



ارائه چارچوبی جامع و کارا برای ارزیابی موتورهای جستجوی فارسی

سبحان موسوی^۱، معصومه عظیم‌زاده^۱، مریم محمودی^۱ و علیرضا یاری^۱

^۱ پژوهشگاه ارتباطات و فناوری اطلاعات، تهران، ایران
{s.moosavi, azim_ma, mahmoudy, a_yari}@itrc.ac.ir

چکیده

با افزایش حجم داده‌ها در سطح وب، دست‌یابی به اطلاعات مورد نیاز بدون استفاده از موتورهای جستجو تقریباً ناممکن است. به این ترتیب موتورهای جستجوی بسیاری در سال‌های اخیر ارائه شده‌اند. مهمترین عامل در تعیین محبوبیت موتورهای جستجو، کارایی آنها در ارائه خدمات مناسب به کاربران می‌باشد. اما چالش اصلی در اینجاست که عوامل متعددی در تعیین کارایی یک موتور جستجو نقش دارند. نپرداختن به تمامی جنبه‌ها برای تعیین کارایی و هزینه‌های بالا را می‌توان عمده‌ترین مشکلات فعالیت‌هایی دانست که پیش از این در حوزه سنجش کارایی موتورهای جستجو مطرح گردیده‌اند. در این پژوهش، به ارائه چارچوبی مناسب برای سنجش کارایی موتورهای جستجو پرداخته شده است که در عین کم‌هزینه بودن، جنبه‌های گوناگونی همچون کیفیت نتایج بازگشتی، واسط کاربری، محبوبیت و پوشش در آن لحاظ گردیده است. تمرکز اصلی در این چارچوب بر روی زبان فارسی می‌باشد. همچنین بر اساس چارچوب پیشنهادی، به بررسی و مقایسه موتورهای جستجویی پرداخته شده است که امکان جستجو در سطح وب فارسی را فراهم آورده‌اند. نتایج بررسی‌ها، نشان از کارایی قابل ملاحظه موتورهای جستجوی عمومی (همچون گوگل) و فراجویشگرهای فارسی (همچون سلام و بیاب) داشته است. از سوی دیگر، موتور جستجوی پارسی‌جو نیز به عنوان مناسب‌ترین موتور جستجوی فارسی معرفی می‌گردد.

کلمات کلیدی

چارچوب ارزیابی، موتور جستجو، فراجویشگر، زبان فارسی، ارزیابی خودکار، واسط کاربری

۱- مقدمه

می‌گردند. برخی پژوهش‌ها عوامل مذکور را به دسته‌هایی همچون کیفیت نمایه^۲، کیفیت نتایج، کیفیت ویژگی‌های جستجو و نیز قابلیت کاربردپذیری^۴ موتورهای جستجو تقسیم‌بندی می‌نمایند [۵، ۶]. اما مقالات موجود در زمینه مقایسه موتورهای جستجو، معمولاً جنبه‌های محدودی را جهت انجام مقایسه انتخاب کرده و لزوماً به همه عوامل پرداخته نمی‌شود.

پژوهش‌هایی همچون [۵، ۷-۱۱]، تنها به جنبه سنجش کیفیت نتایج بازگشتی پرداخته‌اند. به این معنی که معمولاً با در نظر گرفتن فهرست مشخصی از پرس و جوها، نتایج بازگشتی را با استفاده از عوامل انسانی مورد بازبینی قرار می‌دهند. سپس نتایج این بازبینی به کمک معیارهایی مطرح در علم ارزیابی اطلاعات (همچون دقت، یادآوری و یا موارد مشابه دیگر) گزارش می‌گردد. از سوی دیگر، پژوهش‌هایی نیز وجود دارند که جنبه‌هایی غیر از کیفیت نتایج بازگشتی را مورد بررسی قرار داده‌اند. از جمله آنها می‌توان به مواردی همچون [۱۲-۱۵] اشاره کرد؛ در این پژوهش‌ها با بررسی ویژگی‌های

افزایش روز افزون استفاده از موتورهای جستجو در راستای برآورده نمودن نیاز اطلاعاتی کاربران، نمایانگر نقش مهم آنها در حوزه فناوری اطلاعات است. به طوری که هر چه میزان کارایی آنها بالاتر باشد، به همان میزان در میان کاربران از محبوبیت بیشتری برخوردار خواهند بود. اما مساله سنجش کارایی موتورهای جستجو در دهه‌های اخیر، موضوع بسیاری از پژوهش‌های علمی و نیز طرح‌های تجاری بوده است [۱]. این مساله برای اولین بار و به صورت جامع در سال ۱۹۹۵ مطرح گردیده است [۲]. همایش‌هایی همچون TREC^۱ [۳] و یا CLEF^۲ [۴] نیز بیشتر با هدف توسعه بسترهای لازم برای سنجش کارایی جنبه‌های مختلف ارزیابی اطلاعات در حوزه موتورهای جستجو و به صورت سالانه برگزار می‌گردند.

در حالت کلی، مقایسه میان موتورهای جستجو کاری پیچیده است؛ چرا که جنبه‌ها و عوامل گوناگونی جهت انجام مقایسه مطرح

۲- چارچوب پیشنهادی برای ارزیابی موتورهای

جستجو

در این بخش به توصیف چارچوب ارزیابی برای موتورهای جستجوی فارسی خواهیم پرداخت. رویکرد مورد استفاده برای ارائه چارچوب پیشنهادی به این شکل خواهد بود که ابتدا برخی معیارهای مشخص برای انجام ارزیابی معرفی می‌گردند؛ سپس شاخص‌های مورد استفاده برای ارزیابی هر معیار توصیف می‌شود. در پایان نیز در صورت لزوم، روش مورد استفاده برای اندازه‌گیری هر شاخص ارائه می‌گردد. معیارهای مورد استفاده در این چارچوب عبارتند از پوشش^۷، کارایی^۸، واسط کاربری^۹ و محبوبیت^۹. در زیربخش‌های آتی به توصیف این معیارها پرداخته می‌شود.

۲-۱- پوشش

پوشش به معنای نسبت تعداد صفحات نمایه‌سازی شده به تعداد کل صفحات موجود در سطح وب می‌باشد [۲۲]. این معیار برای موتورهای جستجو و فراجویشگرها به صورت جداگانه قابل محاسبه خواهد بود. به همین منظور، شاخص‌های زیر برای معیار پوشش معرفی می‌گردند:

- پوشش موتور جستجو: در حالت کلی، تعیین اندازه دقیق پوشش یک موتور جستجو امکان‌پذیر نیست؛ اما راه‌هایی برای تخمین آن وجود دارد [۲۳]. در این پژوهش از روش کلمات توقف^{۱۰} [۲۴] برای تخمین پوشش وب فارسی توسط موتورهای جستجوی مختلف استفاده شده است. در عین حال، بر اساس بررسی‌های به انجام رسیده، پوشش برخی از جستجوگرها نیز به صورت دقیق مشخص شده است. بنابراین میزان پوشش یا بر اساس آمار دقیق و یا بر اساس روش تقریبی ارائه می‌گردد.
- پوشش فراجویشگر: از آنجایی که فراجویشگرها خود اقدام به نمایه‌سازی محتوای وب نمی‌کنند و با ترکیب نتایج موتورهای جستجوی مختلف پوشش اطلاعاتی خود را بالا می‌برند، بنابراین برای محاسبه پوشش آنها، تعداد موتورهای جستجو مورد استفاده در ایجاد نتایج در نظر گرفته خواهد شد.

۲-۲- واسط کاربری

یک رابط کاربری مناسب علاوه بر برآوردن نیازمندی‌های گرافیکی، باید تعادلی میان قدرت بیان ساده و بیان منطقی را فراهم آورد. رابط کاربری که درک و پیمایش آسانی دارد، موجب رضایتمندی کاربر خواهد شد. جهت برآورد و اندازه‌گیری این معیار، از دو زیر معیار قابلیت‌های جستجو و نمایش نتایج استفاده می‌گردد. در جدول (۱) شاخص‌های مورد استفاده برای هر یک از این زیرمعیارها مشخص گردیده‌اند. لازم به ذکر است که شاخص‌های مربوط به بخش واسط کاربری که در جدول (۱) ارائه شده‌اند، به صورت شهودی و توسط کاربر انسانی اندازه‌گیری می‌شوند.

مختلف موتورهای جستجو که غالباً در دسته‌هایی همچون کیفیت ویژگی‌های جستجو و یا قابلیت‌های کاربردپذیری جای می‌گیرند، مقایسه‌ای تحلیلی به انجام رسیده است.

در کنار موارد فوق، مقالاتی ارائه شده‌اند که در آنها به صورت همزمان به بررسی ویژگی‌ها و قابلیت‌های موتورهای جستجو و نیز کارایی آنها از دیدگاه نتایج بازگشتی پرداخته شده است [۲، ۶، ۱۶، ۱۷]. برخی پژوهش‌ها نیز چارچوبی را برای مقایسه موتورهای جستجو ارائه کرده‌اند [۶، ۱۶]. چنین چارچوبی جنبه‌های مختلفی را جهت انجام مقایسه و تحلیل مناسب موتورهای جستجو فراهم می‌آورد.

با توجه به اینکه ارزیابی انسانی به عنوان یک دسته از روش‌های کیفیت‌سنجی بسیار هزینه‌بر، محدود (بر اساس تعداد پرس و جوهای در نظر گرفته شده و نیز تعداد نتایج) که در ازای هر پرس و جو قضاوت می‌گردند) و نیز وابسته به تفکر اشخاص^۵ می‌باشد [۱]، به این ترتیب و برای نخستین بار در سال ۲۰۰۱ ایده ارزیابی خودکار موتورهای جستجو مطرح گردید [۱۸]. مقالات متعددی در یک دهه اخیر در زمینه ارزیابی خودکار ارائه شده‌اند که از جمله مهمترین آنها می‌توان به مواردی همچون [۱۹-۲۱] اشاره کرد. در این مقالات سعی شده است تا با بهره‌گیری از تکنیک‌های مختلف، نحوه استدلال و شهود انسانی در جهت ارزیابی اسناد وب از دیدگاه ارتباط آنها با پرس و جوهای مطرح شده شبیه‌سازی گردد. بنابراین، اهداف اصلی در این پژوهش‌ها را می‌توان کاهش هزینه‌های مربوط به ارزیابی انسانی و نیز دستیابی به دقتی قابل قبول در راستای سنجش کارایی موتورهای جستجو به حساب آورد.

اما محدود نمودن بررسی‌ها به تنها یک جنبه از ویژگی‌ها و قابلیت‌های موتورهای جستجو (معمولاً سنجش کیفیت نتایج بازگشتی)، نتیجه‌گیری‌ها و تحلیل‌های مرتبط را به چالش خواهد کشید [۱۶]. بنابراین نیاز است تا چارچوبی ارائه گردد که در قالب آن دو هدف مهم برآورده گردد: (۱) در نظر گرفتن ویژگی‌ها و عوامل مختلف جهت انجام مقایسه‌ای جامع میان موتورهای جستجو و (۲) استفاده از رویکردی کارا در راستای کاهش هزینه‌های مربوط به فرآیندهای مقایسه و ارزیابی. به این ترتیب، در این مقاله به دنبال ارائه چارچوبی خواهیم بود که بر اساس آن بتوان اهداف تعیین شده را تا حد امکان برآورده ساخت. تمرکز اصلی در این پژوهش، مقایسه میان موتورهای جستجویی خواهد بود که امکان جستجو در سطح زبان فارسی را فراهم آورده‌اند. این موتورهای جستجو می‌توانند عمومی (همچون گوگل)، مختص زبان فارسی (همچون پارسی‌جو) و یا فراجویشگر (همچون سلام) باشند.

ادامه این مقاله به این شکل سازماندهی خواهد شد: ابتدا و در بخش دوم، چارچوب پیشنهادی برای ارزیابی موتورهای جستجوی فارسی ارائه می‌گردد. سپس در بخش سوم، بر اساس چارچوب پیشنهادی به ارائه نتایج حاصل از ارزیابی و مقایسه موتورهای جستجوی انتخابی پرداخته خواهد شد. بخش چهارم یا همان بخش پایانی نیز به انجام نتیجه‌گیری و تحلیل اختصاص خواهد داشت.

۳-۲- محبوبیت

تعداد مراجعات انجام گرفته به یک وبسایت، نشان دهنده میزان محبوبیت و اهمیت آن می‌باشد. بنابراین، میزان محبوبیت یک وبسایت با استفاده از شاخص‌های زیر قابل محاسبه است:

- میزان ترافیک وبسایت: برای محاسبه ترافیک، از اطلاعات وبسایت الکسا^{۱۱} استفاد می‌گردد. به این معنی که بر اساس آمارهای ارائه شده از سوی الکسا، می‌توان میزان بازدید یک وبسایت را در بازه‌های زمانی مشخص استخراج نمود.
- تعداد پیوندهای ورودی به یک وبسایت از سوی دیگر وبسایت‌ها: تعداد ارجاعات داده شده به یک وبسایت بر اساس اطلاعات و آمارهای موجود در وبسایت الکسا قابل اندازه‌گیری می‌باشد.

۳-۱- موتورهای جستجوی منتخب برای انجام

مقایسه و تحلیل

موتورهای جستجویی که در این پژوهش مورد ارزیابی قرار خواهند گرفت، شامل موتورهای جستجوی عمومی، موتورهای جستجوی فارسی و نیز فراجوشگرهای فارسی می‌باشد. در جدول (۱)، فهرست این موتورهای جستجو به تفکیک نوع قابل مشاهده است. لازم به ذکر است که سایر موتورهای جستجوی فارسی، به دلایلی همچون محبوبیت پایین، سطح کیفی نامناسب در ارائه سرویس به کاربران و یا قطع بودن سرویس، در این پژوهش مورد بررسی قرار نگرفته‌اند.

۳-۲- نتایج بررسی معیارهای پوشش، محبوبیت و

واسط کاربری

نتایج بررسی برای معیارهای پوشش، محبوبیت و واسط کاربری در جدول (۱) قابل مشاهده است. همانطوری که دیده می‌شود، برای هر یک از این معیارها، با توجه زیرمعیارها و شاخص‌های تعریف شده، نتایج ارزیابی در جدول مذکور به تفکیک بیان گردیده است.

۳-۳- نتایج بررسی معیار کارایی

همانطوری که در بخش ۲ گفته شد، در این قسمت با بهره‌گیری از روشی خودکار، کارایی موتورهای جستجوی مختلف از دیدگاه نتایج بازگشتی مورد بررسی قرار می‌گیرد. در ادامه، ابتدا مجموعه پرس و جوهای مورد استفاده در این بررسی تشریح می‌گردند؛ سپس نتایج حاصل از این بررسی ارائه خواهند شد.

۳-۳-۱- مجموعه پرس و جوهای مورد استفاده

در این بررسی، از دو مجموعه پرس و جو مختلف استفاده شده است: (۱) مجموعه همشهری که حاوی ۱۰۰ پرس و جو مختلف است و در سال‌های ۲۰۰۸ و ۲۰۰۹ در کنفرانس CLEF ارائه شده است [۲۶] و (۲) مجموعه برگرفته شده از وبسایت^{۱۲} Google trend که حاوی ۳۸ پرس و جوی است؛ این پرس و جوها نشان دهنده پرکاربردترین موارد در میان کاربران ایرانی در بین سال‌های ۲۰۰۷ الی ۲۰۱۲ می‌باشند. نکته حائز اهمیت در اینجا این است که مجموعه اول، حاوی ۱۰۰ پرس و جو استاندارد است. این پرس و جوها بر اساس مجموعه استاندارد همشهری طرح شده‌اند و به طور متوسط هر پرس و جو ۳.۵ کلمه را شامل می‌گردد. این پرس و جوها را بسته به طول پرس و جو (تعداد کلمات)، می‌توان از طیف ساده تا دشوار دسته‌بندی نمود. اما بیشتر این پرس و جوها متوسط و یا دشوار می‌باشند. از سوی دیگر، مجموعه پرس و جوهای پرکاربرد معمولاً تک کلمه‌ای می‌باشند؛ و به دلیل فراوانی اسناد مرتبط با آنها در سطح وب، می‌توان آنها را پرس و جوهای ساده به شمار آورد. به این ترتیب، ترکیبی از پرس و جوهای

۳-۲-۴- کارایی

در این پژوهش، با استفاده از رویکردی خودکار به سنجش کارایی موتورهای جستجوی فارسی از دیدگاه کیفیت نتایج بازگشتی پرداخته خواهد شد. این رویکرد که شمارش مرجع^{۱۳} [۱۹، ۲۵] نام دارد، بر پایه استدلالی ساده بنا نهاده شده است: اگر در ازای پرس و جوی مشخصی همچون q، سند خاصی همچون d توسط بیشتر از یک موتور جستجو بازگردانده شود، سند d مرتبط در نظر گرفته خواهد شد. روش شمارش مرجع دارای نسخه‌های گوناگونی است؛ در کاملترین نسخه از این روش، میزان ارتباط یک سند با یک پرس و جوی خاص، علاوه بر تعداد موتورهای جستجویی که آن را بازگردانده‌اند، به جایگاه سند مذکور در بین نتایج بازگردانده شده نیز بستگی خواهد داشت.

اما جهت سنجش کارایی از شاخص‌هایی همچون Mean-Average-Precision (MAP) و Precision-At-N (P@N) به ازای مقدار مشخص برای N استفاده می‌گردد. برای محاسبه این شاخص‌ها، تنها کافی است که مرتبط بودن و یا نبودن یک سند خاص در ازای پرس و جوی مطرح شده تعیین گردد. سپس، هر یک از شاخص‌های مذکور به سادگی قابل اندازه‌گیری خواهند بود.

۳- نتایج حاصل از ارزیابی و مقایسه موتورهای

جستجوی فارسی

در این بخش به بیان نتایج حاصل از ارزیابی و مقایسه موتورهای جستجوی فارسی پرداخته خواهد شد. پیش از پرداختن به نتایج، ابتدا موتورهای جستجوی تحت بررسی معرفی خواهند شد. سپس نتایج حاصل از بکارگیری هر یک از معیارهای معرفی شده در بخش‌های قبلی بیان خواهند شد. همچنین در صورت نیاز، شرایط انجام آزمایش و منابع داده‌ای مورد استفاده نیز تشریح خواهند گردید.

جدول (۱). معرفی معیارهای مورد استفاده برای ارزیابی موتورهای جستجوی فارسی به همراه شاخص‌های اندازه‌گیری و نیز نتایج ارزیابی

فراجویشرهای فارسی				موتورهای جستجوی فارسی		عمومی			نوع موتور جستجو		نام زیرمعیار	معیار	
ت:ج	سلام	خا:اساس	پارسین	آ:ا	پ:ا	ز:ا	خ:ا	پ:ا	شاخص مورد استفاده برای اندازه‌گیری				
۴ موتور جستجو (گوگل، بینگ، یاهو و اسک)	۴ موتور جستجو (گوگل، بینگ، یاهو و زال)	۲ موتور جستجو (گوگل و بینگ)	۲ موتور جستجو (گوگل و بینگ)	کمتر از ۱ میلیون صفحه	حدود ۱۲۰ میلیون صفحه	کمتر از ۲۰۰ میلیون صفحه	استفاده از نتایج بینگ	حدود ۷۰۰ میلیون صفحه			-	پوشش	
x	x	x	x	x	✓	x	✓	✓	جستجوی پیشرفته		قابلیت‌های جستجو	واسط کاربری	
x	✓	x	x	x	✓	x	x	✓	تصحیح خطاهای نوشتاری				
x	✓	x	x	x	x	x	✓	✓	پیشنهاد تکمیل پرس‌وجو				
x	x	x	x	x	✓	✓	✓	✓	تعیین عبارات در پرس‌وجوها				
✓	✓	✓	✓	x	x	x	x	✓	استفاده از عملگرهای منطقی در پرس و جو				
x	✓	x	x	x	x	✓	✓	✓	جستجو بر اساس	نوع سند			
x	✓	x	x	x	x	✓	x	✓		تاریخ			
x	x	x	x	x	x	✓	x	✓		عنوان			
x	x	x	x	x	x	✓	✓	✓		آدرس			
✓	✓	✓	✓	x	✓	✓	✓	✓	محدود به یک آدرس خاص				
✓	✓	✓	✓	x	✓	✓	✓	✓	جستجو کلمات عمومی				
x	✓	x	x	x	✓	✓	✓	✓	گروه‌بندی نتایج		نمایش نتایج		
x	✓	x	x	x	x	✓	✓	✓	پیشنهاد پرس‌وجوی مرتبط				
x	x	✓	✓	✓	✓	✓	✓	✓	نمایش تعداد نتایج				
x	x	x	x	✓	✓	x	x	✓	نمایش مدت زمان پاسخگویی				
x	x	x	✓	✓	✓	✓	✓	✓	مشخص کردن لغات پرس‌وجو در نتایج				
✓	✓	✓	✓	✓	✓	✓	✓	✓	قابلیت جستجوی انواع محتوا	متن			
x	x	x	✓	x	✓	✓	✓	✓		تصویر			
x	x	x	x	x	x	✓	✓	✓		ویدئو			
x	x	x	✓	x	✓	✓	✓	✓		علمی			
x	x	✓	x	x	x	✓	✓	✓	نقشه				
x	✓	x	✓	✓	✓	x	x	x	قابلیت‌های پردازشی زبان فارسی				
۲۳۳۸	۹۹۹	۱۸۱۲۸	۱۴۸	۱۵۰۳	۸۶۲	۳۰	۲	۱	میزان ترافیک (برگرفته شده از الکسا)		محبوبیت		
۵۱	۱۷۲	۱۹۶	۳۱۶۸	۶	۱۶۱	۱۶۸۸	۱۹۸۴۸	۱۹۲۳۸	تعداد پیوندهای ورودی از سوی دیگر وبسایت‌ها (برگرفته شده از الکسا)				

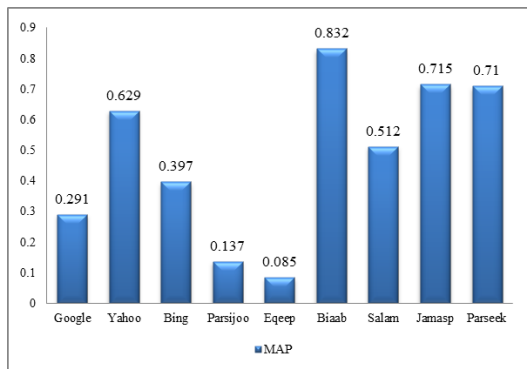
است. حال می‌توان برای تعیین مرتبط بودن و یا نبودن اسناد، حد مشخصی را به عنوان آستانه‌ی ارتباط^{۱۴} تعیین نمود. در ساده‌ترین حالت، این آستانه برابر با صفر در نظر گرفته می‌شود. در آینده می‌توان با استفاده از رویکردهای یادگیری ماشین^{۱۵}، مقادیر دقیق‌تری را به عنوان حد آستانه تعیین نمود.

• نتایج ارزیابی بر اساس شاخص MAP: نتایج ارزیابی با استفاده از معیار MAP در نمودارهای ۱- الف و ۱- ب به ترتیب به ازای مجموعه‌های همشهری و Google trend به نمایش در آمده است. در اینجا، به ازای هر پرس و جو ۱۰۰ نتیجه برتر بازگردانی شده است، اما لزوماً تمامی موتورهای جستجو تعداد ۱۰۰ نتیجه را به ازای هر پرس و جو بازنگردانده‌اند؛ به طور مثال فراجویشرگر

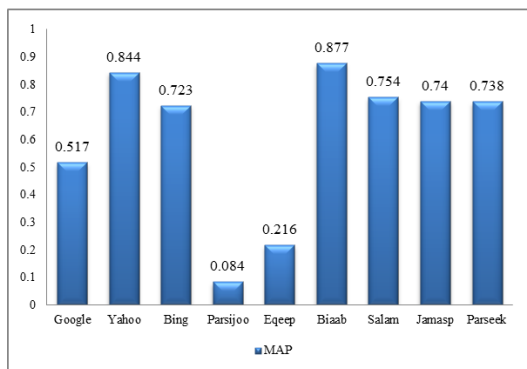
ساده، متوسط و دشوار برای ارزیابی موتورهای جستجو بکار گرفته خواهند شد.

۳-۳-۲- نتایج حاصل از ارزیابی خودکار به روش شمارش مرجع

در این پژوهش، از نسخه چهارم الگوریتم شمارش مرجع [۱۹] برای تعیین مرتبط بودن و یا نبودن اسناد بازگشتی توسط یک موتور جستجو به ازای پرس و جویی مشخص در مقایسه با سایر موتورهای جستجو استفاده می‌گردد. بر اساس این الگوریتم، هر سند امتیازی مبنی بر میزان ارتباط با پرس و جوی مطرح شده کسب می‌نماید؛ جایگاه اسناد در میان نتایج بازگشتی نیز در محاسبه این امتیاز موثر

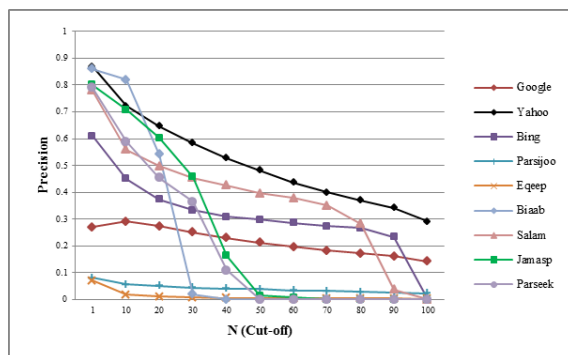


الف-۱

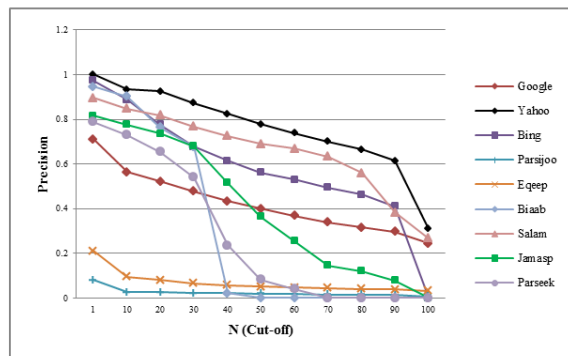


ب-۱

نمودار (۱). نتایج کارایی بر اساس معیار MAP؛ الف) بهره‌گیری از مجموعه همشهری. ب) بهره‌گیری از مجموعه Google trend



الف-۲



ب-۲

نمودار (۲). نتایج کارایی بر اساس معیار P@N؛ الف) بهره‌گیری از مجموعه همشهری. ب) بهره‌گیری از مجموعه Google trend

پارسیک، معمولاً تعداد کمتر از ۴۰ نتیجه را به ازای هر دو مجموعه پرس و جوی مورد بررسی بازگردانده است.

نتایج ارزیابی بر اساس شاخص P@N: همانطوری که پیش از این نیز اشاره گردید، معیار Precision در ۱۱ سطح مختلف برای N مورد محاسبه قرار می‌گیرد. نتایج حاصل از این بررسی با استفاده از مجموعه‌های همشهری و Google trend به ترتیب در نمودارهای الف-۲ و ب-۲ نمایش داده شده است. بر اساس نمودارهای مذکور دیده می‌شود که مقدار دقت بدست آمده برای فراجویشگرها، معمولاً با افزایش مقدار N با کاهش شدید مواجه می‌شود. دلیل اصلی این کاهش، تعداد نتایج کمی است که به ازای پرس و جوهای مختلف (خصوصاً پرس و جوهای متوسط و یا دشوار) بازمی‌گردانند.

۴- نتیجه‌گیری و تحلیل

در این پژوهش به ارائه چارچوبی کارا در راستای ارزیابی موتورهای جستجوی فارسی پرداخته شد. طراحی این چارچوب بر مبنای دو هدف مهم شکل گرفت: بررسی همه جانبه ویژگی‌ها و قابلیت‌های موتورهای جستجو؛ و نیز کاهش هزینه‌های مربوط به ارزیابی و مقایسه. نتایج استفاده از چارچوب پیشنهادی در راستای مقایسه و ارزیابی موتورهای جستجوی فارسی که در بخش قبلی ارائه شد، در چندین قسمت قابل تحلیل است. نتایج ارزیابی محبوبیت نشان می‌دهد که به طور کلی موتورهای جستجوی عمومی از محبوبیت بیشتری برخوردار هستند. اما در میان جویشگرهای فارسی، می‌توان دید که فراجویشگر پارسیک محبوبتر از سایر موارد است.

از نقطه نظر مقایسه ویژگی‌های واسط کاربری، می‌توان مدعی بود که موتور جستجوی گوگل در تمامی شاخص‌ها سرآمد است؛ و این خود می‌تواند دلیلی بر محبوبیت این جستجوگر در میان کاربران فارسی‌زبان باشد. تنها انتقادی که می‌توان در رابطه با گوگل مطرح ساخت، ضعف در پردازش زبان فارسی در مقایسه با سایر زبان‌های شناخته شده است. اما موتورهای جستجوی همچون پارسی جو و اکیپ و همچنین فراجویشگرهای پارسیک و سلام، در راستای برآوردن این نیاز اقدامات مناسب‌تری را انجام داده‌اند. قابل ذکر است که جویشگرهای پارسی جو و سلام امکان پیشنهاد پرس‌وجوی صحیح را نیز در هنگام بروز خطای املائی فراهم می‌نمایند.

در قسمت کارایی، دیده می‌شود که معمولاً موتورهای جستجوی عمومی همچون یاهو، کارایی بالاتری را ارائه می‌دهند. بخصوص اینکه این کارایی بر اساس نمودارهای P@N تقریباً در بسیاری از سطوح N مشخص است. از سوی دیگر، فراجویشگرهای فارسی نیز کارایی مناسبی را بر اساس شاخص‌های معرفی شده ارائه می‌دهند. بطوریکه فراجویشگرهای بیاب و سلام مناسب‌ترین نتایج را به ترتیب بر اساس معیارهای MAP و P@N بدست آورده‌اند. با این وجود، باید در نظر داشت که فراجویشگرها صرفاً به ترکیب نتایج موتورهای جستجو

- Information Science and Management (IJISM), vol. 8, pp. 75-85, 2012.
- [16] F. C. Johnson, J. R. Griffiths, and R. J. Hartley, DEVISE: a framework for the evaluation of Internet search engines vol. 100, 2001.
- [17] K. Srinivas, P. Srinivas, and A. Govardhan, "A survey on the performance evaluation of various meta search engines," International Journal of Computer Science Issues, vol. 8, pp. 359-364, 2011.
- [18] I. Soboroff, C. Nicholas, and P. Cahan, "Ranking retrieval systems without relevance judgments," in Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, 2001, pp. 66-73.
- [19] S. Wu and F. Crestani, "Methods for ranking information retrieval systems without relevance judgments," in Proceedings of the 2003 ACM symposium on Applied computing, 2003, pp. 811-816.
- [20] F. Can, R. Nuray, and A. B. Sevdik, "Automatic performance evaluation of Web search engines," Information processing & management, vol. 40, pp. 495-514, 2004.
- [21] Y. Liu, Y. Fu, M. Zhang, S. Ma, and L. Ru, "Automatic search engine performance evaluation with click-through data analysis," in Proceedings of the 16th international conference on World Wide Web, 2007, pp. 1133-1134.
- [22] Y. Kim, S. Lee, B. Kang, and R. Williams, "Coverage and delay forecast modeling of search engine services," in PKAW, 2008, pp. 62-73.
- [23] A. Pirkola, "The Effectiveness of Web Search Engines to Index New Sites from Different Countries," Information Research: An International Electronic Journal, vol. 14, 2009.
- [24] R. Pasley, P. Clough, R. S. Purves, and F. A. Twaroch, "Mapping geographic coverage of the web," in Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems, 2008, p. 19.
- [25] S. Wu and F. Crestani, "Data fusion with estimated weights," in Proceedings of the eleventh international conference on Information and knowledge management, 2002, pp. 648-651.
- [26] A. AleAhmad, H. Amiri, E. Darrudi, M. Rahgozar, and F. Oroumchian, "Hamshahri: A standard Persian text collection," Knowledge-Based Systems, vol. 22, pp. 382-387, 2009.

پردازنده و معمولاً کار اضافه‌ای را در زمینه خزش، نمایه‌سازی و یا جستجو انجام نمی‌دهند. از سوی دیگر، قابل مشاهده است که عملکرد پارسی‌جو بر اساس پرس و جوهای متوسط و یا دشوار (مجموعه همشهری) بهتر از اکیپ می‌باشد. اما در قبال پرس و جوهای ساده (مجموعه Google trend)، اکیپ عملکرد مناسب‌تری را ارائه می‌دهد. بنابراین در کل می‌توان پارسی‌جو را مناسب‌ترین موتور جستجوی فارسی به لحاظ کارایی به شمار آورد.

مراجع

- [1] R. Ali and M. M. S. Beg, "An overview of Web search evaluation methods," Computers & Electrical Engineering, vol. 37, pp. 835-848, 2011.
- [2] H. Chu and M. Rosenthal, "Search engines for the World Wide Web: A comparative study and evaluation methodology," in PROCEEDINGS OF THE ANNUAL MEETING-AMERICAN SOCIETY FOR INFORMATION SCIENCE, 1996, pp. 127-135.
- [3] D. Harman, "The First Text Retrieval Conference (TREC-1)," 1993.
- [4] C. Peters, Cross-Language Information Retrieval and Evaluation: Workshop of Cross-Language Evaluation Forum, CLEF 2000, Lisbon, Portugal, September 21-22, 2000, Revised Papers vol. 2069: Springer, 2001.
- [5] D. Lewandowski, "The retrieval effectiveness of web search engines: considering results descriptions," Journal of Documentation, vol. 64, pp. 915-937, 2008.
- [6] D. Lewandowski, "A Framework for Evaluating the Retrieval Effectiveness of Search Engines," 2012.
- [7] J. Griesbaum, "Evaluation of three German search engines: Altavista. de, Google. de and Lycos. de," Information Research, vol. 9, pp. 9-4, 2004.
- [8] B. J. Jansen and P. R. Molina, "The effectiveness of Web search engines for retrieving relevant ecommerce links," Information Processing & Management, vol. 42, pp. 1075-1098, 2006.
- [9] J. Véronis, "A comparative study of six search engines," URL: <http://www.up.univmrs.fr/veronis/pdf/2006-comparative-study.pdf>, 2006.
- [10] B. J. Jansen, "The comparative effectiveness of sponsored and nonsponsored links for Web e-commerce queries," ACM Transactions on the Web (TWEB), vol. 1, p. 3, 2007.
- [11] W. Tawileh, J. Griesbaum, and T. Mandl, "Evaluation of five web search engines in Arabic language," Proceedings of LWA2010, 2010.
- [12] M. Buzzi, P. Andronico, and B. Leporini, "Accessibility and usability of search engine interfaces: preliminary testing," in Proc. of 8th ERCIM UI4ALL Workshop, 2004.
- [13] A. Isfandyari-Moghaddam and Z. Bahari-Movaffagh, "Evaluating and comparing search features of Web metasearch engines: A checklist-based approach," Malaysian Journal of Library & Information Science, vol. 15, pp. 1-17, 2010.
- [14] M. Kaur, N. Bhatia, and S. Singh, "Web Search Engines Evaluation Based on Features And End-User Experience," International Journal of Enterprise Computing and Business Systems, vol. 1, 2011.
- [15] M. Erfanmanesh and F. Didegah, "Evaluating Function of Persian Search Engines on the Web Using Correspondence Analysis," International Journal of

زیرنویس‌ها

¹ Text REtrieval Conference

² Conference and Labs of the Evaluation Forum

³ Index quality

⁴ Usability

⁵ Subjective

⁶ Coverage

⁷ Effectiveness

⁸ User interface

⁹ Popularity

¹⁰ Stop words

¹¹ www.alexam.com

¹² Reference count

¹³ <http://www.google.com/trends/>

¹⁴ Relevancy threshold

¹⁵ Machine learning