگان اینترنت است و آنچه برای جست‌وجوگران اینترنت اهمیت دارد، صرفه‌جویی در وقت آنان است. به نقل از Can و همکاران یک موتور کاوش وب یک نظام بازیابی اطلاعات است که برای تعیین محل صفحات وب مرتبط با سؤال کاربر به کار می‌رود (۳). موتور جست‌وجو برنامه‌ای نرم‌افزاری است که به کاربر کمک می‌کند تا اطلاعات ذخیره شده بر کامپیوتر یا شبکه‌ای از کامپیوترها مانند اینترنت را بیابد (۴). موتورهای جست‌وجوی وب با جمع‌آوری صفحات وب و ایجاد پایگاه‌های اطلاعاتی، فرآیند مرور و جست‌وجو را برای استفاده‌کنندگان تسهیل می‌کنند. به کمک این ابزارها و از

طریق کلیدواژه، عبارت و سایر قابلیت‌ها می‌توان به جست‌وجوی اطلاعات در اینترنت پرداخت. هر کدام از موتورهای جست‌وجو از برخی جنبه‌ها با هم تفاوت دارند و نقاط قوت و ضعف و نیز گستره‌ی اطلاعات خاص خود را دارا می‌باشند که باید بر اساس نیاز اطلاعاتی و نوع مدرک مورد جست‌وجو، هر یک از این ابزارها را برای جست‌وجوی منابع وب انتخاب کرد.

در واقع، موتورهای جست‌وجو از نظر تنوع، محتوا و استراتژی‌های جست‌وجو، تهیه‌ی منابع و مجموعه ابزارهایی که برای کمک به استفاده‌کنندگان ارایه می‌دهند، تفاوت‌ها و پیچیدگی‌های خاص خود را دارا هستند (۵). این موتورهای کاوش، حجم عظیمی از اطلاعات را با سرعت مشخص جست‌وجو می‌کنند و به طور تقریبی تمام ابزارهای کاوش، خود را بهترین، کارآمدترین و قدرتمندترین بستر برای جست‌وجو و بازیابی اطلاعات در محیط وب معرفی می‌کنند (۶).

از طرفی، به وسیله‌ی هیچ یک از موتورهای جست‌وجو نمی‌توان در تمام منابع اطلاعاتی موجود در اینترنت جست‌وجو کرد. بنابراین، شناخت و درک خصوصیات میزان کارایی انواع موتورهای جست‌وجو در تأمین نیازهای اطلاعاتی متقاضیان از اهمیت ویژه‌ای برخوردار است (۵)؛ تا کاربران بتوانند با جست‌وجو در چند موتور کاوش، اطلاعات مورد نیاز خود را بیابند.

به نقل از Sullivan بیش از ۸۰ درصد استفاده‌کنندگان وب، از موتورهای کاوش به عنوان نقطه‌ی شروع استفاده می‌کنند (۷)؛ اما از آنجا که اغلب کاربران برای یافتن اطلاعات مورد نیاز خود به چند موتور کاوش مراجعه می‌کنند و نتایج به دست آمده در همه‌ی آن‌ها را بررسی می‌نمایند (۸)، در این بین نتایجی مشترک نیز دیده می‌شود که نشان دهنده‌ی هم‌پوشانی (Overlap) نتایج موتورهای کاوش است. در واقع، به میزان حضور عینی منابع بازیابی شده در نخستین صفحات کلیه‌ی موتورهای کاوش مورد بررسی در بازیابی اطلاعات مربوط به یک عبارت جست‌وجو شده، هم‌پوشانی گفته می‌شود. هر چه نتایج بازیابی شده توسط موتورهای کاوش اشتراک بیشتری با هم داشته باشند، هم‌پوشانی بیشتری رخ خواهد داد. در این گستره‌ی وسیع

جست‌وجوی اطلاعات به این نتیجه دست یافتند که ۸۴/۹۰ درصد از موارد بازیابی شده، مختص به یکی از چهار موتور کاوش بودند. ۱۱/۴ درصد از نتایج بازیابی شده در دو موتور کاوش مشترک بودند. ۲/۶ درصد نتایج در سه موتور کاوش و در نهایت فقط ۱/۱ درصد از نتایج بازیابی شده در هر چهار موتور کاوش تحت بررسی به صورت مشترک بازیابی شده بود (۱۲).

اسفندیاری مقدم و پریخ در مقاله‌ای تحت عنوان «مقایسه‌ی تطبیقی هم‌پوشانی نتایج در ابرموتورهای کاوش و موتورهای کاوش تحت پوشش آن‌ها»، به مقایسه و ارزیابی کارایی این ابزارهای جست‌وجو در حل مسایل کاربران پرداختند. آن‌ها تعدادی کلیدواژه را در پنج ابرموتور جست‌وجوی رایگان موجود در وب شامل Widow، InfoGrid، Ez2find، IcySpicy و نیز موتورهای کاوش تحت پوشش آن‌ها شامل Yahoo، Google، WiseNut و MSN مورد جست‌وجو قرار دادند و ده نتیجه‌ی ابتدایی بازیابی شده توسط این دو گروه ابزار کاوش را با هم مقایسه کردند. مهم‌ترین نتیجه‌ی این پژوهش آن بود که کاربران شبکه باید آگاه باشند که محدود کردن جست‌وجوی خود به یک موتور کاوش مانند Google منجر به از دست دادن رکوردهای مهمی خواهد شد که در دیگر موتورهای کاوش و یا راهنماهای وب دارای رتبه‌ی بالایی هستند. در عوض، استفاده از ابرموتورهای کاوش باعث دسترسی کارتر و مؤثرتر به اطلاعات مرتبط خواهد شد (۱۳).

در پژوهشی که در دانشگاه ایالتی پنسیلوانیا (Pennsylvania state university) و دانشگاه فن‌آوری کوئینزلند (Queensland university of technology) انجام شد، هم‌پوشانی نتایج اولین صفحه‌ی ۴ موتور کاوش Google، Ask jeeves، Yahoo و Live Search مورد بررسی قرار گرفت و مشخص گردید از مجموع ۷۷۶۴۳۵ نتیجه‌ی بازیابی شده، تنها ۰/۶ درصد از نتایج در صفحه‌ی اول موتورهای کاوش با هم هم‌پوشانی داشتند. به طور کلی، ۸۸/۳ درصد از کل نتایج به یک موتور کاوش اختصاص داشت. ۸/۹ درصد از کل نتایج بین دو موتور کاوش و ۰/۶ درصد از کل نتایج بین چهار موتور کاوش مشترک بودند و

اطلاعاتی، حجم عظیمی از اطلاعات حوزه‌ی پزشکی اعم از علوم بالینی و پایه وجود دارد. همچنین، مشاهده می‌شود که روز به روز تعداد سایت‌های تخصصی رو به افزایش است (۹) و به نظر می‌رسد که کاربران در دستیابی به اطلاعات مورد نیاز خود با ابزارهای متنوعی رو به رو می‌باشند. از این‌رو لزوم شناسایی و دستیابی سریع به اطلاعات مرتبط بهداشتی در شبکه‌ی اینترنت، کتابداران و متخصصان را به یک ارزیابی واداشته است (۱۰) و از آنجا که اطلاعات بهداشتی در بین اطلاعات وب در رتبه‌ی چهارم قرار گرفته است و بیشتر کاربران و بیماران برای دریافت اطلاعات بهداشتی از موتورهای کاوش استفاده می‌کنند (۱۱)، هدف اصلی این تحقیق، تعیین میزان هم‌پوشانی نتایج بازیابی شده‌ی اطلاعات پزشکی در صفحات اولیه‌ی نتایج از طریق کلیدواژه‌های تخصصی برگرفته از سرعنوان‌های موضوعی پزشکی مش (MESH یا Medical subject headings) در موتورهای کاوش بود.

بنابراین، با توجه به تولید روزافزون اطلاعات در محیط وب و دشواری در بازیابی اطلاعات مرتبط، آشنایی با موتورهای ابرموتورهای کاوش اجتناب‌ناپذیر به نظر می‌رسد. از طرف دیگر، پژوهش حاضر که این ابزارها را در رابطه با کلیدواژه‌های تخصصی پزشکی مورد ارزیابی قرار می‌دهد، از اهمیت خاص خود برخوردار می‌باشد. از آنجایی که قسمت مهمی از اطلاعات موجود در وب، اطلاعات پزشکی و بهداشتی را شامل می‌شود، شناسایی موتورهایی که به صورت بهینه‌تر به امر بازیابی اطلاعات پزشکی در این حیطه می‌پردازند، ضروری به نظر می‌رسد؛ و از نتایج حاصل از آن می‌توان در راهنمایی جامعه‌ی استفاده‌کنندگان اطلاعات پزشکی شامل کتابداران، دانشجویان پزشکی و کلیه‌ی رشته‌های مرتبط با علوم بهداشتی، پزشکان، پژوهشگران و مؤسسات تحقیقاتی در زمینه‌ی پزشکی و ... بهره جست.

Spink و همکاران طی پژوهشی گسترده در مورد میزان هم‌پوشانی و منحصر به فرد بودن نتایج بازیابی شده توسط چهار موتور کاوش اصلی Yahoo، Google، MSN و Ask Jeeves برای مجموعه‌ای عظیم شامل بیش از ۱۲۰۰۰

اکثریت نتایج جست‌وجو در صفحه‌ی اول، مختص به همان موتور کاوش بودند (۱۴).

Spink و همکاران در تحقیقی با عنوان «هم‌پوشانی نتایج موتورهای کاوش در بازیابی سایت‌های تبلیغاتی (Sponsored links) در بین موتورهای کاوش Ask، Google، Yahoo، MSN» به این نتیجه رسید که ۸۵ درصد نتایج بازیابی شده منحصر به یک موتور کاوش بود و هم‌پوشانی در کل موتورها به میزان ۱ درصد نتایج محدود می‌شد (۱۱).

Rather و همکاران در تحقیقی با عنوان «هم‌پوشانی نتایج جست‌وجو شده در وب: مطالعه‌ای بر پنج موتور کاوش»، به بررسی میزان هم‌پوشانی نتایج بازیابی شده در صفحه‌ی نخست پنج موتور کاوش BioWeb، Google، HotBot، Scirus و AltaVista پرداختند و به این نتیجه دست یافتند که هیچ هم‌پوشانی در بازیابی نتایج استخراج شده در پنج موتور کاوش دیده نمی‌شود. اما میزان هم‌پوشانی در چهار موتور کاوش Google، HotBot، Scirus و AltaVista به میزان ۲ درصد به دست آمد (۱۵).

قاضی میرسعید و همکاران در تحقیقی با عنوان «بررسی مقایسه‌ای موتورها و ابرموتورهای کاوش (Meta search engines) منتخب در بازیابی اطلاعات فیزیوتراپی از شبکه‌ی جهانی وب و تعیین هم‌پوشانی آن‌ها»، به بررسی کلیدواژه‌های تخصصی فیزیوتراپی برگرفته از مش در هفت موتور کاوش و هفت ابرموتور کاوش منتخب - که از نظر سایت Searchengineswatch.com پرستفاده‌ترین موتور و ابرموتورها بودند - پرداختند. نتایج این مطالعه نشان داد که موتورهای کاوش Google، AltaVista و AlltheWeb به ترتیب، بیشترین نتایج بازیابی را داشتند و در بین ابرموتورهای کاوش، Ixquick دارای بیشترین نتایج بازیابی بود. در ضمن نتایج بازیابی شده بین ۴۰ الی ۶۰ درصد تعیین گردید که بدیهی است این امر به دلیل تفاوت در شیوه‌ی رتبه‌بندی نتایج در هر موتور و ابرموتور کاوش می‌باشد (۸).

از مطالعه‌ی متون و پیشینه‌های موجود چنین بر می‌آید که تحقیق و کنکاش در ابزارهای بازیابی اطلاعات در محیط

وب به مانند تاریخچه‌ی این ابزارها از سابقه‌ی چندانی برخوردار نیست و هم‌گام با پیشرفت و ارتقای فهرست‌های موضوعی و موتورها و ابرموتورهای کاوش، ضرورت بررسی و ارزیابی آن‌ها از جنبه‌های مختلف به ویژه هم‌پوشانی نتایج بیش از پیش به نظر می‌رسد. همچنین، از آنجا که اطلاع از میزان هم‌پوشانی نتایج بازیابی شده در ابزارهای بازیابی اطلاعات، در جلوگیری از اتلاف وقت کاربران مهم تلقی می‌گردد و سهم اطلاعات پزشکی - بهداشتی در میان انبوه اطلاعات موجود در وب از جایگاه ویژه‌ای برخوردار است، این تحقیق سعی بر شناسایی کاراترین موتورها در بازیابی اطلاعات پزشکی داشت.

هدف از این پژوهش، تعیین میزان هم‌پوشانی نتایج بازیابی شده‌ی کلیدواژه‌های تخصصی پزشکی در موتورهای کاوش وب بود. با این امید که بهترین موتورهای کاوش را به جامعه‌ی استفاده‌کنندگان و جست‌وجوگران اطلاعات پزشکی و بهداشتی در محیط وب معرفی نماید.

روش بررسی

تحقیق حاضر، کاربردی و از نوع پیمایشی بود که با رویکرد توصیفی - تطبیقی در سال ۱۳۸۹ انجام شد. کاربرد تحقیق پیمایشی در مطالعاتی است که در پی توصیف کمی یا چند جنبه از امور هستند و همچنین مطالعاتی که به کشف نوعی رابطه‌ی هم‌خوانی و نه به طور الزامی همبستگی بین دو یا چند متغیر توجه دارند. استفاده از تحقیق پیمایشی به مطالعه‌ی انسان‌ها محدود نمی‌شود و در مورد سایر پدیده‌ها نیز به کار گرفته می‌شود. در این روش، محقق به گردآوری اطلاعات درباره‌ی واقعیت‌ها می‌پردازد و نتایج حاصل را به طور کمی تفسیر می‌نماید (۱۶).

از آنجا که در پیمایش تطبیقی یا مقایسه‌ای، مقایسه‌ی دو یا چند محیط پژوهشی بر پایه‌ی ضوابط و نشان دادن نقاط قوت و ضعف این نوع محیط‌ها است (۱۷)، می‌توان گفت روش تحقیق حاضر از نوع توصیفی - تطبیقی است. روایی آن بر اساس مشاهدات پژوهشگران، مشورت با برخی استادان صاحب‌نظر در این حوزه، و به ویژه استفاده از متون مرتبط

صفحه‌های نخست مجموع بازیافته‌های جست‌وجوگرها می‌پردازد (۱۳) و نیز دلیل اصلی مرور صفحه‌های نخست، ارایه‌ی مرتبط‌ترین نتایج در صفحه‌ی اول نتایج بازیابی شده می‌باشد (۱۸)، ده نتیجه‌ی اول هر موتور کاوش به صورت سیاهه‌ای، لیست و با استفاده از نرم‌افزار Excel، پس از بررسی میزان اشتراکات رکوردهای بازیابی شده در موتورها محاسبه و به عنوان میزان هم‌پوشانی در نظر گرفته شد.

کلیدواژه‌ها و عبارات‌های پزشکی برگرفته از سرعنوان‌های پزشکی مش:

"Bone Demineralization Technique"
 "Cardiology Service, Hospital"
 "Dental leakage"
 "Diagnostic Imaging" and "Clinical Care"
 Dexamethasone and Pharmacokinetics
 Laparotomy or Laparoscopy
 Preleukemia
 Radiotherapy and Neoplasms– Surgery
 Stomatitis
 Tuberculosis or Malaria– Immunology

برای تعیین هم‌پوشانی نتایج بازیابی شده‌ی کلیدواژه‌های تخصصی پزشکی در چهار موتور کاوش Live search، Google، Ask و Yahoo، ده کلیدواژه و عبارت مشترک در آن‌ها مورد کاوش قرار گرفتند. در این پژوهش، به میزان حضور عینی منابع بازیابی شده در نخستین صفحات موتورهای کاوش مورد بررسی در بازیابی اطلاعات پزشکی در ارتباط با یک عبارت جست‌وجو شده هم‌پوشانی گفته می‌شود و هر چه نتایج بازیابی شده توسط موتورهای کاوش اشتراک بیشتری با هم داشته باشند، هم‌پوشانی بیشتری رخ خواهد داد. از این‌رو، ده نتیجه‌ی اول (نتایج به دست آمده در صفحه‌ی نخست موتورهای کاوش) در یک سیاهه، گردآوری و با تطبیق و بررسی نتایج به دست آمده و استفاده از معادله‌ی مربوط، میزان هم‌پوشانی نتایج بازیابی شده‌ی صفحه‌ی نخست موتورهای کاوش محاسبه شد. در همین راستا، برای هر کلیدواژه جدولی تنظیم گردید (برای نمونه و پرهیز از حجیم شدن مقاله، جدول ۲ ارایه شده است). ستون اول هر جدول، نشان دهنده‌ی نتایج به دست آمده در صفحه‌ی

(۱۸، ۱۵، ۱۳-۱۲) مورد تأیید قرار گرفت. همچنین با توجه به رویکرد توصیفی- تطبیقی اتخاذ شده در این پژوهش، آمار توصیفی با محوریت یک معادله‌ی آماری به کار گرفته شد. جامعه‌ی مورد بررسی در این پژوهش شامل موتورهای کاوش عمومی وب بود. بر اساس متون موجود و مشاهدات محققین، ابتدا مجموعه‌ای مشتمل بر ۲۶ موتور کاوش جمع‌آوری و از این بین پس از حذف ۲۲ مورد زیر، چهار موتور کاوش Yahoo، Live search (Live)، Google و Ask مورد بررسی قرار گرفتند (جدول ۱).

جدول ۱: دلایل حذف سایر موتورهای کاوش از جامعه‌ی پژوهش

دلایل	موتور کاوش
عدم دسترسی رایگان	Britannica Directory
تلفیق با سایر موتورهای کاوش	Yahoo → Overture Froogle → Google Google → Go MSN → Live search
مرده بودن	Direct Hit MessageKing Deja.com Openfind iWon FlipperWiseNut NBCi Northern Light WebTopTeoma Magellan InvisibleWeb.com
فیلتر در ایران	Lycos Monstercrawler AlltheWeb Infoseek
صرف فهرست موضوعی بودن	LookSmart Open Directory

در جهت نیل به هدف پژوهش و تعیین میزان هم‌پوشانی نتایج بازیابی شده در چهار موتور کاوش Ask، Yahoo، Live search و Google، ابتدا ده کلیدواژه یا عبارت پزشکی از سرعنوان‌های موضوعی پزشکی مش، که در زیر به آن‌ها اشاره می‌شود، به صورت تصادفی انتخاب شد. از آن‌جا که کاربر برای رفع نیاز اطلاعاتی خود فقط به مرور

کاوش Google؛ سومین مؤلفه یعنی عدد ۱ نشان دهنده‌ی بازیافت عینی نتیجه‌ی X۲۱۳ در موتور کاوش Live و چهارمین مؤلفه یعنی عدد ۰ نشان دهنده‌ی عدم بازیافت عینی نتیجه‌ی X۲۱۳ در موتور کاوش Yahoo است. برای مثال، نمونه‌ی استخراج شده‌ی نتایج بازیابی کلیدواژه‌ها در مورد کلیدواژه‌ی ۱ (Bone Demineralization Technique) در جدول ۲ آمده است.

یافته‌ها

جهت محاسبه‌ی هم‌پوشانی و استخراج درصد هم‌پوشانی نتایج بازیابی شده در صفحات نخست موتورهای کاوش با سایر موتورها از معادله‌ی $(Xikm) \times \frac{1}{T}$ استفاده شد. در این معادله، # نشان دهنده‌ی تعداد بردارهای موجود در نتایج بازیابی شده‌ی یک موتور کاوش و مخرج این معادله، T، نشان دهنده‌ی تعداد کل نتایج بازیابی شده در صفحه‌ی نخست هر موتور کاوش در بازیابی ۱۰ کلیدواژه است. در نتیجه، مقدار T برابر با ۱۰ نتیجه‌ی بازیابی شده در صفحات نخست موتورها ضربدر ۱۰ کلیدواژه‌ی جست‌وجو شده در آن‌ها ($T = k \times i$) است. بنابراین، در اینجا T برابر با $10 \times 10 = 100$ است. به طور مثال، اگر بخواهیم میزان

نخست به ترتیب رتبه‌ی هر بازیافت از یک الی دهمین نتیجه است و در ستون‌های بعدی، هر ستون مختص به نتایج صفحه‌ی نخست یک موتور کاوش تحت بررسی در نظر گرفته شده است. هر خانه‌ی جدول، متناظر با یک نتیجه‌ی بازیابی شده است که آن با متغیر Xikm نشان داده شده است. به عبارت دیگر، هر Xikm نشان دهنده‌ی آمین نتیجه‌ی بازیابی شده‌ی صفحه‌ی نخست کلیدواژه‌ی K در موتور کاوش M است. در جدول، هر Xikm با یک بردار سطری ۴ بعدی از ۰ و ۱‌ها نشان داده می‌شود. اگر موتور کاوش Ask را اولین موتور کاوش در نظر بگیریم؛ آن‌گاه اولین مؤلفه‌ی بردار عدد ۱ خواهد بود و مؤلفه‌های بعدی در صورت بازیافت عینی در سایر موتورها با عدد ۱ و در صورت عدم حضور در بین نتایج بازیافت شده‌ی صفحه‌ی نخست موتورهای کاوش، با عدد ۰ نشان داده خواهد شد. برای نمونه، دومین نتیجه‌ی بازیابی شده‌ی اولین کلیدواژه‌ی ۱ (Bone Demineralization Technique) در موتور کاوش Live (۳) با بردار X۲۱۳ نشان داده شده است؛ که در جدول ۱، بردار فوق (X۲۱۳) برابر است با ۰۱۱۰؛ از سمت راست اولین مؤلفه یعنی عدد ۰ نشان دهنده‌ی عدم بازیافت عینی نتیجه‌ی X۲۱۳ در موتور کاوش Ask؛ دومین مؤلفه یعنی عدد ۱ نشان دهنده‌ی بازیافت عینی نتیجه‌ی X۲۱۳ در موتور

جدول ۲: بازیافت عینی کلیدواژه‌ی ۱ "Bone Demineralization Technique" در موتورهای کاوش تحت بررسی

۱۰ نتیجه‌ی صفحه‌ی نخست	Ask (۱)	Google (۲)	Live (۳)	Yahoo (۴)
نتیجه‌ی اول	۱۱۱۱	۱۱۱۱	۱۱۱۱	۱۰۰۱
نتیجه‌ی دوم	۱۰۰۱	۱۱۰۰	۰۱۱۰	۰۰۱۱
نتیجه‌ی سوم	۱۱۰۰	۰۱۰۰	۰۰۱۰	۱۱۱۱
نتیجه‌ی چهارم	۱۱۰۰	۰۱۰۰	۰۰۱۰	۱۰۰۱
نتیجه‌ی پنجم	۱۰۰۰	۱۱۰۰	۰۰۱۰	۰۰۰۱
نتیجه‌ی ششم	۱۱۰۱	۰۱۰۰	۰۰۱۰	۱۱۰۱
نتیجه‌ی هفتم	۱۰۰۰	۰۱۰۰	۰۰۱۰	۰۰۰۱
نتیجه‌ی هشتم	۱۰۰۰	۰۱۱۰	۰۰۱۰	۰۰۰۱
نتیجه‌ی نهم	۱۰۰۱	۰۱۰۰	۰۰۱۰	۰۰۰۱
نتیجه‌ی دهم	۱۰۰۰	۱۱۰۱	۰۰۱۱	۰۰۰۱

درصد نتایج بین Ask، Google، Live و Yahoo مشترک بوده‌اند که با جمع این مقادیر (۲۲ + ۶ + ۷ + ۱۱ = ۴۶) دریافت می‌شود که ۴۶ درصد نتایج Google با موتور کاوش Ask هم‌پوشانی دارد (نمودار ۱).

بر اساس نمودار ۱، مشخص گردید که ۳ درصد نتایج در بین موتورهای کاوش Ask و Live؛ ۷ درصد نتایج بین Ask، Google و Live؛ ۲ درصد نتایج بین Ask، Google، Live و Yahoo مشترک بوده‌اند که با جمع این مقادیر (۳ + ۷ + ۲ + ۱۱ = ۲۳) دریافت می‌شود که ۲۳ درصد نتایج Ask با موتور کاوش Live هم‌پوشانی دارد.

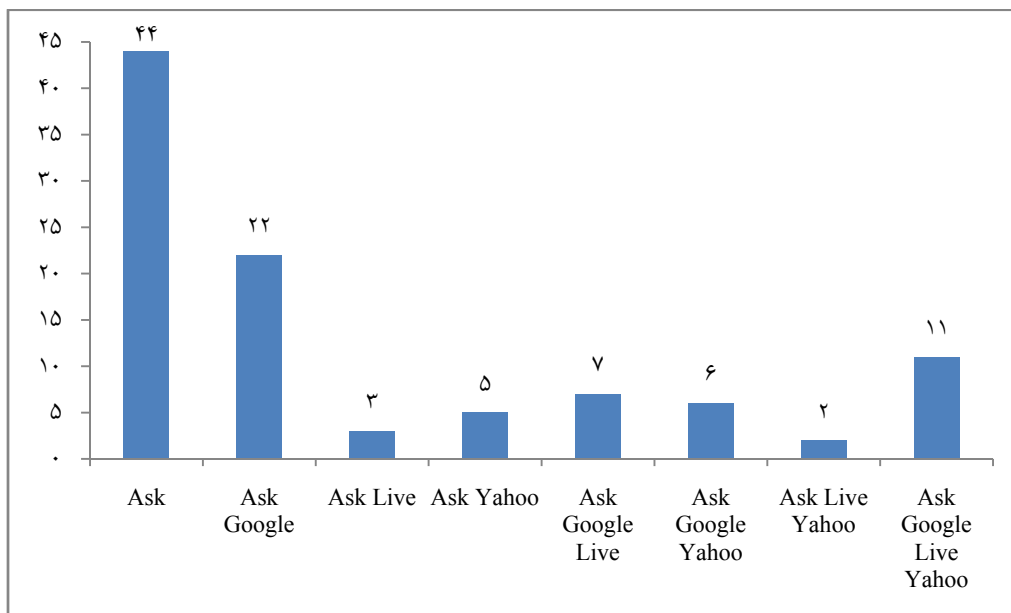
همچنین ۵ درصد نتایج در بین موتورهای کاوش Ask و Yahoo؛ ۶ درصد نتایج بین Ask، Google و Yahoo؛ ۲ درصد نتایج بین Ask، Live، Yahoo و ۱۱ درصد نتایج بین Ask، Google، Live و Yahoo مشترک بوده‌اند که با جمع این مقادیر (۵ + ۶ + ۲ + ۱۱ = ۲۴) دریافت می‌شود که ۲۴ درصد نتایج Ask با موتور کاوش Yahoo هم‌پوشانی دارد.

هم‌پوشانی نتایجی را که در موتور کاوش Ask با موتور کاوش دیگر یعنی Google هم‌پوشانی دارند استخراج نماییم، کافی است Xikm‌های موتور کاوش Ask را در بین نتایج موتور Google محاسبه نماییم و در معادله‌ی پیش‌گفت به کار ببندیم. از آنجا که مشترکات بین دو موتور کاوش با ۴ بردار متفاوت (۱۱۰۰، ۱۱۱۰، ۱۱۱۱ و ۱۱۰۱) برآورد شده‌اند، عدد به دست آمده از ۴ بردار موجود پس از محاسبه به معادله‌ی مزبور و به طریق زیر با هم جمع بسته می‌شوند:

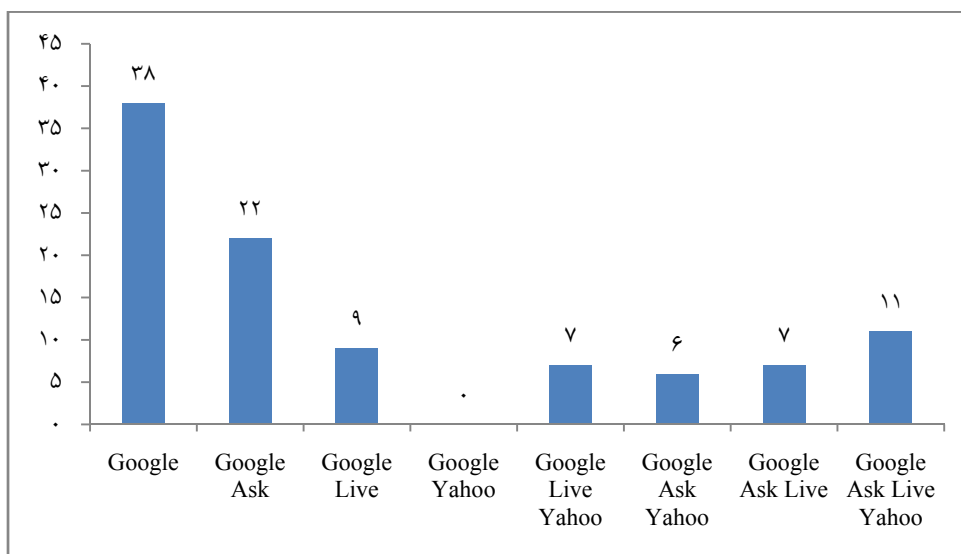
$$\frac{1}{T_{\#(1111) \times}} + \frac{1}{T_{\#(1101) \times}} + \frac{1}{T_{\#(1110) \times}} + \frac{1}{T_{\#(1100) \times}}$$

$$\%46 = \frac{1}{100} \times 11 + \frac{1}{100} \times 7 + \frac{1}{100} \times 6 + \frac{1}{100} \times 22$$

بر این اساس، ۲۲ درصد نتایج در بین موتورهای کاوش Ask و Google؛ ۶ درصد نتایج بین Ask، Google و Yahoo؛ ۷ درصد نتایج بین Ask، Google، Live و ۱۱ و



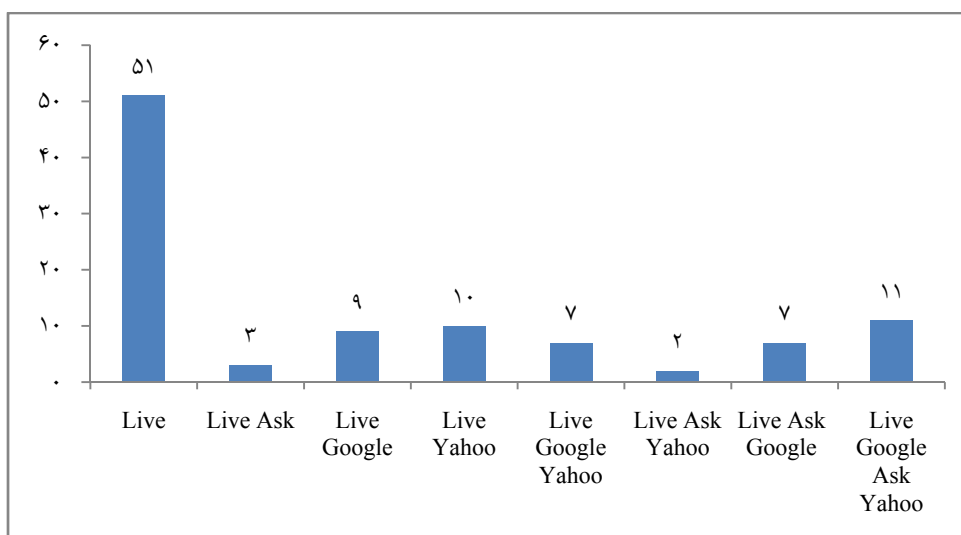
نمودار ۱: بررسی میزان حضور نتایج بازیابی شده‌ی ۱۰ کلیدواژه در نخستین صفحه‌ی موتور کاوش Ask و هم‌پوشانی آن‌ها در سایر موتورهای تحت بررسی



نمودار ۲: بررسی میزان حضور نتایج بازیابی شده‌ی ۱۰ کلیدواژه در نخستین صفحه‌ی موتور کاوش Google و هم‌پوشانی آن‌ها در سایر موتورهای تحت بررسی

می‌شود که ۴۶ درصد نتایج Google با موتور کاوش Ask هم‌پوشانی دارد. ۹ درصد نتایج در بین موتور کاوش Live و Google؛ ۷ درصد نتایج بین Google، Ask و Live؛ ۷ درصد نتایج بین Google، Yahoo و Live؛ و ۱۱ درصد نتایج بین Ask، Google، Live و Yahoo مشترک بوده‌اند که با جمع این مقادیر (۳۴ = ۱۱ + ۷ + ۷ + ۹) دریافت می‌شود که ۳۴ درصد نتایج Google با موتور کاوش Live هم‌پوشانی دارد.

بر اساس نمودار ۲، مشخص گردید که ۳۸ درصد نتایج در موتور کاوش Google تنها در آن بازیابی شده و در هیچ یک از سایر موتورها دیده نشده‌اند. ۲۲ درصد نتایج در بین موتور کاوش Google و Ask؛ ۶ درصد نتایج بین Google، Ask و Yahoo؛ ۷ درصد نتایج بین Ask، Google، Live و Yahoo؛ ۱۱ درصد نتایج بین Ask، Google، Live و Yahoo مشترک بوده‌اند که با جمع این مقادیر (۴۶ = ۱۱ + ۷ + ۶ + ۲۲) دریافت



نمودار ۳: بررسی میزان حضور نتایج بازیابی شده‌ی ۱۰ کلیدواژه در نخستین صفحه‌ی موتور کاوش Live و هم‌پوشانی آن‌ها در سایر موتورهای تحت بررسی

نتایج با سایر موتورها بود. تعداد کم نتایج منحصر به موتور Google می‌تواند نشان دهنده‌ی پایگاه داده‌های وسیع این موتور کاوش باشد. همچنین، مشخص گردید که موتورهای کاوش Google و Ask بیشترین هم‌پوشانی را به میزان ۴۶ درصد نتایج در بین موتورها دارا بودند. میزان هم‌پوشانی Google و Live نیز به میزان ۳۴ درصد و هم‌پوشانی Google و Yahoo به میزان ۲۴ درصد بود. موتور کاوش Yahoo بیشترین هم‌پوشانی را با Live به میزان ۳۰ درصد داشت. در مجموع، می‌توان اظهار کرد که موتورهای کاوش دارای هم‌پوشانی کمی نسبت به هم بودند. صفحه‌ی نخست نتایج هر موتور کاوش، صفحه‌ای است که در مقایسه با نتایج صفحه‌ی نخست سایر موتورها بی‌مانند به نظر می‌آید.

پیشنهادها

به تمامی پژوهشگران، کتابداران و کاربران متخصص در امور پزشکی و عموم افراد پیشنهاد می‌گردد در هنگام کاوش اطلاعات پزشکی در دنیای اینترنت جست‌وجوی خود را در چند موتور کاوش پیگیری کنند تا به مدارک مرتبط از میان انبوه منابع موجود در وب، دسترسی یابند و در صورتی که از موتور کاوش Google استفاده می‌نمایند، کاوش خود را در موتورهای کاوش دیگر به خصوص، در موتور کاوش Yahoo نیز تکرار نمایند و در صورتی که از موتور Live و Yahoo جهت جست‌وجو استفاده می‌گردد، همان جست‌وجوها را در موتور کاوش Google یا Ask نیز انجام دهند. از طرف دیگر، به کاربران توصیه می‌شود در جست‌وجوی اطلاعات پزشکی از موتورهای جست‌وجوی تخصصی نیز استفاده کنند.

تشکر و قدردانی

بدینوسیله نویسندگان بر خود فرض می‌دانند از داوران محترمی که با ارایه‌ی دیدگاه‌های ارزنده‌ی خود بر غنای مقاله افزودند، سپاس‌گزاری نمایند.

References

1. Akbari A. Comparing search engines and met search engines in terms of the retrieval of physiotherapeutic information from the web and overlapping rate [Thesis]. Tehran: School of Medical Informatics & Management,

نتایج بازیابی شده‌ی کلیدواژه‌های پزشکی برگرفته از مش در صفحه‌ی نخست موتورهای کاوش Ask، Google، Live و Yahoo به میزان ۱۱ درصد بود؛ این در حالی است که نتایج تحقیقات Rather و همکاران نشان داد که هیچ هم‌پوشانی در بازیابی نتایج استخراج شده در موتورهای کاوش دیده نمی‌شود (۱۵). همچنین، Spink و همکاران طی تحقیقی با هدف بررسی هم‌پوشانی نتایج موتورهای کاوش در بازیابی سایت‌های تبلیغاتی (Sponsored links) در بین موتورهای کاوش Ask، Google، Yahoo و MSN به این نتیجه رسید که ۸۵ درصد نتایج بازیابی شده منحصر به یک موتور کاوش است و هم‌پوشانی در کل موتورها به میزان ۱ درصد نتایج محدود می‌شود (۱۱). یافته‌های این مطالعه با پژوهش حاضر مغایر است. به نظر می‌رسد هم‌پوشانی بیشتر نتایج بازیابی شده‌ی کلیدواژه‌های پزشکی در موتورهای کاوش، که در این پژوهش مورد تأیید قرار گرفت، نسبت به هم‌پوشانی نتایج موتورهای کاوش در بازیابی سایت‌های تبلیغاتی، نشان دهنده‌ی اهمیت بیشتر اطلاعات حوزه‌ی پزشکی و سلامت باشد؛ زیرا افزایش مخاطبان این حوزه و تلاش در جهت جذب آن‌ها از سوی موتورهای کاوش، هر یک از این ابزارها را وا داشته است که در نمایه کردن اطلاعات حوزه‌ی پزشکی و سلامت، همت و جدیت بیشتری از خود نشان دهند.

نتیجه‌گیری

نتایج حاصل از این تحقیق مشخص ساخت که موتورهای کاوش در بازیابی کلیدواژه‌های پزشکی در نخستین صفحه‌ی نتایج، ۴۸ درصد نتایج بازیابی شده را در یک موتور بازیابی می‌نمایند که نیمی از مرتبط‌ترین نتایج را شامل می‌شوند. موتور کاوش Yahoo دارای بیشترین نتیجه‌ای بود که تنها در خود آن بازیابی شد و موتور کاوش Google با کمترین نتیجه‌ی منحصر به خود، دارای بالاترین میزان هم‌پوشانی در

- Iran University of Medical Sciences; 2006. [In Persian].
2. Bazac D. Search the Web More Efficiently: Tips, Techniques and Strategies. *Scientific Communication, Monthly Journal of Irandoc* 2004; 2(4): 12-7. [In Persian].
 3. Can F, Nuray R, Sevdik AB. Automatic performance evaluation of Web search engines. *Information Processing & Management* 2004; 40(3): 495-514.
 4. Encarta encyclopedia. Search engine [Online]. 2011 [Cited 2011 Apr 1]; Available from: URL: http://encarta.msn.com/encyclopedia_761582857/Search_Engine.html/
 5. Mohammadi SH. Internet search tools. *Scientific Communication, Monthly Journal of Irandoc* 2003; 1(3): 18-22. [In Persian].
 6. Kousha K. Internet search tools: Principles, skills and search capabilities in the Web. Tehran: Ketabdar Publication; 2003. [In Persian].
 7. Sullivan D. Nielsen NetRatings search engine ratings [Online]. 2006 Aug 21 [Cited 2011 Apr 1]; Available from: URL: <http://searchenginewatch.com/2156451/>
 8. Ghazi-Mirsaeed J, Haghani H, Akbari A. A comparative study of search engines and metasearch engines in retrieving physiotherapy information from WWW and determining their overlapping. *Health Information Management* 2007; 4(1): 11-21. [In Persian].
 9. Tabatabaei A. Information technology in Medical Sciences. Tehran: Jamenegar Publication; 2006. [In Persian].
 10. Wu G, Li J. Comparing Web search engine performance in searching consumer health information: evaluation and recommendations. *Bull Med Libr Assoc* 1999; 87(4): 456-61.
 11. Spink A, Jansen BJ, Kathuria V, Koshman S. Overlap among major web search engines. *Internet Research*, 2006; 16(4): 419-26.
 12. Spink A, Jansen BJ, Blakely C, Koshman S. A study of results overlap and uniqueness among major web search engines. *Information Processing and Management* 2006; 42(5): 1379-91.
 13. Isfandyari-Moghaddam A, Parirokh M. A comparative study on overlapping of search results in metasearch engines and their common underlying search engines. *Library Review* 2006; 55(5): 301-6. [In Persian].
 14. Less Than 1% Overlap in Major Search Engine First-Page Results [Online]. 2007 Jun 11 [cited 2011 Apr 1]; Available from: URL: <http://www.marketingcharts.com/interactive/less-than-1-overlap-in-major-search-engine-first-page-results-611/>
 15. Rather RA, Lone FA, Jeelani SG. Overlap in Web search results: A study of five search engines. *Library Philosophy and Practice* [Online]. 2008 [Cited 2011 Apr 1]; Available from: URL: <http://www.webpages.uidaho.edu/~mbolin/rather-lone-shah.htm/>
 16. Dayyani MH. Research methods in Librarianship. Mashhad: Computer Library Publication; 2006. [In Persian].
 17. Krishan K. Research methods in Library and Information Science. Trans. Rahadoost F, Khosravi F. Tehran: National Library Publication; 1995. [In Persian].
 18. Hochstotter N, Lewandowski D. What users see-Structures in search engine results pages. *Information Sciences* 2009; 179(12): 1796-812.

The Overlap Rate of Searching Medical Keywords in General Search Engines*

Alireza Isfandyari Moghaddam, PhD¹; Zohreh Bahari Movaffagh²

Abstract

Introduction: Knowing the rate of overlapping in results obtained from information retrieval tools is important to prevent waste of time. Medical and health-related information has an outstanding place among the variety of information available on the Internet. Hence, this research aimed to study the rate of overlap among medical search results derived from MESH (medical subject headings) in general search engines.

Methods: This applied, descriptive, comparative approach survey was conducted in 2010. The research sample included 4 general search engines, i.e. Yahoo, Live Search, Google, and Ask. In order to determine the rate of overlapping among the searches, 10 medical keywords were first selected randomly. The first 10 results retrieved by each search engine were then listed. Finally, the overlap rate was calculated by means of Microsoft Excel. To determine the overlapping search results, 10 common terms were searched in the mentioned search engines. Then, 10 results retrieved in the first page of each search engine were collected and compared. The overlap rate among the retrieved results was finally calculated using a checklist and a researcher-made equation whose validity was confirmed based on observations of the researchers, comments of some expert professors, and especially referring to related literature. In addition, due to the descriptive, comparative approach of the research, descriptive statistics revolving around a statistical equation was utilized.

Results: It was found that search engines can retrieve 48% of relevant searches retrieved by other search engines. Yahoo had the highest number of unique results while Google had the highest number of results overlapping with other search engines. The overlap rate of the 4 studied search engines was 11%.

Conclusion: The results obtained from different search engines hardly overlap. Therefore, users are suggested to search more than one search engine simultaneously to access more relevant documents. Moreover, they are advised to use specialized search engines to search medical information.

Keywords: Internet; Information Retrieval; Keywords; Search Engine.

Type of article: Original Article

Received: 7Jan, 2011

Accepted: 12 Sep, 2011

Citation: Isfandyari Moghaddam A, Bahari Movaffagh Z. **The Overlap Rate of Searching Medical Keywords in General Search Engines.** Health Information Management 2012; 9(2): 214.

* This paper was derived from an MSc thesis.

1. Assistant Professor, Library and Information Sciences, Islamic Azad University, Hamadan Branch, Hamadan, Iran. (Corresponding Author) Email: ali.isfandyari@gmail.com

2. MSc Student, Library and Information Sciences, Islamic Azad University, Hamadan Branch, Hamadan, Iran.