

بررسی میزان ربط مدارک بازیابی شده با استفاده از ابزار بازیابی بین زبانی google در رشته کشاورزی

فاطمه جمشیدی قهفرخی^۱ | کارشناس ارشد،
کتابداری و اطلاع‌رسانی

دریافت: ۱۳۹۰/۰۸/۳۰ | پذیرش: ۱۳۹۰/۱۱/۲۶

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۸۲۲۳-۲۲۵۱
شاپا(الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در SCOPUS و ISC
http://jipm.irandoc.ac.ir
دوره ۲۸ | شماره ۳ | ص ۷۶۱-۷۶۰
بهار ۱۳۹۲
نوع مقاله: پژوهشی

1. Jam.fatema@gmail.com

چکیده: در این پژوهش، به بررسی میزان ربط مدارک پرداخته شده است که در حوزه موضوعی کشاورزی با استفاده از ابزار بازیابی زبانی google، بازیابی می‌شوند. بدین منظور، ابتدا با استفاده از مقالات موجود در مجلات فارسی دارای چکیده انگلیسی، به استخراج عبارت‌ها و اصطلاحات تخصصی فارسی حوزه موضوعی کشاورزی و معادل انگلیسی آنها اقدام شد. تعداد ۳۰ عبارت و اصطلاح تخصصی در قالب سه دسته استخراج گردید: دسته اول، عبارت‌های موضوعی که فقط در رشته کشاورزی کاربرد دارند. دسته دوم، اصطلاحات تخصصی کشاورزی که در سایر رشته‌ها نیز کاربرد دارند و دسته سوم، اصطلاحات تخصصی کشاورزی که در خارج از این حوزه تخصصی، اصطلاح عمومی محسوب می‌شوند. سپس به جستجوی مدارک با استفاده از این عبارات و اصطلاحات پرداخته شد و نتایج جستجو از نظر میزان ربط موضوعی با حوزه کشاورزی مورد بررسی قرار گرفت. یافته‌های پژوهش حاکی از آن است که ابزار بازیابی بین زبانی google در بازیابی بین زبانی مدارک مرتبط برای دو دسته عبارت و اصطلاح تخصصی کشاورزی با موفقیت عمل نمی‌کند، یک دسته عبارت و اصطلاحاتی که علاوه بر رشته کشاورزی در رشته‌های دیگر نیز کاربرد دارند و دسته دیگر، عبارات و اصطلاحاتی که در خارج از حوزه موضوعی کشاورزی جزء واژگان عمومی محسوب می‌شوند. اما در مورد عبارات و اصطلاحاتی که فقط در رشته کشاورزی کاربرد دارند، عملکرد به نسبت مطلوب‌تری دارد.

کلیدواژه‌ها: اصطلاح تخصصی، اصطلاح عمومی، بازیابی بین زبانی، ترجمه ماشینی، حوزه موضوعی کشاورزی، میزان ربط، ابزار بازیابی زبانی google

۱. مقدمه

اساس بازیابی اطلاعات بر جدا کردن مدارک مرتبط از نامرتب است و ربط جزء جدانشدنی این فرایند محسوب می‌شود. ربط از جمله مفاهیم بنیانی در جریان تبادل اطلاعات و ارتباطات است. ربط، هسته طراحی نظام‌های بازیابی و ارزیابی آنها را تشکیل می‌دهد و جزء اصلی در ارزیابی کارایی نظام بازیابی اطلاعات به‌شمار می‌رود. طبق تعریف واژه‌نامه وبستر ربط به معنای داشتن نسبتی مهم با موضوع تحت بررسی است (داورپناه ۱۳۸۳). پدیده ربط در حوزه اطلاع‌رسانی، به‌خصوص در جریان ذخیره و بازیابی اطلاعات از برجستگی و اهمیت خاصی برخوردار است. در اهمیت ربط همین بس که به اعتقاد بعضی، اطلاعاتی که رد و بدل می‌شود، اگر مرتبط نباشد در اصل، اطلاعات نیست (خالویی ۱۳۸۷). به علاوه، با گسترش روزافزون استفاده از اینترنت و غلبه بر محدودیت‌های فنی و شبکه‌ای که به مدد توسعه فناوری اطلاعات و ارتباطات حاصل شده است، کاربران و جستجوگران اطلاعات دیگر فقط به منابع اطلاعاتی که به زبان آنها نوشته شده است، اکتفا نمی‌کنند. دسترسی به همه اطلاعات مرتبط در دیگر زبان‌ها، اکنون نه آرزو، بلکه حق طبیعی کاربران شناخته می‌شود. این تنوع زبانی اگرچه در ابتدا مفید به نظر می‌رسد، اما می‌تواند مانعی برای دسترسی به اطلاعات تلقی شود (علیزاده ۱۳۸۳). بنابراین، امروزه بازیابی اطلاعات به فرایندهای سنتی آن خلاصه نمی‌شود، بلکه اهدافی بزرگ‌تر نیز در این حوزه مطرح شده است. ادواردز تعداد زبان‌های زنده دنیا را چیزی حدود ۴۵۰۰ زبان تخمین می‌زند که از میان آنها در حدود ۳۰ زبان هستند که هر یک توسط دست کم ۳۰ میلیون نفر استفاده می‌شوند و بدیهی است که برای تبادل اطلاعات در این جامعه نیز مورد استفاده قرار می‌گیرند (Edwards 1994).

در بازیابی اطلاعات چندزبانه، دیگر نمی‌توان به یک زبان خاص محدود شد. اینترنت به عنوان محل تلاقی این زبان‌ها، بیشترین نمود این گوناگونی را در خود جا داده است. آمارها نشان می‌دهد که استفاده از اینترنت در چندسال اخیر رشد قابل ملاحظه‌ای داشته است (Wang and Oard 2005). این تنوع جغرافیایی با تنوع زبانی نیز همراه است، به طوری که با رشد منابع اینترنتی، مشکلات زبانی در دسترسی و بهره‌گیری از این منابع نیز بیشتر شده است. راه حل غلبه بر این مشکلات، بهره‌گیری از بازیابی اطلاعات بین زبانی است (علیزاده ۱۳۸۸). در بازیابی بین زبانی کاربر ناگزیر از استفاده از نرم‌افزار یا ابزارهای ترجمه ماشینی است، اما استفاده از این ابزارها مشکلاتی را نیز به دنبال خود دارند از جمله بروز ابهام در ترجمه‌های ماشینی و ...

ایجاد نظام‌های بازیابی اطلاعات در مقیاس وسیع بسیار پرهزینه است. هزینه‌های کاوش قابل ملاحظه هستند و زمان قابل توجهی برای جستجوی اطلاعات در پایگاه‌ها توسط متخصصان اطلاع‌رسانی و کاربران نهایی صرف می‌شود و از همه مهم‌تر اینکه یک کاوش در بهترین حالت جستجو می‌تواند آنچه را که مورد جست و جو است بیابد. در حالی که اطلاعات بالقوه دیگر حذف می‌شود و در بدترین حالت جستجو می‌تواند اطلاعات بی‌ارزش را ارائه دهد و موارد مورد نیاز و مرتبط را حذف کند. بنابراین مهم است بدانیم کدام نظام‌ها و کاوش‌ها عملکرد موثرتری دارند.

مورد توجه قراردادادن و درک مفهوم ربط دست‌کم در ساخت و به‌کارگیری نظام‌های اطلاعاتی - که امروزه در محیط وب پیچیدگی بیش از پیش یافته‌اند - هنوز هم به‌جا و لازم است تا این نظام‌ها بتوانند اطلاعاتی متناسب با مسائل و نیازهای اطلاعاتی کاربران فراهم سازند (مختاری ۱۳۸۹). بنابراین، پرداختن به پژوهش در حوزه ارزیابی میزان ربط منابعی که از طریق نظام‌های بازیابی اطلاعات در اختیار کاربر قرار می‌گیرد، مهم است و گامی در جهت بهبود کیفیت نظام‌های بازیابی اطلاعات است، که می‌تواند به طراحی و ایجاد نظام‌های اطلاعاتی بهتر بیانجامد.

از آنجا که ابزار بازیابی بین‌زبانی google، یک ابزار ترجمه درخواست در یک نظام بازیابی اطلاعات پرکاربرد است؛ و در مرحله ترجمه درخواست، از زبان مبدا به زبان یا زبان‌های مقصد، امکان بروز ابهام وجود دارد. ممکن است این مسأله منجر به بازیابی مدارک غیرمرتبط با حوزه موضوعی مدنظر گردد. از طرفی، ترجمه اصطلاحات به صورت تخصصی نیز می‌تواند بر ابهام ترجمه‌ای که توسط این ابزار صورت می‌گیرد، بیافزاید. بنابراین ممکن است مدارک بازیابی شده از طریق این ابزار، در بازیابی بین‌زبانی، از ربط کافی با حوزه‌های موضوعی تخصصی برخوردار نباشد. بنابراین، در این مقاله سعی شده است تا ابزار بازیابی بین‌زبانی google، به عنوان یک ابزار ترجمه ماشینی مورد بررسی قرار گیرد و مشخص شود که آیا با توجه به اینکه در ترجمه ماشینی احتمال بروز ابهام در ترجمه درخواست از زبان مبدا به زبان یا زبان‌های مقصد وجود دارد، مدارک بازیابی شده در بازیابی بین‌زبانی از طریق ابزار بازیابی بین‌زبانی google، از ربط موضوعی کافی برخوردار است یا خیر. همچنین، تا حد امکان نقایص این ابزار مشخص گردد. تمامی این تلاش‌ها با هدف بهبود نظام‌های بازیابی اطلاعات چندزبانه و کمک به طراحی نظام‌های بازیابی اطلاعات کارآمدتر و متناسب‌تر با نیاز کاربران صورت پذیرفته است.

۲. اهداف پژوهش

هدف اصلی

تشخیص میزان ربط موضوعی مدارکی که با استفاده از ابزار بازیابی زبانی google، در حوزه موضوعی کشاورزی بازیابی می شوند.

اهداف فرعی

۱. تشخیص میزان ربط موضوعی مدارکی که با استفاده از ابزار بازیابی زبانی google- به عنوان یک ابزار ترجمه ماشینی- برای اصطلاحات تخصصی که فقط در رشته کشاورزی کاربرد دارد، بازیابی می شود.
۲. تشخیص میزان ربط موضوعی مدارکی که با استفاده از ابزار بازیابی زبانی google- به عنوان یک ابزار ترجمه ماشینی- برای اصطلاحات تخصصی که علاوه بر رشته کشاورزی در سایر رشته‌ها نیز کاربرد دارد، بازیابی می شود.
۳. تشخیص میزان ربط موضوعی مدارکی که با استفاده از ابزار بازیابی زبانی google- به عنوان یک ابزار ترجمه ماشینی- برای اصطلاحات عمومی که در رشته کشاورزی اصطلاح تخصصی به حساب می آید، بازیابی می شود.

۳. سؤالات پژوهش

۱. مدارکی که از طریق ابزار ترجمه ماشینی ابزار بازیابی زبانی google برای اصطلاحات تخصصی - که فقط در رشته کشاورزی کاربرد دارد- بازیابی می شود، تا چه حد از ربط موضوعی با حوزه کشاورزی برخوردار است؟
۲. مدارکی که از طریق ابزار بازیابی زبانی google، برای اصطلاحات تخصصی - که علاوه بر رشته کشاورزی در سایر رشته‌ها نیز کاربرد دارد - بازیابی می شود، تا چه حد از ربط موضوعی با حوزه کشاورزی برخوردار است؟
۳. مدارکی که از طریق ابزار بازیابی زبانی google برای اصطلاحات عمومی - که در رشته کشاورزی اصطلاح تخصصی محسوب می شود - بازیابی می شود، تا چه حد از ربط موضوعی با حوزه کشاورزی برخوردار است؟

۴. تعاریف

۴-۱. معرفی ابزار بازیابی بین زبانی google

گوگل، یکی از قدرتمندترین و پرمراجعه کننده ترین موتورهای جستجو در وب است. این

موتور جستجو، امکان دسترسی آسان به بیش از ۳ میلیارد صفحه وب را فراهم می کند و هر روز پذیرای میلیون ها کاربر از سراسر دنیاست. برای دسترسی به خدمات بازیابی بین زبانی در گوگل باید مراحل زیر را طی شود. قبل از هر چیز لازم به اشاره است که :

- درخواست خود را به زبان مادری در صفحه اصلی جستجو در گوگل وارد می کنیم تا کاوشگر بازیابی اولیه را برای ما انجام دهد.
- در کادر کناره چپ صفحه از بین گزینه های موجود، گزینه More search tools را انتخاب می کنیم.
- از منوی باز شده گزینه Translated foreign pages را برمی گزینیم.



- به این ترتیب کادر مربوط به بازیابی بین زبانی در کاوشگر گوگل بر روی صفحه نمایش داده می شود.

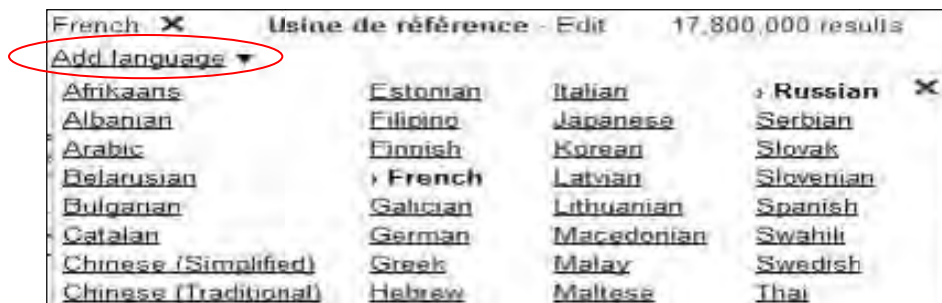


- در این کادر (برای بازیابی بین زبانی) به کاربر امکان داده می شود عبارت درخواست خود را به زبان بومی وارد کند، سپس مشخص کند که مایل است مدارک مرتبط با عبارت درخواست وارد شده، از بین مدارک چه زبان‌هایی بازیابی شوند و در نهایت مایل است مدارک بازیابی شده به چه زبانی ترجمه شود و در اختیار او قرار گیرد.
- با زدن گزینه My language باز می شود که ۵۲ زبان برای انتخاب در اختیار می گذارد. با انتخاب از بین آنها کاربر اعلام می کند که مایل است مدارک بازیابی شده، به چه زبانی ترجمه شود. به عنوان مثال کاربر فارسی زبان، اگر زبان فارسی را برگزیند، مدارک بازیابی شده از هر زبانی که باشد، برای او به زبان فارسی ترجمه و ارائه می شود.



- با زدن گزینه Add language کاربر تعیین می کند که موتور جستجو از بین مدارک چه زبان یا زبان‌هایی نتایج را بازیابی کند. در واقع کاربر از طریق این گزینه، زبان یا زبان‌های مقصد را انتخاب می کند. امکان انتخاب دو حالت وجود دارد: یک حالت انتخاب خود کار زبان‌ها توسط موتور جستجو و حالت دیگر انتخاب زبان‌ها توسط کاربر. در حالت اول، موتور جستجو، زبان‌هایی را به طور خود کار انتخاب می کند و از بین

مدارک نوشته‌شده به این زبان‌ها، به جستجویی پردازد. اما در حالت دوم، کاربر امکان می‌یابد که از بین ۵۲ زبان مشخص شده دست به انتخاب بزند.



- پس از انتخاب زبان مقصد، عبارت درخواست مورد نظر به زبان انتخابی، ترجمه و تعداد مدارک بازرگانی شده در مقابل آن نشان داده می‌شود. اگر بخواهیم مدارک بازرگانی شده از یک زبان (به عنوان مثال، زبان انگلیسی) نمایش داده شود، باید بر روی تعداد مدارک همان زبان کلیک کنیم.



- در این ابزار ترجمه ماشینی، دو بار عمل ترجمه انجام می‌گیرد، بار اول، ترجمه عبارت درخواست کاربر به زبان‌هایی که در کادر Add language انتخاب می‌کند و بار دوم ترجمه مدارک بازرگانی شده به زبانی که کاربر در کادر my language انتخاب می‌کند. همچنین می‌توان مدارک ترجمه‌شده را به زبان اصلی نیز مشاهده کرد. آنچه در این پژوهش بر آن تمرکز شده است، اولین مرحله ترجمه‌ای است که در این ابزار صورت می‌گیرد، یعنی ترجمه عبارت درخواست از زبان مبدا به زبان یا زبان‌های مقصد. چرا که ابهام در این مرحله از ترجمه است که منجر به بازرگانی مدارک نامربوط با نیاز کاربران خواهد شد.

۲-۴. بازیابی بین زبانی

نوعی از بازیابی اطلاعات است که در آن دست کم دو زبان وجود دارد: زبان عبارت جستجو، و زبان مجموعه مدارک. زبان عبارت جستجو را زبان مبدا (زبان اصلی)، و زبان مجموعه مدارک را زبان هدف یا مقصد می‌نامند. یک نظام بازیابی اطلاعات بین زبانی، مدارک را در زبانی که با زبان عبارت جستجو متفاوت است، بازیابی می‌کند. البته کاربرد نظام بازیابی اطلاعات بین زبانی، عبارت جستجو را به زبان بومی خویش ارائه می‌کند، اما مدارک دریافتی بر اساس زبان مجموعه مدارک خواهد بود. این نظام، کار جستجوگرانی را که به چند زبان تسلط دارند، ساده می‌کند و در عین حال، جستجوگرانی را که فقط به یک زبان تسلط دارند قادر می‌سازد که عبارت جستجو را به زبان خود وارد کند و آنگاه با استفاده از دانش خود یا با بهره‌گیری از کمک دیگران، بین مدارک بازیابی شده تمایز قائل شوند. مدارکی که مربوط تشخیص داده شده است، سپس با استفاده از عامل انسانی یا ماشینی، ترجمه و استفاده می‌شود (Ballesteros and Croft 1998).

۳-۴. ربط

ربط به عنوان مفهومی ادراکی دربردارنده ساختاری تعاملی و پویاست که ضمن داشتن قصد در قالب یک زمینه از طریق استنباط حاصل می‌شود (داورپناه ۱۳۸۳). ساراسویک، ربط را مقیاسی برای سنجش اثربخشی تماس بین یک منبع و یک مقصد (گیرنده) در فرایند ارتباط دانسته است. یعنی اگر پیامی از منبعی ارسال گردد و منبع آن را دریافت کند و در ساختار دانش وی تغییر ایجاد کند، بدین معنی است که آن پیام مرتبط است. این تعریف ربط را از دیدگاه ارتباط و اثربخشی آن سنجیده است (Sarasevic 1996).

در بافت علم اطلاع‌رسانی، ربط شاخص یا معیاری است که کارایی تبادل اطلاعات میان افراد و نظام‌های بازیابی اطلاعات را در فرایند ارتباطی و بر اساس ارزش‌گذاری افراد نشان می‌دهد. مفهوم ربط در علم اطلاع‌رسانی به طور معمول توسط پرسشگر برای قضاوت درباره این به کار می‌رود که آیا پیام ارائه شده کمکی به رفع خواست اطلاعاتی می‌کند یا نه. ساراسویک چند عامل را در امر قضاوت ربط دخیل می‌داند: ربط سیستمی (ارتباط میان سؤال و شاخص‌های پیام بازیابی شده)، ربط موضوعی (ارتباط میان مباحث مطرح شده در سؤال و موضوعات پیام‌های بازیابی شده)، ربط شناختی (ارتباط میان خواست اطلاعاتی و اطلاعات به دست آمده از پیام‌های بازیابی شده) و ربط انگیزشی (ارتباط میان انگیزه و هدف پرسشگر و

اطلاعات بازیابی شده) (اخوتی ۱۳۸۳). لازم به اشاره است که در این پژوهش از ربط موضوعی برای قضاوت ربط مدارک بازیابی شده استفاده شده است.

۵. پیشینه پژوهش

در بازیابی‌های بین‌زبانی از آنجا که به‌طور معمول، عبارت‌های درخواست کوتاه هستند، نمی‌توانند بافتار مناسبی را برای بازیابی فراهم آورند. همچنین، به جای استفاده از جملات باید از ترکیب چندین کلیدواژه استفاده کرد. بنابراین، ابزارهای بازیابی بین‌زبانی در تشخیص و ترجمه بامشکل مواجه می‌شوند و به دنبال آن، ابهام در ترجمه عبارت‌های درخواست پیش می‌آید. در ادامه، به پژوهش‌هایی اشاره می‌شود که هر یک به نوعی، بازیابی بین‌زبانی و ابزارها و فنون آن را مورد بررسی قرار داده‌اند و به بیان مشکلات و راهکارهایی برای بهبود در ترجمه درخواست پرداخته‌اند.

مای‌انگ و جان مطالعه‌ای در ترجمه درخواست مبتنی بر واژه‌نامه در بازیابی اطلاعات بین‌زبانی انجام داده‌اند. پژوهش آنها بر این فرض استوار است که داشتن یک روش ترجمه درخواست ثمربخش، یکی از عناصر اساسی برای یک نظام کاربردی بازیابی اطلاعات بین‌زبانی با کیفیت قابل قبول، محسوب می‌شود. این مطالعه پس از بیان مختصر روش‌های ترجمه درخواست و برآورد نتایج برای روش‌های ترجمه درخواست مبتنی بر واژه‌نامه، به تشریح یک روش ساده اما ثمربخش پرداخته است که در آن وابستگی اطلاعاتی برای حل مشکل ابهام به کار گرفته می‌شود که به عنوان عمده‌ترین عامل برای عملکرد پایین موقعیت تک‌زبان ساخته شده است. نتایج تجربی حاصل از این پژوهش بر روی حجم عظیمی از عبارت‌های جستجو در مجموعه "ترک" نشان داده است که رویکرد این مطالعه در مقایسه با رویکردهای بکار گرفته شده قبلی، برای بهبود ثمربخشی بازیابی، بسیار موفقیت‌آمیز بوده است (Myaeng and Jang, 1999).

سی‌او، کیم، و ریم در مطالعه‌ای با عنوان "بهبود ترجمه درخواست در بازیابی اطلاعات بین‌زبانی انگلیسی - کره‌ای"، برای غلبه بر ابهام در ترجمه عبارت‌های درخواست کاربران، روشی ارائه می‌کند که در آن بهترین ترجمه برای عبارت درخواست از بین گزینه‌هایی که برای هر عبارت درخواست وجود دارد (کاندیدها)، انتخاب می‌شود. یعنی تمامی ترکیبات ترجمه عبارت درخواست که با اصطلاحات درخواست به زبان مبدا مطابقت دارند ساخته شده و سپس، بهترین آنها برای بازیابی اطلاعات مورد استفاده قرار می‌گیرد. این روش با استفاده از

مجموعه آزمون بازیابی اطلاعات بین زبانی انگلیسی - کره‌ای (NTCIR-3)، مورد سنجش قرار گرفت، نتایج حاکی از بهبودهایی در ترجمه درخواست است (Seo, Kim, and Rim 2004). کیم، سی‌اُو و ریم در پژوهشی دیگر با عنوان "بازیابی اطلاعات با استفاده از معانی کلمات: رویکرد برچسب‌زنی اصل معنی" به ارائه الگویی با استفاده از معانی کلمات برای بازیابی اطلاعات بین‌زبانی پرداخته‌اند. این الگو، معانی صحیح کلمه را که در وردنت به عنوان عبارت‌های درخواست و عبارت‌های مدرک تعیین شده است، از طریق یک روش کنترل‌نشده، با استفاده از اطلاعات هم‌رخدادی که به صورت خودکار ساخته می‌شود، تعیین می‌کند. در واقع، این روش امکان رفع مشکلاتی را که منجر به ابهام در ترجمه عبارت‌ها می‌گردد، فراهم می‌کند. نتایج تجربی حاصل از این پژوهش بر روی حجم عظیمی از عبارت‌های جستجو در مجموعه "ترک" نشان داده است که رویکرد این مطالعه در مقایسه با رویکردهای به کار گرفته شده قبلی، برای بهبود ثمربخشی بازیابی، بسیار موفقیت‌آمیز بوده است (Seo, Kim, and Rim 2004).

کیشیدا پژوهشی با عنوان "فنون رفع ابهام بر اساس مجموعه مدارک هدف برای بازیابی اطلاعات بین زبانی: سنجش تجربی عملکرد فنون مختلف" ارائه داده است. از آنجا که ترجمه درخواست مبتنی بر واژه‌نامه برای بازیابی اطلاعات بین‌زبانی، بیشتر منجر به ایجاد ابهام می‌گردد و ترجمه‌های متفاوت با معانی گوناگون را برای یک عبارت درخواست مبدأ ارائه می‌دهد. در این پژوهش روش‌هایی برای رفع ابهام ترجمه‌هایی که فقط بر اساس مجموعه‌های مدارک هدف انجام می‌گیرد، مورد بررسی قرار گرفته است. در ابتدا، دو روش برای رفع ابهام شرح داده شده است: (۱) روش استفاده از آمارهای هم‌رخدادی کلمه در مجموعه؛ (۲) روشی که بر اساس بازخورد ربط نما انجام می‌گیرد. سپس، این فنون با استفاده از مجموعه آزمون CLEF 2003 برای جستجوهای دوزبانه آلمانی - ایتالیایی، به صورت تجربی مورد مقایسه قرار گرفتند. این بررسی‌ها نشان داد که فنون هم‌رخدادی کلمات از این جهت که بهترین الگوریتم برای انتخاب ترجمه‌ها را فراهم می‌کند، برتر است و از طرف دیگر، روش بازخورد ربط نما در زمینه جستجو کارایی بالایی از خود نشان داده است، اگر چه آزمون‌های آماری به قدر کافی این نتایج را حمایت نمی‌کند. همین بررسی‌ها در مورد جستجوهای بین‌زبانی در زبان‌های فرانسه - ایتالیایی و انگلیسی - ایتالیایی با استفاده از مجموعه آزمون CLEF 2003 نیز تکرار شده است و در نهایت، همین نتایج به تأیید رسیده‌اند (Kishida 2007).

نگیوان و همکاران در پژوهش خود با عنوان "ویکی ترنسلیت: ترجمه درخواست برای

بازیابی اطلاعات بین‌زبانی، فقط با استفاده از ویکی‌پدیا"، به معرفی ویکی‌ترنسلیت به عنوان سامانه‌ای که ترجمه درخواست را برای بازیابی اطلاعات بین‌زبانی را انجام می‌دهد، پرداختند. آنها از طریق جستجوی مباحثی از زبان‌های هلندی، فرانسه و اسپانیایی در بین مجموعه‌ای از اطلاعات زبان انگلیسی، به ارزیابی ویکی‌ترنسلیت پرداختند و در پایان به این نتیجه رسیدند که این ابزار، از کارایی ۶۷ درصدی در مقایسه با یک نظام جستجوی تک‌زبان برخوردار است (Nguyen et al. 2009).

با نگاهی به پژوهش‌های پیشین مشخص می‌شود که به عقیده پژوهشگران قبلی "رفع ابهام در ترجمه درخواست اساسی‌ترین راه‌حل در بازیابی‌های بین‌زبانی است. بنابراین، داشتن یک روش ترجمه درخواست ثمربخش، یکی از عناصر اساسی برای یک نظام کاربردی بازیابی اطلاعات بین‌زبانی با کیفیت و قابل قبول محسوب می‌شود" (Myaeng and Jang 1999; Seo, Kim, and Rim 2004; Kishida 2006). این پژوهشگران روش‌های مختلفی را به منظور غلبه بر ابهام در ترجمه درخواست به آزمون گذاشته و به این نتیجه رسیده‌اند که این روش‌ها در بهبود ثمربخشی بازیابی، بسیار موفقیت‌آمیز بوده است (Myaeng and Jang 1999; Seo, Kim, and Rim 2004; Kishida 2006).

در این پژوهش، ابزار بازیابی بین‌زبانی Google مورد بررسی قرار گرفته و قصد بر آن است که مشخص شود آیا بر اساس پژوهش‌های پیشین و با توجه به اینکه در ترجمه ماشینی احتمال بروز ابهام در ترجمه درخواست وجود دارد؛ مدارک کپکه در بازیابی بین‌زبانی از طریق این ابزار بازیابی می‌شود، از ربط موضوعی برای رشته کشاورزی برخوردار است یا خیر.

۶. روش گردآوری اطلاعات

۱. در مرحله اول، به تعدادی کلیدواژه فارسی در زمینه موضوعی کشاورزی نیاز بود. بنابراین، برای استخراج این کلیدواژه‌ها، از مقاله‌های فارسی موجود در مجلات تخصصی کشاورزی کمک گرفته شد. برای پیدا کردن مجلات تخصصی کشاورزی از بانک اطلاعات نشریات کشور و مجلات ایرانی (مگیران) استفاده شد. به این ترتیب که با وارد کردن واژه کشاورزی، به جستجوی مجلات این حوزه پرداخته شد، مجلات فارسی دارای چکیده انگلیسی (از جمله فصلنامه پژوهش و سازندگی، مجله دانش کشاورزی پایدار) انتخاب شدند. این دست مجلات نسبت به سایر مجلات حوزه کشاورزی، برای انتخاب در پژوهش حاضر ارجحیت داشتند، چراکه در مجلات فارسی

دارای چکیده انگلیسی، نویسنده مقاله، که متخصص این حوزه موضوعی است، اصطلاحات تخصصی را ترجمه کرده است، و دیگر نیازی به استفاده از اصطلاحنامه برای یافتن معادل انگلیسی کلمات وجود ندارد. از جهتی معادلی که نویسنده مقاله (متخصص موضوعی) ارائه می‌دهد به صحت نزدیک‌تر است. سپس، در آرشو این مجلات به جستجوی مقالات حوزه‌های موضوعی متفاوت کشاورزی (همچون زراعت، اقتصاد کشاورزی، آبیاری زهکشی، جنگل و مراتع، باغبانی، آبخیزداری، منابع طبیعی، و...) پرداخته شد.

در مجموع ۱۰ مقاله با خصوصیات گفته‌شده انتخاب گردید. همچنین، از نمایه اصطلاحات فارسی-انگلیسی موجود در پایان دو کتاب تخصصی این حوزه نیز استفاده شد. در نهایت، ۳۰ اصطلاح و عبارت موضوعی تخصصی از مجموع مقالات و کتاب‌های منتخب، استخراج شد.^۱ این ۳۰ اصطلاح موضوعی در قالب سه دسته گنجانده شدند: دسته اول، عبارت‌های موضوعی که فقط در رشته کشاورزی کاربرد دارند (مثل پرآبیاری، شاخص سطح برگ، ظرفیت زراعی و...). دسته دوم، اصطلاحات تخصصی کشاورزی که در سایر رشته‌ها نیز کاربرد دارند (مثل حدآستانه، جزء مولی، خاکدانه و...).^۲ دسته سوم، اصطلاحات تخصصی کشاورزی که خارج از این حوزه تخصصی، اصطلاح عمومی محسوب می‌شوند (مثل رطوبت‌زایی، بازیافت و...). معادل انگلیسی این عبارت‌ها و اصطلاحات نیز از همان منبع استخراج کلمات گرفته شد.

۲. پس از استخراج کلیدواژه‌ها، به بازیابی بین‌زبانی پرداخته شد. در این مرحله، از کاوشگر google به عنوان ابزار جستجوی بین‌زبانی مدارک استفاده شد. در این پژوهش، منظور از بازیابی بین‌زبانی آن است که پژوهشگر یک عبارت درخواست در حوزه موضوعی کشاورزی را به زبان خود، برای نظام بازیابی اطلاعات بیان کند و نظام از بین منابع موجود در زبان یا زبان‌های دیگر مدنظر پژوهشگر، مدارک را بازیابی کند. در واقع، درخواست به زبان مادری وارد می‌شود و نظام بازیابی اطلاعات، پس از ترجمه

۱. (Kim, Seo, and Rim 2004) از ۲۵ کلمه برای بازیابی اطلاعات با استفاده از معانی کلمات استفاده کرده‌اند. در پژوهش (Myaeng and Jang 1999) نیز در یک مرحله، ۴۸ واژه کره‌ای و در مرحله دیگر، ۲۴ واژه کره‌ای در بازیابی بین‌زبانی مورد استفاده قرار گرفته‌اند.

۲. به‌طور معمول اصطلاحات دسته دوم، علاوه بر رشته کشاورزی در رشته‌هایی مانند فیزیک، علوم زمین، زیست‌شناسی نیز به کار گرفته می‌شوند.

عبارت درخواست، در بین منابع موجود در سایر زبان‌ها مدارک را جستجو می‌کند. بدین منظور:

۱) وارد صفحه اصلی موتور جستجوی گوگل می‌شود و واژه مورد نظر را وارد می‌کند، سپس دکمه Search را انتخاب می‌کند تا بازیابی اولیه توسط کاوشگر google انجام شود.

۲) در کادر کناره چپ صفحه گزینه More search tools را انتخاب کرده.

۳) از منوی باز شده گزینه Translated foreign pages را برگزیده، تا کادر مربوط به بازیابی بین‌زبانی در کاوشگر google بر روی صفحه نمایش داده شود.

۴) در کادر باز شده گزینه my language را انتخاب و بر روی زبان فارسی در منوی باز شده زیر آن کلیک می‌کند تا مشخص شود که ابزار، مدارک بازیابی شده را در نهایت به چه زبانی ترجمه کند.

۵) گزینه Add language را انتخاب و بر روی زبان انگلیسی در منوی باز شده زیر آن کلیک می‌کند؛ در واقع، زبان انگلیسی را به عنوان زبان مقصد (زبان مدارک) برگزیده است.

۶) سپس، معادل عبارت درخواستی به زبان یا زبان‌های مقصد همراه با تعداد مدارک بازیابی شده نمایش داده می‌شود. چون فقط مدارک بازیابی شده به زبان انگلیسی مدنظر پژوهشگر است، بر روی تعداد مدارک بازیابی شده در این زبان کلیک می‌کند تا در صفحه بعد فقط مدارک بازیابی شده از این زبان نمایش داده شود.

۳. مرحله سوم، برآورد میزان ربط^۱ مدارک بازیابی شده توسط این ابزار است. در این پژوهش، منظور از ربط، ربط موضوعی است؛ بدین معنی که نتایج بازیابی شده از طریق بازیابی بین‌زبانی در کاوشگر google با مباحث رشته کشاورزی چه میزان ارتباط دارد و میزان ربط آن نیز بدین معنی است که از بین ۱۰ مدرک اول بازیابی شده با استفاده از ابزار بازیابی بین‌زبانی google، چند درصد مدارک از ربط موضوعی با مباحث رشته کشاورزی برخوردار است (تشخیص مربوط بودن یا نبودن مدارک بازیابی شده توسط

۱. در اینجا ارزیابی میزان ربط بر اساس مدل سنتی سنجش ربط صورت گرفته است. در مدل سنتی، ربط در نتیجه مطابقت پرسش و منابع بازیابی شده حاصل می‌شود و از طریق بررسی این مطابقت مورد ارزیابی قرار می‌گیرد (Sarasevice 1996).

پژوهشگر انجام گرفته است). در نهایت، به دسته‌بندی نتایج به دست آمده و ارائه توصیفی و نتیجه‌گیری از آنها پرداخته شد.

۷. ارائه توصیفی داده‌ها و پاسخ به سؤالات پژوهش

پاسخ سؤال اول پژوهش

جدول ۱. میزان ربط مدارک بازیابی شده برای اصطلاحات و عبارات مختص رشته کشاورزی

میزان ربط (درصد)	معادلی که ابزار بازیابی بین زبانی Google برای واژه در نظر گرفته است.	معادل تخصصی واژه در مقالات کشاورزی	عبارت‌ها و اصطلاحات
۲۰	Reference plant	Reference crop	گیاه مرجع
۵۰	Leaching	leaching	آبشویی
۹۰	Field capacity	Field capacity	ظرفیت زراعی
۹۰	Hydrological regime	Hhydrologic Regime	رژیم هیدرولوژیکی
۶۰	Full irrigation	Over-irrigation	پر آبیاری
۱۰۰	Seed inoculation	Seed inoculation	تلقیح بذر
۲۰	The moisture curve	absorbtion curves	منحنی رطوبت‌گیری
۱۰۰	Wilting point	Wilting point	نقطه پژمردگی
۹۰	Plant density	Plant density	تراکم گیاه
۰	LAI	LAI	شاخص سطح برگ

در مجموع، ابزار بازیابی بین زبانی Google، در بازیابی بین زبانی مدارک مرتبط، در مورد اصطلاحات و عبارات مختص رشته کشاورزی، به طور متوسط ۶۲ درصد با موفقیت^۱ عمل می‌کند. بالاترین میزان ربط، ۱۰۰ درصد است که در ۳۰ درصد موارد رخ داده است، اما پایین‌ترین میزان ربط صفر درصد است که در ۱۰ درصد موارد مشاهده شده است، در بقیه موارد نیز میزان ربط به طور متوسط ۴۶/۶ درصد می‌باشد. با اینکه google یک موتور جستجوی

۱. در اینجا موفقیت بدین معنی است که آیا موضوع مدارک بازیابی شده با درخواست مطرح شده که در حوزه موضوعی کشاورزی است، ارتباط دارد یا خیر. این همان ربط موضوعی از دیدگاه ساراسویک است که می‌گوید: ربط موضوعی ارتباط میان مباحث مطرح شده در سؤال و موضوعات پیام‌های بازیابی شده است.

عمومی است، اما در مورد بازیابی اصطلاحات مختص رشته کشاورزی به نسبت خوب عمل می‌کند و ۶۲ درصد ربط در بازیابی مدارک نتیجه به نسبت خوبی است.

پاسخ سؤال دوم

جدول ۲. میزان ربط مدارم بازیابی شده برای اصطلاحات و عبارات تخصصی رشته کشاورزی که در رشته‌های دیگر نیز کاربرد دارند.

میزان ربط (در صد)	سایر رشته‌هایی که اصطلاح را به کار می‌برند.	معادلی که ابزار بازیابی بین‌زبانی Google برای واژه در نظر گرفته است.	معادل تخصصی واژه در مقالات کشاورزی	اصطلاحات تخصصی
۰	فیزیک، علوم اجتماعی	Threshold	Threshold level	حد آستانه
۰	روانشناسی	Tension	stress	تنش
۶۰	فیزیک	The amount of light received	light interception	میزان دریافت نور
۱۰	فیزیک	Depreciation factor, light	Light extinction Coefficient	ضریب استهلاک نور
۰	علوم زمین	Aggregate	Soil aggregate	خاکدانه
۰	فیزیک	Diffusivity	Diffution	پخشیدگی
۰	فیزیک، علوم زمین، سیستم‌های رایانه‌ای	Infiltration	Infiltration	نفوذ پذیری
۰	شیمی	Mole fraction	Moli fraction	جزء مولی
۶۰	زیست	Rhizobium	Rhizobia	ریزوبیوم
۱۰	علوم زمین	Weathering	Weathering	هوازدگی

ابزار بازیابی بین‌زبانی Google در بازیابی مدارک مرتبط با اصطلاحات تخصصی کشاورزی که در سایر رشته‌ها نیز کاربرد دارند، به طور متوسط ۱۴ درصد با موفقیت عمل می‌کند. بالاترین میزان ربط، ۶۰ درصد و پایین‌ترین آن صفر درصد است. در ۶۰ درصد موارد، میزان ربط در بازیابی مدارک، صفر است. اما در مورد ۴۰ درصد باقی‌مانده، میزان ربط ۳۵ درصد است. شاید بتوان بزرگ‌ترین دلیل پایین بودن میزان ربط در بازیابی مدارک برای این دست اصطلاحات را ابهام در ترجمه دانست. به طور کلی، ابزار بازیابی بین‌زبانی Google در بازیابی مدارک مرتبط با این دست اصطلاحات، ضعیف عمل می‌کند.

پاسخ سؤال سوم

جدول ۳. میزان ربط مدارک بازیابی شده برای اصطلاحات رشته کشاورزی که خارج از این حوزه تخصصی اصطلاح عمومی محسوب می شوند.

میزان ربط (درصد)	معادلی که ابزار بازیابی بین زبانی Google برای واژه در نظر گرفته است.	معادل تخصصی واژه در مقالات کشاورزی	اصطلاحات تخصصی
۲۰	Bud	Seedling	جوانه
۱۰	Content creation	Adsorbtion	رطوبت زایی
۰	Dehydration	Diffic-it-irrigation	کم آبی
۳۰	Root	root	ریشه
۲۰	Recycling	Recycling	بازیافت
۰	Species	Kind	گونه
۱۰	Porosity	Pore	تخلخل
۰	Collar	collar	یقه
۰	Flooding	Water logging	غرقاب
۰	Adherence	Adhersion	چسبندگی

ابزار بازیابی بین زبانی Google، در بازیابی بین زبانی مدارک مرتبط با اصطلاحات تخصصی کشاورزی که در خارج از حوزه تخصصی، واژگان عمومی محسوب می شوند، به طور متوسط ۹ درصد با موفقیت عمل می کند. بالاترین میزان ربط، ۳۰ درصد و پایین ترین آن صفر درصد است. در ۵۰ درصد موارد میزان ربط، صفر است و در ۵۰ درصد باقی مانده مدارک بازیابی شده از ۱۸ درصد ربط برخوردار هستند. به طور کلی، ابزار بازیابی بین زبانی Google در بازیابی مدارک مرتبط با این دسته از اصطلاحات، بسیار ضعیف عمل می کند.

۸. بحث و نتیجه گیری

با توجه به داده های موجود در سه جدول ۱، ۲ و ۳ می توان چنین نتیجه گرفت که ابزار بازیابی بین زبانی Google در بازیابی های بین زبانی مدارک مرتبط با اصطلاحات و عبارات تخصصی که فقط مختص رشته کشاورزی هستند، بهتر عمل می کند (۶۳ درصد مدارک بازیابی شده، مدارک مربوط هستند)، علت این امر آن است که واژگان و عبارتهایی که فقط مختص رشته کشاورزی هستند، همان طور که از اسمشان پیداست فقط در همین رشته کاربرد

دارند. بنابراین، با وارد کردن آنها نمی‌توان انتظار داشت که مدارکی در حوزه‌های دیگر برای ما بازیابی شود. در بازیابی مدارک مرتبط با اصطلاحاتی که در سایر رشته‌ها نیز کاربرد دارند و یا اصطلاحاتی که در خارج از حوزه تخصصی، اصطلاح عمومی محسوب می‌شوند، عملکرد موفقیت‌آمیزی ندارد. در مورد دسته اول می‌توان گفت چون این اصطلاحات در رشته‌های دیگر نیز به کار می‌روند، بدیهی است که در بازیابی، مدارک مربوط به رشته‌های دیگر نیز بازیابی شوند. ضعیف‌ترین میزان موفقیت در بازیابی مدارک مرتبط، مربوط به دسته سوم اصطلاحات مد نظر در این پژوهش است؛ یعنی اصطلاحاتی که در واقع اصطلاح عمومی هستند، اما در رشته کشاورزی کاربرد تخصصی دارند. در مورد این کلمات چون کاربردی عمومی دارند احتمال اینکه مدارک بازیابی شده با رشته‌ای تخصصی مثل کشاورزی مرتبط باشند، اندک است. بنابراین، برای جلوگیری از ریزش کاذب در مورد گروه دوم و سوم عبارات‌ها پیشنهاد می‌شود در کنار عبارات‌ها و اصطلاحاتی که کاربرد عمومی دارند یا در رشته‌هایی غیر از کشاورزی نیز کاربرد دارند، واژه‌هایی وارد شوند که مفهوم مد نظر کاربر را به نحوی دقیق‌تر مشخص کنند. به عنوان مثال، در کنار عبارات‌هایی مانند "جزء مولی، نفوذپذیری، پخشیدگی و ..." که کاربرد زیادی در سایر رشته‌ها دارند، واژه‌هایی مثل "خاک" یا هر واژه دیگری که نمایانگر ارتباط عبارت با مباحث کشاورزی و منظور کاربر است، وارد شوند.

اما دلایلی که می‌توان برای عدم کارایی این ابزار در بازیابی‌های بین‌زبانی مرتبط برشمرد از این قرار است:

ابهام در ترجمه یکی از دلایل عدم کارایی این ابزار در بازیابی بین‌زبانی مدارک مرتبط است. به عنوان مثال، عبارت "منحنی رطوبت‌گیری" که معادل انگلیسی صحیح آن در حوزه تخصصی کشاورزی *absorbtion curves* است را *moisture curves making* ترجمه می‌کند. یا به عنوان مثال، عبارت جوانه را *Bud* ترجمه می‌کند، در حالی که ترجمه صحیح آن در حوزه تخصصی کشاورزی *seedling* است و این ایراد باعث بازیابی مدارک نامربوط خواهد شد. اما اگر به دقت به جدول‌ها بنگریم متوجه می‌شویم که ابهام در ترجمه درخواست نقش چندانی در بروز بازیابی‌های نامربوط ندارد. در مورد ۵۰ درصد عبارات‌های درخواست (آبشویی، ظرفیت زراعی، رژیم هیدرولوژیکی، تلقیح بذر، نقطه پژمردگی، تراکم گیاه، شاخص سطح برگ، پخشیدگی، جزء مولی، ریزوبیوم، هوازدگی، ریشه، بازیافت، یقه، نفوذپذیری)، ابزار ترجمه به طور دقیق، همان ترجمه‌ای را ارائه داده که متخصص موضوعی در نظر داشته و در مورد ۲۰ درصد از عبارات‌ها (گیاه مرجع، پرآبیاری، حد آستانه، خاکدانه، تخلخل، چسبیدگی)، ابزار

ترجمه، معادلی نزدیک به ترجمه مدنظر متخصص موضوعی ارائه داده است. بنابراین، در حوزه بازیابی های تخصصی (رشته کشاورزی) نمی توان ابهام در ترجمه درخواست را مهم ترین علت بازیابی های نامربوط دانست. این بر خلاف نظر پژوهشگرانی است که اساسی ترین علت بازیابی های بین زبانی ناموفق را ابهام در ترجمه درخواست می دانند (Myaeng and Jang 1999; Seo, Kim, and Rim 2004; Kishida 2006).

لازم به اشاره است که نمی توان فقط کاوشگر و ابزارهای ترجمه بین زبانی را علت بازیابی های نامربوط دانست، بلکه عملکرد کاربر در بازیابی نیز در این بین نقشی اساسی ایفا می کند. باید توجه داشت که کاوشگر بر اساس الگوریتم های بازیابی که برایش تعریف شده است (مانند همانندی) اقدام به بازیابی می کند و این کاربر است که باید منظور خود را هر چه دقیق تر برای کاوشگر بیان کند. نمی توان انتظار داشت که با وارد کردن عبارتی با کاربرد عمومی یا عبارتی که در چند حوزه کاربرد دارد به بازیابی های بی ربط دست یافت، بلکه لازم است کاربر از عبارات دقیق برای بیان منظور خود استفاده کند.

دلیل دیگر اینکه در مورد عبارات هایی که دارای علامت اختصاری معادل به زبان انگلیسی هستند، ابزار بازیابی بین زبانی Google عبارت را به علامت اختصاری انگلیسی مختص آن ترجمه می کند، اما تمامی مدارکی را بازیابی می کند که این علامت اختصاری با معانی دیگر را نیز شامل می شوند. به عنوان مثال، در مورد عبارت "شاخص سطح برگ" علامت اختصاری انگلیسی معادل آن LAI است که از Leaf Area Index گرفته شده است. ماشین علامت اختصاری را درست تشخیص می دهد، اما در مرحله بعدی که بازیابی مدارک به زبان مقصد است، تمامی مدارکی را بازیابی می کند که این علامت اختصاری را حتی با معانی متفاوت دارند. این دلیلی برای بازیابی مدارک غیر مرتبط با اصطلاح یا عبارت مورد نظر است.

دیگر اینکه ابزار نمی تواند معادل انگلیسی برخی از اصطلاحات را تشخیص دهد. به عنوان مثال، عبارت "هدایت موئینگی" به هر شکلی که نوشته شود مانند هدایت موئینگی، هدایت موئینگی، و... ماشین نمی تواند معادلی برای آن پیدا کند. همین طور است در مورد واژه خمیرائی. بنابراین می توان گفت ویژگی های زبان فارسی نیز در این بین دخیل هستند و به نوبه خود مشکلاتی را ایجاد می کنند.

بعضی از واژه ها را به طور کامل اشتباه ترجمه می کند، به عنوان مثال، واژه "همروند" که معادل صحیح آن isotrope است را simultaneous ترجمه می کند. یا در مورد واژه خاکدانه نیز همین گونه عمل می کند.

اما به طور کلی، اگر میزان ربط مدارک بازیابی شده برای هر سه دسته اصطلاح مد نظر را محاسبه کنیم به این نتیجه خواهیم رسید که ابزار بازیابی بین‌زبانی Google در بازیابی مدارک مرتبط، در حوزه موضوعی کشاورزی به طور متوسط با ۲۱ درصد موفقیت عمل می‌کند. یعنی عملکردی ضعیف در بازیابی مدارک مرتبط در این حوزه تخصصی دارد.

۹. پیشنهادات

- از آنجا که استفاده از ابزارهای بازیابی بین‌زبانی، به کاربران (به خصوص کاربران متخصص) کمک می‌کند که بدون نیاز به تسلط کامل بر زبان‌های دیگر و با صرف مدت زمان کمتر به بازیابی مدارک موجود در سایر زبان بپردازند، بنابراین پیشنهاد می‌شود در مورد وجود و امکان استفاده از چنین ابزارهایی اطلاعات لازم در اختیار کاربران قرار داده شود.
- آگاهی از نحوه جستجوی صحیح برای انجام یک بازیابی موفق امری ضروری است، بنابراین حتی اگر کاربران از وجود بهترین ابزارهای بازیابی مطلع باشند، بدون آگاهی از نحوه اطلاع‌یابی صحیح از طریق این ابزارها نمی‌توانند به شکل مؤثر از آنها بهره‌وری کنند. بنابراین، پیشنهاد می‌شود راهبردهای جستجوی صحیح و نحوه استفاده از ابزارهای کاوش به کاربران آموزش داده شود، چرا که اگر کاربر بتواند درخواست خود را در قالب عبارت‌های دقیق و صحیح برای ابزار کاوش بیان کند، بخش زیادی از مشکل حل شده است.
- پیشنهاد می‌شود در پژوهش‌های دیگری سایر ابزارهای رایگان بازیابی بین‌زبانی شناسایی، معرفی و عملکرد آنها در بازیابی مدارک مورد بررسی قرار گیرد.
- پیشنهاد می‌شود در پژوهشی، دیگر ابزارهای جستجو در کاوشگر گوگل مورد بررسی قرار گیرد و نواقص آنها به بحث گذاشته شود.

۱۰. منابع

- اخوتی، مریم. ۱۳۸۳. مفهوم ربط در نظام‌های بازیابی اطلاعات: مروری بر نظریه‌ها و ادبیات موجود. *مجله اطلاع‌شناسی* ۲ (۱): ۲۳-۴۵.
- خالویی، مرضیه. ۱۳۸۷. ربط و مفهوم آن در بازیابی اطلاعات. *علوم و فناوری اطلاعات* ۲۳ (۳): ۱۰۵-۱۱۸.
- داورپناه، محمدرضا، و محمد رضایی. ۱۳۸۳. مفهوم ربط. *مجله اطلاع‌شناسی* ۲ (۱): ۴۷-۶۲.

علیزاده، حمید، رحمت‌الله فتاحی، و محمدرضا داورپناه. ۱۳۸۸. بررسی کارآمدی روشهای موجود در بازیابی اطلاعات بین زبانی فارسی-انگلیسی با استفاده از واژه‌نامه دو زبانه ماشین خوان. *علوم و فناوری اطلاعات* ۲۵ (۱): ۵۳-۶۹.

علیزاده، حمید. ۱۳۸۳. مشکلات دسترسی به اطلاعات در جهان شبکه‌ها. *فصلنامه کتاب* ۱۵ (۲): ۱۱۵-۱۲۱. ج. مختاری، حیدر، و عباس میرزایی. ۱۳۸۹. *رابط در علم اطلاع‌رسانی*. تهران: چاپار.

Ballesteros, L., B. Croft. "Resolving Ambiguity for Cross- Language Retrieval (Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval)". ACM: New York. pp:64-71. (1998). <http://dl.acm.org/citation.cfm?id=290958> (accessed 07 Aug. 2013).

Edwards, J. *Multilingualism*. London: Penguin. 1994.

Kim, Sang-Bum., Seo, Hee-cheol., Rim, Hae-chang. 2004. Information Retrieval Using Word Sense: Root Sense Tagging Approach (in proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval)". ACM: New York 258-265. <http://dl.acm.org/citation.cfm?id=1009038> (accessed 07 Aug. 2013).

Kishida, Kazuaki. 2007. "Term disambiguation techniques based on target document collection for cross-language retrieval: An empirical comparison of performance between techniques". *Information Processing and Management* 43:103-120.

Masuchi, Hiroshi., Kaufman, Stefan., Flournory, Raymond., Peters, Stanley. 1999. Query Translation Method for Cross Language Information Retrieval (In: Proceedings of the Workshop on Machine Translation for Cross Language Information Retrieval, Singapore)". *MT Summit VII* pp: 30-34.

Myaeng, Sung Hyon., Jang, Myung-Gil. 1999. Complementing Dictionary- Based Query Translation With Corpus Statistic for Cross- Language IR. *MT Summit VII* pp: 165-171.

Nguyen, D., Overwijk, A., Hauff, C., Trieschnigg, R.B., Hiemstra, D., G.de jong, F.M. 2009. "Wiki Translate: Query Translation for Cross Language Information Retrieval using only Wikipedia (In: Proceedings of the 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008)". Springer Berlin Heidelberg. pp: 58-65. http://link.springer.com/chapter/10.1007%2F978-3-642-04447-2_6 (accessed 07 Aug. 2013).

Sarasevic, Tefko. 1996. Relevance reconsidered. *Information science: Integration in perspectives*". Proceedings of the Second Conference on Conceptions of Library and Information Science. Copenhagen (Denmark), 14-17 Oct pp: 201-218.

Savoy, Jacques. 2002. "Cross-Language Information Retrieval: Experiments Based on CLEF 2000 Corpora". *Information Processing and management* 39:75-115.

Seo, Hee-Cheol., Kim, Sang-Bum., Rim, Hae-Chang., Myaeng, Sung-Hyon. 2004. "Improving Query Translation in English-Korean Cross-Language Informaton Retrieval". *Information Processing and management* 41:507-522.

Wang, Jianqiang., W. Oard, Douglas. 2005. "Document and Query Expansion Using Side Collections and Thesauri (In Proceedings of the CLEF 2005 Workshop, Vienna, Austria)". Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007%2F11878773_88 (accessed 07 Aug. 2013).

Yibo, Zhang., Le, Sun., Lin, Du., Yufang, Sun. 2000. "Query Translation in Chines-English Cross Language Information Retrieval (Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics)". Association for Computational Linguistics. 13. <http://dl.acm.org/citation.cfm?id=1117807> (accessed 07 Aug. 2013).

The Measurement of Relevance Amount of Documents That by Using of Google Cross-Language Retrieval about Agriculture Subject Area Are Retrieved

Fatemeh Jamshidi Ghahfarokhi*
Master in Library and Information Science

Iranian Journal of
**Information
Processing &
Management**

Abstract: In this study, the relevance amount of documents has been investigated by using google cross-language retrieval tools about a agriculture subject area in cross-language retrieval form. For this purpose, by using Persian journals articles that have had English abstracts, Persian phrases and subject terms with their English equivalent were extracted. In three classes us, thirty number of phrases and subject terms of agriculture area were extracted: First class, subject phrases that only in agriculture are used; Secondary, agriculture subject terms that in other fields are used too; Third class, agriculture subject terms that out of this field are considered as public term. Then by these phrases and terms, documents were searched, and relevance amount of search results are investigated. Results of study showed that google cross-language retrieval tools for two classes of phrases and terms, in cross-language retrieval of relevance document about agriculture subject area, aren't succeed: one class, agriculture subject terms that in other fields are used too. other class, agriculture subject terms that out of agriculture field are considered as public term. Google cross-language retrieval tools about subject phrase and terms that only in agriculture field are used, are performance rather desirable than other two class of phrase and terms.

Keywords: machine translation, google cross-language retrieval tools, agriculture subject area, public term, subject term, relevance amount, cross-language retrieval

Iranian Research Institute
For Science and Technology

ISSN 2251-8223

eISSN 2251-8231

Indexed in LISA, SCOPUS & ISC

Vol.28 | No.3 | pp: 741-760

Spring 2013

*Corresponding author: Jam.fatema@gmail.com