

تحلیل استنادی و ترسیم نقشهٔ تاریخ‌نگاشتی تولیدات علمی کشاورزی در نمایهٔ استنادی علوم در سال‌های ۲۰۰۰ تا ۲۰۰۸

میترا پشوتنی‌زاده

دانشجوی دکتری کتابداری و اطلاع‌رسانی

دانشگاه شهید چمران اهواز

فریده عصاره*

دکترای کتابداری و اطلاع‌رسانی

استاد دانشگاه شهید چمران اهواز

اطلاعات
علوم و فناوری

دریافت: ۱۳۸۸/۰۴/۲۰ پذیرش: ۱۳۸۸/۰۶/۳۱ مقاله برای اصلاح به مدت ۱۵ روز نزد پدیدآوران بوده است.

چکیده: در این پژوهش تلاش شده با بررسی تعداد ۲۲,۶۱۷ رکورد حوزه کشاورزی که در پایگاه «وب‌آوساینس» در محدودهٔ زمانی ۲۰۰۰ تا ۲۰۰۸ نمایه شده‌اند و با استفاده از فنون تحلیل استنادی، نویسندگان و مؤسسات کلیدی و میزان همکاری گروهی بین آنان، مجلات هسته، نرخ رشد تولیدات علمی، قالب و زبان انتشارات، و کشورهای پیشرو در این حوزه مشخص شوند و نقشهٔ تاریخ‌نگاشتی علم کشاورزی ترسیم گردد. محاسبهٔ مقدار متوسط نرخ رشد سالانهٔ انتشارات برای این سال‌ها، عدد ۷٪ را نشان داد. بررسی نوع مدارک نشان داد که مجموع ۲۲,۶۱۷ رکورد، در ۱۵ نوع قالب مختلف ارائه شده‌اند. این رکوردها به ۲۵ زبان زندهٔ دنیا و با همکاری ۱۷۳ کشور به رشتهٔ تحریر درآمده‌اند. تمامی مقالات کشاورزی بازاریابی شده از پایگاه «وب‌آوساینس» در ۳۲۵۶ مجله منتشر گردیده و از سوی ۱۴,۸۵۲ مؤسسهٔ مختلف ارسال شده‌اند. تعداد ۵۱,۶۵۵ نویسنده از سال ۲۰۰۰ تا ۲۰۰۸، مقالات خود را در حوزه کشاورزی به رشتهٔ تحریر درآورده و در پایگاه «آی‌اس‌آی» به ثبت رسانده‌اند و در کل تعداد ۶۵۸,۳۸۱ منبع را مورد استناد قرار داده‌اند. خوشه‌های شکل گرفته در نقشهٔ تاریخ‌نگاشتی بر اساس LCS و GCS نیز شامل ۵ خوشه بود.

کلیدواژه‌ها: علم‌سنجی؛ تحلیل استنادی؛ تولیدات علمی کشاورزی؛ ساختار علم کشاورزی؛ پایگاه «وب‌آوساینس»؛ نقشهٔ تاریخ‌نگاشتی

فصلنامهٔ علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا (چاپی) ۵۲۰۶-۱۷۳۵
شاپا (الکترونیکی) ۵۵۸۳-۲۰۰۸
نمایه در LISA و SCOPUS
<http://jst.irandoc.ac.ir>
دوره ۲۵ | شماره ۱ | صص ۲۳-۵۲
پاییز ۱۳۸۸

نوع مقاله: علمی پژوهشی

* پدیدآور رابط fosareh@yahoo.com

۱. مقدمه

کتابسنجی، علم‌سنجی و اطلاع‌سنجی از جمله روش‌های کمی معتبری هستند که در ارزیابی علمی و مقایسه کشورهای، دانشگاه‌ها، مؤسسات علمی و حتی دانشمندان به طور انفرادی و بر اساس انتشارات علمی آنان، کاربرد دارند. تحلیل استنادی یکی از روش‌های معتبر در این سه حوزه (کتابسنجی، علم‌سنجی و اطلاع‌سنجی) است که با شمارش تعداد مدارک تولیدشده و تعداد استنادهای تعلق‌گرفته به آن‌ها و اطلاعات کتابشناختی آن‌ها در تعیین نویسندگان کلیدی و مجلات هسته، مطالعه تاریخ علم، شناسایی رشته‌های نوظهور احتمالی، و سیاست‌گذاری‌های علمی کمک می‌کند (عصاره ۱۳۸۴، ۲۷۳).

در گذشته انجام مطالعاتی که از روش تحلیل استنادی استفاده می‌کردند بسیار پرزحمت بودند و نتایج نیز دقت کم‌تری نسبت به زمان حال داشتند؛ ولی در زمان حاضر با گسترش و خودکارسازی فنون گردآوری اطلاعات، این گروه از مطالعات با دقت، کیفیت و سرعت بالاتری انجام می‌گیرند و نرم‌افزارهای مختلف طراحی شده نیز به ارائه تفاسیر و تحلیل‌های پیچیده‌تر از این داده‌های کتابشناختی و استنادی می‌پردازند. یکی از این نرم‌افزارها «هیست‌سایت»^۱ است که با استفاده از داده‌های استنادی پایگاه «وب‌آوساینس»^۲ نتایج ارزشمندی را در قالب جداول و نمودارهای مختلف با شاخص‌های متعدد، در اختیار پژوهشگران و علاقه‌مندان به این حوزه قرار می‌دهد.

«هیست‌سایت» نرم‌افزاری است که در نتیجه ارزیابی طولانی‌مدت نیازهای کاربران پایگاه‌های استنادی به وجود آمده است. کتابداران و کاربران نیاز به شناخت آثار مهم در یک رشته یا موضوع خاص دارند؛ پژوهشگران نیز به مرور تاریخ علم و ظهور موضوعات جدید علاقه‌مندند؛ بنابراین «هیست‌سایت» برای پاسخگویی به نیاز هر دو گروه به وجود آمد. این نرم‌افزار از یک روش داده‌کاوی بر مبنای تحلیل پیوندهای استنادی بین مدارک مختلف استفاده می‌کند. کتابداران مرجعی که ابزارهای کتابشناختی را در یک موضوع ناآشنا تهیه می‌کنند و کتابدارانی که به ترسیم دوبعدی داده‌های استنادی، تعیین آثار مهم و اثرگذار، و شناخت مجلات هسته و مهم در یک موضوع می‌پردازند، می‌توانند از این نرم‌افزار استفاده کنند (Garfield, Pudovkin, and Istom 2003a).

¹ HistCite™² Web Of Science (WOS)

با توجه به اعتبار روش تحلیل استنادی و قابلیت‌های نرم‌افزار «هیست‌سایت»، در این نوشته تلاش شده است به بررسی تحلیلی و تاریخی‌نگاشتی تولیدات علمی حوزه کشاورزی و میزان رشد و توسعه آن آثار در طول سال‌های ۲۰۰۰ تا ۲۰۰۸ پرداخته شود.

۲. بیان مسئله

این اندیشه که علم می‌تواند به صورت فضایی نمایش داده شود، دیرزمانی است که قابل ردیابی است. برای مثال «بوش» در سال ۱۹۴۵، ترسیم سه‌بعدی رشته‌های علمی را توصیف کرد. همچنین در دهه ۱۹۷۰ به اندیشه ترسیم ساختار علم در علوم اجتماعی و جغرافیای انسانی و نیز در جامعه‌شناسی اشاره شده است. تعداد زیادی از دانشمندان، ساختار علم در حوزه‌های علمی مختلف را بر اساس متون علمی آن حوزه، و با استفاده از تحلیل خوشه‌ها^۱، تحلیل عاملی^۲ و مقیاس چندبعدی^۳ ترسیم کرده‌اند (گولد و وایت^۴، ۱۹۷۴ به نقل از عصاره ۱۳۷۷؛ کرین^۵، ۱۹۷۴ به نقل از عصاره ۱۳۷۷).

ترسیم ساختار علم برای رشته‌های مختلف و پیگیری آخرین تغییرات آن‌ها، موضوع مورد توجه دانشمندان، کتابداران، فیلسوفان، دولتمردان و ناشران است و متون علمی، ماده‌های اصلی برای این ترسیم محسوب می‌شوند. در ترسیم ساختار علم سه جزء در نظر گرفته می‌شوند: عناصر فردی، عناصر مرتبط با یکدیگر که یک شبکه را به وجود آورده‌اند، و تفسیر روابط بین عناصر (Paul 2001, 65).

دانشمندان مختلف با استفاده از ابزارها و روش‌های متفاوت اقدام به مصورسازی و ترسیم ساختار علم در رشته‌های علمی مختلف پرداخته‌اند، ولی ویژگی خاص نرم‌افزار «هیست‌سایت» در ترسیم نقشه علمی بر اساس تقدم زمانی است (Garfield, Pudovkin and Istomin 2002). با توجه به مزیت‌های برشمرده، برخی از فواید تاریخ‌نگاری علمی به کمک این نرم‌افزار را می‌توان به شرح زیر دانست:

- ◇ مشخص شدن جریان اطلاعات از گیرنده به فرستنده و بالعکس؛
- ◇ نمایش توسعه و پیشرفت علم؛

¹ Cluster Analysis

² Factor Analysis

³ Multidimensional Scalings

⁴ Gold & White

⁵ Crane

- ◇ مقایسه رشد علم در سال‌های مختلف؛
- ◇ نمایش تاریخ علم؛
- ◇ مشخص شدن آثار مهم و اثرگذار در هر رشته؛
- ◇ نشان دادن موضوعات جدید و زمان مطرح شدن و رشد آن‌ها.

یکی از رشته‌هایی که سابقه طولانی در ایران دارد کشاورزی است. گفته می‌شود که ایران یکی از اولین کشورهای دنیا است که در آن، کشاورزی و تمدن شروع شده است. حتی در حفاری‌های اطراف کاشان، آثاری از یک سیستم زراعی پیشرفته مربوط به شش هزار سال پیش یافت شده است. از دوران‌های مختلف تاریخ نیز کتاب‌های مختلف کشاورزی بر جای مانده (درخشانی ۱۳۸۴) و امروزه نیز مطالعات و فعالیت‌های علمی، آگاهی از موضوعات و مسائل مهم رشته و کشف روش‌های علمی مختلف، نقش مهمی در صنعت کشاورزی هر کشور دارد. بنابراین در این پژوهش تلاش شده با شناسایی نویسندگان کلیدی، مجلات هسته، موضوعات و مقالات مطرح، مؤسسات مهم، کشورهای پیشرو و زبان غالب نوشته‌های علمی در این حوزه، شناسایی و معرفی شوند، زیرا با شناخت نویسندگان کلیدی و مجلات هسته می‌توان اقدام به برقراری ارتباط علمی با آن‌ها نمود و از نتایج فعالیت‌هایشان بهره برد. همچنین مراکز و کتابخانه‌های تخصصی این رشته خواهند توانست با دیدی آگاهانه‌تر در گزینش مجلات، آثار نویسندگان، و شناسایی موضوعات جدید در سال‌های اخیر، مجموعه جامع‌تری را تهیه کنند و در اختیار کاربران قرار دهند. به علاوه نقشه علمی و تاریخ‌نگاشتی آن به مسئولین در سیاست‌گذاری‌های علمی برای این رشته کمک فراوانی خواهد کرد.

هدف این مقاله بررسی تحلیلی انتشارات بازایی شده در حوزه کشاورزی در طول سال‌های ۲۰۰۰ تا ۲۰۰۹ (یک بازه ده ساله) از پایگاه استنادی «وب‌آوساینس» و ترسیم ساختار این علم با استفاده از نرم‌افزار «هیست‌سایت» به منظور شناسایی مقالات و نویسندگان مهم و اثرگذار در این رشته، بررسی تاریخی آن، و نیز شناسایی برخی از خوشه‌های علمی تشکیل شده در این حوزه در بازه زمانی مورد مطالعه بود. ولی چون انتشارات سال ۲۰۰۹ به طور کامل در این پایگاه وارد نشده، بازه زمانی ۲۰۰۰ تا ۲۰۰۸ در نظر گرفته شد تا نتایج و مقایسه‌ها منطقی و صحیح انجام شوند.

در این مقاله سعی می‌شود به منظور دستیابی به اهداف فوق به سؤالات زیر پاسخ داده شود:

۱. روند رشد تولیدات علمی در رشته کشاورزی چگونه بوده است؟
۲. در ترسیم علم‌نگاشتی تولیدات علمی کشاورزی، خوشه‌های مهم کدام‌اند؟
۳. آثار مهم و اثرگذار این رشته در پایگاه و بازه زمانی مورد بررسی کدام‌اند؟
۴. نویسندگان فعال (پرتولید) در رشته کشاورزی در پایگاه و بازه زمانی مورد بررسی کدام‌اند؟
۵. ضریب همکاری نویسندگان این بخش از علم در پایگاه و بازه زمانی مورد بررسی به چه میزان است؟
۶. دانشگاه‌ها و مؤسسات دارای بیش‌ترین تولیدات علمی در حوزه کشاورزی در پایگاه و بازه زمانی مورد بررسی کدام‌اند؟
۷. کدام کشورها در تولید آثار مورد بررسی، فعالیت بیش‌تری داشته‌اند؟
۸. مجلات مهم و هسته در پایگاه و بازه زمانی مورد بررسی در رشته کشاورزی کدام‌اند؟
۹. فعالیت‌های علمی یادشده در چه قالب‌هایی ارائه شده‌اند؟
۱۰. متون علمی تولیدشده در حیطه کشاورزی به چه زبان‌هایی نوشته شده‌اند؟

۳. پیشینه پژوهش

در اینجا با توجه به هدف اصلی این پژوهش که ترسیم نقشه تاریخ‌نگاشتی تولیدات علمی در حوزه کشاورزی است و نیز با توجه به نتایج بررسی‌ها که نشان داد تاکنون چنین پژوهشی در این رابطه در حوزه کشاورزی صورت نگرفته، تلاش شد مدارکی که تحلیل استنادی را به عنوان روش، در حوزه کشاورزی یا حوزه‌های نزدیک به آن به کار برده‌اند و نیز مدارکی که به ترسیم ساختار علم در یکی از حوزه‌های علمی پرداخته‌اند، مورد بررسی واقع شوند.

«شو کین» و «کوشیک» ۶۱ مقاله یک مجله را مورد تحلیل استنادی قرار دادند و نشان دادند که ۸۱٪ پژوهشگران در تألیف متون خود از نشریات ادواری استفاده می‌کنند و نیم عمر استنادها نیز کم‌تر از ۲۰ سال محاسبه گردید (Shokeen and Kaushik 2004). «عصاره» و «مک کین» برای ترسیم ساختار تحقیقات شیمی ایران (۱۹۹۰ تا ۲۰۰۶) از میان ۷,۶۸۲ مقاله شیمی نمایه‌شده در نمایه استنادی علوم و با استفاده از روش نویسندگان

هم‌استناد^۱ در پایگاه دایالوگ^۲، اطلاعات خود را گردآوری و نسبت به ترسیم ساختار شیمی ایران اقدام نمودند. نتایج حاصل از پژوهش نشان داد که دانشمندان ایرانی و بین‌المللی در ۷ خوشهٔ مختلف در زیرساخت شیمی ایران فعالیت دارند (Osareh and McCain 2008).

«کیم» و «لی» نیز برای ترسیم ساختار مطالعات آرشویی از روش داده‌کاوی استفاده کردند و داده‌های خود را از ۴۳۲ مقاله در سال‌های ۲۰۰۱ تا ۲۰۰۴ جمع‌آوری نمودند. آن‌ها با استفاده از روش تحلیل عاملی، ۴۳ خوشه تشکیل دادند. در نهایت مشخص شد که پژوهشگران ۴۳ خوشه، در ۷ گروه موضوعی به فعالیت پژوهشی اشتغال دارند (Kim and Yan Lee 2008).

در ایران نیز «تصویری قمصری» و «جهان‌نما» پایان‌نامه‌های ارائه‌شده توسط مهندسان پژوهشکدهٔ مهندسی جهاد کشاورزی در فاصلهٔ سال‌های ۱۳۶۶ تا ۱۳۸۳ را مورد تحلیل استادی قرار دادند و به تعیین الگوهای رفتار علمی پژوهشگران در چگونگی استفاده از منابع پرداختند (تصویری قمصری و جهان‌نما ۱۳۸۵).

«کریمیان مزیدی» و «اشرفی» نیز به بررسی استادی سه نشریه- علوم کشاورزی، دانش کشاورزی، و تحقیقات کشاورزی ایران- به منظور تعیین مجلات هسته، میزان استناد به منابع اطلاعاتی، روحیهٔ همکاری نویسندگان، و زبان منابع مورد استفاده پرداختند (کریمیان مزیدی و اشرفی ۱۳۸۶).

«حمیدی»، «اصنافی» و «عصاره» به بررسی تحلیلی و ترسیم ساختار انتشارات علمی تولیدشده در حوزه‌های کتابسنجی، علم‌سنجی، اطلاع‌سنجی و وب‌سنجی در پایگاه «وب‌آوساینس» در سال‌های ۱۹۹۰ تا ۲۰۰۵ پرداختند. آنان ۶۱۸ رکورد بازیابی‌شده را از نظر موضوع، زبان، کشور، نویسنده، سال انتشار، و ... بررسی کردند و سپس برای ترسیم ساختار علم، رکوردها را وارد نرم‌افزار «هیست‌سایت» نمودند و روابط استادی و آثار مهم و مؤثر در این حوزه‌های علمی را مشخص کردند (حمیدی، اصنافی، و عصاره ۱۳۸۷).

از مرور پیشینه‌ها می‌توان چنین نتیجه گرفت که برخی از پیشینه‌ها به تحلیل استادی مدارک حوزهٔ کشاورزی، و برخی به ترسیم نقشهٔ علمی رشته‌های مختلف پرداخته‌اند و

¹ Co-citation

² DIALOG

نتایج قابل توجهی در زمینه الگوهای رفتار علمی پژوهشگران، موضوع، مجلات هسته، روحیه همکاری گروهی، و نویسندگان مؤثر یا هم‌استناد به دست آورده‌اند. به علاوه، ساختار برخی از حوزه‌های مورد بررسی ترسیم شده و نویسندگان مهم و مقالات اثرگذار آن حوزه‌ها معرفی شده‌اند.

۴. روش انجام پژوهش

این پژوهش از نوع توصیفی است و برای انجام آن در مجموع ۲۲,۶۱۷ رکورد کتابشناختی در تاریخ ۳ دی ماه ۱۳۸۷، از پایگاه «وب آوساینس» در زمینه موضوعی کشاورزی (با جستجوی کلیدواژه Agriculture) با اعمال محدودیت زمانی ۲۰۰۰ تا ۲۰۰۸ بارگذاری، و به حالت متن ساده^۱ ذخیره شدند. سپس تمامی این داده‌ها در فایل‌های ۵۰۰ رکوردی به نرم‌افزار «هیست‌سایت» منتقل شدند و تحلیل‌های لازم بر روی آن‌ها با استفاده از این نرم‌افزار انجام گرفت. تحلیل‌های انجام شده در هر قسمت (نویسنده، مجله، کلمات، مؤسسات، و...) به صورت فایل‌های مجزا بر روی رایانه شخصی در صفحه گسترده نرم‌افزار اکسل^۲ ذخیره شدند و نتایج آن‌ها یا به صورت مستقیم یا با انجام پاره‌ای تحلیل‌های دیگر، در قسمت تحلیل داده‌ها گزارش می‌گردند. همچنین بر اساس این داده‌ها، نقشه علم‌نگاشتی کشاورزی نیز توسط این نرم‌افزار ترسیم شد.

محدودیت‌های این پژوهش بیش‌تر مربوط به اطلاعات کتابشناختی انتشارات پایگاه استنادی علوم^۳، و عدم مستندسازی آن‌ها هستند:

◇ استاندارد نبودن برخی روش‌های استناددهی و یکدست نبودن آن‌ها در مقالات «وب آوساینس»؛

◇ ناهماهنگی ناشی از به کارگیری روش‌های مختلف در مختصرنویسی نام مجلات؛

◇ ثبت نام یک نویسنده با صورت‌ها و املاهای متفاوت؛

◇ تغییر نام خانوادگی نویسندگان زن بعد از ازدواج (Garfield, Pudvokin, and Istomin 2003b)

◇ تفاوت در املاهای کلمات؛

¹ Plain Text

² Excel

³ WOS

◇ تأکید «وب آوساینس» بر زبان انگلیسی و توجه کم تر به مقالاتی که به دیگر زبانها منتشر می شوند.

۵. تحلیل دادهها

با جستجوی انجام شده در پایگاه «وب آوساینس»، تعداد ۲۲,۶۱۷ رکورد در زمینه کشاورزی و در محدوده زمانی مورد نظر بازمیابی شدند (جدول شماره ۱).

جدول ۱. میزان انتشارات علمی کشاورزی در نمایه استنادی علوم به تفکیک سال

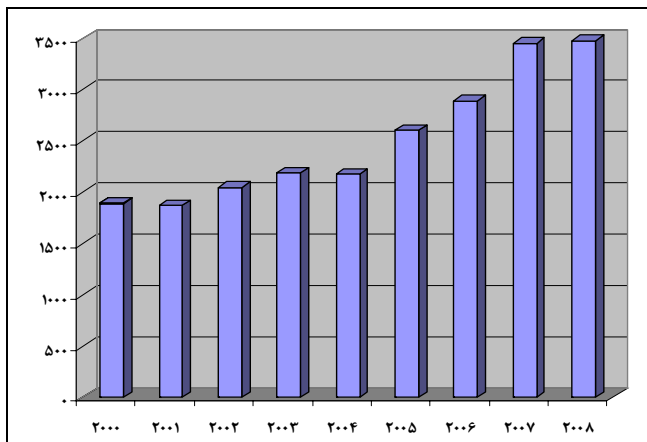
ردیف	سال انتشار	تعداد انتشارات	درصد از کل	«تی ال سی اس»	«تی جی سی اس»
۱	۲۰۰۰	۱۸۹۳	۸/۲	۳۷۶۴	۲۵۴۳۵
۲	۲۰۰۱	۱۸۷۱	۸/۱	۳۶۰۹	۲۱۵۵۰
۳	۲۰۰۲	۲۰۴۸	۸/۹	۳۴۵۷	۲۱۵۸۶
۴	۲۰۰۳	۲۱۸۹	۹/۵	۳۳۵۱	۲۰۹۶۵
۵	۲۰۰۴	۲۱۷۶	۹/۴	۲۴۶۳	۱۶۳۲۳
۶	۲۰۰۵	۲۶۱۰	۱۱/۳	۲۴۴۳	۱۴۳۷۶
۷	۲۰۰۶	۲۸۹۳	۱۲/۵	۱۴۴۹	۹۶۳۸
۸	۲۰۰۷	۳۴۵۵	۱۵	۹۰۹	۴۸۱۱
۹	۲۰۰۸	۳۴۸۱	۱۵/۱	۱۷۸	۹۳۱
۱۰	نامشخص	۱	۰	۰	۰

با مشاهده این جدول مشخص است که بیشترین میزان «تی ال سی اس»^۱ و «تی جی سی اس»^۲ در سال ۲۰۰۰ وجود دارد. پس از این سال، سال ۲۰۰۱ دارای بیشترین «تی ال سی اس» و سال ۲۰۰۲ با ۳۶ اختلاف از سال ۲۰۰۱ دارای بیشترین «تی جی سی اس» هستند. با توجه به دوره باروری استنادی که حداقل ۲ سال است، می توان استدلال کرد که

^۱ (TLCS= Total Local Citation Score) تعداد کل استادهایی که در مجموعه بازمیابی شده، به مقالات بازمیابی شده در هر سال بخصوص، تعلق گرفته است.

^۲ (TGCS= Total Global Citation Score) تعداد کل استادهایی که در پایگاه «وب آوساینس» به مقالات بازمیابی شده در هر سال بخصوص، تعلق گرفته است.

تولیدات علمی مربوط به سال‌های اولیه (۲۰۰۰ تا ۲۰۰۲) فرصت بیش‌تری برای دریافت استناد داشته‌اند و شاید علت تعداد استنادهای دریافتی بیش‌تر آن‌ها، به این دلیل باشد. به منظور محاسبه متوسط نرخ رشد انتشارات حوزه کشاورزی در طی سال‌های مورد بررسی، از میانگین هندسی استفاده شده است. محاسبه مقدار متوسط نرخ رشد سالانه انتشارات برای سال‌های ۲۰۰۰ تا ۲۰۰۸ عدد ۷٪ را نشان می‌دهد، که این نرخ رشد در مقایسه با نرخ رشد متون شیمی در ایران که در پژوهش «عصاره» و «مک‌کین» ۲۶٪ محاسبه شده، قابل تأمل است (Osareh and McCain 2008). به طور کلی روند رشد تولیدات علمی حوزه کشاورزی در تمامی سال‌ها دارای نرخ رشد صعودی بوده است، اگرچه بر طبق نمودار شماره ۱، وقفه‌هایی بسیار مختصر در سال‌های ۲۰۰۱ و ۲۰۰۴ در آن دیده می‌شود.



نمودار ۱. میزان انتشارات در حوزه کشاورزی در هر سال

بررسی نوع مدارک نشان داد که مجموع ۲۲,۶۱۷ رکورد، در ۱۵ نوع قالب مختلف ارائه شده‌اند. همان‌طور که انتظار می‌رفت قالب «مقاله» بیش‌ترین تعداد را به خود اختصاص داده و دارای بالاترین میزان «تی‌ال‌سی‌اس» و «تی‌جی‌سی‌اس» نیز می‌باشند (جدول ۲).

جدول ۲. قالب رکوردهای بازیابی شده

ردیف	نوع مدرک	تعداد رکوردها	«تی ال سی اس»	«تی جی سی اس»
۱	مقاله	۱۶۶۹	۱۶۰۹۷	۹۷۰۸۶
۲	مقاله سمینار	۲۷۷۵	۲۹۳۷	۱۵۹۸۴
۳	مقاله مروری	۱۵۴۲	۲۶۴۳	۲۲۵۷۰
۴	مرور کتاب	۶۹۹	۷	۱۰
۵	منابع ویرایشی	۴۹۳	۲۹۵	۲۱۱۲
۶	چکیده گردهمایی‌ها	۲۰۵	۰	۲
۷	مقولات خبری	۱۴۸	۱۲	۸۹
۸	نامه‌ها	۴۵	۲۳	۶۷
۹	تصحیحات	۱۹	۰	۰
۱۰	مقولات کتابشناختی	۹	۳	۸
۱۱	چاپ مجدد	۹	۹	۳۸
۱۲	کتابشناسی	۱	۰	۲
۱۳	مرور پایگاه‌ها	۱	۰	۰
۱۴	متون داستانی و سرگرم کننده	۱	۰	۰
۱۵	مرور نرم افزار	۱	۰	۰

این رکوردها با این تنوع در نوع مدرک، به ۲۵ زبان زنده دنیا نیز نوشته شده‌اند که به منظور جلوگیری از طولانی شدن جدول، تنها ۵ مورد از آن‌هایی که دارای بیشترین درصد بودند در جدول ۳ نشان داده شده‌اند. بر طبق این جدول، میزان انتشارات و استنادات تعلق گرفته به زبان انگلیسی با اختلاف قابل توجهی بیش تر از دیگر زبان‌ها است و در رتبه ۱ قرار دارد و بیانگر آن است که این زبان، زبان غالب انتشارات علمی در این حوزه است. سه زبان اول در این جدول از نظر میزان استناد از سوی مجموعه بازیابی شده «تی ال سی اس» و کل «وب آوساینس» («تی جی سی اس») نیز به همین ترتیب در بالاترین رتبه‌ها قرار دارند. این تفاوت چشمگیر در تعداد استنادها به مقالات انگلیسی زبان در مقایسه با دیگر زبان‌ها، بار دیگر غالب بودن زبان انگلیسی به عنوان زبان علمی، و نیز در

پایگاه «آی اس آی» را نشان می دهد. شایان ذکر است که هرچند یک مقاله به زبان عربی هم نوشته شده، ولی هیچ رکوردی به زبان فارسی گزارش نشده است.

جدول ۳. پنج زبان برتر در تولید اطلاعات علمی مورد بررسی

ردیف	زبان	تعداد رکوردها	درصد از کل	«تی ال سی اس»	«تی جی سی اس»
۱	انگلیسی	۲۱۷۲۹	۹۴/۶	۲۱۸۸۲	۱۳۷۱۱۵
۲	آلمانی	۳۷۱	۱/۶	۷۱	۲۵۴
۳	فرانسوی	۲۶۲	۱/۱	۲۹	۱۶۰
۴	اسپانیایی	۱۸۰	۰/۰۷	۷	۹۶
۵	پرتغالی	۱۴۵	۰/۰۶	۱۸	۱۰۳

تحلیل داده ها نشان داد که مدارک نمایه شده در حوزه کشاورزی در پایگاه «آی اس آی»، با همکاری ۱۷۳ کشور به رشته تحریر درآمده اند. از فعال ترین کشورها، ۵ تای اول آن ها در جدول ۴ آورده شده اند. همان طور که در جدول ۳ نشان داده شد زبان های اسپانیایی و پرتغالی در زمره زبان هایی بودند که تعداد قابل توجهی از رکوردها به آن زبان ها منتشر شده اند، ولی این کشورها (اسپانیا و پرتغال) با اختلاف چشمگیری از کشور دهم، در مرتبه های ۱۱ و ۱۲ قرار گرفته اند.

جدول ۴. پنج کشور برتر مشارکت کننده در تولید اطلاعات علمی مورد بررسی

ردیف	کشور	تعداد رکوردها	درصد از کل	«تی ال سی اس»	«تی جی سی اس»
۱	آمریکا	۷۹۸۹	۳۴/۷	۹۷۳۲	۶۴۸۸۷
۲	انگلیس	۲۲۷۴	۹/۹	۳۷۵۹	۲۱۲۹۸
۳	کانادا	۱۲۹۳	۵/۶	۱۲۶۶	۸۰۱۴
۴	آلمان	۱۲۴۲	۵/۴	۱۲۱۷	۸۶۷۳
۵	استرالیا	۱۰۸۴	۴/۷	۱۲۲۲	۹۲۱۲

مقاله های کشورهای آمریکا و انگلستان نیز از نظر تعداد کل استادهایی که از سوی مجموعه بازبایی شده («تی ال سی اس») و از سوی کل پایگاه «وب آوساینس»

«تی‌جی‌سی‌اس») گرفته‌اند، همچنان در رتبه‌های ۱ و ۲ قرار دارند. با توجه به این که در قسمت قبلی دیده شد که زبان انگلیسی، زبان غالب در متون علمی است و این زبان، زبان رسمی در دو کشور فوق نیز می‌باشد، شاید این نتیجه چندان دور از انتظار نباشد: ولی نسبت ۳/۵ برابر انتشارات آمریکا به انگلستان قابل تأمل است. در میان این ۱۷۳ کشور، ایران در رتبه ۴۰ با ۹۸ مقاله («تی‌ال‌سی‌اس» برابر با ۲۲ و «تی‌جی‌سی‌اس» برابر با ۲۱۸) قرار گرفته است. این تعداد مقاله، با توجه به سابقه‌ای که کشاورزی در ایران دارد قابل قبول به نظر نمی‌رسد و تلاش بیش‌تر پژوهشگران این عرصه را می‌طلبد.

تحلیل بیش‌تر داده‌ها نشان داد که حدود ۲۰٪ کشورهای مورد بررسی (۳۴ کشور)، حدود ۸۹٪ تولیدات علمی را به عهده داشته‌اند. از میان این ۳۴ کشور، ۸ کشور جهان سوم (چین، هند، ترکیه، آفریقای جنوبی، نیجریه، جمهوری چک، کنیا، و تایلند) دیده می‌شوند که سهمی کم‌تر از ۲۵٪ داشته‌اند.

۱-۵. مجلاتی که مقالات مورد بررسی را به چاپ رسانده‌اند

یکی از مهم‌ترین کانال‌های رسمی تبادل اطلاعات در رشته‌های علمی، مجلات علمی هر رشته هستند. بنابراین شناسایی مجلات مهم و معتبر هر زمینه موضوعی از اهمیت خاصی برخوردار است. تمامی مقالات کشاورزی بازاریابی شده از پایگاه «وب‌آوساینس» در ۳۲۵۶ مجله منتشر شده‌اند. عنوان پنج مجله‌ای که به ترتیب بیش‌ترین تعداد مقالات را منتشر کرده‌اند، در جدول ۵ به نمایش گذاشته شده‌اند.

جدول ۵. پنج عنوان مجله برتر و دارای بیش‌ترین تولیدات علمی

ردیف	مجله	تعداد مقاله‌ها	«تی‌جی‌سی‌اس» «تی‌ال‌سی‌اس»	
۱	AGRICULTURE ECOSYSTEMS & ENVIRONMENT	۳۸۹	۱۳۰۱	۴۱۴۷
۲	WATER SCIENCE AND TECHNOLOGY	۱۷۶	۸۸	۵۷۱
۳	AGRICULTURAL ECONOMICS	۱۶۸	۲۱۱	۵۱۴
۴	AGRICULTURAL SYSTEMS	۱۶۶	۴۸۷	۱۳۱۵
۵	JOURNAL OF SUSTAINABLE AGRICULTURE	۱۶۵	۱۳۰	۲۲۶

رشد روزافزون انتشارات در سطح جهان، افزایش سالانه قیمت مجلات و کاهش بودجه باعث شده که کتابداران بخش مجموعه‌سازی با وسواس و دقت بیش‌تری اقدام به سفارشات، بخصوص در مورد نشریات کنند؛ زیرا نشریات از یک سو دارای روزآمدترین اطلاعات هستند و از سوی دیگر سفارش آن‌ها بودجه قابل توجهی را به خود اختصاص می‌دهد و انتخاب نامناسب هر عنوان می‌تواند به منزله هدر دادن بودجه حداقل یک سال اشتراک نشریه باشد. بنابراین کتابداران از قواعد و روش‌های مختلفی استفاده می‌کنند تا بتوانند با انتخاب تعداد کم‌تر نشریات و صرف هزینه کم‌تر، حجم بیش‌تری از اطلاعات و دانش را در اختیار کاربران قرار دهند.

یکی از این قواعد، قاعده برادفورد^۱ است که در سال ۱۹۳۴ منتشر شد و مبین آن است که تعداد کمی از مجلات، درصد بالایی از کل مقالات هم‌موضوع را منتشر می‌کنند. به صورت دقیق‌تر می‌توان گفت که طبق این قاعده، اگر مجلات علمی را بر اساس بازدهی نزولی مقالات آن‌ها در ارتباط با موضوع خاصی مرتب کنیم، می‌توانیم آن مجلات را به دو گروه تقسیم نماییم:

۱. یک گروه مجلات هسته مرکزی که بیش‌ترین مقالات آن‌ها به آن موضوع اختصاص دارد؛

۲. چندین گروه که در اطراف هسته مرکزی قرار می‌گیرند و در این گروه‌ها نیز تعداد مقالات برابر با هسته مرکزی است، ولی در هر مدار، تعداد نشریات از عدد ثابتی تبعیت می‌کند (میرشمسی ۱۳۶۹، ۳۰۴-۳۰۵).

بنابراین برای تعیین مجلات هسته کشاورزی، ابتدا سیاهه‌ای از عناوین مجلات بر اساس میزان استناد به آن‌ها (یک بار بر اساس «تی‌ال‌سی‌اس» و بار دیگر بر مبنای «تی‌جی‌سی‌اس») و به ترتیب نزولی تهیه گردید. در کل ۲۲,۰۲۶ استناد در خود مجموعه بازایی شده («تی‌ال‌سی‌اس») به ۳,۲۵۶ مجله بازایی شده وجود داشت. محاسبات انجام‌شده بر اساس قاعده برادفورد برای تعیین مجلات هسته بر اساس «تی‌ال‌سی‌اس»، تعداد ۵ عنوان را نشان داد و با انتخاب آن‌ها از ابتدای سیاهه، مجلات معرفی شده در جدول ۶ به‌عنوان مجلات هسته شناسایی شدند.

^۱ Bradford

جدول ۶. مجلات هسته رشته کشاورزی بر اساس «تی‌ال‌سی‌اس»

رتبه	مجله	تعداد مقاله‌ها	«تی‌ال‌سی‌اس»
۱	AGRICULTURE ECOSYSTEMS & ENVIRONMENT	۳۸۹	۱۳۰۱
۴	AGRICULTURAL SYSTEMS	۱۶۶	۴۸۷
۶۲	SCIENCE	۶۶	۴۳۱
۴۰	JOURNAL OF APPLIED ECOLOGY	۷۸	۴۱۴
۱۷	JOURNAL OF RURAL STUDIES	۱۰۰	۳۸۹

برای استنادهایی نیز که این مجلات از سوی کل مقالات «وب‌آوساینس» گرفته‌اند قاعده برادفورد به صورت جداگانه محاسبه شد و بر اساس این محاسبات، ۱۲ مجله هسته بر اساس استنادهای جهانی شناخته شدند. مقالاتی که در جدول شماره ۷ در ردیف‌های خاکستری رنگ قرار گرفته‌اند، در هر دو قسمت «تی‌ال‌سی‌اس» و «تی‌جی‌سی‌اس» مشترک بوده‌اند.

جدول ۷. مجلات هسته رشته کشاورزی بر اساس «تی‌جی‌سی‌اس»

رتبه	مجله	تعداد مقاله‌ها	«تی‌جی‌سی‌اس»
۱	AGRICULTURE ECOSYSTEMS & ENVIRONMENT	۳۸۹	۴۱۴۷
۳۳	ECOLOGICAL APPLICATIONS	۸۳	۲۵۳۴
۲۳	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	۹۶	۲۴۸۸
۴۰	JOURNAL OF APPLIED ECOLOGY	۷۸	۱۹۴۷
۱۱۹	NATURE	۴۰	۱۹۳۵
۶۲	SCIENCE	۶۶	۱۸۷۱
۴	AGRICULTURAL SYSTEMS	۱۶۶	۱۳۱۵
۱۲	SCIENCE OF THE TOTAL ENVIRONMENT	۱۲۱	۱۲۷۹
۲۴	JOURNAL OF ENVIRONMENTAL QUALITY	۹۳	۱۱۷۱
۱۶۲	AMERICAN JOURNAL OF CLINICAL NUTRITION	۳۰	۱۱۳۳
۷۷	CONSERVATION BIOLOGY	۵۶	۱۱۰۹
۲۸	BIOLOGICAL CONSERVATION	۸۶	۱۰۸۶

۲-۵. فراوانی کلمات در مدارک مورد مطالعه

نرم افزار «هیست سایت» قادر است کلمات عنوان مقالات را نیز بر اساس تعداد رخدادها با استفاده از سیاهه^۱ کلمات بازدارنده^۱ مشخص کند و آن‌ها را نمایش دهد. بررسی داده‌های این قسمت نشان می‌دهد که چه جنبه‌هایی از یک موضوع، بیش‌تر اهمیت دارند یا به تازگی مطرح شده‌اند. جدول شماره ۸، ده کلمه پربسامد از کل ۲,۵۳۸ کلمه در عنوان مقالات بازیابی شده را نشان می‌دهد. ستون «تی‌ال‌سی‌اس» بیانگر تعداد کل استنادهایی است که در خود مجموعه بازیابی شده، به عناوین مقالات بازیابی شده‌ای که شامل این کلمات هستند، تعلق گرفته و «تی‌جی‌سی‌اس» نشان می‌دهد که مقالات بازیابی شده‌ای که در عنوان آن‌ها این کلمات وجود داشته، چند بار از سوی دیگر مقالات در پایگاه «وب‌آوساینس» مورد استناد قرار گرفته‌اند.

جدول ۸. کلمات پراستفاده در عنوان مقالات

ردیف	کلمه	تعداد مقاله‌ها	«تی‌ال‌سی‌اس»	«تی‌جی‌سی‌اس»
۱	AGRICULTURE	۴۰۱۴	۵۲۳۵	۱۵۰۴۵
۲	AGRICULTURAL	۱۷۴۲	۲۷۶۱	۱۱۴۱۲
۳	SOIL	۱۴۱۸	۱۶۹۴	۹۳۴۱
۴	WATER	۱۱۷۴	۱۰۸۷	۶۱۹۳
۵	USE	۱۱۰۸	۱۳۲۹	۷۱۷۸
۶	MANAGEMENT	۱۱۰۷	۱۴۰۸	۶۵۸۶
۷	LAND	۹۳۷	۱۱۹۱	۶۰۳۱
۸	PRODUCTION	۸۹۵	۱۲۱۱	۵۶۳۰
۹	ANALYSIS	۸۶۸	۸۰۹	۵۰۷۲
۱۰	EFFECTS	۸۲۵	۷۳۰	۴۹۲۷

همان‌طور که انتظار می‌رود کلمات کشاورزی (در حالت اسم و صفت) بیش‌ترین بسامد را در مقالات دارند، ولی بسامد کلمات «خاک»، «آب»، «استفاده»، «مدیریت»، و ...

^۱ Stop List

نشان‌دهنده موضوعات مهمی است که در حال حاضر در این حوزه مورد توجه بوده و بالطبع بیش‌ترین حجم مقالات را به خود اختصاص داده‌اند. شایان ذکر است که موارد ۱، ۲، ۳، ۵ و ۶ در هر دو شاخص «تی‌ال‌سی‌اس» و «تی‌جی‌سی‌اس» نیز دارای بیش‌ترین بسامد هستند و این نکته تأیید‌کننده اهمیت این کلمات در این حوزه است، زیرا این شاخص‌ها بیانگر آن‌اند که مقالاتی که دارای این واژگان در عنوان خود بوده‌اند چه در خود مجموعه و چه در پایگاه «وب‌آوساینس»، بیش‌تر مورد استناد و استفاده قرار گرفته‌اند.

۳-۵. مؤسسات پر تولید در حوزه اطلاعات کشاورزی

مجموعه مقالات بازایی شده، از سوی ۱۴۸۵۲ مؤسسه مختلف ارسال شده‌اند که ۵ مورد فعال‌ترین این مؤسسات در جدول ۹ آورده شده‌اند. البته برای تعداد ۷۹۸ مقاله، مؤسسه‌ای در قسمت آدرس مقاله نوشته نشده، ولی پس از آن، مؤسسه‌ای از ایالات متحده با داشتن ۴۹۱ مقاله ثبت شده در پایگاه استنادی «آی‌اس‌آی»، فعال‌ترین سازمان در این حوزه به‌شمار می‌آید. استفاده از تجربیات و راهنمایی‌های چنین سازمان‌هایی، می‌تواند حوزه کشاورزی را در هر منطقه‌ای رشد و ارتقا دهد.

جدول ۹. مؤسسات فعال در زمینه تولید مقالات کشاورزی

ردیف	مؤسسه	تعداد مقاله‌ها	«تی‌ال‌سی‌اس»	«تی‌جی‌سی‌اس»
۱	United States Department of Agriculture, Agricultural Research Service	۴۹۱	۷۶۱	۴۶۳۳
۲	University of California. Davis	۲۹۲	۶۲۱	۳۱۷۱
۳	United States Department of Agriculture	۲۸۰	۲۸۵	۱۸۳۱
۴	Chinese Academy of Agriculture Science	۲۷۹	۳۱۹	۲۰۳۲
۵	Wageningen University and Research Centre	۲۲۹	۵۸۰	۲۳۳۳

مؤسسات ردیف ۱، ۲ و ۵ از نظر میزان کل تعداد استنادهای تعلق گرفته در مجموعه به مقالاتی که از سوی این مؤسسات نوشته شده‌اند («تی‌ال‌سی‌اس») در بالاترین رده‌ها قرار داشتند. مؤسسات ردیف ۱ و ۲ از نظر کل تعداد استنادها در پایگاه «وب‌آوساینس»

«تی جی سی اس») نیز با تفاوت قابل توجهی از دیگر مؤسسات، در بالاترین مرتبه قرار گرفته‌اند. این امر به اهمیت و اعتبار این دو مؤسسه در حوزه موضوعی کشاورزی اشاره دارد.

۵-۴. پراشتارترین نویسندگان

تعداد ۵۱,۶۵۵ نویسنده از سال ۲۰۰۰ تا ۲۰۰۸ مقالات خود را در حوزه کشاورزی به رشته تحریر درآورده و در پایگاه «آی اس آی» به ثبت رسانده‌اند. از این میان، ۵ نویسنده که دارای بیشترین تعداد مقالات بوده‌اند در جدول ۱۰ معرفی شده‌اند و ملیت آنان نیز مشخص شده است.

جدول ۱۰. پراشتارترین نویسندگان در حوزه کشاورزی

ردیف	نویسندگان با کشور و محل اشتغال	تعداد مقاله‌ها	«تی ال سی اس»	«تی جی سی اس»
۱	Lal R, Ohio State Univ, USA	۳۰	۷۰	۲۷۰
۲	Schnug E, Fed Agr Res Ctr, Germany	۲۴	۱۶	۴۷
۳	Sudduth KA, Univ missouri, USA	۲۴	۱۴۴	۲۶۸
۴	McCarl BA, Texas A & M Univ, USA	۲۲	۸۷	۲۲۰
۵	Rozelle S, Univ Calif Davis, USA	۲۲	۴۲	۹۵

همان‌طور که مشاهده می‌شود غیر از «سودوث»^۱ که جزو ۵ نویسنده برتر نیز هست، دیگر نویسندگان دارای مقالاتی اندک ولی بسیار اثرگذار بوده‌اند، زیرا استادهای زیادی را به خود اختصاص داده‌اند.

میزان همکاری گروهی نویسندگان مقالات نیز مورد بررسی قرار گرفت. در ابتدا مقالات بر اساس تعداد نویسندگانشان رتبه‌بندی شدند و ضریب همکاری گروهی نویسندگان^۲ برابر با ۰/۵ محاسبه گردید که ضریب خوبی بود و نشان می‌دهد که همکاری در نویسندگی از سطح قابل قبولی برخوردار است.

^۱ Sudduth

^۲ Collaboration Coefficient (CC)

۵-۵. منابع مورد استناد^۱

نرم افزار «هیستسایت» علاوه بر تحلیل های بالا، قابلیت دیگری نیز دارد که بسیار مهم است. این نرم افزار قادر است منابع مورد استناد در مقالات بازیابی شده را بازیابی کند و نشان دهد که کدامیک از آن ها در مجموعه بازیابی شده، وجود دارد و کدامیک در جستجوی انجام شده، بازیابی نشده یا اصلاً در پایگاه «وب آوساینس» حضور ندارد. این قابلیت مدارکی را مشخص می کند که با حوزه بازیابی شده در ارتباط اند، ولی در مجموعه بازیابی شده و در پایگاه «وب آوساینس» قرار ندارند. با استفاده از این امکان می توان به دریای عظیمی از مدارک مرتبط خارج از پایگاه «وب آوساینس» نیز دست یافت. این مدارک احتمالاً از نوع مدارک زیر هستند:

◇ مقالات قدیمی و چاپی که تحت پوشش پایگاه «وب آوساینس» قرار نمی گیرند؛
◇ مدارکی همچون کتاب، دستنامه ها، و ... که در پایگاه استنادی «آی اس آی» نمایه نمی شوند؛

◇ مقالات مجلاتی که در «آی اس آی» حضور ندارند (Garfield, Paris, and Stock 2006, 395).

در پژوهش حاضر نیز تلاش شده است با استفاده از این قابلیت، مدارک مهمی که مورد استفاده نویسندگان مقالات بازیابی شده قرار گرفته اند شناسایی شوند. در مقالات بازیابی شده در حوزه کشاورزی، تعداد ۶۵۸,۳۸۱ منبع مورد استناد قرار گرفته اند که از میان آن ها، ۱۰ مقاله که بیش ترین استناد به آن ها تعلق گرفته است به ترتیب در جدول ۱۱ آورده شده اند و همان طور که مشاهده می شود تنها ۳ مورد از آن ها (ردیف های ۶، ۹، و ۱۰) در مجموعه بازیابی شده حضور داشته اند. این مقالات که ۷ مورد آن ها جزو مقالات پایگاه «وب آوساینس» نمی باشند، با میزان استنادشان بیانگر بیش ترین میزان ارتباط با حوزه بازیابی شده می باشند و می توان آن ها را در زمره مقالات اصلی در حوزه کشاورزی به حساب آورد. با توجه به سال نشر این مقالات که همگی قبل از سال ۲۰۰۲ به چاپ رسیده اند به نظر می رسد که این مقالات در زمره مقالات پایه در کشاورزی هستند و به مباحث مهمی در مقالات خود پرداخته اند. بررسی موضوعی این مقالات از حوصله این مقاله خارج است و جا دارد که در مقاله جداگانه ای به این مسئله پرداخته شود.

^۱ Cited References (CR)

جدول ۱۱. مهم ترین مدارک مورد استناد

ردیف	نویسنده، سال انتشار، مشخصات مجله	تعداد مقاله های استناد دهنده
۱	BOSERUP E, 1965, CONDITIONS AGR GROWT	۱۳۲
۲	MURPHY J, 1962, ANAL CHIM ACTA, V27, P31	۱۲۰
۳	CARPENTER SR, 1998, ECOL APPL, V8, P559	۱۱۸
۴	VITOUSEK PM, 1997, SCIENCE, V277, P494	۱۱۳
۵	MATSON PA, 1997, SCIENCE, V277, P504	۱۰۷
۶	TILMAN D, 2001, SCIENCE, V292, P281	۹۵
۷	VITOUSEK PM, 1997, ECOL APPL, V7, P737	۹۳
۸	MYERS N, 2000, NATURE, V403, P853	۸۸
۹	ROBERTSON GP, 2000, SCIENCE, V289, P1922	۸۳
۱۰	TILMAN D, 2002, NATURE, V418, P671	۸۱

۵-۶. ترسیم نقشه تاریخی نگاشتی تولیدات علمی حوزه کشاورزی

ترسیم ساختار هر شاخه ای از علم، مبتنی بر نویسندگان و آثار مؤثر آن حوزه است و با توجه به اهمیت آن، امروزه نرم افزارهای گوناگونی همچون «اس پی اس اس»، «پث فایندر»^۱، و «هیست سایت» برای انجام این امر در دسترس هستند. هر یک از این نرم افزارها در ترسیم ساختار علم از جنبه ای خاص به مسئله می پردازند. مثلاً «هیست سایت» سیر تحول تاریخی یک حوزه علمی را نشان می دهد؛ زیرا این نرم افزار قابلیت رسم نقشه را به ترتیب سال نشر آثار (ترتیب تاریخی) دارد. در این نوع ترسیم ها دو گروه از آثار وجود دارند: آثاری که به لحاظ اهمیتشان، بسیار مورد استناد قرار گرفته اند (دایره های بزرگ در نمودارهای ۲ و ۳) و آثاری که به منابع دیگر استناد داده اند (که در نمودارهای ۲ و ۳ با خطوط جهت دار و پیکان، این روابط را نشان داده اند).

از میان ۲۲,۶۱۷ رکورد بازایی شده در بخش کشاورزی، تعداد ۱۵۰ مدرک (۰/۶ درصد) بر مبنای شاخص امتیاز استناد محلی («ال سی اس») در ترسیم ساختار علم کشاورزی لحاظ شده اند. نمودار شماره ۲ نمایانگر خوشه های اصلی این ساختار است.

^۱ PathFinder

شایان ذکر است که به علت طولانی بودن تصویر، ساختار در دو قسمت مجزا که دربرگیرنده تعداد ۵ خوشه بودند (با استفاده از خط چین، خوشه‌ها مشخص شده‌اند)، نمایش داده شده. هر خوشه متشکل از تعدادی مدرک است. در هر خوشه، دایره‌ای که با پیکان پررنگ مشخص شده، به معنای آن است که این مدرک بیش‌ترین میزان استناد را به خود اختصاص داده.

در خوشه اول که در سال‌های ۲۰۰۰ تا ۲۰۰۵ شکل گرفته، «رابرتسون»^۱ (در سال ۲۰۰۰) مدرک شماره ۱۷۸۸ را که بیش‌ترین استناد را به خود اختصاص داده در رابطه با گازهای گلخانه‌ای نوشته و با گروهی از نوشته‌ها که همگی از دایره‌های بزرگی تشکیل شده‌اند (مقالات پراستناد)، پیوند دارد (جدول شماره ۱۲).

در همین خوشه مدرک شماره ۱۲۹۷ متعلق به «اسمیت»^۲ (سال ۲۰۰۰) به بیان «الزام در تغییرات آب‌وهوایی اروپا به علت کاهش کمی کربن در خاک آنجا برای کشاورزی» پرداخته و همان‌طور که مشاهده می‌شود تعداد قابل توجهی از مدارک با آن در ارتباط‌اند که همگی بر مسئله «تجزیه و آزادسازی کربن» بخصوص در اروپا تأکید دارند. مدرک قابل توجه دیگر در این خوشه، مدرک شماره ۶۱۴۶ از «کاسمن»^۳ (سال ۲۰۰۳) است که با پیوندهای مختلف خود، ارتباط قوی بین خوشه یک و دو برقرار کرده (جدول شماره ۱۲).

بررسی‌ها نشان داد که مدرک اخیر در باره «حفاظت از محیط زیست و رشد کیفیت محیطی» است. بنابراین با توجه به عام‌بودن نسبی موضوع، ارتباط آن با دو مقاله «تیلمن»^۴ (سال ۲۰۰۱ و ۲۰۰۲)، یعنی پراستنادترین مدارک خوشه ۲، (مدرک شماره ۳۰۳۲ با ۹۵ استناد و مدرک شماره ۵۴۶۱ با ۸۱ استناد) که به ترتیب در رابطه با «پیش‌بینی تغییرات محیطی آینده در کشاورزی» و «مباحث عملی در مورد بهره‌وری» بحث کرده‌اند، و مدرک شماره ۲۵۲۱ از «هانسن»^۵ در رابطه با «تأثیرات محیطی کشاورزی ارگانیک» و دیگر مدارک، قابل توجه می‌باشد. در خوشه دوم علاوه بر دو مقاله «تیلمن»، مقاله دیگری در سال ۲۰۰۱ توسط «دونالد»^۶ (شماره ۲۶۴۷) در مورد «تأثیر کشاورزی بر از بین رفتن گونه‌ای از پرندگان» به چاپ رسیده که پیوندهای زیادی با آن برقرار شده است. در سال

¹ Robertson

² Smith

³ Cassman

⁴ Tilman

⁵ Hansen

⁶ Donald

۲۰۰۵ «شارنتکه»^۱ (مدرک شماره ۱۲، ۲۲۳ در جدول ۱۲) به بررسی «اکوسیستم‌های متنوع بیولوژیکی» پرداخته و بسیاری از مدارک این خوشه را به یکدیگر پیوند داده است. در همین سال مقاله‌ای توسط «پرتی»^۲ (شماره ۱۱، ۱۲۲) به چاپ رسیده که با استنادهایش دو خوشه ۲ و ۳ را که هر دو در سال‌های ۲۰۰۰ تا ۲۰۰۵ شکل گرفته‌اند به یکدیگر پیوند داده است. این مقاله که در مورد «هزینه‌های مواد غذایی برای خانواده‌های انگلیسی با بررسی مسافت طولانی مسیر تولید و عرضه» است، با مدرک شماره ۱۵۱۹ در سال ۲۰۰۰ از خوشه ۲ که نوشته‌شده همین نویسنده و در مورد «ارزیابی هزینه‌های خارجی کشاورزی در انگلستان» است، و مدرک شماره ۶۵۳۸ (اثر «وینتر»^۳) در باب «محلی‌گرایی و اقتصاد جدید غذایی» پیوند دارد. «هینریچز»^۴ (مدرک شماره ۱۴۷۷) نیز که نویسنده‌ی پر استنادترین مدرک خوشه ۳ است، در رابطه با «بازارهای محلی غذایی» صحبت کرده. پس از این خوشه، خوشه کوچک دیگری در فاصله سال‌های ۲۰۰۱ تا ۲۰۰۴ دیده می‌شود که خوشه‌ای خود-استنادی است. این خوشه متشکل از چهار مقاله است که سه مقاله آن مربوط به «سودوث» است. مدرک شماره ۳۰۷۹ از «سودوث» (۲۰۰۱) پر استنادترین مدرک این خوشه است و به همراه مدارک شماره ۲۶۴۱ و ۷۰۴۰ (مربوط به سال‌های ۲۰۰۱ و ۲۰۰۳) که از همین نویسنده می‌باشند، در مورد «رسانایی الکترونیکی خاک» هستند. ضمناً این نویسنده، سومین نویسنده پراتشار و مؤثر در حوزه کشاورزی است. مدرک شماره ۹۵۹۸ (اثر «آدام شوک»^۵) نیز در این خوشه در مورد «سنسورهای خاکی» است. پس از این خوشه، مدرک شماره ۲۴۷۹ (اثر «ویلسون»^۶) با ۶۵ استناد و پرداختن به موضوع «بهره‌وری بیشتر در تولیدات»، با دیدی انتقادی به خاطر تغییراتی که در طبیعت در حال رخ دادن است، با چند مقاله دیگر (مثلاً مدرک شماره ۵۱۴۰ از «اونز»^۷) یک خوشه را در سال‌های ۲۰۰۱ تا ۲۰۰۴ تشکیل داده‌اند. شاید بتوان در یک گروه‌بندی کلی گفت که خوشه یکم به مسائل «حفاظت از محیط زیست و رشد کیفیت محیط»، خوشه دوم به بررسی «تغییرات و تأثیرات محیطی بر کشاورزی و بالعکس»، خوشه سوم به موضوع «اقتصاد کشاورزی و بازارهای محلی

¹ Tscharnkte

² Pretty

³ Winter

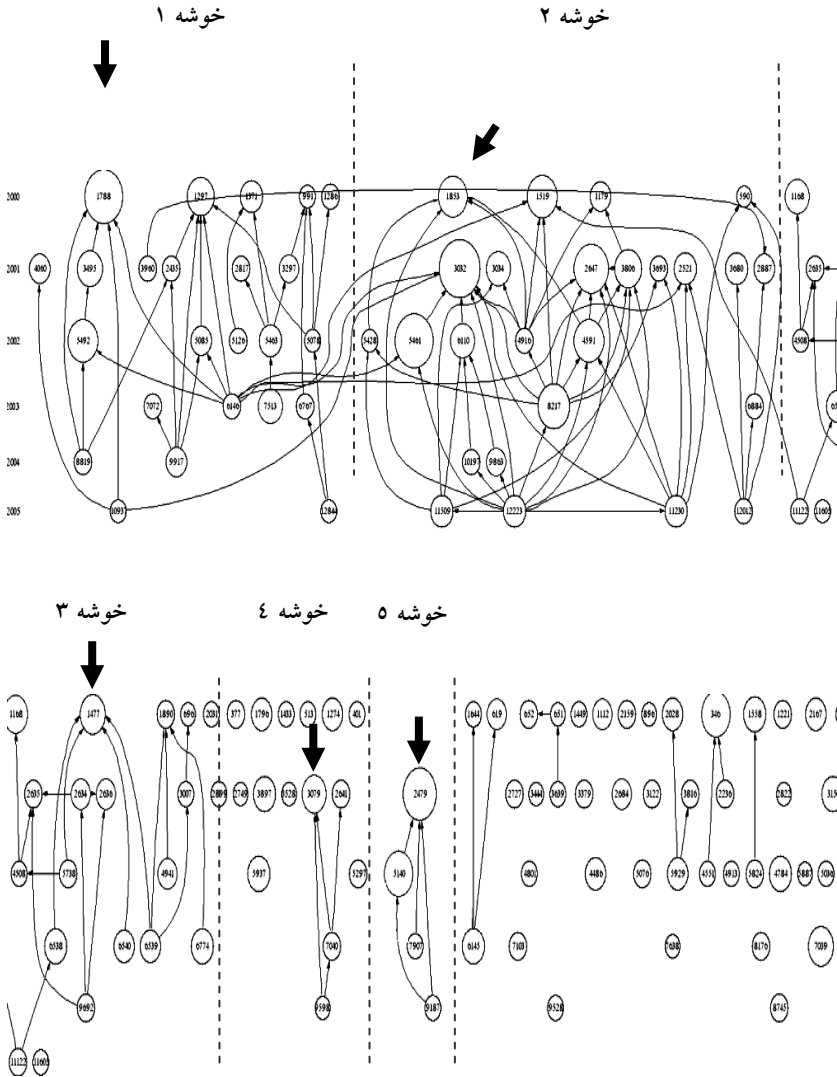
⁴ Hinrichs

⁵ Adamchuk

⁶ Wilson

⁷ Evans

غذایی»، خوشه چهارم به مسائل «رسانایی الکترونیکی خاک» و خوشه پنجم به «بررسی و نقد بهره‌وری بیش‌تر در تولیدات کشاورزی» پرداخته‌اند (نمودار شماره ۲).



نمودار ۲. ترسیم نقشه علم‌نگاشتی ۱۵۰ مدرک برتر در حوزه کشاورزی بر اساس شاخص

«ال سی اس»

در یک جمع‌بندی کلی می‌توان استدلال کرد که بر اساس الگوریتم «هیست‌سایت»، ۱۵۰ مقاله برتر به لحاظ استنادهای «ال‌سی‌اس» در میان ۲۲،۶۱۷ مدرک که در طول سال‌های ۲۰۰۰ تا ۲۰۰۸ در رشته کشاورزی و در نمایه استنادی علوم، نمایه شده‌اند، مقالات منعکس در نقشه علم‌نگاشتی نمودار شماره ۲ هستند. همچنین می‌توان نتیجه گرفت که موضوعات شکل گرفته در بازه زمانی مورد مطالعه، مهم‌ترین موضوعاتی بوده‌اند که در سطح جهان، نظر دانشمندان این حوزه را به خود معطوف داشته‌اند و هر ۱۵۰ مقاله و نویسندگان آن را می‌توان به عنوان مقالات اثرگذار و نویسندگان کلیدی این حوزه در پایگاه و بازه زمانی مورد بررسی معرفی نمود. اطلاعات کتابشناختی تمامی مدارک معرفی شده در این قسمت، همراه با شماره مدرک آن‌ها در جدول شماره ۱۲ آورده شده است.

به منظور ترسیم نقشه علم‌نگاشتی بر اساس استنادهای «جی‌سی‌اس»، به علت کمبود تعداد پیوندها، نمونه به ۳۰۰ مدرک افزایش داده شد تا تشکیل خوشه‌ها امکان‌پذیرتر شود. در بررسی موضوعات مقالات خوشه اول که در سال‌های ۲۰۰۰ تا ۲۰۰۵ شکل گرفته است، مشخص شد که اکثر مدارک مهم و پر استناد این خوشه (مدارک شماره ۳۸۰۶، ۱۱۲۳۰، ۱۸۵۳، ۱۲۲۲۳، ۴۵۹۱) در رابطه با «تغییرات زیست‌محیطی و تنوع بیولوژیکی در اکوسیستم» هستند (جدول شماره ۱۳). اثرگذارترین مقاله این خوشه (شماره ۳۰۳۲) همان مقاله «تیلمن» است که همان‌طور که در نقشه علمی بر اساس «ال‌سی‌اس» نیز ذکر شد، به موضوع «پیش‌بینی تغییرات محیطی جهانی» پرداخته است. حضور این مقاله در هر دو نقشه علم‌نگاشتی در این پژوهش، نشان‌دهنده تأثیر آن در این حوزه است. خوشه دوم که در سال‌های ۲۰۰۰ تا ۲۰۰۶ شکل گرفته به واسطه استناد «کلارک»^۱ (مدرک شماره ۳۴۲۹) به همین مقاله از «تیلمن»، به خوشه ۱ مرتبط شده است. مقاله «کلارک» در مورد «پیش‌بینی‌های اکولوژیکی» است و وجه مشترک آن با مقاله «تیلمن» در مورد «بررسی تغییرات در آینده محیط زیست» است. «رابرتسون» (مدرک شماره ۱۷۸۸) نیز به بررسی «گازهای گلخانه‌ای و اثرات آن‌ها بر اتمسفر» پرداخته است.

پر استنادترین مدرک خوشه بعد (شکل‌گیری در سال‌های ۲۰۰۰ تا ۲۰۰۳)، مدرک شماره ۱۱۱۲ و متعلق به «هنسن»^۲ در سال ۲۰۰۰ که به «پوشش گیاهی جهانی» پرداخته، و

¹ Clark

² Hansen

مدرک شماره ۱۶۴۴ از «مرتز»^۱ در همین سال و دارای دو پیوند در مورد «تغییرات پوشش گیاهی» می‌باشند. بررسی دیگر مدارک این خوشه نیز موضوعات مشابهی را نشان می‌دهد.

جدول ۱۲. اطلاعات مربوط به مدارک مهم خوشه‌ها بر اساس «ال سی اس»

شماره مدرک	نویسنده، سال انتشار و مشخصات مجله	«ال سی اس» «ال سی اس»	
۱۲۹۷	Smith P, 2000, GLOBAL CHANGE BIOLOGY, V6, P525	۴۲	۱۰۶
۱۴۷۷	Hinrichs CG, 2000, JOURNAL OF RURAL STUDIES, V16, P295	۳۷	۷۶
۱۵۱۹	Pretty JN, 2000, AGRICULTURAL SYSTEMS , V65, P113	۵۱	۱۰۰
۱۷۸۸	Robertson GP, 2000, SCIENCE, V289, P1922	۸۳	۲۳۶
۲۴۷۹	Wilson GA, 2001, TRANSACTIONS OF THE INSTITUTE OF BRITISH GEOGRAPHERS, V26, P77	۳۳	۱۱۳
۲۵۲۱	Hansen B, 2001, AGRICULTURE ECOSYSTEMS & ENVIRONMENT, V83, P11	۳۰	۵۶
۲۶۴۱	Fraisse CW, 2001, TRANSACTIONS OF THE ASAE, V44, P155	۲۰	۳۵
۲۶۴۷	Donald PF, 2001, PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES B-BIOLOGICAL SCIENCES, V268, P25	۱۶	۴۱
۳۰۳۲	Tilman D, 2001, SCIENCE, V292, P281	۹۵	۳۳۹
۳۰۷۹	Sudduth KA, 2001, COMPUT ELECTRON AGRIC, V31, P239	۳۶	۶۱
۵۱۴۰	Evans N, 2002, PROGRESS IN HUMAN GEOGRAPHY, V26, P313	۴۷	۶۲
۵۴۶۱	Tilman D, 2002, NATURE, V418, P671	۸۱	۲۳۸
۶۱۴۶	Cassman KG, 2003, ANNU REV ENVIRON RESOUR, V28, P315	۱۸	۴۰
۶۵۳۸	Winter M, 2003, JOURNAL OF RURAL STUDIES , V19, P23	۳۰	۷۳
۷۰۴۰	Sudduth KA, 2003, AGRONOMY JOURNAL , V95, P472	۲۰	۳۷
۹۵۹۸	Adamchuk VI, 2004, COMPUT ELECTRON AGRIC, V44, P71	۱۵	۳۹
۱۱۱۲۲	Pretty JN, 2005, FOOD POLICY, V30, P1	۲۰	۳۶
۱۲۲۲۳	Tscharntke T, 2005, ECOLOGY LETTERS, V8, P857	۳۰	۱۲۳

¹ Mertens

پراستنادترین مدرک خوشه چهارم که به طور کامل از خوشه قبل جدا بوده و در طی سال‌های ۲۰۰۰ تا ۲۰۰۵ تشکیل شده، متعلق به «کاوون»^۱ (شماره ۹۴۷) در رابطه با «تحلیل عملکرد گیاهان در مواقع سختی» می‌باشد. دیگر مدارک مهم و اثرگذار این خوشه (به شماره‌های ۲۷۳۲، ۸۰۷۸ و ۱۰,۲۴۱) نیز به مسائلی همچون «تغییرات نمکی در گیاهان»، «عکس‌العمل گیاهان در برابر خشکی، نمک، و درجه حرارت بالا» و «مقابله در برابر کمبود آب» پرداخته‌اند (جدول شماره ۱۳). خوشه‌ای کوچک در کنار این خوشه قرار دارد که زمان شکل‌گیری آن مربوط به سال‌های ۲۰۰۰ تا ۲۰۰۴ بوده و پراستنادترین و مهم‌ترین مدرک آن، مدرک شماره ۱۲۹۷ است که به بیان «الزام در تغییرات آب‌وهوایی اروپا به علت کاهش کمی کربن در خاک آنجا برای کشاورزی» پرداخته و دیگر مدارک این خوشه نیز همه بر «مسئله کربن در خاک اروپا» تأکید دارند. بعد از این خوشه، خوشه‌های کوچکی دیده می‌شوند که چندان مورد توجه نیستند. در کل می‌توان موضوع خوشه یکم را در رابطه با «تغییرات و تنوع بیولوژیکی در اکوسیستم»، خوشه دوم را «پیش‌بینی‌هایی در رابطه با تغییرات آینده در محیط زیست»، خوشه سوم را مربوط به «پوشش گیاهی زمین»، خوشه چهارم را «بررسی عملکرد گیاهان و نحوه مقابله آنان با تغییرات شدید اکولوژیکی» و خوشه پنجم را «مسئله کربن در خاک اروپا» معرفی کرد.

همچنین می‌توان نتیجه گرفت که بر اساس الگوریتم «هیست‌سایت» و استنادهای «جی‌سی‌اس»، سیصد مقاله برتر در میان ۲۲,۶۱۷ مدرک که در سال‌های ۲۰۰۰ تا ۲۰۰۸ توسط متخصصان رشته کشاورزی به رشته تحریر درآمده و در پایگاه نمایه‌استنادی علوم، نمایه شده‌اند، مهم‌ترین و مؤثرترین مقالات در این پایگاه و بازه زمانی می‌باشند. موضوعات ۵ خوشه تشکیل‌شده نیز از جمله مهم‌ترین موضوعاتی بودند که مورد توجه دانشمندان رشته کشاورزی در سطح جهان واقع شده‌اند. نویسندگان هر ۳۰۰ مقاله مورد بررسی را می‌توان جزو نویسندگان کلیدی این حوزه در پایگاه نمایه‌استنادی علوم به حساب آورد.

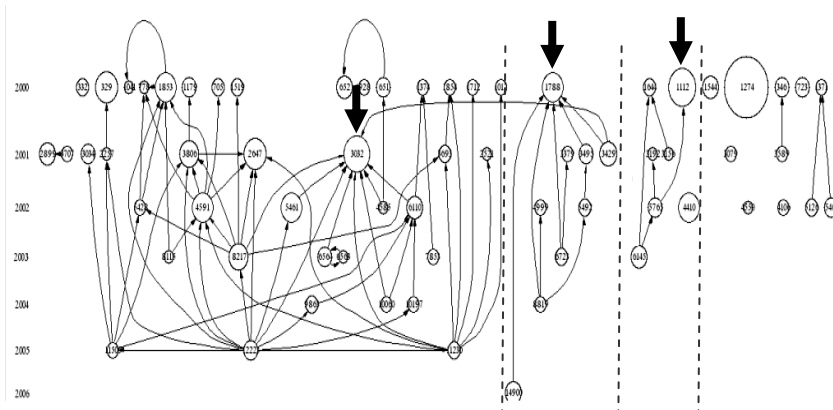
اطلاعات کتابشناختی تمامی مدارک معرفی شده در این قسمت در جدول ۱۳ آمده است.

^۱ Kovtun

خوشه ۱

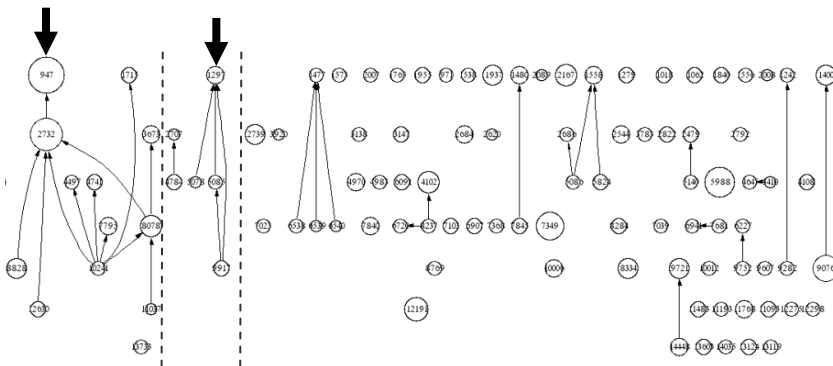
خوشه ۲

خوشه ۳



خوشه ۴

خوشه ۵



نمودار ۳. ترسیم نقشه علم‌نگاشتی ۳۰۰ مدرک برتر در حوزه کشاورزی بر اساس شاخص

«جی‌سی‌اس»

۵-۷. بررسی مقالات حوزه کشاورزی ایران

از میان کل ۲۲,۶۱۷ مدرک در این مطالعه، ایران جمعاً دارای ۹۸ مدرک بود. این مقالات از سوی مقالات ایرانی حوزه کشاورزی ۶ بار و در کل پایگاه «وب‌آوساینس»، ۲۱۸ بار مورد استناد قرار گرفته‌اند. ایرانیان نوشته‌های خود را در قالب مقاله (۸۸ مورد)،

مقاله مروری (۸ مورد) و مقاله سمینار (۲ مورد) ارائه کرده‌اند. توزیع مقالات بر اساس تاریخ نشر، نشان داد که این پژوهشگران بیشترین میزان انتشارات را در سال ۲۰۰۸ (۳۰ مقاله) داشته‌اند. کل ۹۸ مقاله را ۳۳۱ پژوهشگر ایرانی تحریر نموده‌اند. از میان ۳۳۱ پژوهشگر ایرانی این حوزه در پایگاه «وب آوساینس»، «جلالی» با ۴ مقاله و سپس «پور»، «صامتی» و «سپاسخواه» هر یک با ۳ مقاله فعالترین نویسندگان ایرانی محسوب می‌شوند.

جدول ۱۳. اطلاعات مربوط به مدارک مهم خوشه‌ها بر اساس «جی‌سی‌اس»

شماره مدرک	نویسنده، سال انتشار و مشخصات مجله	«جی‌سی‌اس» «آل‌سی‌اس»	
۹۴۷	Kovtun Y, 2000, PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, V97, P2940	۷	۴۴۰
۱۱۱۲	Hansen MC, 2000, INTERNATIONAL JOURNAL OF REMOTE SENSING, V21, P1331	۲۳	۳۹۱
۱۲۹۷	Smith P, 2000, GLOBAL CHANGE BIOLOGY, V6, P525	۴۲	۱۰۶
۱۶۴۴	Mertens B, 2000, ANNALS OF THE ASSOCIATION OF AMERICAN GEOGRAPHERS, V90, P467	۱۶	۹۱
۱۷۸۸	Robertson GP, 2000, SCIENCE, V289, P1922	۸۳	۲۳۶
۱۸۵۳	Chamberlain DE, 2000, JOURNAL OF APPLIED ECOLOGY, V37, P771	۵۰	۲۲۲
۲۷۳۲	Zhu JK, 2001, TRENDS IN PLANT SCIENCE, V6, P66	۱۰	۳۵۴
۳۰۳۲	Tilman D, 2001, SCIENCE, V292, P281	۹۵	۳۳۹
۳۴۲۹	Clark JS, 2001, SCIENCE, V293, P657	۳	۱۵۷
۳۸۰۶	Kleijn D, 2001, NATURE, V413, P723	۴۴	۱۹۰
۴۵۹۱	Robinson RA, 2002, JOURNAL OF APPLIED ECOLOGY, V39, P157	۵۲	۲۱۵
۸۰۷۸	Wang WX, 2003, PLANTA, V218, P1	۱۰	۱۸۰
۱۰۲۴۱	Chaves MM, 2004, JOURNAL OF EXPERIMENTAL BOTANY, V55, P2365	۶	۶۰
۱۱۲۳۰	Hole DG, 2005, BIOLOGICAL CONSERVATION, V122, P113	۲۹	۸۹
۱۲۲۲۳	Tscharntke T, 2005, ECOLOGY LETTERS, V8, P857	۳۰	۱۲۳

نویسندگان ایرانی بیشترین میزان همکاری خود را ابتدا با کشور آمریکا (۸ مقاله) و سپس به ترتیب با سریلانکا (۶)، انگلستان (۴)، استرالیا (۳)، هند (۳) و ۲۱ کشور دیگر

داشته‌اند. از میان مؤسسات نیز، دانشگاه تهران با ۹ مقاله و دانشگاه شیراز و علوم پزشکی تهران هر یک با ۸ مقاله، فعال‌ترین مؤسسات در رشته کشاورزی بوده‌اند. به علت تعداد کم پیوندها، در نقشه علم‌نگاشتی این حوزه در ایران، خوشه مشخصی مشاهده نشد و از این‌رو از ترسیم نقشه آن خودداری شده است.

۶. بحث و نتیجه‌گیری

در این پژوهش با جستجوی انجام‌شده در پایگاه «وب‌آوساینس»، تعداد ۲۲,۶۱۷ رکورد در زمینه کشاورزی و در محدوده زمانی ۲۰۰۰ تا ۲۰۰۸ بازیابی شدند که محاسبه مقدار متوسط نرخ رشد سالانه انتشارات برای این سال‌ها، عدد ۷ درصد را نشان داد. بررسی نوع مدارک نشان داد که مجموع ۲۲,۶۱۷ رکورد، در ۱۵ قالب مختلف ارائه شده‌اند. همان‌طور که قابل انتظار بود قالب «مقاله» بیش‌ترین تعداد را به خود اختصاص داده است. این رکوردها به ۲۵ زبان زنده دنیا و با همکاری ۱۷۳ کشور به رشته تحریر درآمده‌اند. تمامی مقالات کشاورزی بازیابی‌شده از پایگاه «وب‌آوساینس» در ۳۲۵۶ مجله منتشر، و از سوی ۱۴,۸۵۲ مؤسسه مختلف ارسال شده‌اند که مؤسسه United States Department of Agriculture, Agricultural Research Service با داشتن ۴۹۱ مقاله ثبت‌شده در پایگاه استنادی «آی‌اس‌آی»، فعال‌ترین سازمان در این حوزه به نظر می‌رسد. تعداد ۵۱,۶۵۵ نویسنده از سال ۲۰۰۰ تا ۲۰۰۸ مقالات خود را در حوزه کشاورزی به رشته تحریر درآورده و در پایگاه «آی‌اس‌آی» به ثبت رسانده‌اند و در کل، تعداد ۶۵۸,۳۸۱ منبع را مورد استناد قرار داده‌اند.

مجلات هسته در این پژوهش، با مجلات هسته در پژوهش «کریمیان مزیدی» و «اشرفی» در این حوزه متفاوت بود. شاید بتوان چنین نتیجه گرفت که حوزه‌های پژوهشی نویسندگان کشاورزی ایران با نویسندگان جهان در این حوزه متفاوت است. با توجه به این که این مجلات، مجله هسته این رشته شناخته شدند، شناسایی و در دسترس قرار دادن آن‌ها برای محققین ایرانی در این رشته بی‌فایده نیست؛ زیرا این مجلات حاوی مطالبی بسیار معتبر و مهم از این حوزه هستند و بیانگر مسائل و مشکلات پژوهشی فراگیر اکثر کشورهای جهان می‌باشند (کریمیان مزیدی و اشرفی ۱۳۸۶).

از تحلیل خوشه‌های شکل گرفته در نقشه‌های تاریخ‌نگاشتی، به‌طور کلی می‌توان چنین برداشت کرد که موضوع «تغییرات زیست‌محیطی و تنوع زیستی در اکوسیستم» یکی از مسائل مهم و مطرح در این حوزه است، زیرا در هر دو ساختار علمی بر اساس استنادهای «ال‌سی‌اس» و «جی‌سی‌اس»، خوشه بزرگی به این موضوعات اختصاص یافته بود.

خوشه‌های شکل گرفته در نقشه تاریخ‌نگاشتی بر اساس «ال‌سی‌اس» شامل ۵ خوشه بود که خوشه یکم به مسائل «حفاظت از محیط زیست و رشد کیفیت محیط»، خوشه دوم به بررسی «تغییرات و تأثیرات محیطی بر کشاورزی و بالعکس»، خوشه سوم به موضوع «اقتصاد کشاورزی و بازارهای محلی غذایی»، خوشه چهارم به «مسائل رسانایی الکترونیکی خاک»، و خوشه پنجم به «بررسی و نقد بهره‌وری بیش‌تر در تولیدات کشاورزی» پرداخته‌اند. نتایج بررسی نقشه تاریخ‌نگاشتی بر مبنای «جی‌سی‌اس» نیز شامل ۵ خوشه بود که خوشه یکم موضوع «تغییرات و تنوع زیستی در اکوسیستم»، خوشه دوم «پیش‌بینی‌هایی در رابطه با تغییرات آینده در محیط زیست»، خوشه سوم «پوشش گیاهی زمین»، خوشه چهارم «بررسی عملکرد گیاهان و نحوه مقابله آنان با تغییرات شدید اکولوژیکی»، و خوشه پنجم «مسئله کربن در خاک اروپا» را تحت پوشش قرار داده بودند.

بررسی مقالات مربوط به ایران، نتایج چندان مطلوبی را نشان نداد. تعداد این مدارک اندک بود و تلاش بیش‌تر از سوی پژوهشگران این رشته را می‌طلبد. حال که نتایج این پژوهش، مؤسسات و نویسندگان کلیدی رشته را مشخص کرده است تلاش در ایجاد ارتباط و همکاری علمی در تألیف مقالات با آنان می‌تواند گامی مؤثر در ارتقاء انتشارات علمی ایرانیان در رشته کشاورزی باشد. به‌علاوه، امکان دسترسی به این مقالات و نویسندگان آنان، می‌تواند فرصت ارزشمندی را در اختیار پژوهشگران کشورمان در این حوزه علمی قرار دهد.

۷. فهرست منابع

تصویری قمصری، فاطمه و محمدرضا جهان‌نما. ۱۳۸۵. تحلیل استنادی پایان‌نامه‌های پژوهشگران پژوهشکده مهندسی جهاد کشاورزی. فصلنامه کتاب ۱۷(۳): ۸۸-۷۵.

- حمیدی، علی، امیررضا اصنافی، و فریده عصاره. ۱۳۸۷. بررسی تحلیلی و ترسیم ساختار انتشارات علمی تولید شده در حوزه‌های کتابسنجی، علم‌سنجی در پایگاه Web of Science طی سال‌های ۱۹۹۰-۲۰۰۵. *کتابداری و اطلاع‌رسانی* ۱۱(۲): ۱۸۲-۱۶۱.
- درخشانی، مسعود. ۱۳۸۴. تاریخچه کشاورزی ایران. *ویلاگ کشاورزی، زراعت و اصلاح نباتات*. <http://keshavarzi.blogfa.com/post-2.aspx> (دسترسی در ۸/۱۲/۴).
- عصاره، فریده. ۱۳۷۷. تحلیل استنادی. *فصلنامه کتاب* ۹(۳ و ۴): ۴۸-۳۴.
- عصاره، فریده. ۱۳۸۴. علم‌سنجی: ابعاد، روشها و کاربردهای آن. در *مجموعه مقالات همایشهای انجمن کتابداری و اطلاع‌رسانی ایران*، به کوشش محسن حاجی زین‌العابدینی، ۲۷۱-۲۸۷. تهران: انجمن کتابداری و اطلاع‌رسانی ایران: سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران.
- کریمیان مزیدی، محمدعلی، و محمد اشرفی. ۱۳۸۶. تحلیل استنادی مقالات سه نشریه علوم کشاورزی واحد علوم و تحقیقات دانشگاه آزاد اسلامی، دانش کشاورزی دانشگاه تبریز و تحقیقات کشاورزی ایران. *فصلنامه کتاب* ۱۸(۱): ۶۸-۵۱.
- میرشمسی، شهرزاد. ۱۳۶۹. قانون بردفورد و کتابسنجی. *فصلنامه کتاب* ۱(۲، ۳ و ۴): ۳۲۴-۲۹۰.
- Garfield, Eugene, Soren W. Paris, & Wolforg G. Stock. 2006. A Software Tool for Informetric Analysis of citation Linkage. *Information* 57(8): 391- 400.
- Garfield, Eugene, A.I. Pudovkin, and V.S. Istomin. 2002. *Algorithmic Citation-Linked Historiography-Mapping the Literature of Science*. Presented the ASIS&T 2002: Information, Connections and Community. 65th Annual Meeting of ASIST, Philadelphia, PA.
- Garfield, Eugene, A.I. Pudovkin, and V.S. Istomin. 2003. Mapping the Output of Topical Searches in the Web of Knowledge and Case of Watson-Crick. *Information Technology and Libraries* 22(4):183- 187.
- Garfield, Eugene, A.I. Pudovkin, & V.S. Istomin. 2003. Why Do We Need Algorithmic Histogramy? *Journal of the American Society for Information Science and Technology* 54(5): 400- 412.
- Kim, Heejung and Jae Yan Lee. 2008. Exploring the emerging intellectual structure of archival studies using text mining: 2001 – 2004. *Journal of Information Science* 34(3): 356- 369.
- Osareh, Farideh and Katherine W. McCain. 2008. The structure of Iranian chemistry research, 1990-2006: An author cocitation analysis, *Journal of the American Society for Information Science and Technology* 59(13): 2146- 2155.
- Paul, Ray J. 2001. Visualizing a Knowledge Domain's Intellectual Structure. *Research Feature*. www.pages.drexel.edu/papers/copruiter.html (accessed Dec. 7, 2008)
- Shokeen, Ashu and Sanjay K. Kaushik. 2004. Indian journal of plant physiology: A citation analysis. *Annals of Library and Information Studies* 51(3): 104- 107.

بررسی کارآمدی روش‌های موجود در بازیابی اطلاعات بین‌زبانی فارسی - انگلیسی با استفاده از واژه‌نامه دوزبانه ماشین‌خوان

حمید علیزاده*

دکترای کتابداری و اطلاع‌رسانی
استادیار مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

رحمت‌الله فتاحی

دکترای کتابداری و اطلاع‌رسانی
استاد دانشگاه فردوسی مشهد

محمد رضا داورپناه

دکترای کتابداری و اطلاع‌رسانی
دانشیار دانشگاه فردوسی مشهد

اطلاعات
علوم و فناوری

دریافت: ۱۳۸۸/۰۶/۱۱ | پذیرش: ۱۳۸۸/۰۸/۱۱ | مقاله برای اصلاح به مدت یک ماه و شش روز نزد پدیدآوران بوده است.

چکیده: در این پژوهش میزان تأثیر انجام پردازش‌های زبان طبیعی بر روی ترجمه عبارت‌های جستجو با آزمون فرضیه‌های پژوهش مشخص گردید. فنون پردازش زبان طبیعی که برای پردازش عبارت‌های جستجو به کار گرفته شد شامل قطعه‌بندی متن، شناخت گونه‌های زبانشناختی، حذف سیاهه‌بازدارنده، تحلیل مورفولوژیک، و برجسب‌زنی انواع نقش دستوری بود. آزمون فرضیه اول نشان داد که استفاده از روش ترجمه اولین برابرنهاده در مقایسه با شیوه انتخاب همه برابرنهاده‌ها موجب کارآمدی بیشتر در بازیابی می‌گردد. آزمون فرضیه دوم نشان داد که اگرچه تحلیل مورفولوژیک واژه‌هایی که به وسیله واژه‌نامه ترجمه نشدند باعث افزایش ضریب دقت بازیافت می‌گردد، اما تفاوت معناداری با عدم انجام این تحلیل ایجاد نمی‌نماید. بررسی فرضیه سوم نیز نشان داد که ترجمه عبارتی در مقایسه با ترجمه واژه به واژه باعث کارآمدی بیش‌تر می‌گردد. یافته‌های دیگر این پژوهش نیز نشان داد که دگرنویسی واژه‌های فارسی ترجمه‌ناپذیر با حروف انگلیسی و قرار دادن آن‌ها در عبارت جستجوی نهایی در مقایسه با حذف آن‌ها از عبارت‌های جستجو، می‌تواند منجر به افزایش کارآمدی گردد.

کلیدواژه‌ها: بازیابی اطلاعات بین‌زبانی؛ پردازش زبان طبیعی؛ واژه‌نامه دوزبانه ماشین‌خوان؛ ارزیابی بازیابی اطلاعات

* پدیدآور رابط halizade@gmail.com

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا (چاپی) ۱۷۳۵-۵۲۰۶
شاپا (الکترونیکی) ۲۰۰۸-۵۵۸۳
نماینده در LISA و SCOPUS
<http://jlist.irandoc.ac.ir>
دوره ۲۵ | شماره ۱ | صص ۵۳-۷۰
پاییز ۱۳۸۸

نوع مقاله: علمی پژوهشی

۱. مقدمه

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 تنها به منابع اطلاعاتی که به زبان آن‌ها نوشته شده 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 هدف‌های بزرگ‌تر (یعنی غلبه بر موانع زبانی در هنگام جستجو و بازیابی اطلاعات) 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 حداقل ۳۰ میلیون نفر استفاده می‌شود (Edwards 1994, 15). بدیهی است که برای تبادل اطلاعات در این جامعه اطلاعاتی 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 چند سال اخیر رشد قابل ملاحظه‌ای داشته است. این نرخ رشد بویژه در خاورمیانه، آمریکای جنوبی، و آفریقا بسیار چشمگیر است (Wang 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 بین‌زبانی^۱ است (در این پژوهش از واژه **بازیابی** برای بیان اختصاری «بازیابی اطلاعات بین‌زبانی» استفاده شده است). بازیابی اطلاعات بین‌زبانی نوعی از بازیابی اطلاعات است که در آن حداقل دو زبان وجود دارد: زبان عبارت جستجو، و زبان مجموعه مدارک. زبان عبارت جستجو را زبان اصلی^۲، و زبان مجموعه مدارک را زبان هدف یا مقصد^۳ می‌نامند. یک نظام بازیابی اطلاعات بین‌زبانی (بازیابی)، مدارک را در زبانی که با زبان عبارت جستجو متفاوت است

¹ Cross Language Information Retrieval (clir)

² Source Language

³ Target Language

بازیابی می‌کند. البته کاربر نظام بازیبن، عبارت جستجو را به زبان بومی خویش ارائه می‌کند، اما مدارک دریافتی بر اساس زبان مجموعه مدارک خواهد بود. نظام بازیبن، کار جستجوگرانی که به چند زبان تسلط دارند را ساده می‌کند و در عین حال، جستجوگرانی را که تنها به یک زبان تسلط دارند، قادر می‌سازد عبارت جستجو را به زبان خود ارائه کنند 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 شده‌اند، سپس با استفاده از عامل انسانی یا ماشینی، ترجمه و استفاده می‌شوند (Ballesteros and Croft 1998).

۲. بیان مسئله

در حال حاضر، نظام بازیابی اطلاعات بین‌زبانی فارسی- انگلیسی وجود ندارد که کاربران گوناگون از آن در جهت برآورده ساختن نیاز خود در هنگام جستجو در محیط‌های جدید استفاده کنند. از آنجا که تاکنون تحقیقی نیز در زمینه کاربرد زبان فارسی در این حوزه به عمل نیامده، تصویر روشنی از ظرفیت‌ها و امکانات بالقوه این زبان در فرایند بازیبن وجود ندارد. پس از گذشت دو دهه از به‌کارگیری نظام‌های بازیبن در دنیا، هنوز در حوزه زبان فارسی حرکتی برای شناخت، طراحی و بهره‌گیری از مزایای این نظام انجام نشده و به همین علت پیش‌زمینه‌ها و ملزومات آن نیز شناخته نشده است. امروزه اگر از فعالیت‌های انبوه انجام‌شده در زبان‌هایی چون انگلیسی و فرانسه بگذریم، حتی از زبان‌هایی چون عربی، تایلندی و اندونزیایی هم در این حوزه، سال‌ها عقب هستیم. بر اساس آنچه که در ارتباط با اهمیت چنین نظامی مطرح شد، این تحقیق می‌کوشد راه‌حلی برای مسئله‌های زیر بیابد:

مشخص نیست که استفاده از فنون پردازش زبان طبیعی از قبیل تحلیل مورفولوژیک، برچسب‌زنی انواع نقش‌های دستوری، و دگرنویسی به چه میزان در ارتقاء کارآمدی نظام بازیبن تأثیر دارد. همچنین مشخص نیست که در هنگام ترجمه عبارت‌های جستجو، اتخاذ چه رویکردی در انتخاب برابرنهاده‌های ارائه‌شده در واژه‌نامه، منجر به دستیابی به نتیجه مناسب‌تری برای عملکرد نظام بازیبن می‌گردد. بدیهی است که بسیاری از نام‌های خاص، در واژه‌نامه موجود نیست؛ نحوه رفتار با کلمات ترجمه‌ناپذیر و خارج از واژه‌نامه نیز مسئله‌ای است که در مقابل یک نظام بازیبن قرار دارد. انسان در ارتباط‌های کلامی از

الگوی عبارتی برای انتقال بهتر مفاهیم بهره می‌گیرد؛ اما مشخص نیست که در ترجمه عبارات‌های جستجو، شناسایی عبارات‌ها و ترجمه آن‌ها به صورت عبارتی، تا چه میزان بر کارآمدی نظام بازیابی اطلاعات بین‌زبانی فارسی-انگلیسی تأثیر دارد. در نهایت، فراهم آوردن داده‌های تجربی برای بررسی این نادانسته‌ها می‌تواند زوایای ناشناخته بازیابی اطلاعات بین‌زبانی فارسی-انگلیسی را روشن نماید.

هدف کلی این پژوهش، توسعه نظری و عملی دانش بازیابی فارسی با تمرکز خاص بر ابزار ترجمه عبارات‌های جستجو است. اهداف دیگر این پژوهش شامل موارد زیر است:

- ◇ انتخاب مناسب‌ترین شیوه از بین رویکردهای مختلف در ترجمه عبارات جستجو با واژه‌نامه دوزبانه ماشین‌خوان. وجود رویکردهای مختلف، تصمیم‌گیری برای انتخاب مناسب‌ترین شیوه را منوط به انجام پژوهشی نظام‌مند می‌سازد. در این پژوهش سعی می‌شود با بررسی رویکردهای مختلف به ترجمه عبارات‌های جستجو در یک محیط اطلاعاتی و با استفاده از مجموعه عبارات‌های جستجوی آزمون‌شده، مناسب‌ترین شیوه ترجمه انتخاب، و با استفاده از معیارهایی چون قضاوت ربط، کارآمدی آن ارزیابی گردد.

- ◇ یک نظام بازیابی اطلاعات بین‌زبانی دارای اجزایی است که یکی از مهم‌ترین آن‌ها، قسمتی است که برای ترجمه عبارات‌های جستجو استفاده می‌شود. این تحقیق در پی بررسی لزوم بهره‌گیری از ابزار پردازش زبان طبیعی (از قبیل ابزار حذف واژه‌های سیاهه‌بازدارنده، و تحلیل مورفولوژیک) در کنار نظام ترجمه است.
- ◇ اگرچه این تحقیق به زبان‌های فارسی و انگلیسی محدود شده، اما نتایج آن و رویکردهای به کار گرفته شده می‌تواند نقشه راهی برای انجام پژوهش‌هایی مشابه در دیگر زبان‌ها ایجاد کند و در ایجاد نظام‌های بازیابی فارسی به دیگر زبان‌ها نیز راهگشا باشد.

فرضیه‌های این پژوهش به شرح زیر هستند:

۱. در هنگام استفاده از واژه‌نامه دوزبانه ماشین‌خوان، رویکرد ترجمه اولین برابر نهاده در مقایسه با رویکرد همه برابر نهاده‌ها، باعث کارآمدی بیش‌تر (افزایش ضریب دقت) در نتایج بازیابی اطلاعات بین‌زبانی فارسی-انگلیسی می‌گردد.

۲. پردازش مورفولوژیک واژه‌های عبارت جستجوی فارسی (که در واژه‌نامه مورد بررسی وجود ندارد) پیش از ترجمه آن‌ها باعث کارآمدی بیش‌تر (افزایش ضریب دقت) در نتایج بازیابی اطلاعات بین زبانی فارسی- انگلیسی می‌گردد.
۳. در هنگام ترجمه عبارت‌های جستجوی فارسی، شیوه ترجمه عبارتی در مقایسه با ترجمه کلمه به کلمه باعث کارآمدی بیش‌تر (افزایش ضریب دقت) در نتایج بازیابی اطلاعات بین زبانی فارسی- انگلیسی می‌گردد.
۴. استفاده از روش دگرنویسی اصطلاح‌های ترجمه‌ناپذیر یا خارج از واژه‌نامه فارسی به حروف انگلیسی و بازیابی بر اساس آن (در مقایسه با حذف این گونه واژه‌ها از عبارت جستجو) باعث کارآمدی بیش‌تر در بازیابی اطلاعات بین زبانی فارسی- انگلیسی می‌گردد.

۳. پیشینه پژوهش

همان‌گونه که پیش از این ذکر شد، یکی از انگیزه‌های اصلی پژوهش در حوزه بازیابی اطلاعات بین زبانی فارسی- انگلیسی، کمبود منابع فارسی در اینترنت در مقایسه با دیگر زبان‌های بزرگ است. کاستی‌های نرم‌افزاری را می‌توان یکی از بزرگ‌ترین دلایل غیبت زبان فارسی در اینترنت دانست. راه غلبه بر چنین مشکلی، توجه به طراحی و راه‌اندازی نرم‌افزارها و نظام‌های اطلاعاتی خاص زبان فارسی است. اکنون که توجه زیادی به تولید محتوای فارسی در وب می‌شود، سرمایه‌گذاری در پشتیبانی فنی این محتوا و ایجاد نظام‌هایی همچون نظام بازیابی نیز ضروری به نظر می‌رسد. با جستجو در نوشته‌های داخلی، تنها مقاله‌ای که در آن به بازیابی اطلاعات بین زبانی فارسی اشاره شده، نوشته «آزادنیا» است که در طرح پیشنهادی خود، برای سند توسعه بومی‌سازی در ایران و به عنوان یکی از محورهای اصلی این سند، به بازیابی چندزبانی اشاره نموده و اینترنت را محملی دانسته که برای بهره‌گیری بهینه از آن، چاره‌ای جز توسعه بازیابی بین زبانی نیست (آزادنیا ۱۳۸۳). تا آنجا که پژوهشگر بررسی نموده، این گفته در حد پیشنهاد باقی مانده و هنوز اقدامی در عملی‌سازی این طرح انجام نشده است.

پژوهش‌های غیرفارسی در حوزه بازیابی بسیار زیاد است و در اینجا برای نمونه به مواردی اشاره می‌شود. اولین تجربه بازیابی با رویکرد واژه‌نامه نشان داد که ترجمه واژه به

واژه عبارت‌های جستجو می‌تواند منجر به کاهش بین ۴۰٪ تا ۶۰٪ کارآمدی بازیابی در مقایسه با بازیابی اطلاعات یک‌زبانه بر اساس همان عبارت‌های جستجو گردد (Hull and Grefenstette 1996). در این پژوهش که درباره بازیابی فرانسوی-انگلیسی انجام شد، با استفاده از محاسبه متوسط دقت بازیافت، نتایج زیر به دست آمد: ترجمه عبارت‌های جستجو به روش واژه به واژه منجر به دستیابی به متوسط دقت بازیافت برابر با ۲۳۵/۰ می‌گردد که این مقدار، معادل ۶۰٪ میزان کارآمدی در بازیابی یک‌زبانه در این پژوهش بود. این میزان برای روش ترجمه عبارتی عبارت‌های جستجو افزایش پیدا کرد و به ۳۵۷/۰ (یعنی ۹۱٪ عملکرد بازیابی یک‌زبانه که برابر با ۳۹۳/۰ بود) افزایش یافت. آن‌ها نتیجه گرفتند که ترجمه عبارتی در مقایسه با ترجمه واژه به واژه عبارت‌های جستجو، منجر به نتیجه بهتری می‌شود.

«چن» نیز با انجام پژوهشی به بررسی کارآمدی ترجمه عبارتی در بازیابی چینی-انگلیسی پرداخت. وی که با استفاده از روش تحقیق ارزیابی برنامه، این پژوهش را انجام داد، ترجمه عبارتی را موفق‌تر از ترجمه واژه به واژه دانست. یافته‌های او نشان داد که در مقایسه با میانگین دقت بازیافت در بازیابی یک‌زبانه، ترجمه عبارتی به ۵۳٪ کارآمدی دست می‌یابد. این در حالی است که این میزان برای ترجمه واژه به واژه به ۴۲٪ رسید. وی اشاره می‌کند که با بهره‌گیری از منابع اضافی و کامل‌تر برای ترجمه عبارت‌ها، می‌توان این میزان از کارآمدی را افزایش داد و به ۸۳٪ کارآمدی به دست آمده برای بازیابی یک‌زبانه رسید (Chen 2002).

تلاش برای رفع ابهام در ترجمه نیز از جمله مواردی است که در پژوهش‌ها به آن توجه شده است. «بالستروس» و «کرافت» (Ballesteros & Croft 1998) برای رفع ابهام در ترجمه، روش آمار هم‌رخدادی^۱ را مطرح کرده‌اند. رویکرد هم‌رخدادی بر این اندیشه استوار است که ترجمه‌های صحیح واژه‌های عبارت جستجو، گرایش به رخداد همزمان در متن‌های زبان هدف دارند، ولی ترجمه‌های نادرست این گرایش را ندارند. آن‌ها در پژوهش خود از یک ابزار برجسب‌زنی نوع نقش دستوری برای انتخاب برابر نهاده‌های زبان هدف، که دارای برجسب‌های نوع نقش دستوری مشابه با اصطلاح‌های عبارت جستجو در

^۱ co-occurrence

زبان اصلی باشد، استفاده کردند. در نهایت با انجام محاسبه‌های مختلف، آمار هم‌رخدادی مشخص، و برابر نهاده‌هایی که در سیاهه رتبه‌بندی بر اساس آن، بالاتر قرار گرفته بودند، به عنوان ترجمه مناسب انتخاب شدند (Ballesteros and Croft 1998).

۴. روش پژوهش

این پژوهش کاربردی است و برای انجام آن از روش پژوهش نیمه تجربی استفاده شده. جامعه آماری این پژوهش، رکوردهای بازیابی شده از موتور جستجوی گوگل بر اساس جستجوی عبارت‌های جستجوی این پژوهش است. روش نمونه‌گیری این پژوهش، روش نمونه‌گیری «کنفرانس بازیابی متن (ترک)»^۱ برای ارزیابی نظام‌های بازیابی اطلاعات است که امروزه در اغلب پژوهش‌های مربوط به ارزیابی نظام‌های بازیابی اطلاعات از آن استفاده می‌شود (Voorhees 1998). این شیوه، مخزن‌سازی^۲ نام دارد. شیوه مخزن‌سازی که در این پژوهش انجام شد به این شرح است: ابتدا عبارت‌های جستجو در اختیار تعدادی جستجوگر قرار گرفت. هر جستجوگر پس از انجام جستجو در سامانه بازیابی اطلاعات (که در این تحقیق، موتور جستجوی گوگل بود) سیاهه نتایج بازیابی شامل ۱۰۰ رکورد اول بازیابی شده برای هر عبارت جستجو را ارائه نمود. سپس با استفاده از این سیاهه‌ها برای هر عبارت جستجو، مخزنی تشکیل شد. رکوردهای خارج از مخزن نیز نامربوط تلقی شدند (و به دلیل بزرگی مجموعه‌های مدارک، چاره‌ای جز این نیست).

با توجه به حجم بالای پردازش‌های زبان طبیعی و مراحل مختلف بازیابی، تعداد ۴۰ عبارت جستجو از مجموعه عبارت‌های جستجوی «ترک» انتخاب شد. هر عبارت جستجو از سه بخش تشکیل شده: بخش اول شامل یک عنوان کوتاه است. بخش دوم توصیف یک جمله‌ای از موضوع (جمله نیاز اطلاعاتی) است. بخش سوم از چند جمله که معیارهای ربط را مشخص می‌کنند، تشکیل شده. در زیر متن انگلیسی یکی از عبارت‌های جستجوی «ترک» مشاهده می‌شود:

Title: War and Radio.

Description: What role do radios play during war or armed conflict?

Narrative: The radio plays an important role by broadcasting propaganda and messages of hope or warning. Relevant documents talk about this role.

¹ Text Retrieval Conference (TREC)

² Pooling

در ساخت عبارت جستجوی نهایی برای ارائه به سامانه اطلاعاتی، از فیلد «عنوان» استفاده شد و از اطلاعات موجود در فیلد «توصیف» نیز در صورت نیاز برای بسط جستجو استفاده گردید.

۵. یافته‌های پژوهش

۱. در هنگام استفاده از واژه‌نامه دوزبانه ماشین‌خوان، رویکرد «ترجمه اولین برابرنهاد» در مقایسه با رویکرد «همه برابرنهادها» باعث کارآمدی بیش‌تر (افزایش ضریب دقت) در نتایج بازیابی اطلاعات بین‌زبانی فارسی-انگلیسی می‌گردد.

ابهام در ترجمه هنگامی رخ می‌دهد که برابرنهادهای انتخاب‌شده، همخوانی مناسبی با واژه موجود در عبارت جستجو نداشته باشند. موضوع ابهام‌زدایی در ترجمه، از سوی پژوهشگرانی چون «دیویس» و «دانینگ» و «پیرکولا» و همکارانش مطرح، و راهکارهایی در مورد آن ارائه شده است (Davis and Dunning 1995; Pirkola et al 2001). این روش به این دلیل که اصطلاح‌های بعضاً نامربوط را به عبارت جستجوی ترجمه‌شده به زبان هدف اضافه می‌کند، می‌تواند منجر به بروز ابهام در ترجمه گردد که این، خود منجر به کارآیی ضعیف در بازیابی می‌گردد. «مهرداد» و «ناصری» معتقدند که زبان‌های طبیعی، خواص ویژه‌ای دارند که از سودمندی نظام‌های بازیابی اطلاعات متنی می‌کاهد. این خواص عبارت‌اند از اختلاف زبانی، و ابهام. منظور از اختلاف زبانی، استفاده از واژه‌ها یا عبارت‌های مختلف برای انتقال یک ایده واحد می‌باشد. ابهام زبانی نیز زمانی رخ می‌دهد که واژه یا عبارت دارای بیش از یک تفسیر است (مهرداد و ناصری ۱۳۸۷).

در پژوهش حاضر برای بررسی شیوه غلبه بر این مشکل، انتخاب اولین برابرنهاد مطرح گردید و این شیوه ترجمه با شیوه دیگر (یعنی استفاده از همه برابرنهادها) مقایسه شد. از آنجاکه در هنگام استفاده از رویکرد «همه برابرنهادها» کلماتی به عبارت جستجوی نهایی وارد می‌شوند که ممکن است در معنای مورد نظر نویسنده مدرک کاربرد نداشته باشند (که این خود منجر به بازیابی ناخواسته تعدادی مدرک نامرتب می‌گردد) و این امر در مقایسه با رویکرد «اولین برابرنهاد» (که معمولاً رایج‌ترین معادل هر کلمه را برای ترجمه انتخاب می‌کند) می‌تواند باعث ناکارآمدی در بازیابی گردد، این فرضیه مطرح شده است.

در این پژوهش، کارآمدی عملکرد بازیابی اطلاعات بین زبانی، با اندازه‌گیری میانگین متوسط دقت بازیافت^۱ مشخص شده است. برای محاسبه این آماره، ابتدا باید متوسط دقت بازیافت برای همه مخزن‌های مدارک، محاسبه شود. نحوه محاسبه این مقیاس به این صورت است که بررسی ربط مدارک موجود در یک مخزن، از بالای سیاهه شروع می‌شود و بلافاصله که یک مدرک مربوط شناخته شد، مقدار دقت بازیافت محاسبه می‌شود. وقتی به پایان سیاهه مدارک موجود در مخزن رسیدیم، همه مقادیرهای دقت بازیافت را جمع می‌زنیم و تقسیم بر تعداد کل مدارک‌های مرتبط موجود در مخزن می‌کنیم. بدین ترتیب متوسط دقت بازیافت، برای آن عبارت جستجو محاسبه می‌شود. در صورت وجود مجموعه‌ای از عبارت‌های جستجو، میانگین متوسط دقت بازیافت برای کل عبارت‌های جستجوی موجود، از طریق جمع کردن متوسط دقت بازیافت همه عبارت‌های جستجو و تقسیم آن بر تعداد کل عبارت‌های جستجو، محاسبه می‌شود (Wang and Oard 2005).

جدول ۱. مقایسه میانگین متوسط دقت بازیافت رویکردهای انتخاب اولین برابر نهاده، و همه برابر نهاده‌ها

روش ترجمه	میانگین متوسط دقت بازیافت	درصد تغییر نسبت به رویکرد اولین برابر نهاده
استفاده از اولین برابر نهاده	۰/۲۲۳	-
استفاده از همه برابر نهاده‌ها	۰/۱۶۸	٪۷۵

همان‌گونه که در جدول ۱ مشاهده می‌شود، از میان دو روش ترجمه مختلف، استفاده از اولین برابر نهاده با دستیابی به عدد ۰/۲۲۳، میانگین متوسط دقت بازیافت بیش‌تری داشته است. رویکرد انتخاب همه برابر نهاده‌ها با دستیابی به ٪۷۵ کارآمدی رویکرد اولین برابر نهاده، باعث کاهش میزان کارآمدی شده است. مقایسه رویکردهای مختلف ترجمه، کارآمدی بیش‌تر رویکرد انتخاب اولین برابر نهاده در فرایند ترجمه عبارت‌های جستجوی نظام‌بازین را نشان می‌دهد. نتایج فوق با یافته‌های «الجلایل» و

^۱ mean average precision

«فرایدر» که به بررسی رویکردهای مختلف ترجمه در بازمین عربی - انگلیسی پرداخته بودند مشابهت دارد. آنان نیز به این نتیجه رسیده بودند که انتخاب اولین برابر نهاده در مقایسه با انتخاب همه برابر نهاده‌ها، موجب کارآمدی بیش تر نظام بازمین می گردد (Aljalayil and Frieder 2001).

مقیاس دیگری که در این پژوهش برای ارزیابی کارآمدی نظام بازمین استفاده شد، «میانگین دقت بازیافت در سطوح مختلف بازیابی»^۱ بود. براساس این مقیاس، دقت بازیافت در سطوح مختلف بازیابی یک سیاهه رتبه بندی شده مشخص می شود. در این پژوهش، دقت بازیافت در سطوح مختلف بازیابی (یعنی در سطح ۵، ۱۰، ۲۰، و ۳۰ مدرک بازیابی شده) محاسبه شد. این مقیاس ابتدا برای هر عبارت جستجو، جداگانه محاسبه، و آنگاه با محاسبه میانگین آن‌ها در کل مجموعه عبارت‌های جستجو، شمای کلی متوسط عملکرد نظام مشخص گردید. محاسبه دقت بازیافت در سطوح مختلف بازیابی می تواند تصویر کامل تری از نحوه عملکرد یک نظام بازیابی اطلاعات را مشخص کند. این مقیاس، موقعیت‌هایی را مدل سازی می کند که در آن، کاربران ترجیح می دهند تنها مدارکی را بررسی کنند که در سیاهه‌های نتایج بازیابی، رتبه بالاتری دارند. این امر بویژه در جستجوهای وبی که در آن، موتورهای جستجو مدارک بسیاری را بازیابی می کنند و کاربران، اغلب صفحه‌های اولیه این نتایج را مرور می کنند، نمود پیدا می کند.

جدول ۲. محاسبه میانگین دقت بازیافت در سطوح مختلف بازیابی
بر اساس انتخاب برابر نهاده‌های مختلف

دقت بازیافت	اولین برابر نهاده	همه برابر نهاده‌ها
در ۵ مدرک	۰/۴۸۵	۰/۳۸۴
در ۱۰ مدرک	۰/۴۴۱	۰/۳۶۱
در ۲۰ مدرک	۰/۳۹۹	۰/۳۱۶
در ۳۰ مدرک	۰/۳۱۸	۰/۲۸۷

¹ mean precision at various cut-off levels

اطلاعات موجود در جدول ۲ نیز مؤید این نکته است که رویکرد انتخاب اولین برابرنهاده در سطوح مختلف بازیابی، برتری محسوسی نسبت به رویکرد انتخاب همه برابرنهاده‌ها دارد و باعث کارآمدی بیش‌تر بازیابی اطلاعات بین‌زبانی می‌گردد. میزان دقت بازیافت چنانکه مشاهده می‌شود روندی نزولی دارد و با حرکت به سمت انتهای سیاهه مدارک و افزایش مدارکی که مورد قضاوت ربط قرار می‌گیرد، کاهش می‌یابد.

۲. پردازش مورفولوژیک واژه‌های عبارت جستجوی فارسی (که در واژه‌نامه مورد بررسی وجود ندارد) پیش از ترجمه آن‌ها در مقایسه با عدم انجام این پردازش، باعث کارآمدی بیش‌تر (افزایش ضریب دقت) در نتایج بازیابی اطلاعات بین‌زبانی فارسی- انگلیسی می‌گردد.

در این قسمت به بررسی میزان اهمیت به کارگیری تحلیل مورفولوژیک اصطلاح‌های مورد ترجمه در نظام بازیابی اطلاعات بین‌زبانی پرداخته شده است. پژوهشگران در مورد نحوه و سطح انجام چنین تحلیل‌هایی و میزان سودمندی آن‌ها در ارتقاء عملکرد بازیابی اطلاعات، نظرات مختلفی داشته‌اند. یکی از مهم‌ترین کارها در این حوزه، توسط «پرت» ارائه شده است (Porter 1980).

در تشریح تحلیل مورفولوژیک (ساختواژه) باید گفت که بسیاری از اصطلاح‌های موجود در واژه‌نامه‌ها شکل مفرد کلمات، بدون پیشوند و پسوند و ادات جمع است. از آنجا که بسیاری از شکل‌های مختلف صرفی کلمات در واژه‌نامه وجود ندارند، در هنگام ترجمه عبارت‌های جستجو، با انبوهی از کلمات ترجمه‌نشده روبرو می‌شویم. تحلیل مورفولوژیک در حل این مشکل، ضروری به نظر می‌رسد. با این تحلیل، اشکال مختلف کلمات پردازش می‌شوند، پیشوند و پسوندها حذف، و کلمات جمع به مفرد تبدیل می‌شوند و سپس ترجمه انجام می‌گیرد. طرح این فرضیه، برای آزمون سودمند بودن و ضرورت انجام تحلیل مورفولوژیک در هنگام ترجمه صورت گرفته است.

در این پژوهش تحلیل مورفولوژیک بر روی واژه‌هایی انجام شد که به شکل اولیه موجود در عبارت جستجو، در واژه‌نامه یافت نشد. با مقایسه بازیابی بر اساس به کارگیری این تحلیل و بازیابی عبارت‌های جستجویی که واژه‌های ترجمه‌نشده از آن حذف شده است، میزان کارآمدی این روش مشخص گردید.

جدول ۳. مقایسه بازیبن با استفاده از پردازش مورفولوژیک واژه‌های ترجمه‌نشده و عدم انجام این پردازش

نوع اجرا	میانگین متوسط دقت بازیافت	درصد تغییر
بازیبن بدون تحلیل مورفولوژیک	۰/۱۸۰	-
بازیبن با تحلیل مورفولوژیک	۰/۲۲۳	٪۲۳

در جدول ۳ مشاهده می‌شود که میانگین متوسط دقت بازیافت در شیوه اول که بدون پردازش مورفولوژیک بوده برابر با ۰/۱۸۰ است، در حالی که پردازش مورفولوژیک واژه‌هایی که ترجمه نشده و در مرحله قبلی حذف شدند (و آنگاه ترجمه آن‌ها با واژه‌نامه) منجر به افزایش این رقم به میزان ٪۲۳ گردید. همین مسئله این بار در سطوح مختلف بازیابی آزمایش شد که نتایج به دست آمده، در جدول ۴ به نمایش گذاشته شده است.

جدول ۴. محاسبه میانگین دقت بازیافت در سطوح مختلف بازیابی بر اساس مقایسه انجام پردازش مورفولوژیک واژه‌های ترجمه‌نشده و عدم انجام این پردازش

دقت بازیافت	بدون پردازش مورفولوژیک	با پردازش مورفولوژیک
در ۵ مدرک	۰/۳۹۶	۰/۴۸۵
در ۱۰ مدرک	۰/۳۷۲	۰/۴۴۱
در ۲۰ مدرک	۰/۳۱۰	۰/۳۹۹
در ۳۰ مدرک	۰/۲۶۳	۰/۳۱۸

نتایج مندرج در جدول ۴ نشان می‌دهد که انجام پردازش مورفولوژیک در سطوح مختلف بازیابی باعث افزایش میانگین دقت بازیافت و در نتیجه کارآمدی بیش‌تر این روش می‌گردد.

۳. در هنگام ترجمه عبارت جستجوهای فارسی، شیوه ترجمه عبارتی در مقایسه با ترجمه واژه به واژه، باعث کارآمدی بیش‌تر (افزایش ضریب دقت در بازیافت) در نتایج بازیابی اطلاعات بین زبانی فارسی- انگلیسی می‌گردد.

وجود «عبارت» در بعضی از جملات جستجو این سؤال را در ذهن پدید می‌آورد که یک عبارت را باید یک ترکیب واژگانی در نظر گرفت و کلمات آن را با هم ترجمه کرد، یا این که باید واژه‌های عبارت به صورت جدا از هم ترجمه گردد. پژوهشگران رویکردهای مختلفی را در مواجهه با این مورد به کار گرفته‌اند. برای مثال «بالستروس» و «کرافت» در پژوهشی به بررسی تأثیر رویکردهای «آماری» و «پردازش زبان طبیعی» در ترجمه عبارتی پرداخته‌اند (Ballesteros and Croft 1998). این فرضیه به دلیل استفاده زیاد از عبارت‌ها در ساخت جملات جستجوی فارسی، و در جهت بررسی کارآمدی روش ترجمه عبارتی در بازیابی، مطرح گردید.

شناسایی عبارت‌های موجود در جملات جستجو و ترجمه آن‌ها به صورت عبارت (و نه به صورت واژه‌های مجزا) از چالش‌های پیش روی بازیابی است. در این پژوهش با استفاده از برچسب‌زنی انواع نقش دستوری (یعنی مشخص کردن نوع دستوری واژه‌ها از قبیل اسم، صفت، قید، و ...) تلاش شد عبارت‌های موجود، مشخص و در صورت موجود بودن در واژه‌نامه، به صورت عبارتی ترجمه گردند. این شیوه ترجمه سپس با شیوه ترجمه واژه به واژه مقایسه گردید.

جدول ۵. مقایسه کارآمدی شیوه ترجمه عبارتی با شیوه ترجمه واژه به واژه

نوع اجرا	میانگین متوسط دقت بازیافت	درصد تغییر
بازیابی با استفاده از ترجمه واژه به واژه	۰/۲۲۳	-
بازیابی با استفاده از ترجمه عبارتی	۰/۳۱۹	٪۴۳

همان‌گونه که در جدول ۵ مشاهده می‌شود، شیوه ترجمه عبارتی با دستیابی به میانگین متوسط دقت بازیافتی برابر با ۰/۳۱۹ در مقایسه با ترجمه واژه به واژه کارآمدتر بوده و باعث ارتقاء آن به میزان ٪۴۳ می‌گردد. این یافته‌ها با نتایج پژوهش‌های انجام‌شده

توسط «هال» و «گرفنستت» که در مورد بازبین فرانسوی-انگلیسی انجام شد و «چن» که به بررسی بازبین چینی-انگلیسی پرداخت همخوانی دارد (Hull and Grefenstette 1996; Chen 2002). در آن پژوهش‌ها نیز کارآمدی روش ترجمه عبارتی در مقایسه با ترجمه واژه به واژه مشخص گردیده بود.

کارآمدی ترجمه عبارتی در مقایسه با ترجمه واژه به واژه در سطوح مختلف بازیابی نیز بررسی شد که نتایج این بررسی در جدول ۶ نشان داده شده است.

جدول ۶. نتایج بررسی میانگین دقت بازیافت در سطوح مختلف بازیابی بر اساس ترجمه عبارتی و ترجمه واژه به واژه

دقت بازیافت	ترجمه واژه به واژه	ترجمه عبارتی
در ۵ مدرک	۰/۴۸۵	۰/۵۹۲
در ۱۰ مدرک	۰/۴۴۱	۰/۵۴۱
در ۲۰ مدرک	۰/۳۹۹	۰/۵۱۲
در ۳۰ مدرک	۰/۳۱۸	۰/۴۸۱

نتایج اطلاعات موجود در جدول ۶ نشانگر برتری نسبی ترجمه عبارتی در مقایسه با ترجمه واژه به واژه در سطوح مختلف بازیابی است. این یافته‌ها در تکمیل نتایج پیشین، مبین سودمندی بهره‌گیری از برجسب‌زنی انواع نقش‌های دستوری و شناسایی و ترجمه عبارت‌ها در فرایند بازیابی اطلاعات بین‌زبانی است.

۴. استفاده از روش دگرنویسی اصطلاح‌های ترجمه‌ناپذیر یا خارج از واژه‌نامه فارسی با حروف انگلیسی و بازیابی بر اساس آن، در مقایسه با حذف آن‌ها از عبارت‌های جستجو، باعث کارآمدی بیش‌تر بازیابی اطلاعات بین‌زبانی فارسی-انگلیسی می‌گردد.

واژه‌هایی که ترجمه‌ناپذیر و خارج از واژه‌نامه هستند (مثل نام‌های خاص) اغلب از مهم‌ترین اجزای یک عبارت جستجو به‌شمار می‌روند. «پیرکولا» مشکل ترجمه‌ناپذیری آن‌ها را از بزرگ‌ترین چالش‌های پیش روی نظام بازبین می‌داند (Pirkola et al 2001). در این پژوهش برای حل مشکل عدم ترجمه این گونه واژه‌ها، شیوه دگرنویسی (نوشتن

واژه‌های ترجمه‌ناپذیر به الفبای زبان دیگر) پیشنهاد شده است. این شیوه در ادامه این قسمت با شیوه حذف این واژه‌ها از عبارت جستجو، مقایسه شده است.

جدول ۷. مقایسه میانگین متوسط دقت بازیافت بازیابی با استفاده از دگرنویسی، و بازیابی بدون استفاده از دگرنویسی

نوع اجرا	میانگین متوسط دقت بازیافت	درصد تغییر
بازیابی با استفاده از دگرنویسی	۰/۲۲۳	٪۷۴
بازیابی بدون استفاده از دگرنویسی	۰/۱۲۸	-

همان‌گونه که در جدول فوق مشاهده می‌شود، استفاده از روش دگرنویسی واژه‌های ترجمه‌ناپذیر در مقایسه با حذف آن‌ها از عبارت‌های جستجو باعث ارتقاء عملکرد به میزان ٪۷۴ می‌شود. این یافته‌ها با نتایج پژوهش‌های «کو» و «گرفنستت»، و «چنگ» و همکاران همخوانی دارد (Qu and Grefenstette 2003; Cheng et al 2004). در پژوهش «چنگ» و همکاران نیز دگرنویسی واژه‌های ترجمه‌ناپذیر در مقایسه با حذف آن‌ها، باعث افزایش کارآمدی نظام بازیابی به میزان ٪۸۰ گردیده بود. مقایسه این دو روش در سطوح مختلف بازیابی نیز مؤید همین برتری است. جدول ۸ این نکته را نشان می‌دهد.

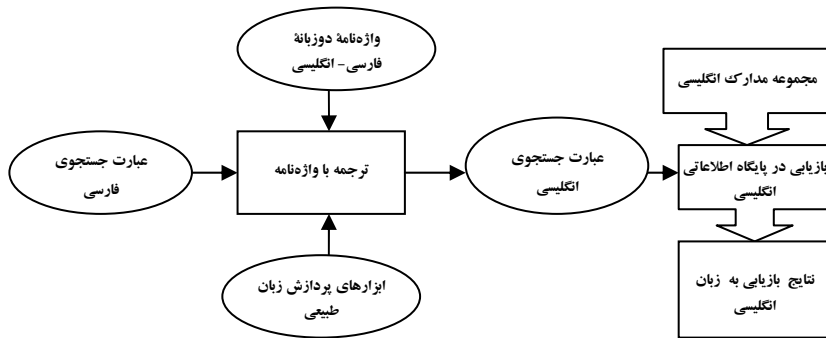
جدول ۸. مقایسه میانگین دقت بازیافت بازیابی با استفاده از دگرنویسی، و بدون استفاده از دگرنویسی در سطوح مختلف بازیابی

دقت بازیافت	بازیابی با دگرنویسی	بازیابی بدون دگرنویسی
در ۵ مدرک	۰/۴۸۵	۰/۲۶۵
در ۱۰ مدرک	۰/۴۴۱	۰/۲۲۱
در ۲۰ مدرک	۰/۳۹۹	۰/۲۰۲
در ۳۰ مدرک	۰/۳۱۸	۰/۱۸۵

مجموع یافته‌ها در این مرحله از بازیابی مؤید این است که روش دگرنویسی واژه‌های ترجمه‌ناپذیر و خارج از واژه‌نامه، در مقایسه با عدم استفاده از این شیوه و حذف این گونه واژه‌ها از عبارات‌های جستجو، باعث کارآمدی نظام بازیابی اطلاعات بین زبانی می‌گردد.

۶. مدل پیشنهادی نظام بازیابی فارسی- انگلیسی

بر اساس مفاهیم نظری که پیش‌تر ذکر گردید و با تکیه بر دانش برگرفته از یافته‌های این پژوهش، مدل پیشنهادی نظام بازیابی فارسی- انگلیسی به شرح زیر ارائه می‌گردد:



شکل ۱. مدل پیشنهادی نظام بازیابی فارسی- انگلیسی

در تحلیل این مدل باید گفت که نظام بازیابی فارسی- انگلیسی نظام خودکاری است که با استفاده از واژه‌نامه الکترونیکی و هم‌افزایی ابزارهای پردازش زبان طبیعی، پس از پردازش و ترجمه عبارات‌های جستجوی زبان اصلی (فارسی)، مدارک را از مجموعه زبان هدف (انگلیسی) بازیابی می‌کند. مدل پیشنهادی، انعطاف‌پذیری مناسبی برای استفاده در جفت‌های زبانی دیگر نیز دارد و بر اساس آن می‌توان از زبان فارسی به مجموعه مدارک در زبان‌های دیگر دست یافت. این مدل از سویی کاربردی و از سوی دیگر، ارتباطی است. ارتباطی از آن جهت که اجزای کاربردی نظام در جهت عملکرد بهینه، باید ارتباط نزدیکی داشته باشند. نظام بازیابی پیشنهادی، کارکردهای زیر را پشتیبانی می‌نماید:

◇ ترجمه عبارت جستجوی فارسی

- ◇ تحلیل‌های پردازش زبان طبیعی
- ◇ بازیابی مدارک در زبان انگلیسی

ویژگی بارز این مدل، استفاده از ابزار پردازش زبان طبیعی است که موجب افزایش دقت در ترجمه و کارآمدتر شدن بازیابی می‌گردد. این مدل اولین مدلی است که برای فرایند بازیابی اطلاعات بین‌زبانی فارسی مطرح شده و می‌تواند با انجام پژوهش‌های دیگر تکمیل‌تر گردد.

۷. نتیجه‌گیری

دلایل بسیاری وجود دارد که یک نظام بازیابی به همان کارآمدی که بازیابی یک‌زبانه نائل می‌شود، دست پیدا نمی‌کند. اولین و مهم‌ترین دلیل، وجود دو زبان در بازیابی و ساختارهای متفاوت آن‌ها است. این دوگانگی موجب بسیاری ابهام‌ها می‌گردد که بازیابی اطلاعات یک‌زبانه هرگز با آن مواجه نمی‌شود. دلیل دیگر آن است که متن و زمینه عبارت‌های جستجو در هنگام ترجمه، چندان در نظر گرفته نمی‌شوند؛ چه بسا در مواردی مهجورترین برابرنهاده یک واژه، با توجه به مفهوم عبارت جستجو مناسب‌ترین انتخاب باشد. در این پژوهش با استفاده از فنون پردازش زبان طبیعی از قبیل تحلیل مورفولوژیک، برچسب‌زنی انواع نقش دستوری و دگرنویسی، مشخص گردید که استفاده از ابزار زبان‌شناسی در فرایندهای بازیابی اطلاعات می‌تواند منجر به کارآمدتر شدن این نظام‌ها گردد. شناسایی الگوهای زبانی و ساختار واژه‌ها، یک نظام خودکار بازیابی را قادر می‌سازد که مدارک را با دقت بیش‌تری بازیابی کند و رضایتمندی کاربر را افزایش دهد. فونمی که در این پژوهش استفاده شدند اغلب فنون نحوی بودند و علی‌رغم اهمیت محدودیت‌هایی نیز دارند. بهره‌گیری از فنون پیشرفته‌تر هوش مصنوعی و استفاده از تحلیل‌های معناشناسی می‌تواند در آینده، سامانه‌های بازیابی را نسبت به نمونه‌های موجود، سامانه‌های موفق‌تری نشان دهد.

۸. فهرست منابع

آزادنیا، محمد. ۱۳۸۳. طرح مقدماتی سند توسعه بومی‌سازی در ایران. در مجموعه مقالات اولین کارگاه پژوهشی زبان فارسی و رایانه. تهران: طیف‌آرا.

- علیزاده، حمید. ۱۳۸۳. مشکلات دسترسی به اطلاعات در جهان شبکه‌ها. فصلنامه کتاب ۱۵(۲): ۱۲۱-۱۱۵.
- مهراد، جعفر و مریم ناصری. ۱۳۸۷. پردازش زبان طبیعی و بازیابی اطلاعات. تهران: چاپار، شیراز: مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری.
- Adriani, M. 2000. Using statistical term similarity for sense disambiguation in cross- language information retrieval. *Information retrieval* 38(2): 69-80.
- Aljlal, Mohammad, and Phir Frieder. 2001. *Effective Arabic- English Cross- Language Information Retrieval via Machine- Readable Dictionaries and Machine Translation*. Oral presented at the ACM Tenth Conference on Information and Knowledge Management, Atlanta.
- Ballesteros, L, and B. Croft. 1998. *Resolving Ambiguity for Cross- Language Retrieval*. Presented at the SIGIR, Australia, Melbourne.
- Chen, H. H. 2002. *Chinese information extraction techniques*. Presented at the SSIMIP, Singapore.
- Cheng, Pu-Jen, Pan, Yi-Cheng Lu, Lu Wen-Hsiang and Lee Feng Chien. 2004. Creating Multilingual Translation Lexicons with Regional variations Using Web Corpora. In *Proceedings Of ACL*. www.aclweb.org/anthology/P/P04/P04-1068.pdf. (accessed may 15, 2008).
- Davis, M, and T. Dunning. 1995. *Query Translation Using Evolutionary Programming for Multilingual Information Retrieval*. Presented at the 4th Evolutionary Programming Conference.
- Edwards, J. 1994. *Multilingualism*. London, England: Penguin.
- Hull, D. and G. Grefenstette. 1996. Querying Across Languages; A Dictionary –Based Approach to Multilingual Information Retrieval. In *Proceedings of the 19th Annual International ACM Sigir*, 49-57. Zurich, Switzerland.
- Pirkola, A., H. Hedlund, T. Keskustaloh and K. Jarvelin. 2001. Dictionary – Based Cross-language Information Retrieval: Problems, Methods and Research Findings. *Information Retrieval* 4C3/4: 209-230.
- Porter, M. F. 1980. An algorithm for suffix stripping. *Program* 14: 130-137.
- Qu, Yan, and Gregory Grefenstette. 2003. *Automatic transliteration for Japanese-to-English text retrieval*. Presented at SIGIR, Canada, Toronto. <http://dblp.uni-trier.de/db/conf/sigir/sigir2003.html#QuGE03> (accessed August 23, 2009).
- Voorhees, Ellen M. 1998. Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness. In C. J. Van rijs Bergen, W. Bruce Croft and Alistair Moffat, *Proceedings of the 21st Annual International ACM Sigir Conference on Research and Development in Information Retrieval*, 315-323. ACM Press.
- Voorhees, Ellen M. 2003. Overview of TREC2002. <http://trec.nist.gov/pubs/trec11/papers/OVERVIEW.11.pdf> (accessed Jan. 21, 2008)
- Wang, Jianqiang, and Douglas W. Oard. 2005. *Document and Query Expansion Using Side Collections and Thesauri*. Presented at the CLEF 2005, Vienna, Austria. <http://www.informatik.uni-trier.de/~ley/db/conf/clef/clef2005.html#WangO05> (accessed August 23, 2009).