

بررسی کارآمدی ویژگی محدودیت زبان موتورهای کاوش: مطالعه موردی بازیابی نتایج فارسی و عربی در گوگل، بینگ و یاهو

مریم یقظین^۱

عبدالرسول جوکار^۲

چکیده

هدف: هدف پژوهش حاضر بررسی کارآمدی ویژگی محدودیت زبان در سه موتور کاوش گوگل، بینگ و یاهو و تعیین کارآمدترین موتور کاوش در ارائه نتایج مرتبط با پرس‌وجوهایی با محدودیت زبان فارسی و عربی است.

روش / رویکرد پژوهش: این پژوهش با روش پیمایشی-توصیفی انجام شده است. نمونه مورد پژوهش، ۳۸ کلمه هم‌نویسه میان زبان فارسی و عربی، اما با معانی متفاوت میان دو زبان است.

یافته‌ها: نتایج حاصل از پژوهش نشان داد که پس از محدودیت زبان به زبان فارسی و عربی، موتور کاوش گوگل به لحاظ بازیابی تعداد و رتبه‌بندی نتایج مرتبط با زبان هدف، عملکرد ضعیف‌تری نسبت به موتور کاوش بینگ و یاهو دارد. علاوه بر این سه موتور کاوش در بازیابی نتایج غیرمرتبط با زبان هدف تفاوت معنی‌داری با یکدیگر ندارند. اما این سه موتور کاوش در رتبه‌بندی نتایج غیرمرتبط با زبان هدف تفاوت معنی‌داری با یکدیگر نشان دادند و این تفاوت نیز به موتور کاوش گوگل بازمی‌گردد. همچنین عملکرد موتورهای کاوش در بازیابی نتایج غیرمرتبط با زبان هدف به ترافیک و یا محبوبیت صفحات غیرمرتبط بازمی‌گردد.

نتیجه‌گیری: جستجوگرانی که از ویژگی محدودیت زبان موتورهای کاوش به منظور محدود کردن نتایج جستجو استفاده می‌کنند، در بینگ و یاهو از بخت بالاتری برای دستیابی به نتایج مرتبط برخوردارند.

کلیدواژه‌ها: موتور کاوش، محدودیت زبان، کارآمدی، بازیابی اطلاعات، رتبه بندی نتایج

^۱ دانشجوی دکتری علم اطلاعات و دانش‌شناسی، دانشگاه شیراز (نویسنده مسئول) yaghtin.maryam@gmail.com

^۲ عضو هیأت علمی گروه علم اطلاعات و دانش‌شناسی، دانشگاه شیراز ajowkar2003@yahoo.com

با ظهور اینترنت، موتورهای کاوش به ابزار مهمی برای دستیابی به اطلاعات تبدیل شده‌اند. اغلب کاربران وب از موتورهای کاوش برای جستجو و دستیابی به اطلاعات مرتبط با نیازهای اطلاعاتی خود استفاده می‌کنند. اما با اینکه موتورهای کاوش بسیاری در دسترس است، گاهی آنان نمی‌توانند به سرعت به اطلاعات مطلوب خود دست یابند. از سوی دیگر، هر موتور کاوش نیز، نتایج متفاوت برای پرس‌وجویی^۱ یکسان ارائه می‌دهد (کاشیک^۲، ۲۰۱۲). بنابراین ارزیابی و تعیین موتور کاوش مناسب برای دریافت مرتبط‌ترین اطلاعات در وب از اهمیت بسیاری برخوردار است. در واقع بررسی کیفیت و مقایسه موتورهای کاوش مختلف به کاربران وب کمک می‌کند تا موتور کاوش مناسب با نیاز اطلاعاتی خود را شناسایی کنند.

بررسی و ارزیابی کیفیت موتورهای کاوش شامل چهار بعد کیفیت نمایه، کیفیت نتایج، قابلیت استفاده موتور کاوش و کیفیت ویژگی‌های جستجو، است. منظور از کیفیت ویژگی‌های جستجو، عملکرد صحیح مجموعه‌ای از ویژگی‌های جستجو مانند جستجوی عنوان، جستجوی عبارت خاص، جستجوی بولی، محدود کردن جستجو به صفحات در یک زبان خاص، یک کشور یا ناحیه خاص، صفحات مشابه، محدودیت نوع فایل، جستجوی آدرس وبی، محدودیت تاریخ و ... است. استفاده از ویژگی‌های جستجو به منظور دستیابی به اطلاعات مرتبط‌تر و تامین نیازهای اطلاعاتی خاص کاربران وب مفید بوده و سبب می‌شوند تا آنان بازیابی بهتری را تجربه کنند (لواندوسکی، هشتاتر^۳، ۲۰۰۸). اما میزان عملکرد صحیح برخی ویژگی‌های جستجو مانند صفحات مشابه، محدودیت تاریخ و محدودیت زبان در موتورهای کاوش مختلف مشخص نیست (لواندوسکی، ۲۰۰۴؛ ۲۰۰۸).

منظور از ویژگی محدودیت زبان، زبان صفحه هدف است. از آنجا که زبان صفحه هدف تعیین می‌کند که آیا کاربران وب می‌توانند نتایج جستجو را درک کنند یا خیر، از اهمیت بسیاری برخوردار است. کاربرانی که اطلاعات را در یک زبان خاص جستجو می‌کنند، ممکن است نتوانند صفحات دیگر زبان‌ها را بخوانند. از سوی دیگر، در صورتی که پرس‌وجوها، کلمات هم‌نویسه اما با معانی متفاوت میان زبان‌های مختلف باشند، نتایج بازیابی شده از دیگر زبان‌ها حتی در صورت امکان درک توسط کاربران، غیرمرتبط محسوب می‌شوند. با عملکرد صحیح ویژگی محدودیت زبان در موتورهای کاوش، صفحات غیرمرتبط و نامطلوب کاربر از مجموعه نتایج خارج می‌شوند. لازم به ذکر است شناسایی خودکار زبان متون یک مسئله حل شده است (کاونار و ترنکل^۴، ۲۰۰۴؛ دانینگ^۵، ۱۹۹۴)، و تنها نیاز به الگوریتمی برای تعیین زبان متون موجود در وب وجود دارد

¹. Query

². Kaushik

³. Lewandowski & Höchstötter

⁴. Cavnar & Trenkle

⁵. Dunning

(مارتینز و سیلوا^۱، ۲۰۰۵). علاوه بر این، عواملی مانند مکان سرور نیز می‌توانند به تشخیص زبان صفحات وب کمک کنند (لواندوسکی، ۲۰۰۸). بنابراین با عملکرد صحیح ویژگی محدودیت زبان در موتورهای کاوش، نباید صفحاتی به جز صفحات زبان هدف در نتایج ظاهر گردند.

از سوی دیگر، با افزایش روزافزون کاربران و صفحات غیرانگلیسی زبان، عملکرد صحیح موتورهای کاوش به منظور دسترسی به اطلاعات با زبان موردنظر کاربر اهمیت بسیاری یافته است. همچنین، با توجه به اقبال عمومی کاربران فارسی زبان به استفاده از موتورهای کاوش بین‌المللی، مهجوری زبان فارسی در مقایسه با زبان‌هایی چون انگلیسی، پایین بودن شمار گویشوران فارسی در سراسر جهان، کمبود موتور کاوش خاص زبان فارسی و شباهت ظاهری بسیار میان زبان فارسی و عربی، پژوهش در این باره اهمیتی ویژه می‌یابد تا مشخص گردد ویژگی محدودیت زبان موتورهای کاوش بین‌المللی در بازایی صحیح نتایج فارسی و عربی به چه میزان کارآمد است.

از این رو، پژوهش حاضر می‌کوشد تا ویژگی محدودیت زبان را در سه موتور کاوش گوگل، بینگ و یاهو بررسی و مقایسه کند تا از این رهگذر ضمن روشن ساختن کارآمدی موتورهای کاوش در بازایی نتایج فارسی و عربی، مشخص گردد که کدام موتور کاوش برای ارائه نتایج مرتبط با پرس‌وجوهایی با محدودیت زبان مناسب‌تر هستند. به این منظور در این پژوهش، پرس‌وجوهای هم‌نویسه میان زبان فارسی و عربی شناسایی شده و کارآمدی سه موتور کاوش در بازایی نتایج مرتبط با هر پرس‌وجو بررسی و مقایسه گردیده است.

پژوهش‌های متعددی در حوزه موتورهای کاوش صورت گرفته است اما در زمینه بررسی کارآمدی موتورهای کاوش به لحاظ ویژگی‌های جستجو کمتر تلاشی صورت گرفته است. از مهمترین تلاش‌های صورت گرفته در این حوزه می‌توان به پژوهش ژانگ، فی و لی^۲ (۲۰۱۳) اشاره کرد. آنان با بررسی و مقایسه کارآمدی ویژگی‌های جستجو مانند جستجوی عنوان، جستجوی پایه، جستجوی عبارت خاص، جستجوی پی‌دی‌اف و جستجوی آدرس وبی^۳ در موتورهای کاوش گوگل، گوگل چینی، و بایدو^۴ دریافتند که موتور کاوش گوگل دارای بهترین عملکرد بازایی به لحاظ پنج ویژگی جستجو است. علاوه بر این موتور کاوش گوگل بهترین عملکرد رتبه‌بندی صفحات را به خود اختصاص داده است. نتایج پژوهش کیم، فیلد و کارترایت^۵ (۲۰۱۲) پیرامون بررسی رفتار کاربران در جستجوی کتاب دیجیتال در کتابخانه دیجیتال و موتورهای کاوش وب و محدودیت زبان در جستجوی پیشرفته، حاکی از آن است که کاربران اغلب برای محدود کردن نتایج جستجو از فیلدهای عنوان،

¹. Martins & Silva

². Zhang, Fei & Le

³. URL

⁴. Baidu

⁵. Kim, Field & Cartright

نویسندگان، ناشران و زبان به ترتیب استفاده می‌کنند. لواندوسکی (۲۰۰۸) نیز دریافت که موتورهای کاوش گوگل و ام‌اس‌ان هنگام محدود کردن نتایج به زبان خارجی به درستی عمل نمی‌کنند. فتاحی^۱ و همکاران (۲۰۰۸) در پژوهشی با انتخاب دو ویژگی جستجوی پیشرفته (جستجوی عنوان دقیق و جستجوی آدرس وبی) به بررسی تاثیر اصطلاحات غیرموضوعی و نیمه موضوعی در گسترش پرس‌وجو پرداختند. نتایج حاصل از پژوهش آنان نشان داد که در صورت محدود کردن پرس‌وجو به جستجوی عنوان دقیق و جستجوی آدرس وبی، نتایج جستجو بهبود می‌یابد. ژانگ و لین^۲ (۲۰۰۷) نیز معتقدند با توجه به انتشار اطلاعات در زبان‌های مختلف، پشتیبانی موتورهای کاوش از زبان‌های متفاوت اهمیت بسیاری دارد. آنان دریافتند که موتورهای کاوش گوگل، ای‌زدتوفایند^۳ و آن‌لاین‌لینک^۴ به ترتیب، بهترین موتورهای کاوش به لحاظ پشتیبانی از زبان‌های متفاوت هستند. همچنین با اینکه موتورهای کاوش بسیاری از زبان‌های مختلف پشتیبانی می‌کنند، اما تنها تعداد اندکی هستند که ویژگی ترجمه بین‌زبانی را به کار می‌برند. علاوه بر این، ویژگی پشتیبانی از زبان‌های مختلف در موتورهای کاوش در سطح لغوی (و نه معنایی) باقی مانده است. غانی، جونز و لدنیک^۵ (۲۰۰۵) در بخشی از پژوهش خود به بررسی عملکرد محدودیت زبان در صفحات وب اسلوونیایی پرداختند. آنان ۱۰۰ صفحه وب اسلوونیایی را از میان چندین هزار صفحه وب اسلوونیایی زبان که توسط فیلتر زبان محدود شده است، به طور تصادفی انتخاب کرده و مورد بررسی قرار دادند. نتایج حاصل از پژوهش آنان نشان داد که ۹۹ درصد صفحات به زبان اسلوونیایی هستند. علاوه بر این، ۹۰ تا ۹۵ درصد صفحاتی که در دسته غیراسلوونیایی زبان دسته‌بندی شده بودند واقعا غیراسلوونیایی بودند. نتایج پژوهش لواندوسکی (۲۰۰۴) پیرامون جستجوی پرس-وجوی‌هایی با محدودیت تاریخی در موتورهای کاوش گوگل، تئوما^۶ و یاهو نشان داد که جستجو با محدودیت محدودیت تاریخی در موتورهای کاوش مذکور به درستی کار نمی‌کنند. همچنین، در مجموع موتور کاوش گوگل نسبت به دیگر موتورهای کاوش مورد بررسی بهتر عمل کرده، اما برای هر پرس‌وجو به تفکیک بهترین عملکرد را نداشت. اسروکا^۷ (۲۰۰۰) نیز در پژوهشی به بررسی توانایی موتورهای کاوش وب برای بازیابی اطلاعات لهستانی پرداخت. موتورهای کاوش مورد بررسی نسخه لهستانی دو موتور کاوش عمومی اینفوسیک^۸ و آلتاویستا^۹ و سه موتور کاوش بومی بودند. نتایج حاصل از پژوهش وی نشان داد که موتورهای کاوش

1. Fattahi

2. Zhang & Lin

3. EZ2Find

4. Onlinelink

5. Ghani, Jones & Mladenic

6. Teoma

7. Sroka

8. Infoseek

9. AltaVista

اینفوسیک و اونت. پی‌ال^۱ دارای بیشترین میزان دقت در بازیابی اطلاعات مرتبط بودند. علاوه بر این موتور کاوش اینفوسیک، بیشترین سرعت بازیابی را به خود اختصاص داد. رجبی و نوروزی (۱۳۹۴) در پژوهشی به شناسایی و ارزیابی امکانات جستجو در موتورهای جستجوی فارسی، تعیین کارآمدترین موتور جستجو به لحاظ بازیابی اطلاعات مرتبط، شناسایی موتورهای جستجوی کارآمد از حیث محاسبه میزان جامعیت و مانعیت و تعیین میزان همپوشانی میان آن‌ها پرداختند. آنان دریافتند که موتورهای جستجوی فارسی از امکانات جستجوی مناسبی برخوردارند و می‌توان تا حدودی از عملکرد آن‌ها اطمینان حاصل کرد. اما به لحاظ بازیابی نتایج مرتبط و میزان جامعیت، بین موتورهای جستجو فاصله و تفاوت معناداری وجود دارد. نتایج پژوهش دری (۱۳۹۳) نشان داد که موتورهای کاوش معنایی هاکیا^۲، داک داک گو^۳، کلاز^۴، لکس^۵ و فکت بایتز^۶ در شاخص شاخص ویژگی‌های عادی و معنایی، عملکرد مطلوب و کارایی مورد انتظار را نداشتند. وی معیار جستجو در زبان خاص را نیز مورد بررسی قرار داد و نشان داد که هیچیک از موتورهای کاوش مورد بررسی دارای این ویژگی نیستند. شریف (۱۳۹۱) نیز با استفاده از روش تجربی، تغییرات عملکرد یاهو و گوگل از نظر پوشش کمی-زمانی نمایه‌سازی و نقش عناصر فراداده‌ای در رتبه‌بندی صفحه‌های وب را بررسی کرد و نشان داد که این دو موتور کاوش در بعد کمی-زمانی و بعد کیفی عناصر ابرداده‌ای تفاوت چندانی در عملکرد خود نداشته‌اند. محمداسمعیل و منصور کیایی (۱۳۹۰) نیز دریافتند که موتور کاوش یاهو و ابرموتور کاوش کری‌گاید^۷ بیشترین مدارک فیزیک را بازیابی کرده و موتور کاوش عمومی آ-اوال^۸ و ابر موتور کاوش اینفو بیشترین درصد همپوشانی را با سایر موتورهای کاوش عمومی و ابرموتورها دارند.

در مجموع، با توجه به پژوهش‌های صورت گرفته در حوزه ارزیابی موتورهای کاوش، می‌توان چنین نتیجه گرفت که موتور کاوش گوگل در برخی ویژگی‌های جستجو مانند جستجوی عنوان، جستجوی عبارت خاص و ... بهترین عملکرد را نسبت به سایر موتورهای کاوش مورد بررسی داشته است. اما در ویژگی جستجوی محدودیت زبان (زبان انگلیسی و آلمانی)، موتورهای کاوش گوگل و ام‌اس‌ان به خوبی عمل نکرده‌اند. با این حال، تا کنون پژوهشی به بررسی کارآمدی ویژگی محدودیت زبان موتورهای کاوش بین‌المللی در تشخیص میان نتایج فارسی و عربی نپرداخته است. بنابراین نیاز به انجام چنین پژوهشی جهت ارزیابی کیفی موتورهای

1. Onet.pl

2. Hakia

3. DuckDuckGo

4. Cluuz

5. Lexxe

6. Factbites

7. Curry Guide

8. AOL

کاوش و تعیین میزان کارآمدی آنان در تشخیص میان نتایج فارسی و عربی، ضروری به نظر می‌رسد. نتایج حاصل از آن سبب می‌شود که کاربران وب به ویژه کاربران فارسی زبان بتوانند در تامین نیازهای اطلاعاتی خود که نیاز به محدودیت زبان دارند، موتور کاوش مناسب را انتخاب کرده و به اطلاعات مرتبط تر دست یابند.

پرسش‌ها

۱. پس از محدودیت زبان به زبان هدف (فارسی یا عربی)، چه درصدی از نتایج موتور کاوش گوگل، بینگ و یاهو غیرمرتبط با زبان هدف (فارسی یا عربی) است؟
۲. وضعیت رتبه‌بندی نتایج غیرمرتبط با زبان هدف در موتور کاوش گوگل، بینگ و یاهو چگونه است؟
۳. آیا تفاوت معنی‌داری میان موتورهای کاوش مورد مطالعه به لحاظ میانگین نتایج غیرمرتبط با زبان هدف وجود دارد؟
۴. آیا تفاوت معنی‌داری میان موتورهای کاوش مورد مطالعه به لحاظ میانگین رتبه‌های نتایج غیرمرتبط با زبان هدف وجود دارد؟
۵. آیا ترافیک صفحات غیرمرتبط با زبان هدف بر رتبه‌بندی آنان در سه موتور کاوش گوگل، بینگ و یاهو تاثیرگذار است؟

روش‌شناسی

پژوهش حاضر با استفاده از روش پیمایشی-توصیفی انجام شده است. در گام اول، ۳ موتور کاوش گوگل، بینگ و یاهو انتخاب شدند. دلیل استفاده از موتورهای کاوش مذکور در وهله اول، پوشش بین‌المللی و پشتیبانی آن‌ها از زبان فارسی و عربی بوده است. علاوه بر این، سه موتور کاوش مذکور، در سال ۲۰۱۵ به ترتیب در صدر موتورهای کاوش محبوب قرار داشته‌اند^۱. فرض پژوهش حاضر این است که این گرایش جهانی برای ایران نیز صادق است.

در گام بعد، سیاهه‌ای از پرس‌وجوها با جستجو در واژه‌نامه‌های فارسی و عربی متعددی گردآوری گردید (به طور مثال مسعود، ۱۳۸۲). نمونه مورد پژوهش، ۳۸ کلمه هم‌نویسه میان زبان فارسی و عربی، اما با معانی متفاوت میان دو زبان است. به عبارت دیگر، پرس‌وجوهای انتخاب شده، کلماتی بودند که هم در زبان فارسی و هم در زبان عربی به کار می‌روند، اما در هر یک از زبان‌ها دارای معانی متفاوتی هستند. به طور مثال، کلمه "بهار" در دو زبان فارسی و عربی به کار می‌رود، اما در زبان فارسی اغلب به معنای "فصل بهار" است اما در

^۱ <http://www.ebizmba.com>, www.alexa.com

زبان عربی به معنای "ادویه" یا "چاشنی" به کار می‌رود. دلیل انتخاب کلمات هم‌نویسه میان زبان فارسی و عربی به عنوان پرس‌وجو این است که نتایجی در هر محدودیت زبان (چه به زبان فارسی و چه به زبان عربی) بازیابی شوند تا امکان ارزیابی ویژگی محدودیت زبان موتورهای کاوش فراهم گردد. از سوی دیگر، اهمیت عملکرد صحیح ویژگی محدودیت زبان موتورهای کاوش زمانی بیش از پیش آشکار می‌گردد که کلمات در زبان‌های مختلف دارای معانی متفاوت باشند (لواندوسکی، ۲۰۰۸). در همین راستا، تاکید بر معنای متفاوت هم-نویسه‌ها در دو زبان فارسی و عربی صورت گرفت. شایان ذکر است، لیست پرس‌وجوها در جدول ۱ آمده است.

جدول ۱- لیست پرس‌وجوها

مخابرات	بهار	زبون	حوصله	تقلیدی	صورت	خسیس
منظور	ناظر	اداره	شراب	مجتمع	مکتب	نفر
دغدغه	برق	ملت	ادویه	تخلف	عکس	نشاط
مجرم	موظف	مصرف	تورم	اسباب	جمعیت	ملی
ملی	رقیب	غرور	لحیم	وجه	معلومات	ساری
میمون	رسوم	شوکت				

در گام بعد، هر پرس‌وجو، دو بار در صفحه جستجوی پیشرفته هر موتور کاوش، یک بار با محدودیت زبان فارسی و بار دیگر با محدودیت زبان عربی جستجو گردید. با توجه به اینکه، پرس‌وجو در میان زبان فارسی و عربی مشترک است، نتایجی در هر محدودیت زبان بازیابی می‌شوند. سپس ۲۰ نتیجه اول هر جستجو، به منظور تعیین فارسی یا عربی بودن نتایج بررسی گردید^۱. قضاوت ربط بدین نحو صورت گرفت که پس از محدودیت زبان به زبان فارسی، نتایج به زبان عربی، و پس از محدودیت زبان به زبان عربی، نتایج به زبان فارسی، غیرمرتبط محسوب می‌شوند.

دلیل بررسی ۲۰ نتیجه اول نیز این است که صفحات اول نتایج، بهترین و مرتبطترین نتایجی است که یک موتور کاوش ارائه می‌دهد. علاوه بر این، مطالعه رفتار کاربران در موتورهای کاوش نشان داده است که کاربران فقط چند نتیجه نخست را بررسی می‌کنند و مابقی نتایج جستجو را مورد بررسی قرار نمی‌دهند (لواندوسکی، ۲۰۰۸).

^۱. در مجموع ۴۵۶۰ نتیجه (۳۸*۳*۲۰) مورد بررسی قرار گرفت.

لازم به ذکر است، از آنجا که تعیین زبان فارسی یا عربی نتایج دارای اطلاعات غیرمتنی امکان‌پذیر نبود، ۲۰ نتیجه اول که دارای اطلاعات متنی بودند، مورد بررسی قرار گرفتند. ارزیابی عملکرد هر موتور کاوش برای هر پرس‌وجو و مقایسه آنان با تعیین فراوانی و رتبه نتایج غیرمرتبط بازیابی شده، بررسی شد. از آنجا که وب‌سایت الکسا^۱، بر اساس ترافیک صفحات وب، به هر صفحه رتبه‌ای اختصاص داده است، آدرس وبی هر نتیجه غیرمرتبط در وب‌سایت الکسا بررسی گردید تا مشخص شود که آیا ترافیک صفحه یا تعداد زیاد بازدیدکنندگان صفحه غیرمرتبط بر بازیابی آن در ۲۰ نتیجه اول تاثیرگذار بوده است یا خیر.

به منظور تامین اعتبار، تمامی شرایط به دقت کنترل گردید. ابتدا سیاهه پرس‌وجوها توسط سه متخصص زبان عربی مورد تایید نهایی قرار گرفت. سپس به منظور اجتناب از تاثیر زبان رابط کاربری بر نتایج، در تمامی موارد جستجو، زبان رابط کاربری موتور کاوش، یکسان (عربی) در نظر گرفته شد (لواندوسکی، ۲۰۰۸). علاوه بر این تمامی مراحل جستجو و گردآوری داده از یک آی‌پی^۲ یکسان صورت گرفت تا از تاثیر جغرافیا بر نتایج اجتناب شود. همچنین، از آنجا که هر روز بر تعداد منابع موجود در وب افزوده می‌شود، به تناسب حجم نمایه و بانک اطلاعاتی موتور کاوش نیز افزوده می‌شود. بنابراین به منظور یکسان‌سازی شرایط برای سه موتور کاوش، هر یک از کلیدواژه‌ها، در بازه زمانی سه روزه، در تاریخ ۱۵ تا ۱۷ مردادماه سال ۱۳۹۴ در سه موتور کاوش جستجو گردید. گردآوری داده‌ها با مشاهده فراوانی و رتبه نتایج غیرمرتبط بازیابی شده در هر دو بار جستجو، در هر موتور کاوش به تفکیک ثبت شد. علاوه بر این، رتبه جهانی^۳ ارائه شده در وب‌سایت الکسا برای هر نتیجه غیرمرتبط ثبت گردید. سپس به کمک بسته نرم‌افزار آماری علوم اجتماعی^۴ و با استفاده از فنون آمار توصیفی (فراوانی و درصد) و استنباطی (آزمون‌های تحلیل واریانس، کروسکال والیس^۵، یو-مان-ویتنی^۶ و همبستگی اسپیرمن^۷) تجزیه و تحلیل صورت گرفت.

یافته‌ها

این بخش با هدف پاسخگویی به پرسش‌های پژوهش، به بررسی و تجزیه و تحلیل هر یک از پرسش‌ها به طور جداگانه می‌پردازد. همچنین به منظور وضوح بیشتر نتایج، تجزیه و تحلیل یافته‌ها در قالب جداولی ارائه می‌شوند.

1. www.alex.com

2. IP

3. Global Rank

4. SPSS

5. Kruskal-wallis

6. Mann-Whitney U

7. Spearman

پرسش اول: پس از محدودیت زبان به زبان هدف (فارسی یا عربی)، چه درصدی از نتایج موتور کاوش گوگل، بینگ و یاهو غیرمرتبط با زبان هدف (فارسی یا عربی) است؟

وضعیت موتورهای کاوش به لحاظ فراوانی نتایج غیرمرتبط با زبان هدف در جدول ۲ به تصویر کشیده شده است. چنان که مشاهده می‌شود، در دو موتور کاوش بینگ و یاهو پس از محدودیت زبان به زبان فارسی، هیچ نتیجه غیرمرتبط با زبان فارسی بازیابی نشده و کلیه نتایج (۱۰۰ درصد) به زبان فارسی است. اما در موتور کاوش گوگل پس از محدودیت زبان به زبان فارسی، ۱۴ نتیجه بازیابی شده (تقریباً ۲ درصد از کل) به زبان عربی بازیابی شد. از سوی دیگر با محدودیت زبان به زبان عربی، در دو موتور کاوش بینگ و یاهو، حدود ۴ درصد از نتایج، به زبان عربی بازیابی نشد^۱. اما موتور کاوش گوگل با بازیابی نادرست ۶۸ مدرک (تقریباً ۹ درصد از کل) به زبان فارسی از دو موتور کاوش دیگر عملکرد ضعیف‌تری ارائه داد.

جدول ۲- وضعیت موتورهای کاوش به لحاظ فراوانی و درصد نتایج غیرمرتبط با زبان هدف

موتور کاوش	نتایج غیرمرتبط با زبان هدف فارسی		نتایج غیرمرتبط با زبان هدف عربی	
	فراوانی	درصد	فراوانی	درصد
گوگل	۱۴	۱/۸۴	۶۸	۸/۹۵
بینگ	۰	۱۰۰	۲۷	۳/۵۵
ياهو	۰	۱۰۰	۲۹	۳/۸۲

پرسش دوم: وضعیت رتبه‌بندی نتایج غیرمرتبط با زبان هدف در موتور کاوش گوگل، بینگ و یاهو چگونه است؟

از آنجا که انتظار می‌رود، پس از محدودیت زبان، کلیه نتایج به زبان هدف بازیابی شوند، رتبه‌بندی نتایج غیرمرتبط از اهمیت بسیاری برخوردار است. جدول ۳ توزیع فراوانی رتبه‌های نتایج غیرمرتبط در موتور کاوش گوگل پس از محدودیت زبان به زبان فارسی را نشان می‌دهد. همان‌طور که پیش‌تر گفته شد، در موتورهای کاوش بینگ و یاهو پس از محدودیت زبان به زبان فارسی، نتایج غیرمرتبط بازیابی نشد و تنها در موتور کاوش گوگل، نتایج غیرمرتبط بازیابی شد. جدول ۳ حاکی از آن است که در موتور کاوش گوگل، رتبه ۸ بیشترین فراوانی نتایج غیرمرتبط را به خود اختصاص داده که تقریباً برابر با ۲۱ درصد از مجموع فراوانی رتبه‌های غیرمرتبط است. پس از رتبه ۸، رتبه‌های ۶، ۱۵ و ۱۹ بیشترین فراوانی را داشته که هر کدام ۱۴/۳ درصد از

^۱. با توجه به اینکه کلمات پرس‌وجو صرفاً در زبان فارسی و عربی (نه زبان دیگر) هم‌نویسه بوده‌اند، نتایج غیرمرتبط تنها به زبان مقابل بوده است.

مجموعه فراوانی رتبه‌های غیرمرتبط را به خود اختصاص می‌دهند. همچنین رتبه‌های ۱ تا ۵، ۷، ۱۲، ۱۳، ۱۷، ۱۸ و ۲۰ با کمترین فراوانی، در هیچیک از پرس‌وجوها نتایج غیرمرتبط در بر نداشته‌اند.

جدول ۳- توزیع فراوانی رتبه‌های نتایج غیرمرتبط در موتور کاوش گوگل با زبان هدف فارسی

رتبه	فراوانی رتبه	درصد فراوانی رتبه	رتبه	فراوانی رتبه	درصد فراوانی رتبه
رتبه ۱	۰	۰	رتبه ۱۱	۱	۷/۱
رتبه ۲	۰	۰	رتبه ۱۲	۰	۰
رتبه ۳	۰	۰	رتبه ۱۳	۰	۰
رتبه ۴	۰	۰	رتبه ۱۴	۱	۷/۱
رتبه ۵	۰	۰	رتبه ۱۵	۲	۱۴/۳
رتبه ۶	۲	۱۴/۳	رتبه ۱۶	۱	۷/۱
رتبه ۷	۰	۰	رتبه ۱۷	۰	۰
رتبه ۸	۳	۲۱/۴	رتبه ۱۸	۰	۰
رتبه ۹	۱	۷/۱	رتبه ۱۹	۲	۱۴/۳
رتبه ۱۰	۱	۷/۱	رتبه ۲۰	۰	۰

جدول ۴ توزیع فراوانی رتبه‌های نتایج غیرمرتبط در موتورهای کاوش گوگل، بینگ و یاهو پس از محدودیت زبان به زبان عربی را نشان می‌دهد. مطابق جدول ۴، در موتور کاوش گوگل، رتبه ۲۰ بیشترین فراوانی نتایج غیرمرتبط را به خود اختصاص داده که تقریباً برابر با ۱۲ درصد از مجموع فراوانی رتبه‌های غیرمرتبط است. پس از آن، رتبه ۱۲ بیشترین فراوانی را دارد که حدود ۹ درصد از مجموع فراوانی رتبه‌های غیرمرتبط محسوب می‌شود. همچنین رتبه‌های ۲، ۳ و ۴ کمترین فراوانی نتایج غیرمرتبط یعنی حدود ۳ درصد از مجموع فراوانی رتبه‌های غیرمرتبط را به خود اختصاص داده‌اند. لازم به ذکر است، تمامی رتبه‌ها حداقل یکبار نتایج غیرمرتبط را دربرداشته و رتبه ۱ به طور خاص، در دو پرس‌وجو، نتیجه غیرمرتبط در برداشت. طبق نتایج جدول ۴، در موتور کاوش بینگ، رتبه ۶ بیشترین فراوانی نتایج غیرمرتبط را به خود اختصاص داده است که تقریباً برابر با ۱۱ درصد از مجموع فراوانی رتبه‌های غیرمرتبط محسوب می‌شود. کمترین فراوانی رتبه‌های غیرمرتبط به رتبه ۱، ۷ و ۱۷ با ۰ درصد از مجموع فراوانی رتبه‌های غیرمرتبط اختصاص دارد. همچنین جدول ۴ نشان می‌دهد که در موتور کاوش یاهو، بیشترین فراوانی به رتبه ۱۵ با حدود ۱۴ درصد از مجموع فراوانی رتبه‌های غیرمرتبط اختصاص دارد. رتبه‌های ۱، ۷، ۹، ۱۰ و ۱۶ نیز کمترین فراوانی را با ۰ درصد از مجموع فراوانی رتبه‌های غیرمرتبط به خود اختصاص داده‌اند.

جدول ۴- توزیع فراوانی رتبه‌های نتایج غیرمرتبط در موتورهای کاوش با زبان هدف عربی

موتور کاوش	رتبه	فراوانی رتبه	درصد فراوانی رتبه	رتبه	فراوانی رتبه	درصد فراوانی رتبه
گوگل	رتبه ۱	۲	۲/۹	رتبه ۱۱	۳	۴/۴
	رتبه ۲	۱	۱/۵	رتبه ۱۲	۶	۸/۸
	رتبه ۳	۱	۱/۵	رتبه ۱۳	۳	۴/۴
	رتبه ۴	۱	۱/۵	رتبه ۱۴	۵	۷/۴
	رتبه ۵	۵	۷/۴	رتبه ۱۵	۳	۴/۴
	رتبه ۶	۲	۲/۹	رتبه ۱۶	۴	۵/۹
	رتبه ۷	۳	۴/۴	رتبه ۱۷	۴	۵/۹
	رتبه ۸	۲	۲/۹	رتبه ۱۸	۴	۵/۹
	رتبه ۹	۳	۴/۴	رتبه ۱۹	۴	۵/۹
	رتبه ۱۰	۴	۵/۹	رتبه ۲۰	۸	۱۱/۸
بینگ	رتبه ۱	۰	۰	رتبه ۱۱	۱	۳/۷
	رتبه ۲	۲	۷/۴	رتبه ۱۲	۲	۷/۴
	رتبه ۳	۲	۷/۴	رتبه ۱۳	۱	۳/۷
	رتبه ۴	۱	۳/۷	رتبه ۱۴	۲	۷/۴
	رتبه ۵	۱	۳/۷	رتبه ۱۵	۲	۷/۴
	رتبه ۶	۳	۱۱/۱	رتبه ۱۶	۱	۳/۷
	رتبه ۷	۰	۰	رتبه ۱۷	۰	۰
	رتبه ۸	۱	۳/۷	رتبه ۱۸	۳	۳/۷
	رتبه ۹	۱	۳/۷	رتبه ۱۹	۱	۳/۷
	رتبه ۱۰	۱	۳/۷	رتبه ۲۰	۲	۷/۴
ياهو	رتبه ۱	۰	۰	رتبه ۱۱	۱	۳/۴
	رتبه ۲	۲	۶/۹	رتبه ۱۲	۱	۳/۴
	رتبه ۳	۲	۶/۹	رتبه ۱۳	۱	۳/۴
	رتبه ۴	۳	۱۰/۳	رتبه ۱۴	۱	۳/۴
	رتبه ۵	۲	۶/۹	رتبه ۱۵	۴	۱۳/۸
	رتبه ۶	۲	۶/۹	رتبه ۱۶	۰	۰
	رتبه ۷	۰	۰	رتبه ۱۷	۱	۳/۴
	رتبه ۸	۳	۱۰/۳	رتبه ۱۸	۱	۳/۴
	رتبه ۹	۰	۰	رتبه ۱۹	۲	۶/۹
	رتبه ۱۰	۰	۰	رتبه ۲۰	۳	۱۰/۳

پرسش سوم: آیا تفاوت معنی‌داری میان موتورهای کاوش مورد مطالعه به لحاظ میانگین نتایج غیرمرتبط با زبان هدف وجود دارد؟

به منظور مقایسه میانگین نتایج غیرمرتبط با زبان هدف در موتورهای کاوش مختلف از آزمون تحلیل واریانس استفاده شد. نتایج به دست آمده در جدول ۵ نشان داده شده است. همان‌گونه که از جدول برمی‌آید، پس از محدودیت زبان به زبان فارسی و عربی، بین سه موتور کاوش گوگل، بینگ و یاهو به لحاظ میانگین نتایج غیرمرتبط با زبان هدف اختلاف معنی‌داری وجود ندارد.

جدول ۵- تحلیل واریانس برای سنجش معنی‌داری اختلاف نتایج غیرمرتبط با زبان هدف در موتورهای کاوش

زبان هدف	مجموع مجذورات	درجه آزادی	میانگین مربعات	مقدار F	سطح معنی‌داری
فارسی	بین‌گروهی	۲	۱/۷۲	۲/۷۷	۰/۰۶۷
	درون‌گروهی	۱۱۱	۰/۶۲		
	مجموع	۱۱۳			
عربی	بین‌گروهی	۲	۱۴/۰۶	۲/۴۵	۰/۰۹۱
	درون‌گروهی	۱۱۱	۵/۷۴		
	مجموع	۱۱۳			

پرسش چهارم: آیا تفاوت معنی‌داری میان موتورهای مورد مطالعه به لحاظ میانگین رتبه‌های نتایج غیرمرتبط با زبان هدف وجود دارد؟

به منظور مقایسه رتبه‌های نتایج غیرمرتبط با زبان هدف در موتورهای کاوش گوگل، بینگ و یاهو از آزمون کروسکال والیس استفاده شد. نتایج در جدول ۶ به تصویر کشیده شده است. همان‌گونه که مشاهده می‌شود، پس از محدودیت زبان به زبان فارسی، نتایج آزمون کروسکال والیس نشان از معنی‌داری اختلاف میان رتبه‌های نتایج غیرمرتبط در سه موتور کاوش مختلف در سطح اطمینان ۹۹ دارد ($X^2=16/71$ ، $P=0/000$). به منظور شناسایی موتور کاوشی که اختلاف معنی‌دار را ایجاد کرده است، از آزمون یو-مان-ویتنی استفاده شد. نتایج حاصل از آزمون نشان داد که اختلاف رتبه‌های نتایج غیرمرتبط میان سه موتور کاوش، به موتور کاوش گوگل باز می‌گردد، به نحوی که میانگین رتبه‌های نتایج غیرمرتبط در موتور کاوش گوگل به نحو معنی‌داری نسبت به موتور کاوش بینگ و یاهو بیشتر است. لازم به ذکر است، دو موتور کاوش بینگ و یاهو به لحاظ میانگین رتبه‌های نتایج غیرمرتبط با یکدیگر اختلاف معنی‌داری ندارند (جدول ۷).

جدول ۶- آزمون کروسکال-والیس برای مقایسه رتبه‌بندی نتایج غیرمرتبط با زبان هدف در موتورهای کاوش مختلف

زبان هدف	خی دو	درجه آزادی	سطح معنی‌داری
فارسی	۱۶۷۱	۲	۰/۰۰۰
عربی	۹/۳۶	۲	۰/۰۰۹

از سوی دیگر، پس از محدودیت زبان به زبان عربی، آزمون کروسکال والیس نشان داد که بین سه موتور کاوش به لحاظ میانگین رتبه‌های نتایج غیرمرتبط اختلاف معنی‌داری در سطح اطمینان ۹۹ درصد وجود دارد ($P=۰/۰۰۹$ ، $X^2=۹/۳۶$) (جدول ۶). به منظور شناسایی موتور کاوشی که این اختلاف را ایجاد کرده است، از آزمون یو-مان-ویتنی استفاده شد. نتایج حاصل از آزمون نشان داد که پس از محدودیت زبان به زبان عربی، اختلاف در میانگین رتبه‌های نتایج غیرمرتبط به موتور کاوش گوگل باز می‌گردد، به نحوی که میانگین رتبه‌های نتایج غیرمرتبط در موتور کاوش گوگل به نحو معنی‌داری از دو موتور کاوش دیگر بیشتر است. لازم به ذکر است که اختلاف میان میانگین رتبه‌های نتایج غیرمرتبط دو موتور کاوش بینگ و یاهو معنی‌دار نشده است (جدول ۷).

جدول ۷- موتورهای کاوش دارای رتبه‌بندی نتایج غیرمرتبط متفاوت معنی‌دار در آزمون یو-مان-ویتنی

زبان هدف	موتور کاوش (الف)	موتور کاوش (ب)	مقدار یو-مان-ویتنی	مقدار Z	سطح معنی‌داری
فارسی	گوگل	بینگ	۲۳۳۷/۵۰	-۲/۹۱	۰/۰۰۴
		یاهو	۲۳۸۰/۰۰	-۲/۹۴	۰/۰۰۳
عربی	گوگل	بینگ	۲۰۶۰/۵۰	-۲/۵۷	۰/۰۱۰
		یاهو	۲۱۳۸/۵۰	-۲/۴۳	۰/۰۱۵

پرسش پنجم: آیا ترافیک صفحات غیرمرتبط با زبان هدف بر رتبه‌بندی آنان در سه موتور کاوش گوگل، بینگ و یاهو تاثیرگذار است؟

از آنجا که پس از محدودیت زبان، نتایجی غیرمرتبط با زبان هدف در ۲۰ نتیجه اول بازیابی می‌شوند، این سوال مطرح می‌گردد که آیا ترافیک و یا بازدید بیشتر نتایج غیرمرتبط سبب کسب رتبه بهتر در موتورهای کاوش گردیده است یا خیر. در این راستا، به منظور دریافت ترافیک هر صفحه به طور مجزا، از رتبه جهانی

ترافیک صفحات که توسط وبسایت الکسا ارائه می‌گردد، استفاده شد. در ادامه به منظور بررسی رابطه میان رتبه ترافیک صفحات غیرمرتبط و رتبه آنان پس از محدودیت زبان به فارسی و عربی در سه موتور کاوش گوگل، بینگ و یاهو از آزمون همبستگی اسپیرمن استفاده شد. نتایج حاصل از آزمون در جدول ۸ به تصویر کشیده شده است. مطابق جدول ۸، رابطه معنی‌داری میان رتبه ترافیک و یا بازدید صفحات غیرمرتبط و رتبه آنان در موتور کاوش گوگل پس از محدودیت زبان به زبان فارسی وجود ندارد. همچنین نتایج حاصل از آزمون نشان از عدم معنی‌داری رابطه میان رتبه ترافیک و بازدید صفحات غیرمرتبط و رتبه آنان در سه موتور کاوش پس از محدودیت زبان به زبان عربی دارد. لازم به ذکر است، در وبسایت الکسا، رتبه ترافیک برخی صفحات غیرمرتبط به دلیل فیلتر بودن صفحات دریافت نشد، این موارد از محاسبه حذف گردیده است^۱.

جدول ۸- نتایج آزمون همبستگی اسپیرمن میان رتبه ترافیک صفحات غیرمرتبط و رتبه آنان در موتورهای کاوش

زبان هدف	موتور کاوش	ضریب همبستگی	تعداد نتایج غیرمرتبط	سطح معنی‌داری
فارسی	گوگل	۰/۳۴	۱۳	۰/۲۵
عربی	گوگل	۰/۱۶	۶۰	۰/۲۳
	بینگ	-۰/۱۲	۲۳	۰/۵۹
	ياهو	-۰/۱۷	۲۰	۰/۴۷

بحث و نتیجه‌گیری

با جستجوی ۳۸ پرس‌وجوی هم‌نویسه (با معانی متفاوت) در زبان فارسی و عربی در سه موتور کاوش گوگل، بینگ و یاهو نشان داده شد که پس از محدودیت زبان به زبان فارسی، تنها دو موتور کاوش بینگ و یاهو عملکرد صددرصد موفق‌تری ارائه دادند. اما موتور کاوش گوگل به لحاظ بازیابی تعداد نتایج مرتبط با زبان هدف، عملکرد ضعیف‌تری نسبت به موتور کاوش بینگ و یاهو دارد. بنابراین دو موتور کاوش بینگ و یاهو می‌توانند نتایج دقیق‌تری نسبت به موتور کاوش گوگل پس از محدودیت زبان به زبان فارسی ارائه دهند. از سوی دیگر، پس از محدودیت زبان به زبان عربی هیچ‌یک از موتورهای کاوش عملکرد صددرصد موفق‌تری ارائه ندادند، اما باز هم موتور کاوش بینگ و یاهو به ترتیب عملکرد بهتری نسبت به موتور کاوش گوگل داشتند. بنابراین دو موتور کاوش بینگ و یاهو می‌توانند نتایج دقیق‌تری نسبت به موتور کاوش گوگل پس از محدودیت زبان به زبان عربی به کاربران ارائه دهند. این یافته همسو با نتایج لواندوسکی (۲۰۰۸) است که نشان داد موتور

^۱ در محدودیت زبان به زبان فارسی، یک مورد از نتایج غیرمرتبط در موتور کاوش گوگل، و در محدودیت زبان به زبان عربی، ۸ مورد نتایج غیرمرتبط در موتور کاوش گوگل، ۴ مورد نتایج غیرمرتبط در موتور کاوش بینگ و ۹ مورد نتایج غیرمرتبط در موتور کاوش یاهو در محاسبه لحاظ نگردید.

کاوش گوگل نسبت به موتورهای کاوش اسک و یاهو عملکرد ضعیف‌تری به لحاظ بازیابی اطلاعات به زبان دیگر دارد.

از آنجا که انتظار می‌رود، پس از محدودیت زبان، کلیه نتایج به زبان هدف بازیابی شوند، رتبه‌بندی نتایج غیرمرتبط نیز از اهمیت بسیاری برخوردار است. در واقع انتظار می‌رود، نتایج غیرمرتبط در رتبه‌های اول ظاهر نگردند. اما نتایج نشان داد که پس از محدودیت زبان به زبان فارسی، بیشترین نتایج غیرمرتبط در رتبه ۸ موتور کاوش گوگل ظاهر گردیده است. اما در رتبه‌های ۱ تا ۵ در هیچیک از پرس‌وجوها نتایج غیرمرتبط ظاهر نگردیده است. از سوی دیگر، پس از محدودیت زبان به زبان عربی، موتور کاوش گوگل در تمامی رتبه‌ها حتی رتبه‌های ۱ تا ۵ حداقل یکبار نتایج غیرمرتبط نشان داد. اما در موتور کاوش بینگ و یاهو رتبه ۱، در هیچیک از پرس‌وجوها نتایج غیرمرتبط نشان نداد، اما اکثریت رتبه‌ها حداقل یکبار نتایج غیرمرتبط در برداشتند. به نظر می‌رسد که علت عملکرد بهتر موتورهای کاوش در زبان هدف فارسی نسبت به زبان هدف عربی نیز این است که جستجو از آی‌پی ایران صورت گرفته است، بنابراین نتایج زبان فارسی در رتبه‌های بهتری نسبت به دیگر زبان‌ها قرار می‌گیرند (لواندوسکی، ۲۰۰۸).

همچنین سه موتور کاوش در بازیابی نتایج غیرمرتبط با زبان هدف تفاوت معنی‌داری با یکدیگر ندارند. در واقع هر سه موتور نتایج ایده‌آلی را ارائه نمی‌دهند و هیچیک از این سه موتور به لحاظ بازیابی نتایج مرتبط با زبان هدف بر یکدیگر برتری ندارند. اما این سه موتور کاوش در رتبه‌بندی نتایج غیرمرتبط با زبان هدف تفاوت معنی‌داری با یکدیگر نشان دادند و این تفاوت نیز به موتور کاوش گوگل بازمی‌گردد. در واقع موتور کاوش گوگل با بازیابی نتایج غیرمرتبط در رتبه‌های بهتر، کارآمدی کمتری نسبت به دیگر رقبای خود دارد. یافته‌های این پژوهش با نتایج حاصل از پژوهش ژانگ، فی و لی (۲۰۱۳) و لواندوسکی (۲۰۰۴) به لحاظ برتری موتور کاوش گوگل در عملکرد ویژگی‌های جستجو هم‌راستا نیست. البته آنان سایر ویژگی‌های جستجو جز محدودیت زبان را مورد بررسی قرار داده و نشان دادند که موتور کاوش گوگل در دیگر ویژگی‌های جستجو مثل محدودیت تاریخ، محدودیت نوع فایل، جستجوی آدرس وبی و ... به طور معنی‌داری بهتر از دیگر موتورهای کاوش مورد بررسی عمل می‌کند.

از آنجا که پس از محدودیت زبان، نتایجی غیرمرتبط با زبان هدف در ۲۰ نتیجه اول بازیابی می‌شوند، این پرسش مطرح می‌شود که آیا ترافیک و یا بازدید بیشتر نتایج غیرمرتبط سبب کسب رتبه بهتر در موتورهای کاوش گردیده است یا خیر. در واقع، آیا موتورهای کاوش به دلیل محبوبیت و یا ترافیک بیشتر صفحات غیرمرتبط، آنان را در رتبه‌های بهتر قرار داده و از محدودیت جستجو به یک زبان خاص صرف نظر کرده‌اند یا خیر. با توجه به اینکه میان رتبه ترافیک و رتبه صفحات غیرمرتبط در موتورهای کاوش مختلف ارتباط معنی-

داری مشاهده نشد، به نظر می‌رسد که عملکرد موتورهای کاوش در بازیابی نتایج غیرمرتبط با زبان هدف به ترافیک و یا محبوبیت صفحات غیرمرتبط بازمنی گردد و به عواملی دیگر مانند ضعف ابزارهای جستجو و الگوریتم‌های تشخیص زبان در موتورهای کاوش مربوط است.

به نظر می‌رسد موتور کاوش گوگل از تشخیص زبان ایستا^۱ استفاده نکرده و به جای آن از تشخیص زبان درجه‌ای^۲ استفاده می‌کند. در تشخیص زبان ایستا، یک مدرک به یک زبان خاص تعلق می‌گیرد اما در تشخیص زبان درجه‌ای، مدرکی که شامل زبان‌های مختلف است به بیش از یک زبان اما با درجه‌های تعلق مختلف اختصاص می‌یابد. از سوی دیگر، زبان‌ها به برخی صفحات نمی‌توانند اختصاص یابند. چون متون صفحه هم دارای زبان فارسی و هم دارای زبان عربی است و یا اینکه گزینه انتخاب زبان در صفحه اول قرار گرفته است. همچنین اگر تعداد کلمات موجود در صفحه بسیار اندک باشد (مثلاً فقط شامل نام، شماره تماس و اطلاعات آدرس باشد)، امکان شناسایی زبان مهیا نمی‌شود (لواندوسکی، ۲۰۰۸).

به طور کلی پژوهش حاضر نشان داد که جستجوگرانی که از ویژگی محدودیت زبان موتورهای کاوش به منظور محدود کردن نتایج جستجو استفاده می‌کنند، در موتور کاوش بینگ و یاهو از بخت بالاتری برای دستیابی به نتایج مرتبط برخوردارند. شناسایی زبان صحیح یک مدرک به لحاظ فنی، مسئله‌ای حل شده محسوب می‌شود (کاونار و ترنکل، ۲۰۰۴؛ دانیگ، ۱۹۹۴) و از آنجا که باید موتورهای کاوش فقط نتایج منطبق با زبان هدف را ارائه دهند، لازم است الگوریتم‌های کارآمد شناسایی زبان مدرک مورد استفاده قرار گرفته و صفحاتی که به بیش از یک زبان تعلق دارند، هنگام استفاده کاربران از ویژگی محدودیت زبان به دقت مورد بررسی قرار گیرند. پیشنهاد می‌گردد پژوهش‌های دیگری به منظور ارزیابی کارآمدی دیگر ویژگی‌های جستجوی موتورهای کاوش در زبان فارسی صورت گیرد تا مشخص گردد که کدام موتور کاوش و ویژگی‌های جستجو برای جستجو و دستیابی به مدارک فارسی مناسب‌تر است. علاوه بر این پیشنهاد می‌گردد، کاربرانی که از ویژگی محدودیت زبان موتورهای کاوش به منظور محدود کردن نتایج جستجو استفاده می‌کنند، موتورهای کاوش بینگ و یاهو را برای دستیابی به نتایج مرتبط‌تر به کار گیرند. همچنین لازم است طراحان موتور جستجو به کارآمدی ویژگی‌های محدودکننده جستجو توجه کرده تا کاربران بتوانند اطلاعات مورد نیاز خود را در کمترین زمان ممکن بازیابی کنند.

1. Static-language detection

2. Graded-language detection

منابع

- دری، راحله (۱۳۹۳). مقایسه و ارزیابی موتورهای جستجوی معنایی. *پژوهشنامه پردازش و مدیریت اطلاعات*، ۳۰(۲): ۴۶۷-۴۸۷.
- رجبی، منصور؛ نوروزی، یعقوب (۱۳۹۴). موتورهای جستجوی فارسی: ارزیابی امکانات جستجو، بازیابی اطلاعات، میزان جامعیت و مانعیت و تعیین همپوشانی میان آنها. *مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۲۶(۳): ۱۳۳-۱۵۰.
- شریف، عاطفه (۱۳۹۱). بررسی تغییرات عملکرد دو موتور کاوش عمومی یاهو و گوگل از نظر پوشش کمی-زمانی نمایه‌سازی و توجه به عناصر ابر داده‌ای در رتبه‌بندی صفحه‌های وب. *پژوهش‌نامه کتابداری و اطلاع-رسانی*، ۲(۱): ۱۷۵-۱۹۴.
- محمداسماعیل، صدیقه؛ منصور کیایی، ربابه (۱۳۹۰). مقایسه موتورهای و ابرموتورهای کاوش عمومی در بازیابی اطلاعات علم فیزیک و میزان همپوشانی آنها. *مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۲۲(۳): ۱۳۰-۱۴۰.
- مسعود، جبران (۱۳۸۲). *فرهنگ رانداالطلاب: عربی به عربی، عربی به فارسی*. ترجمه عبدالستار قمری. تهران: یادوار کتاب.
- Cavnar, W.B. and Trenkle, J. M. (2004). N-gram-based text categorization, pp. 161-176, paper presented at the Third Annual Symposium on Document Analysis and Information Retrieval, Las Vegas, Nevada, USA.
- Dunning, T. (1994). Statistical identification of language, Technical Report MCCS 94-273, New Mexico State University, New Mexico.
- Kaushik, Anna (2012). Judging the capability of search engines and search terms. *International Journal of Information Dissemination and Technology*, 2(1): 6-11.
- Fattahi, R., Wilson, C. S. and Cole, F. (2008). An alternative approach to natural language query expansion in search engines: text analysis of non-topical terms in web documents. *Information Processing & Management*, 44(4): 1503-1516.
- Ghani, R., Jones, R. and Mladenic, D. (2005). Building minority language corpora by learning to generate web search queries. *Knowledge and Information Systems*, 7: 56-83.
- Kim, J. Y., Feild, H., and Cartright, M. (2012, October). Understanding book search behavior on the web. In Proceedings of the 21st ACM international conference on Information and knowledge management (pp. 744-753). ACM.
- Lewandowski, D. (2004). Date-restricted queries in web search engines. *Online Information Review*, 28(6): 420-427.
- Lewandowski, D. (2008). Problems with the use of web search engines to find results in foreign languages. *Online Information Review*, 32 (5): 668 - 672.
- Lewandowski, D., & Höchstötter, N. (2008). *Web searching: A quality measurement perspective* (pp. 309-340). Springer Berlin Heidelberg.
- Martins, B. and Silva, M. J. (2005). Language identification in web pages, pp. 764-768, paper presented at the ACM Symposium on Applied Computing, Santa Fe, New Mexico, USA.

- Sroka, M. (2000). Web search engines for Polish information retrieval: Questions of search capabilities and retrieval performance. *International Information & Library Research*, 32: 87-98.
- Zhang, J. and Lin, S. (2007). Multiple language supports in search engines. *Online Information Review*, 31(4): 516-532.
- Zhang, J., Fei, W., and Le, T. (2013). A comparative analysis of the search feature effectiveness of the major English and Chinese search engines. *Online Information Review*, 37(2), 217-230.

روزدایند ویرایش شده