

نگاشت ساختار و چیدمان تاریخی مفاهیم علم اطلاعات و دانش‌شناسی: با رویکرد متن‌کاوی (۲۰۰۴-۲۰۱۳)

محمد حسن زاده^{*۱}

فاطمه زندیان^۲

سیده سمیه احمدی^۳

چکیده

هدف: پژوهش حاضر با هدف نگاشت و بررسی ساختار مفهومی علم اطلاعات و دانش‌شناسی، دامنه موضوعی مجلات و چیدمان تاریخی مفاهیم بر اساس متن کامل مقالات مجلات هسته صورت گرفت.

روش‌شناسی: روش به‌کاررفته تحلیل محتوای مفاهیم استخراج‌شده بر اساس فنون متن‌کاوی از ۶۸۳۰ مقاله ۱۰ نشریه هسته بر اساس گزارش جی سی آر ۲۰۱۳ (۲۰۰۴-۲۰۱۳) است. خوشه‌بندی واژگان و عبارات بعد از استخراج به‌صورت پارامتریک و انتخاب ۱۵۰ مفهوم بر اساس وزن tf-idf، با استفاده از ماتریس هم‌رخدادی واژگان بر اساس شاخص کسینوس تتا و قرابت مرتبه ثانویه و الگوریتم average-linkage، در نرم‌افزار Provalis Research، انجام شد. همچنین از فراوانی نسبی مفاهیم منتخب جهت تحلیل تناظر، به‌منظور خوشه‌بندی مجلات و دامنه موضوعی آنها و بررسی چیدمان تاریخی مفاهیم استفاده شد.

یافته‌ها: یافته‌های حاصل ۳ خوشه بالغ، ۱ خوشه نیمه‌بالغ و ۳ خوشه نابالغ در تحلیل هم‌واژگانی و چهار خوشه موضوعی در مجلات و دو خوشه زمانی با سیر حرکتی ساعت‌گرد را نشان داد.

نتیجه‌گیری: به‌طور کلی بررسی مفهومی خوشه‌ها نشان داد تولیدات علمی دارای برخی از موضوعات ثابت بوده و نیز موضوعات جدید که تحت تأثیر تغییر پیشرفت علم و فناوری ایجاد شده است. چنانچه این تغییر در چشم‌انداز شناختی در خوشه زمانی نخست مبتنی بر ظهور اینترنت بوده و در خوشه دوم به دلیل استفاده از شاخص‌هاست.

واژگان کلیدی: خوشه‌بندی مفاهیم، علم اطلاعات و دانش‌شناسی، متن‌کاوی، تحلیل محتوا.

۱. دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه تربیت مدرس (نویسنده مسئول)
Email: hasanzadeh@modares.ac.ir
۲. استادیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه تربیت مدرس
۳. دانشجوی کارشناسی ارشد علم اطلاعات و دانش‌شناسی، دانشگاه تربیت مدرس

دریافت: ۱۳۹۶/۱۰/۹

پذیرش: ۱۳۹۷/۳/۵

مقدمه و بیان مسئله

مقدار داده‌هایی که در سطح جهان تولید و ذخیره می‌شود، نه تنها غیرقابل تصور است بلکه همچنان به رشد خود ادامه می‌دهد. مجموعه داده‌هایی که با ویژگی‌هایی مثل حجم بالا، سرعت، تنوع، جامعیت، شاخص‌پذیری و انعطاف مشخص می‌شود و به‌عنوان سرمایه‌های اصلی در حرکت علمی، اقتصادی، اجتماعی و فرهنگی جوامع، سازمان‌ها و شرکت‌های مختلف و توسعه مشارکت افراد در جهان مطرح بوده و با وجود چالش‌های موجود در بهره‌برداری به فرصت‌های رقابتی تبدیل شده‌اند. با این ویژگی‌ها، توجه به علم داده با تنوع داده‌ها روزبه‌روز افزایش یافته و توجه حوزه‌ها و علوم مختلف را در بهره‌برداری از مزیت آن به‌ویژه در مباحث شناختی جلب نموده است. علم اطلاعات و دانش‌شناسی به‌عنوان یکی از حوزه‌هایی که با داده سروکار دارد و در توسعه روش‌های بهره‌برداری به‌خصوص در حوزه نگاشت علم سهیم است.

آنچه که مسئله این تحقیق بوده و در پی کشف الگوی حاکم بر درخت علم اطلاعات و دانش‌شناسی، زیرحوزه‌های تحقیقی بارز و همچنین روابط بین آنهاست. در این تحقیق به‌عنوان یکی از چالش‌های مطرح در تحقیقات پیشین علاوه بر نگاشت ساختار معرفتی حوزه در سطح جهان سعی شده تا به یک نمونه قابل مقایسه با ساختار معرفتی حوزه در ایران دست بیابیم. تحلیل هم‌رخدادی کلمات روش مناسبی برای مشخص کردن الگو و "پویایی حوزه‌های علمی" نویونز^۱، مود^۲ و لول^۳، ۱۹۹۹؛ برونر^۴، چن^۵ و بویاک^۶، ۲۰۰۳) است، که بر زبان طبیعی استوار است. این تحلیل با تکنیک متن‌کاوی روی عنوان، چکیده و متن کامل بروندهای علمی قابل اجراست، البته اجرای آن روی متن کامل با محدودیت‌های سخت‌افزاری و نرم‌افزاری بر حسب داده‌های مورد تحلیل روبه‌رو است. با توجه به نگاشت‌های مختلف بر اساس مجلات و متون استخراجی از پایگاه‌ها در این حوزه، این پژوهش به دنبال اجرای این تکنیک روی متن کامل مقالات نشریات هسته در طی ۱۰ سال اخیر در سطح بین‌المللی است. با ترسیم این نگاشت، الگوی حاکم بر این رشته در سطح بین‌المللی نمایان شده و می‌توان وضعیت کشور را با سطح جهانی مقایسه نمود. همچنین با افزایش انتشارات علمی این حوزه، نسبت به شناسایی گستره موضوعی، تعیین مفاهیم و مباحث پایه و رتبه‌بندی مفاهیم به برنامه‌ریزی مناسب در سطوح مختلف اقدام کرد.

سؤال‌های پژوهش

این پژوهش در کنار دستیابی ساختار و الگوی حاکم در حوزه علم اطلاعات، به بررسی چیدمان تاریخی مفاهیم و خوشه‌بندی مجلات بر اساس مفاهیم می‌پردازد و در پی پاسخ‌گویی به سؤالات زیر است.

۱. مفاهیم کلیدی علم اطلاعات و دانش‌شناسی در نشریات هسته کدام‌اند؟
۲. بر اساس واژگان پایه علم اطلاعات و دانش‌شناسی چه خوشه‌های موضوعی قابل شناسایی هستند؟
۳. دامنه شناختی مجلات هسته بر اساس مفاهیم منتخب چگونه است؟
۴. نحوه چیدمان مفاهیم طی سال‌های (۲۰۱۳-۲۰۰۴) چگونه بوده است؟

1. Noyons
2. Moed
3. Luwel
4. Börner
5. Chen
6. Boyack

پیشینه پژوهش

پیشینه پژوهش در داخل

بررسی پیشینه پژوهش در این حوزه به دلیل ارتباط قوی بین زمینه‌های مختلف تحقیق و فناوری پیچیده است. ارتباط این حوزه با حوزه‌هایی چون زبانشناسی و فرهنگ‌نویسی، متن‌کاوی، مصورسازی، زبان‌شناسی محاسباتی، علوم اجتماعی خاستگاه‌های مختلفی را برای این حوزه رقم می‌زند. در ایران استفاده از تحلیل هم‌رخدادی در کنار مطالعه ساختار حوزه‌های مختلف از جمله حوزه فناوری نانو (محمدی، ۱۳۸۸؛ تیمورپور، سپهری و پزشکی، ۱۳۸۸)، در گرایش‌های حوزه علم اطلاعات و دانش‌شناسی نیز مورد استفاده قرار گرفته است. از این مطالعات می‌توان به مطالعه (موسوی‌زاده، باقری و کربلا آقایی کامران، ۱۳۹۳) در حوزه سازمان‌دهی اطلاعات، مدیریت دانش (کاتبی، ۱۳۹۳؛ حاضری، توکلی زاده راوری و ابراهیمی، ۱۳۹۴)، علم‌سنجی (احمدی و عصاره، ۱۳۹۴)، مدیریت اطلاعات (خادمی و حیدری، ۱۳۹۵) صورت گرفته است.

اما مرتبط‌ترین مطالعات در این حوزه در ایران، شناسایی ساختار واژگان چکیده مقالات "علم اطلاعات و دانش‌شناسی" با استفاده از نگاشت‌های علمی هم‌واژگانی در پایگاه و بگاه علم (۲۰۰۹-۲۰۱۳) است. یافته‌ها نشان می‌دهند واژگان «اطلاعات»، «وب»، «پژوهش»، «تحلیل استنادی»، «دانش»، «کتابخانه»، «مجلات»، و «فناوری» محور اصلی مفاهیم مورد مطالعه در این رشته را تشکیل داده‌اند. همچنین، مفاهیم مورد مطالعه در ۱۳ خوشه قرار گرفته‌اند که محورهای اصلی مطالعات شامل «آموزش و یادگیری؛ سواد اطلاعاتی»، «سازمان‌دهی اطلاعات و دانش»، «منابع اطلاعاتی تحت وب و شبکه‌های اجتماعی»، «اخلاق حرفه‌ای در علم اطلاعات»، «انفورماتیک و خدمات اطلاعات سلامت»، «مدیریت اطلاعات؛ نظام‌های اطلاعاتی؛ مدیریت دانش و نوآوری»، و «مطالعات و شاخص‌های علم‌سنجی و اطلاع‌سنجی» است (مصطفوی، عصاره و توکلی‌زاده راوری، زیر چاپ).

پیشینه پژوهش در خارج

سراغاز استفاده از متن کامل در تهیه نگاشت‌های علمی مبتنی بر هم‌واژگانی برای اولین بار (گلنسون^۱، گلنزل^۲، پرسون^۳، ۲۰۰۵) روش‌های کتابشناختی سنتی را جهت بهینه‌سازی نگاشت مقالات علمی از منظر موضوع خاص علم‌سنجی با مطالعات تحلیل تمام‌متن ترکیب نمودند. داده‌ها شامل تحلیل لغوی متن کامل مقالات و ویژه‌نامه علم‌سنجی متشکل از مقالات ارائه‌شده در نهمین کنفرانس بین‌المللی علم‌سنجی و اطلاع‌سنجی^۴ سال ۲۰۰۳ در چین بود. نتایج حاکی از آن بود که پروفایل وزن‌دهی شده متون، توصیفگر خوبی برای موضوع یک مقاله است. از دیگر اهداف این پژوهش مقایسه قدرت بین تحلیل متن کامل و عنوان-چکیده در رسیدن به محتوای متون بود، که استفاده از متن‌کاوی در ترسیم ساختار فکری حوزه تحقیقاتی تأیید و چون ۲۱ درصد از تمام مقالات در مواقع استفاده از عناوین و چکیده‌ها به‌درستی اختصاص داده نمی‌شوند، استفاده از متن کامل در صورت امکان ارجح‌تر است. اما به دلیل محدود بودن مجموعه متون انتخابی، ساختار روشنی از حوزه ارائه نشد و این مسئله مطرح گردید آیا در متون بیشتر نیز چنین نتیجه‌ای حاصل می‌شود؟ لذا طیف وسیع‌تر و ناهمگن‌تر از متون این حوزه یعنی تمام متون منتشره شده

1 . Glenisson
2 . Glänzel
3 . Persson

4 . 9th International Conference on Scientometrics and Informetrics

نگاشت ساختار و چیدمان تاریخی مفاهیم علم اطلاعات و دانش‌شناسی: با رویکرد متن‌کاوی (۲۰۰۴-۲۰۱۳)

در سال ۲۰۰۳ شماره ۵۶-۵۸ مجله علم‌سنجی شامل ۸۵ مقاله برای مطالعه بعدی انتخاب شد. نتایج در پاسخ به میزان اعتبار استفاده از روش ترکیبی در ارائه ساختار شناختی حوزه به‌عنوان روش مناسب مورد تأیید واقع شد. در پاسخ به هدف اصلی پژوهش ساختار فکری این حوزه برخلاف مطالعه قبلی طیف کامل و ناهمگن‌تری از فعالیت‌های پژوهشی در علم‌سنجی، اطلاع‌سنجی، فناوری‌سنجی^۱ و همچنین موضوعات فراتر از جریان اصلی را نمایان ساخت (گلنسون، گلنزل، جانسنز^۲ و دی مور^۳، ۲۰۰۵).

بار دیگر (جانسنز، لتا^۴، گلنزل و دی مور، ۲۰۰۶) نه با هدف ارائه یک روش‌شناسی بلکه با هدف بسط روش ارائه‌شده بر روی متون ناهمگن علم کتابداری و اطلاع‌رسانی به‌عنوان یک رشته وسیع مورد مطالعه قرار دادند. لذا ۹۳۸ مقاله و یادداشت تمام‌متن منتشرشده بین سال‌های ۲۰۰۲ تا ۲۰۰۴ در پنج نشریه "مدیریت و پردازش اطلاعات"^۵، "مجله انجمن علم و فناوری اطلاعات آمریکا"^۶، "مجله دکوماناسیون"^۷، "مجله علوم اطلاعات"^۸ و "علم‌سنجی"^۹ با محوریت موضوعی روش‌شناختی، مطالعات نظری و رویکردهای کمی رشته انتخاب شدند. نتایج، دو خوشه در کتاب‌سنجی که خوشه بزرگ با کتاب‌سنجی کاربردی یا ارزیابی تحقیقات و کوچک‌تر با موضوعات روش‌شناختی و نظری، خوشه بزرگ دیگر در بازیابی اطلاعات و موضوعات عمومی و گوناگون و درنهایت دو خوشه کوچک نوظهور در وب‌سنجی و مطالعات ثبت اختراعات و فناوری را نشان داد. البته در خوشه بازیابی اطلاعات خوشه کوچک و موقت نیز دیده شده که ناشی از ویژه‌نامه مجله JASIST در این موضوع بود. مطابق انتظارات بازیابی اطلاعات، موضوعات عمومی و وب‌سنجی موضوعات مورد بحث چهار مجله JASIST، JIS، JPM، JDoc بودند؛ درحالی‌که کتاب‌سنجی و ثبت اختراعات موضوع اصلی مجله Sci بود. نتایج حاصل از مطالعات نشان دادند روش خوشه‌بندی ترکیبی متشکل از محتوای متنی و اطلاعات کتاب‌شناختی، عملکرد بهتری نسبت به استفاده از یکی از این روش‌ها دارند. لذا مطالعات جهت توسعه این نوع روش‌ها ادامه یافت و این بار خوشه‌بندی ترکیبی بر اساس معکوس‌کای دو فیشر^{۱۰} جهت ترکیب متن کامل با اسنادها در نگاشت علم اطلاعات مورد استفاده قرار گرفت. در مقایسه با نتایج مطالعه یک بهینه‌سازی کاملاً واضح در این روش دیده شد، اما باز اشتباهاتی در تخصیص مدارک به خوشه‌ها دیده می‌شد (جانسنز، گلنزل و دی مور، ۲۰۰۸).

از دیگر مطالعات می‌توان به مطالعه ساختار حوزه با استفاده از واژگان عناوین ۱۰۳۴۴ مقاله منتشرشده بین سال‌های ۱۹۸۸ و ۲۰۰۷ در ۱۶ مجله این حوزه اشاره نمود. نتایج سه شاخه اصلی شامل دو شاخه قدیمی مرتبط با کتابخانه و اطلاعات و شاخه کاملاً جدای علم‌سنجی و کتاب‌سنجی را نشان دادند. همچنین در نتایج مطالعه پویایی رشته، تکامل مداوم مفاهیم از سال ۱۹۹۲ و توسعه سریع مفاهیم بین سال‌های ۱۹۹۸ و ۲۰۰۱ تحت تأثیر رشد اینترنت دیده می‌شود. به طور کلی تغییر در چشم‌انداز شناختی با ظهور فناوری‌های جدید و منسوخ شدن فناوری‌های قدیمی

1. Technometric
2. Janssens
3. De Moor
4. Leta
5. Information Processing & Management (IPM)
6. Journal of the American Society for Information Science and Technology (JASIST)
7. Journal of Documentation (JDoc)
8. Journal of Information Science (JIS)
9. Scientometrics (SciMetr)
10. Fisher's inverse chi-square

هدایت می‌شود (میلوژویک^۱، ساگیموتو^۲، یان^۳ و دینگ^۴، ۲۰۱۱).

در سال ۲۰۱۳ تحلیل هم‌واژگانی کتابداری و علم اطلاع‌رسانی در کشور چین با هدف نمایاندن ساختار فکری این رشته در طی سال‌های ۲۰۰۸-۲۰۱۲ استفاده گردید. نتایج حاصل نشان دادند که این حوزه در چین: (۱) اغلب دارای موضوعات پژوهشی ثابت و توسعه‌یافته هستند؛ (۲) مقدار کمی از موضوعات نوظهور مثل خدمات اطلاع‌رسانی، مدیریت دانش و اشتراک اطلاعات دارای پتانسیل زیاد برای توسعه هستند و (۳) بیشتر موضوعات حاشیه‌ای و رشدنیافته هستند (هو^۵، هو^۶، دنگ^۷ و لیو^۸، ۲۰۱۳). هم‌زمان مطالعه‌ای مشابه روی رساله‌های دکتری در این رشته در این کشور طی سال‌های ۱۹۹۴ تا ۲۰۱۱ صورت گرفت. نتایج تنوع موضوعی در زمینه‌های تحقیقاتی با تمرکز بر موضوعات منابع اطلاعاتی، هستی‌شناسی، وب معنایی، جستجوی معنایی، دولت الکترونیک، مدیریت منابع اطلاعاتی، مدیریت دانش، نوآوری دانش، اشتراک دانش، سازمان‌دهی دانش، شبکه، خدمات اطلاعاتی، نیازهای اطلاعاتی و کتابخانه‌های دیجیتال نشان داد (زونگ^۹ و همکاران، ۲۰۱۳).

استفاده از تحلیل هم‌رخدادی به منظور توسعه روش‌ها این بار با محاسبه مقادیر اصطلاحات موضوعی بین‌رشته‌ای^۹ همراه با مقادیر شرطی، بسامد واژگان انجام گرفت. نتایج علاوه بر تأیید روش، حوزه‌های مهم بین‌رشته‌ای در آینده را به طور عمده شامل فناوری اطلاعات، سیستم‌های اطلاعاتی، پذیرش کاربران، شبکه‌های اجتماعی، اشتراک دانش و رسانه‌های اجتماعی دانست (زو^{۱۰}، گو^{۱۱}، یو^{۱۲}، رو^{۱۳} و فنگ^{۱۴}، ۲۰۱۶).

مطالعه پیشینه‌ها هر چند اهمیت نگاشت ساختار رشته‌ها بر اساس متن کامل را تأیید می‌کند اما به دلیل موانعی مثل دسترسی به متن کامل مقالات در سطح وسیع، ساختارمند نبودن متون و نبود پایگاه منسجم جهت استخراج ماشینی کمتر مورد توجه بوده است. لذا هدف از این پژوهش رسیدن به ساختار مفهومی حوزه بر اساس متون بیشتر به‌ویژه متون هسته و مطالعه روند گسترش این مفاهیم طی سال‌های مختلف و دستیابی به مفاهیم هسته بر اساس متون هسته در سطح بین‌المللی جهت مقایسه با مفاهیم داخلی در آینده است.

روش‌شناسی پژوهش

در این پژوهش جهت نگاشت علمی حوزه علم اطلاعات و دانش‌شناسی با هدف کشف ساختار شناختی این رشته از روش هم‌واژگانی استفاده شده است. این روش، تکنیک تحلیل محتواست که در مطالعات علم‌سنجی و برنامه‌های تحقیقی و توسعه کاربرد دارد. بدین صورت که از الگوی هم‌رخدادی اقلام (مثل: واژگان/عبارات یعنی مفاهیم) در یک مجموعه از متون به‌منظور شناسایی روابط بین انگاره‌های موجود در حوزه‌های موضوعی ارائه‌شده به شکل متون

1. Milojević
2. Sugimoto
3. Yan
4. Ding
5. Hu
6. Deng
7. Liu
8. Zong
9. Topic Terms Interdisciplinarity
10. Xu
11. Guo
12. Yue
13. Ru
14. Fang

نگاشت ساختار و چیدمان تاریخی مفاهیم علم اطلاعات و دانش‌شناسی: با رویکرد متن‌کاوی (۲۰۰۴-۲۰۱۳)

استفاده می‌شود. برای انجام مراحل مختلف تحقیق از بسته آزمایشی Provalis Research^۱، نرم‌افزارهای Word و Excel، Adobe Acrobat XI 11^۲، ABBYY FineReader 12^۳ و Internet Download Manager^۴ استفاده شده است که در ذیل به این مراحل می‌پردازیم:

۱. گردآوری متون

الف) تهیه و مرتب‌سازی متون: در این مرحله داده‌ها یعنی متون نشریات در سطح بین‌الملل از طریق پایگاه‌های اطلاعاتی مختلف گردآوری و در فایل‌هایی به تفکیک عناوین مجلات، سال نشر و شماره مجله جداگانه ذخیره شد. مطابق جدول ۱ به‌طور کلی بعد از گردآوری منابع و گزینش مقالات با حذف متونی از قبیل مقدمه‌ها، سرمقاله‌ها، نقدها، نمایه‌ها، گزارش‌ها، نامه‌ها و معرفی‌ها تعداد ۶۸۳۰ مقاله جهت متن‌کاوی انتخاب شد. در بین منابع مجله Scientometrics با ۲۰۲۵ درصد و مجله Information and Organization با ۹۰۱ درصد به‌ترتیب بیشترین و کمترین تعداد متون را دارا می‌باشند.

جدول ۱ توزیع فراوانی تعداد مقالات در مجلات بین‌الملل

عنوان	مخفف	تعداد مقالات (۲۰۰۴-۲۰۱۳) درصد
Information and Organization	I&O	۱۳۰ ۱.۹
Information Systems Research	ISR	۳۷۵ ۵.۵
International Journal of Information Management	IJIM	۵۰۷ ۷.۴
Journal of Information Technology	JIT	۲۴۶ ۳.۶
Journal of Informetrics	JoI	۳۸۸ ۵.۷
The Journal of Strategic Information Systems	JSIS	۱۸۶ ۲.۷
Journal of the American Medical Informatics Association	JAMIA	۱۱۷۹ ۱۷.۳
Journal of the Association for Information Science and Technology (Journal of the American Society for Information Science and Technology)	JASIST	۱۷۰۰ ۲۴.۹
MIS Quarterly	MISQ	۳۹۵ ۵.۸
Scientometrics	Sci	۱۷۲۴ ۲۵.۲
جمع		۶۸۳۰ ۱۰۰

ب) پاک‌سازی متون: استخراج واژگان از متن کامل به جای عناوین و چکیده‌ها، مشکلات و پیچیدگی‌های مضاعفی دارد به‌طوری‌که مستلزم موارد پیش‌پردازش خاص بوده و به‌تبع آن زمان‌بر نیز است. یکی از این موارد حذف بخش منابع، قدردانی، یادداشت‌ها، ضمائم و عبارات و واژگان قالبی از مقالات می‌باشد. هم‌زمان با این کار تصحیح مقالات دارای غلط‌نامه و متون تصویری استخراج‌شده با سی‌آر صورت گرفت.

ج) واردکردن داده‌های متنی در نرم‌افزار: داده‌های متنی ذخیره‌شده در فایل‌ها بعد از آماده‌سازی در قالب یک

- 1 . <http://provalisresearch.com/>
- 2 . <http://www.adobe.com/>
- 3 . <http://www.abbyy.com/finereader/>
- 4 . <https://www.internetdownloadmanager.com/>

پیکره^۱ وارد نرم افزار کبیر دی ای ماینر گردید و هریک از متون به صورت یک کیس^۲ و فایل ها با عناوین مجلات در کنار ۵ دوره (۲۰۰۴-۲۰۰۵، ۲۰۰۶-۲۰۰۷، ۲۰۰۸-۲۰۰۹، ۲۰۱۰-۲۰۱۱ و ۲۰۱۲-۲۰۱۳) به عنوان متغیر تعریف شدند.

۲. پیش پردازش متون

در تحلیل هم واژگانی، واژه ها و عبارات پایه اصلی تحقیق محسوب می شوند، که بسته به هدف و روش تحقیق نوع گردآوری واژه ها نیز متفاوت است. همان طور که در بررسی پیشینه روش شناسی پژوهش دیدیم جهت استخراج واحدهای تحلیل هم واژگانی از مقالات مجلات، کنفرانس ها، گزارش ها و حتی فصل های کتاب از دو روش پارامتریک و ناپارامتریک به عنوان روش های اساسی استفاده شده بود. در این پژوهش از روش پارامتریک جهت استخراج واژگان و عبارات با استفاده از نرم افزار ورد استیت در مراحل ذیل صورت گرفت.

الف) حذف کلمات بازدارنده^۳: جهت حذف واژه هایی که دارای ارزش معنایی کمتری هستند، از قبیل: ضمائر، حروف ربط و غیره مثل: "about"، "near"، "and" و غیره از لیست بازدارنده مختص زبان انگلیسی نرم افزار ورد استیت^۴ با تغییراتی جهت بازیابی عبارات استفاده شد.

ب) حذف آدرس های اینترنتی و برخی اسامی خاص: ویژگی اصلی آدرس های اینترنتی و استنادهای درون متنی، هایپر لینک یا لینک دار بودن آنهاست، این ویژگی توسط نرم افزار قابل شناسایی بوده و در متن کاوی حذف می شود. اما در اغلب متون این ویژگی رعایت نشده؛ بنابراین از ویژگی این نوع کلمات مثل داشتن "." استفاده گردید. بدین ترتیب با تعریف این محدودیت در استخراج، واژگان دارای نقطه به صورت یکجا از طریق نرم افزار ورد استیت، بازیابی و سپس از طریق مرور آدرس هایی دارای www یا برخی ویژگی ها مثل دامنه com. شناسایی و در لیست واژگان بازدارنده قرار داده شد.

ج) استخراج عبارات: برای ارائه مفاهیم موجود در متون استفاده از واژگان به تنهایی مناسب نبوده و نیاز به استخراج عبارات و به ویژه لغات ترکیبی مثل "Co-author" می باشد. ورد استیت این امکان را ایجاد کرده تا عبارات متداول در متون استخراج گردد ولی نمی توان آنها را به طور معمول با لیست واژگان ترکیب کرد. برای استخراج هم زمان آنها با واژگان از "X" به جای فاصله یا "-" استفاده شد بدین صورت که با توجه به وزن آستانه واژگان عباراتی مثل "LIBRARY AND INFORMATION" با "LIBRARYXANDXINFORMATION" جایگزین و استخراج گردید.

د) استاندارد سازی لغات: از اساسی ترین کارکردهای استاندار سازی لغات، می توان به یکدستی واژه ها و در نتیجه کاهش حجم واژگان یعنی کاهش افزونگی در لیست واژگان اشاره کرد. در این پژوهش از روش جایگزینی، الگوریتم Lemmatization اجرا شده در ورد استیت استفاده شد که مبتنی بر روش فرهنگ نامه ای تعدیل شده است. بدین صورت که قسمتی مبتنی بر الگوریتم جایگزینی پسوند کروتز^۵ یعنی KSTEM می باشد. همچنین به دلیل استفاده از یک لیست جایگزین به جای برچسب زنی اجزای کلامی واژه ها، در برخی موارد استثنایی از نظر معنایی نامعتبر است، اما افزایش سرعت پردازش و قابلیت بازیابی و تغییر در لیست جایگزین برای متن کاو را فراهم می کند.

1. Corpus
2. Case
3. Stop Words
4. WordStat
5. Krovetz

۳. انتخاب مفاهیم

انتخاب مفاهیم بر پایه روش‌های متنوعی صورت می‌گیرد، از جمله این روش‌ها استفاده از ایده لوهن^۱ (لوهن، ۱۹۵۸)، استفاده از نمودار توزیع فراوانی یا درصد فراوانی مقالات مانند مطالعه (میلوژویک، ساگیموتو، یان و دینگ، ۲۰۱۱)، تحلیل معنایی پنهان (جانسنز و همکاران، ۲۰۰۶)؛ گروه‌بندی و پالایش لغات با استفاده از یک اصطلاح‌نامه، سرعنوان موضوعی، خبرگان و یا استفاده از مفاهیم پایگاه‌های اطلاعاتی (دینگ، چودبری^۲ و فو^۳، ۲۰۰۱؛ لی^۴ و جونگ^۵، ۲۰۰۸؛ محمدی، ۲۰۱۲؛ هو و همکاران، ۲۰۱۳؛ ناصری‌جزه، طباطباییان و فاتح‌راد، ۱۳۹۱) می‌باشد.

در این پژوهش، بنا بر تعداد زیاد لغات به‌ویژه در تحلیل متون تمام‌متن و نبود امکان بازیابی جهت گروه‌بندی ذیل یک واژه، تمایل به انتخاب واژگان خاص در مجموعه مدارک بر حسب ویژگی IDF از روش وزن‌دهی TF-IDF برای ۱۵۰ واژه منتخب استفاده شد. این روش که عمومیت بیشتری دارد خصوصیت وزن‌دهی فراوانی معکوس متن را با فراوانی واژه ترکیب می‌کند، که توسط سالتون^۶ و یانگ^۷ مطرح شده است؛ بنابراین هر چقدر رخداد یک واژه در یک یک مدرک بیشتر باشد، محتوای متن را بهتر نشان می‌دهد. ولی هرچه تعداد مدارکی که واژه در آنها رخ می‌دهد بیشتر باشد، از شاخص بودن آن کاسته می‌شود (بایز-یتس^۸ و ریبریو-نتو^۹، ۲۰۱۱، ص، ۶۸-۷۳).

$$w_{i,j} = tf_{i,j} \times \log \frac{N}{n_i}$$

۴. ماتریس هم‌رخدادی

بعد از انتخاب مفاهیم نوبت به تشکیل ماتریس واژه-واژه است. در این ماتریس مفاهیم ستون‌ها در ردیف‌ها تکرار می‌شود و اعداد حاصل از این رخداد به نوع قرابت مفاهیم در مدارک وابسته است. به این معنی که ماتریس مفاهیم بر اساس چه نوع هم‌رخدادی تشکیل شوند. در واقع دو نوع خویشاوندی وجود دارد، یکی بر اساس قرابت مرتبه اولیه^{۱۰} که کلمات را بر اساس قرار گرفتن کنار هم در یک مدرک و دیگری قرابت مرتبه ثانویه^{۱۱} که مفاهیم بر اساس نزدیکی دو واژه که الزاماً نه به دلیل هم‌رخدادی کنار هم بلکه هم‌رخدادی در یک محیط خوشه‌بندی می‌شوند. از مزایای این روش توانایی آن در گروه‌بندی واژگان است که مترادف و یا جایگزین هم هستند. به‌عنوان مثال زمانی که TUMOR و TUMOUR اغلب و یا اصلاً با هم رخ نمی‌دهند این روش آنها را نزدیک به هم دانسته چراکه هر دوی آنها با کلماتی چون BRAIN و CANCER رخ می‌دهند. جالب‌تر گروه‌بندی مترادف‌های نزدیک و کلمات مرتبط از نظر معنایی^{۱۲} است مثل MILK، JUICE و WINE؛ چراکه آنها گرایش به ترکیب‌شدن با افعال مشابهی مثل DRINK یا POUR و اسامی مثل GLASS دارند (گرفنستته^{۱۳}، ۱۹۹۴).

1. Luhn
2. Chowdhury
3. Foo
4. Lee
5. Jeong
6. Salton
7. Yang
8. Baeza-Yates
9. Ribeiro-Neto
10. First Order Clustering
11. Second Order Clustering
12. Semantically
13. Grefenstette

ماتریس 150×150 از هم‌رخدادی مفاهیم در واقع به‌عنوان داده‌های ورودی برای نگاشت مفاهیم می‌باشد، اما برای تبدیل شدن به نقشه باید با استفاده از سنج‌های مختلف نرمال‌سازی شده و به ماتریس شباهت تبدیل شود. بدین منظور از ضریب کسینوس تتا^۱ استفاده گردید. این ضریب کسینوس زاویه بین دو بردار از مقادیر را اندازه‌گیری می‌کند، که محدوده بین -۱ و +۱ است و نه تنها حضور یک واژه در یک مدرک بلکه میزان حضور واژه را نیز در مدرک محاسبه می‌کند (پرووالیس ریسرچ^۲، ۲۰۱۵) و به‌صورت ذیل محاسبه می‌شود.

$$SIMC(\vec{t}_a, \vec{t}_b) = \frac{\vec{t}_a \cdot \vec{t}_b}{|\vec{t}_a| \times |\vec{t}_b|}$$

۵. مصورسازی و تحلیل داده‌ها

به‌منظور رسیدن به نمای کلی از نگاشت مفهومی^۳ همانند مطالعات هم‌واژگانی، از تحلیل آماری چندمتغیره^۴، تحلیل تحلیل شبکه‌ای^۵ و نقشه‌های حرارتی استفاده گردید. تحلیل آماری چندمتغیره اشاره به روش‌های آماری در تحلیل داده‌هایی بیش از یک متغیر دارد. در واقع تحلیل چندمتغیره شامل مجموعه‌ای از روش‌ها در به‌کارگیری چندین سنج بر روی هر شیء یا فرد در یک یا چند نمونه است. سنج‌ها به‌عنوان متغیر و اشیاء و افراد به‌عنوان اجزای تحقیق، نمونه و یا آزمایش هستند (رنجر^۶، ۲۰۰۳). روش تحلیل چندمتغیره اغلب به‌عنوان یک معیار یا رویکرد اکتشافی کاربرد کاربرد دارد تا روشی برای آزمون فرضیات واقعی (ساساکی^۷ و سوزوکی^۸، ۱۹۸۹). روش‌های تحلیل چندمتغیره متفاوتی در تحلیل داده‌هایی با حجم زیاد وجود دارد، اما مهم‌ترین این تحلیل‌ها در مطالعات هم‌واژگانی با هدف مطالعه ساختار مفاهیم یک رشته، تحلیل خوشه‌ای^۹ و مقیاس‌بندی چندبعدی و تحلیل تناظر^{۱۰} بوده و به‌صورت دندوگرام‌ها، نقشه‌های مختلف مقیاس‌بندی سه‌بعدی و حرارتی مصورسازی شده و مورد تحلیل قرار می‌گیرند.

یافته‌های پژوهش

۱. مفاهیم کلیدی علم اطلاعات و دانش‌شناسی در نشریات هسته کدام‌اند؟

مطابق شکل ۱، لغات و عبارات احصاشده بر حسب TF-IDF مرتب و بسامد لغات بر حسب فونت و رنگ مشخص شده است. چنانچه واژه System دارای بیشترین و Bid دارای کمترین بسامد در بین مفاهیم منتخب می‌باشد.

1. Cosine Theta
2. Provalis Research
3. Concept Maps
4. Multivariate Analysis Methods
5. Link Analysis
6. Rencher
7. Sasaki
8. Suzuki
9. Clustering
10. Correspondence Analysis

CITATION JOURNAL PATENT FIRM PATIENT QUERY USER CLUSTER
PUBLICATION NETWORK CUSTOMER COUNTRY SEARCH SCIENCE CITE INDEX BUSINESS
ARTICLE PROJECT CLINICAL ORGANIZATIONAL DOCUMENT UNIVERSITY SERVICE PARTICIPANT
CONSUMER TASK **MODEL** HEALTH TEAM TRUST NODE COLLABORATION SOCIAL MARKET INNOVATION
ORGANIZATION AUTHOR VARIABLE TAG **KNOWLEDGE** SCIENTIFIC INDEX PRODUCT PAGE
MEDICATION MANAGEMENT IMAGE COST LIBRARY PHYSICIAN **TECHNOLOGY** INDUSTRY COMPANY PRICE THEORY
SCORE RANKING **WEB** PRIVACY ITEM INDICATOR ONLINE WEBSITE SCIENTIST LEARNING TOPIC **PERFORMANCE** DRUG
MANAGER CLIENT ADOPTION VENDOR PERCEIVE ALERT ALGORITHM SECURITY INTERNET HOSPITAL STUDENT RISK COMMUNITY
SYSTEM EMPLOYEE CHINA OUTSOURCING SITE DIGITAL EHR ERP RANK PROVIDER RETRIEVAL TEXT BEHAVIOR CONCEPT
PAPER SELLER IMPLEMENTATION SOFTWARE INTENTION CHINESE GOVERNMENT AUCTION DOMAIN STRATEGY SUPPLIER
MEMBER SEMANTIC CONTENT MEDICAL CODE INTERVIEW HYPOTHESIS MAP WORD DISCIPLINE CAPABILITY CARE MOBILE ACTION
ONTOLOGY RESPONDENT **FEATURE** CATEGORY KEYWORD IMPACTFACTOR CLASS WEIGHT **RESOURCE** MESSAGE SENTENCE CONSTRUCT
CORRELATION OBJECT HEALTHCARE STRATEGIC REFERENCE INTERACTION INSTITUTION **SHARE** PUBLISH COMMUNICATION
DECISION BID **RELATIONSHIP** CONTROL COUNT SIMILARITY BUYER

شکل ۱. نقشه ابری واژگان پایه متون تمام‌متن بین‌الملل

۴. بر اساس واژگان پایه علم اطلاعات و دانش‌شناسی چه خوشه‌های موضوعی قابل شناسایی است؟

بر اساس تحلیل خوشه‌ای و قرابت درجه دو مفاهیم، دندوگرام حاصل از مفاهیم انتخابی در متون تمام‌متن بین‌الملل، بنا به نظر خبرگان به هفت خوشه اصلی تقسیم گردید. مطابق شکل ۲ این هفت خوشه شامل ۳ خوشه بالغ با مجموع ۱۵ زیرخوشه، ۱ خوشه نیمه‌بالغ و ۳ خوشه نابالغ می‌باشد. اولین و بزرگ‌ترین خوشه با تعداد ۸۴ مفهوم از Action شروع و به Price ختم می‌شود، این خوشه یکی از خوشه‌های اصلی و مرتبط با حوزه مدیریت در علم اطلاعات و دانش‌شناسی، شامل دو حوزه با مباحث مدیریت دانش و اطلاعات و مدیریت کسب و کار با ۸ زیرخوشه است.

حوزه مدیریت دانش و اطلاعات به زیرخوشه‌های: (۱) سیستم‌های مدیریت با واژگان عملکرد، کنترل، تعامل، تصمیم‌گیری، استراتژی، خدمات، نرم‌افزار، تولیدات، مدیریت، سازمان‌دهی و هزینه؛ (۲) نظریه‌ها و فرضیه‌ها با مفاهیم رفتار، متغیر، مورد، فرضیه، ارتباطات، اجتماعی، عضو، دانش، اشتراک، فناوری، مدل، ارتباط، نظام، منابع، عملکرد، پیپر، شبکه، ساختار، نظریه، یادگیری، شرکت‌کننده، سایت و کد؛ (۳) وب شامل واژگان دسته، مفهوم، دامنه، خصوصیت، جستجو، وب، واژه، مدرک، محتوا، کاربر، کار، جامعه، گروه، هدف، دیجیتال، اینترنت، پیوسته یا آنلاین و وب‌سایت؛ (۴) پروژه‌های تیمی دولتی؛ (۵) تلفن همراه در کنار واژه‌های پیام و حریم خصوصی تقسیم می‌شوند. حوزه مدیریت کسب و کار با زیرخوشه‌های: (۶) راه‌اندازی کسب و کار شامل مفاهیم اتخاذ، مدیر، سازمانی، کارمند، پیاده‌سازی، ریسک، کسب و کار، شرکت، توانایی، استراتژیک، مؤسسه، صنعت، بازار و نوآوری؛ (۷) مذاکره شامل لغات قصد، درک‌کردن، اعتماد، مذاکره، پاسخ‌گو و امنیت و (۸) خوشه مشتری و فروشنده در کنار مفاهیم تامین‌کننده، مصرف‌کننده، مشتری و قیمت است.

خوشه دوم از Algorithm تا Patent شبکه واژگانی با ۴۲ مفهوم را تشکیل می‌دهد. این خوشه در ارتباط با دو حوزه مهم بازیابی اطلاعات و علم‌سنجی شامل ۵ زیرخوشه است. حوزه بازیابی اطلاعات شامل زیرخوشه‌های: (۱) بازیابی با واژه‌های الگوریتم، کوئری، معنایی و برچسب یا تگ؛ (۲) نگاشت شامل خوشه، کلیدواژه، شباهت، موضوع، نقشه، کتابخانه، تصویر، صفحه، متن، دانشجو و گره و (۳) هستی‌شناسی است. حوزه علم‌سنجی نیز متشکل از زیرخوشه‌های: (۴) تولیدات علمی شامل مفاهیم مقاله، علم، پدیدآور (نویسنده)، مراجع، مؤسسات، انتشار، دانشگاه،

استناد، مجله، انتشارات، علمی و رشته؛ و ۵) شاخص‌های رتبه‌بندی شامل شمارش، شاخص، رتبه، رتبه‌بندی، همبستگی، مقیاس، نمره، وزن و همکاری‌های علمی شامل همکاری، کشور چین، چینی و ثبت اختراع است. خوشه سوم، دیگر خوشه مستقل و مرتبط با نظام سلامت دارای شبکه واژگانی متشکل از ۱۳ واژه است. این خوشه در دل خود با دو زیرخوشه: ۱) خدمات بالینی با مفاهیم هشدار، (تجویز) دارو، پرونده الکترونیک سلامت، بیمار، پزشک، طب بالینی، سلامت و دارو و ۲) ارائه‌دهندگان خدمات شامل مراقبت، مراقبت‌های پزشکی، بیمارستان و ارائه‌دهنده به مباحث پزشکی در این رشته می‌پردازد.

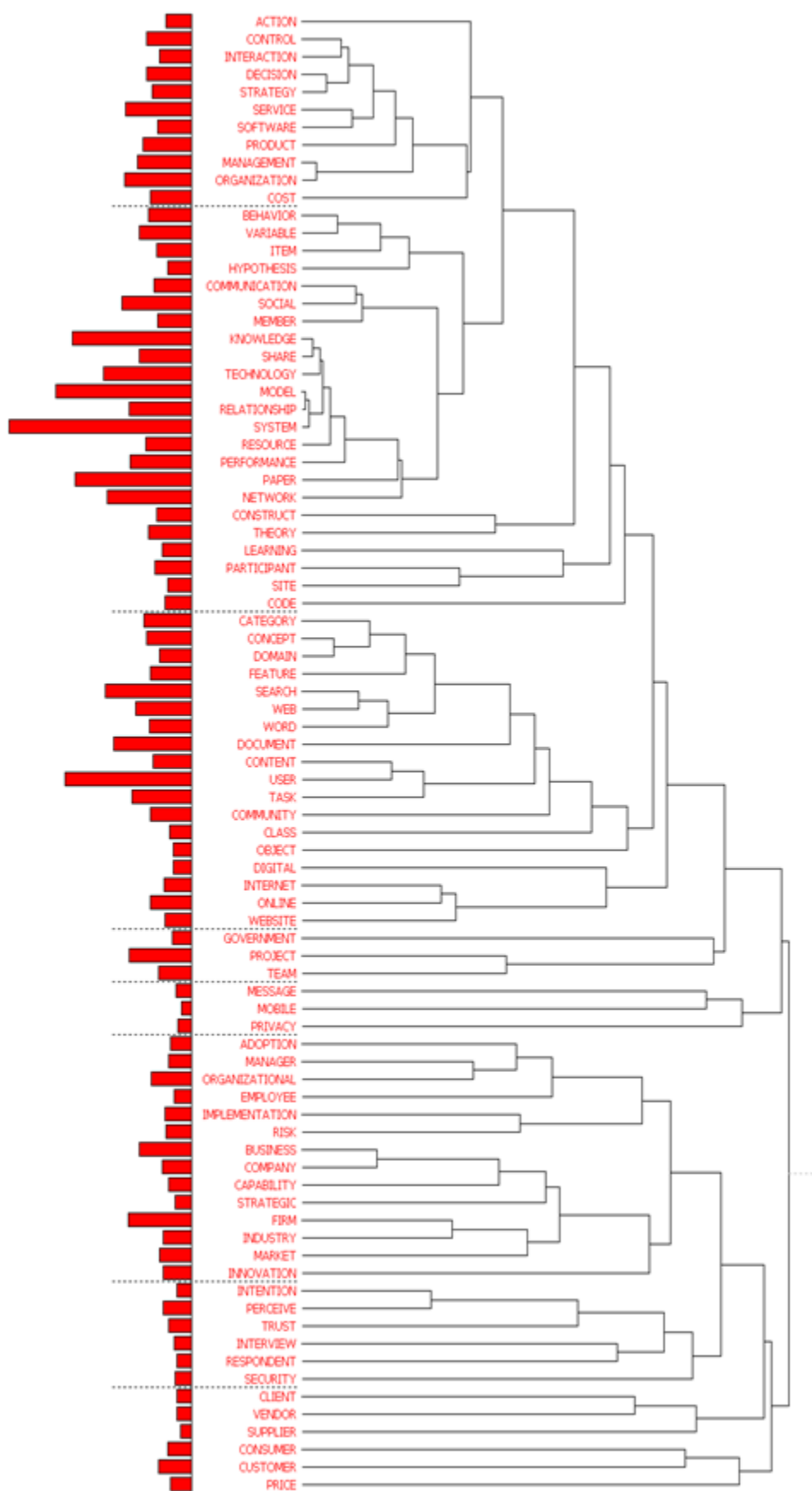
خوشه چهارم با تک عضو خود Sentence از خوشه‌های ضعیف حوزه است و ارتباط آن با مفاهیمی از خوشه دوم به‌ویژه یکی از زیرخوشه‌های آن، وابستگی آن را به حوزه بازبانی اطلاعات نشان می‌دهد. خوشه پنجم با موضوع خرید و فروش با شبکه واژگانی از چهار مفهوم مزایده، پیشنهاد، خریدار و فروشنده یک خوشه کوچک و مستقل را تشکیل می‌دهد. این خوشه هرچند با دیگر خوشه‌ها ارتباط ندارد، از بلوغ کافی برخوردار نیست.

خوشه ششم نیز با دو مفهوم برنامه‌ریزی منابع سازمانی^۲ و برون‌سپاری یکی دیگر از خوشه‌های ضعیف و کوچک است. ارتباط واژگانی این خوشه با مفاهیمی از حوزه کسب و کار گویای وابستگی آن به خوشه یک است. خوشه هفتم متشکل از چهار اصطلاح دانشمند، استناد، ضریب تأثیر و شاخص اچ؛ دیگر خوشه کوچک موجود با واژگان تخصصی حوزه علم‌سنجی است. با توجه به تعداد مفاهیم و ضعیف بودن خوشه و نیز شبکه واژگانی آن با مباحث علم‌سنجی از خوشه دوم دارای ارتباط بوده و به عبارتی خوشه اشتقاق یافته از این خوشه بزرگ می‌باشد.

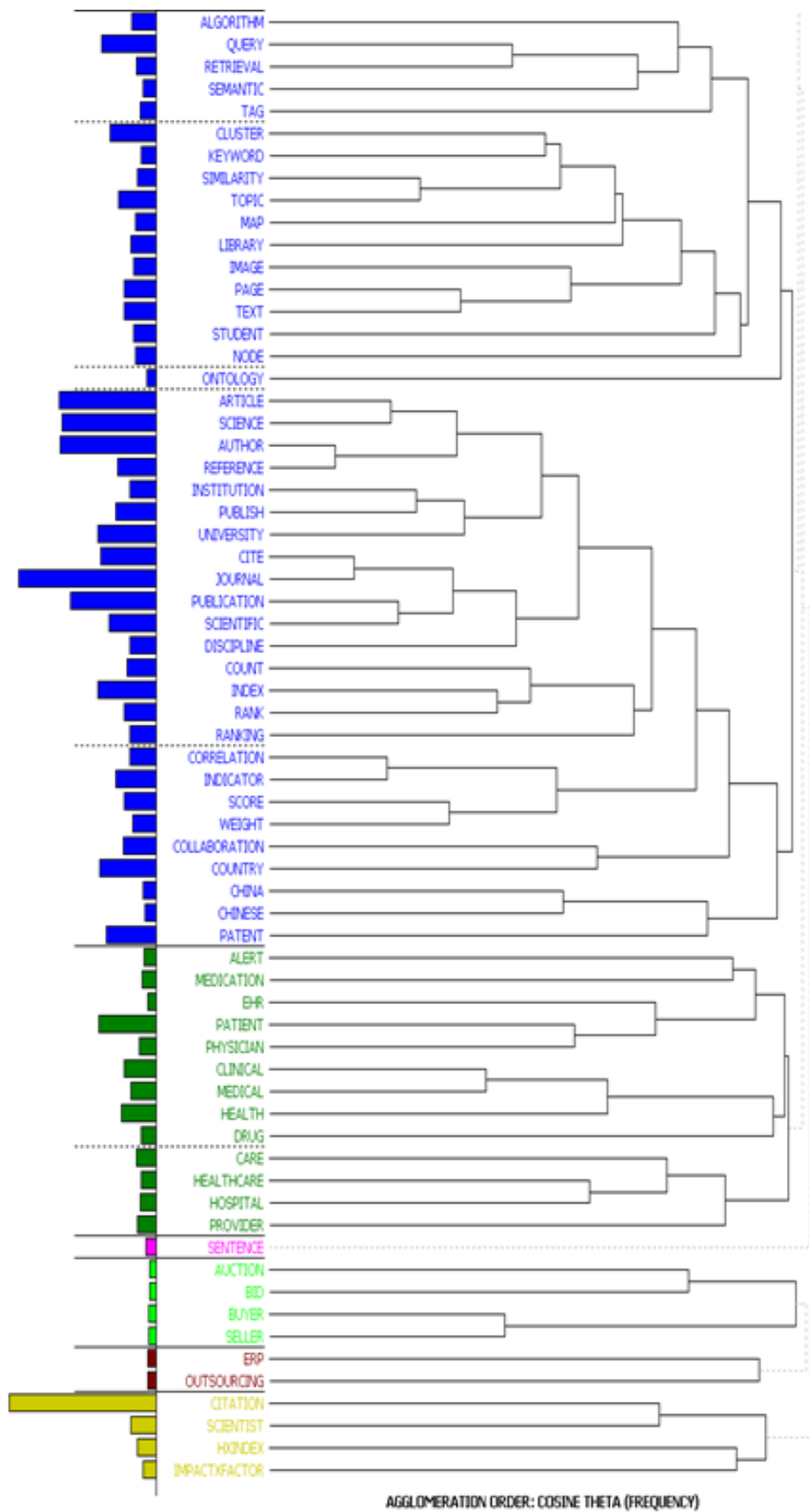
۳. دامنه شناختی مجلات هسته بر اساس مفاهیم منتخب چگونه است؟

تحلیل هم‌رخدادی واژه‌ها امکان شناسایی ساختار مفهومی حوزه را مشخص می‌کند اما درک ساختار نحوه ارتباط این لغات با مجلات نیز مفید خواهد بود. شکل ۳ ساختار سه‌بعدی از ۱۵۰ واژه منتخب در ۱۰ مجله بر اساس حضور آنها را نشان می‌دهد. مجلات نزدیک به مرکز نمودار از طیف لغات عام و مشترک تشکیل شده و هرچه در حاشیه قرار بگیرند، از لغات اخص تشکیل شده‌اند؛ چنانچه مجلات مشابه از نظر مفهومی در شکل کنار هم قرار می‌گیرند. توزیع مجلات در این شکل حاکی از خاص بودن لغات در مجلات و نیز تخصصی بودن مجلات در زیرحوزه‌ها است. به‌طوری‌که چهار خوشه موضوعی قابل شناسایی است. اولین خوشه، با بیشترین تعداد مجله شامل MISQ, I&O, JIM, ISR, JIT, JIS و JSIS می‌باشد. این مجلات دارای حوزه موضوعی مشابه بوده و اغلب با مفاهیمی از خوشه ۱، ۵ و ۶ در خوشه‌بندی هم‌رخدادی واژگان ارتباط دارد. دومین خوشه با مجله JAMIA که دارای لغات خاص در زیرحوزه پزشکی است و با خوشه ۳ ارتباط موضوعی دارد. سومین خوشه نیز با یک عنوان مجله JASIST دارای بیشترین مفاهیم عام رشته بوده و از این منظر طیف موضوعی مشابهی را با سایر خوشه‌ها دارد ولی بیشترین شباهت مفهومی را با خوشه ۱ و ۲ در خوشه‌بندی سلسله‌مراتبی دارد. درنهایت خوشه چهارم با مجلات JoI و Sci دارای مفاهیم خاص در زیرحوزه علم‌سنجی و اطلاع‌سنجی است و با خوشه ۲ و ۷ دارای مفاهیم مشابه می‌باشد. همچنین نکته قابل توجه همبستگی بین خوشه‌هاست، که خوشه سوم با چهارم همبستگی مثبت و بقیه با هم همبستگی منفی دارند. به عبارتی خوشه‌های دارای همبستگی منفی از نظر دارابودن واژگان خاص شباهت کمتری با هم دارند.

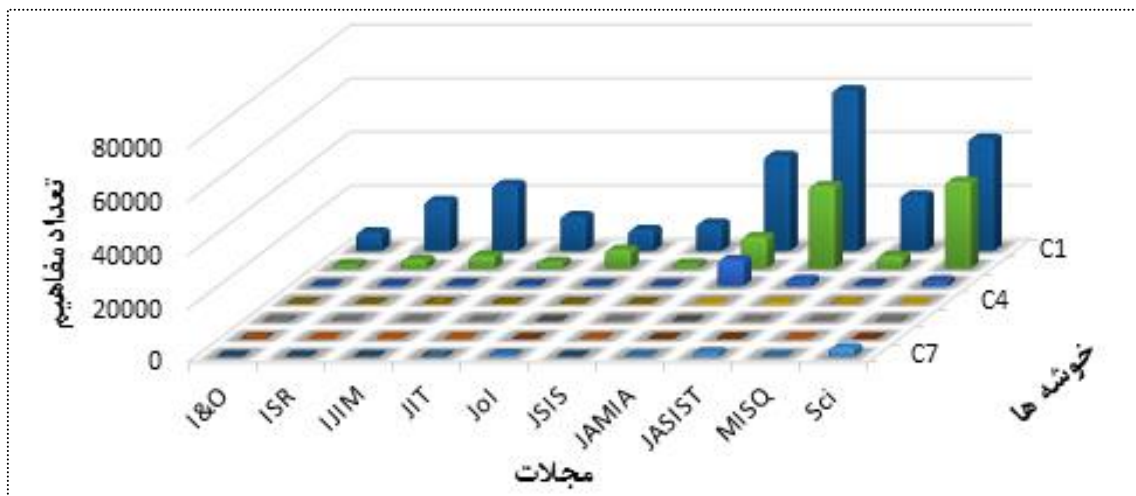
1. Electronic Health Record
2. Enterprise Resource Planning (ERP)



شکل ۲. دندوگرام خوشه‌بندی سطح دو مفاهیم تمام متن بین‌الملل بر اساس کسینوس تتا



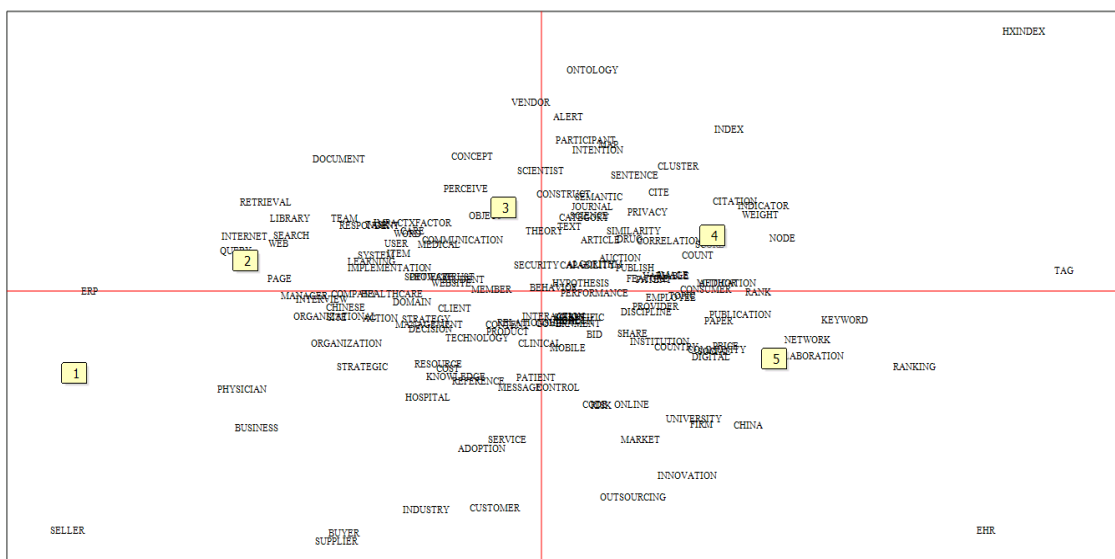
ادامه شکل ۲. دندوگرام خوشه‌بندی سطح دو مفاهیم تمام متن بین‌الملل بر اساس کسینوس تا



نمودار ۱ توزیع مفاهیم (متن کامل بین المللی) هر یک از مجلات در خوشه های مختلف

۴. نحوه چیدمان مفاهیم طی سال های (۲۰۰۴-۲۰۱۳) چگونه بوده است؟

برای درک نحوه تحولات زمانی، بسامد مفاهیم بر اساس سال نشر به صورت دوره ۲ ساله تحلیل و ارتباط و توزیع مفاهیم در سال های مذکور به صورت دویبعدی در شکل ۵ ارائه شد. در این تصویر مفاهیم مشابه از نظر زمانی کنار هم با شماره های مختص هر دوره خوشه بندی گردید. به عبارتی بازه های زمانی مشابه از نظر استفاده مفاهیم نزدیک به هم قرار دارند. همچنین مفاهیم مختص هر بازه در حاشیه و نزدیک به شماره آن و مفاهیم عام نزدیک به مرکز قرار دارند. مطابق شکل رشد تکاملی مشابه با حرکت عقربه های ساعت دیده می شود. چنانچه فاصله زاویه بین مفاهیم سال های نخستین با سال های پایانی یک زاویه منفرجه بوده لذا همبستگی بین آنها منفی و سال های نزدیک به هم نیز همبستگی مثبت دارند. یعنی بین سال های ۲۰۰۴-۲۰۰۵ با سال های ۲۰۰۶-۲۰۰۷، ۲۰۰۶-۲۰۰۷ با ۲۰۰۸-۲۰۰۹ و ۲۰۱۰-۲۰۱۱ با ۲۰۱۲-۲۰۱۳ همبستگی مثبت وجود دارد. نکته قابل توجه نوع همبستگی بین ۲۰۰۸-۲۰۰۹ با ۲۰۱۰-۲۰۱۱ است با زاویه ۹۰ درجه نشان از همبسته نبودن و وجود جهشی قابل توجه با دوره بعدی است و دو خوشه تاریخی را به وجود می آورد.

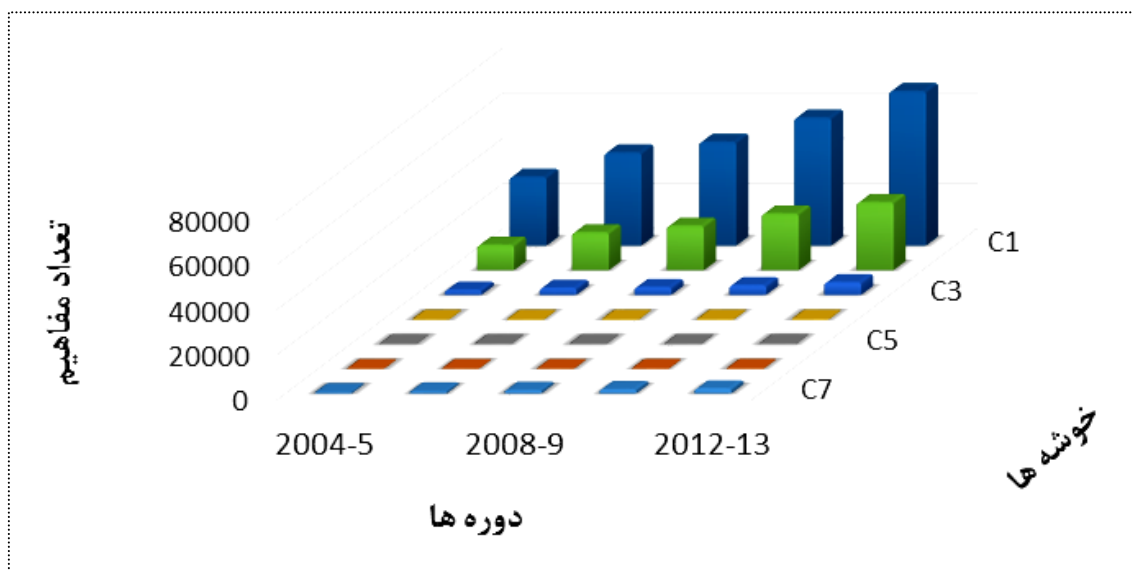


شکل ۵. تحلیل تناظر بین مفاهیم منتخب تمام متن متون بین الملل و دوره های ۲ ساله

نگاشت ساختار و چیدمان تاریخی مفاهیم علم اطلاعات و دانش‌شناسی: با رویکرد متن‌کاوی (۲۰۰۴-۲۰۱۳)

خوشه نخست شامل سه دوره اولیه (۲۰۰۴-۲۰۰۹) و خوشه دوم شامل دو دوره آخر (۲۰۱۰-۲۰۱۳) می‌باشد. مطابق نگاشت در سال‌های نخستین ۲۰۰۴-۲۰۰۵ حالت سکون وجود دارد و مفاهیم مدیریت منابع سازمانی (ERP)، پزشکی، کسب و کار و فروشنده دیده می‌شود. مفاهیم صفحه، کوثری، وب، اینترنت، جستجو، کتابخانه، بازیابی، مدرک، ایده نیز مفاهیمی هستند که در سال‌های ۲۰۰۶ و ۲۰۰۷ مورد توجه بودند. در دوره بعد (۲۰۰۸-۲۰۰۹) نیز توجه به مفهوم مدرک و ایده در ارتباط با مفاهیم هدف، ارتباطات، درک کردن، کارگزار، دانشمند، معنایی، هستی‌شناسی و قصد ادامه یافته و سپس جهشی بین این دوره با دوره بعدی با نمایان شدن لغت شاخص و شاخص اچ صورت گرفته و در کنار مفاهیم هستی‌شناسی، قصد و جمله؛ مفاهیم خوشه، استناد، همبستگی، شمارش، استنادی، وزن، گره، مقیاس و برچسب بروز می‌کند. در آخرین دوره نیز مفاهیمی چون تولیدات، قیمت، جامعه، اجتماع، کلیدواژه، شبکه، همکاری، رتبه‌بندی، دانشگاه، چین و پرونده الکترونیک سلامت (EHR) نمود بیشتری می‌یابد.

برای بررسی دقیق‌تر میزان رشد و توزیع مفاهیم خوشه‌ها در سال‌های مختلف نمودار ۲ ترسیم شد. بنا بر نمودار سیر صعودی مفاهیم در تمامی خوشه‌ها دیده می‌شود. به عبارتی خوشه‌ها به جز خوشه پنجم که در دوره سوم دارای کاهش بوده، رشد متناوب داشته و حاکی از حضور مفاهیم همه خوشه‌ها طی سال‌های مختلف است.



نمودار ۲. توزیع مفاهیم هر یک از دوره‌ها در خوشه‌های مختلف

بحث و نتیجه‌گیری

برای تعیین مفاهیم کلیدی علم اطلاعات و دانش‌شناسی از وزن TF-IDF برای انتخاب ۱۵۰ واژه نخست جهت تحلیل استفاده گردید. خوشه‌بندی مفاهیم در سطح متن کامل مقالات، در ۷ خوشه و ۱۵ زیرخوشه بر اساس ۳ خوشه اصلی نخست صورت گرفت. خوشه نخست با ۸۴ مفهوم و ۸ زیرخوشه به موضوع اصلی مدیریت در دو شاخه مدیریت دانش و اطلاعات و کسب و کار می‌پردازد. شاخه اول به زیرخوشه‌های سیستم‌های مدیریتی شامل سازمان‌دهی تولیدات و هزینه در کنترل عملکرد با استراتژی تصمیم‌گیری و تعامل؛ نظریه‌ها و فرضیه‌های مرتبط با رفتار، ارتباطات اجتماعی، اشتراک دانش، تکنولوژی، شبکه‌ها، یادگیری و غیره؛ وب و اینترنت با مفاهیم دیجیتال، پیوسته، دسته‌بندی ویژگی‌های حوزه و مفاهیم، جستجو، محتوا، کاربران و غیره؛ پروژه‌های تیمی دولتی و درنهایت

تلفن همراه با مفاهیم امنیت و پیام کوتاه تقسیم می‌شود. شاخه مدیریت کسب و کار نیز به با ۳ زیرخوشه به موضوعات راه‌اندازی کسب و کار شامل مدیر، کارمند، ریسک، نوآوری، بازار و غیره؛ مذاکره در کنار مفاهیم درک، قصد، امنیت و اعتماد و پاسخ‌دهنده؛ و مقوله فروشنده، مشتری و قیمت و غیره می‌پردازد.

خوشه دوم با ۴۲ مفهوم به دو مبحث بازیابی اطلاعات و علم‌سنجی در ۵ زیرخوشه می‌پردازد. شاخه اول یعنی بازیابی به سه زیرخوشه بازیابی با مفاهیم الگوریتم، معنایی، کوئری و برجسب؛ نگاشت با مفاهیم خوشه، کلیدواژه، موضوع، کتابخانه و غیره، و درنهایت زیرخوشه سوم به هستی‌شناسی می‌پردازد. شاخه دوم حوزه علم‌سنجی نیز شامل زیرخوشه‌های تولیدات علمی شامل مقاله، پدیدآور، ارجاع، مؤسسات، دانشگاه و غیره؛ شاخص‌های رتبه‌بندی با مفاهیم استناد، نشریات، رتبه، همبستگی، وزن و غیره و درنهایت همکاری‌های علمی شامل کشور، ثبت اختراع، چین و غیره است.

خوشه بعدی با دو زیرخوشه خدمات بالینی و ارائه‌دهندگان این خدمات به مبحث پزشکی در این رشته می‌پردازد. سه خوشه بعدی از جمله خوشه‌های نابالغ و دارای شبکه ارتباطی با دیگر خوشه‌ها و زیرخوشه‌ها می‌باشند. خوشه چهارم با مفهوم جمله به مبحث بازیابی در خوشه دوم، خوشه ششم با برنامه‌ریزی منابع سازمانی و برون‌سپاری به خوشه اول و شاخه دوم یعنی مدیریت کسب و کار و خوشه هفتم با واژگان استنادی، دانشمند، اچ ایندکس و ضریب تأثیر به حوزه علم‌سنجی از خوشه دوم به‌ویژه شاخص‌های علم‌سنجی ارتباط دارند. اما خوشه پنجم با چهار مفهوم مزایده، پیشنهاد، خریدار و فروشنده با موضوع خرید و فروش یک خوشه کوچک و نیمه بالغ را تشکیل می‌دهد.

درواقع مفاهیم علم اطلاعات و دانش‌شناسی در تحلیل هم‌واژگانی بر اساس متن کامل مقالات، متشکل از ۳ خوشه بالغ، ۱ خوشه نیمه‌بالغ و ۳ خوشه نابالغ است. سه خوشه بالغ دارای شبکه واژگان با تعداد مفاهیم بالا هستند و درصد بیشتری از مفاهیم در این خوشه‌ها جا دارد. خوشه نیمه‌بالغ نیز یک خوشه مستقل بوده ولی هنوز بلوغ مفهومی مناسب را ندارد و وضعیت این خوشه‌ها -اینکه به خوشه بالغ تبدیل شوند و یا در سایر خوشه‌ها حل شده و در آنها ادغام شوند- به پژوهش‌های سال‌های آتی بستگی دارد. خوشه‌های نابالغ نیز هنوز وابستگی خود را به خوشه مادر حفظ کرده و دارای وابستگی مفهومی هستند، اما این خوشه‌ها نیز قابلیت تشکیل خوشه مستقل را با احتمال کمتری نسبت به خوشه‌های نیمه‌بالغ داشته و موقعیت آنها را موضوعات پژوهشی آینده مشخص خواهد کرد. به عبارتی این نوع خوشه‌ها در عین وابستگی به خوشه اصلی، استقلال و رشد چندانی در دایره مفاهیم ندارند.

بررسی توازن مفاهیم از دو منظر نشریات و دوره‌ها نیز نشان می‌دهد، خوشه‌ها به‌ویژه مفاهیم دارای نماینده خاصی در مجلات هستند. حضور کم‌رنگ مفاهیم در مرکز نمودار به‌ویژه در متون تمام‌متن حاکی اشتراک مفهومی کم در بین مجلات است؛ یعنی در متون بین‌الملل مجلات به‌صورت تخصصی عمل می‌کنند و مقالات مرتبط با مفاهیم و موضوعات تشکیل‌دهنده خوشه‌های مجلات را پذیرش می‌کنند. همچنین رابطه مستقیم بین تعداد مفاهیم خوشه‌ها و تعداد مجلات و مقالات مختص به آنها وجود دارد، یعنی هر جا مفاهیم بیشتری حضور دارند مجلات نیز تراکم بیشتری دارند. به عبارتی توجه به مفاهیم ناشی از تعداد مجلات مورد تحلیل در موضوع می‌باشد. از نظر روند تکاملی نیز وجود مفاهیم تمامی خوشه‌ها و رشد آنها، نشان‌دهنده تداوم حوزه‌های پژوهشی است. حوزه‌هایی که در تحلیل تناظر زمانی با چیدمان خاص مفاهیم، دو دوره تاریخی را تشکیل می‌دهند. چنانچه اگر دوره اول بر مبنای مفاهیم علم

نگاشت ساختار و چیدمان تاریخی مفاهیم علم اطلاعات و دانش‌شناسی: با رویکرد متن‌کاوی (۲۰۰۴-۲۰۱۳)

اطلاعات و بازیابی اطلاعات (۲۰۰۴-۲۰۰۹) متمرکز بوده، دوره دوم از علم‌سنجی (۲۰۱۰-۲۰۱۳) و استفاده از کارکردهای مفاهیم دوره اول با پیدایش مفاهیم آنتولوژی، نگاشت، تگ و استفاده از شاخص‌ها و الگوریتم‌هاست و می‌توان این دوره را دوره علم‌سنجی و بازیابی اطلاعات نام نهاد. به‌طور کلی بررسی مفهومی خوشه‌ها نشان داد تولیدات علمی دارای برخی از موضوعات ثابت بوده که با تغییر تدریجی مفاهیم جدیدی را ایجاد می‌کنند. چنانچه این تغییر در چشم‌انداز شناختی در خوشه زمانی نخست مبتنی بر ظهور اینترنت بوده درحالی‌که در خوشه دوم به دلیل استفاده از شاخص‌هاست.

پیشنادهای اجرایی پژوهش

- در توسعه رشته‌های زیرمجموعه علم اطلاعات و دانش‌شناسی، به گسترش مفاهیم و واژگان تخصصی این حوزه که به نوعی بیانگر زمینه‌های موضوعی و فعالیت‌های این حوزه هستند توجه شود.
- در تدوین و توسعه نشریات علمی این حوزه، به زمینه‌های پژوهشی این حوزه که با توجه به بررسی تاریخی واژگان این حوزه میسر است توجه ویژه‌ای مبذول داشته شود.

فهرست منابع

- احمدی، ح. و عصاره، ف. (۱۳۹۴). سیر تحول پژوهش‌های علم‌سنجی در ایران. *فصلنامه رهیافت*، ۲۵(۳)، ۶۹-۸۲.
- تیمورپور، ب.، سپهری، م. و پزشک، ل. (۱۳۸۸). روشی نوین برای دسته‌بندی هوشمند متون علمی (مطالعه موردی مقالات فناوری نانو متخصصان ایران). *سیاست علم و فناوری*، ۲(۲)، ۱-۱۴.
- حاضری، ا.، توکلی‌زاده راوری، م. و ابراهیمی، و. (۱۳۹۴). تعیین طبقات اصلی مرتبط با مدیریت دانش در پایگاه وب آو ساینس و مطالعه هم‌پوشانی موضوعی آنها. *پژوهشنامه پردازش و مدیریت اطلاعات*، ۳۰(۴)، ۹۹۷-۱۰۲۳.
- خادمی، ر. ا. و حیدری، غ. (۱۳۹۵). ترسیم ساختار موضوعی مدیریت اطلاعات با استفاده از روش همبندی واژگان طی سال‌های ۱۹۸۶ تا ۲۰۱۲. *فصلنامه علوم و فنون مدیریت اطلاعات*، ۲(۲)، ۵۹-۹۳.
- کاتبی، ف. (۱۳۹۳). *تحلیل هم‌رخدادی واژگان حوزه مدیریت دانش در پایگاه وب آو ساینس در سال‌های ۱۹۹۳ لغایت ۲۰۱۲ به‌منظور دیداری‌سازی واژگان تشکیل‌دهنده زیرساخت این حوزه*. دانشگاه شهید چمران اهواز، دانشکده علوم تربیتی و روانشناسی.
- محمدی، ا. (۱۳۸۸). *حوزه‌های تشکیل‌دهنده فناوری و علم نانو در ایران*. پنجمین همایش دانشجویی فناوری نانو. تهران: دانشگاه علوم پزشکی تهران.
- مصطفوی، ا.، عصاره، ف.، و توکلی‌زاده راوری، م. (زیر چاپ). شناسایی ساختار محتوایی مطالعات علم اطلاعات و دانش‌شناسی بر اساس واژگان و مفاهیم مقالات آن در پایگاه اطلاعاتی وبگاه علم (۲۰۰۹-۲۰۱۳). *پژوهشنامه پردازش و مدیریت اطلاعات*. بازیابی شده در تاریخ ۶ خرداد ۱۳۹۶، از <http://jipm.irandoc.ac.ir/article-1-3234-fa.html>
- موسوی‌زاده، م.، باقری، م. و کربلا آقایی کامران، م. (۱۳۹۳). *مصورسازی حوزه سازمان‌دهی اطلاعات: بررسی ساختار*

گرایش‌های موضوعی مقالات فارسی حوزه سازمان‌دهی اطلاعات. پژوهشنامه کتابداری و اطلاع‌رسانی، ۴(۲)، ۱۹۰-۲۱۱.

ناصری‌جزه، م.، طباطباییان، س. و فاتح‌راد، م. (۱۳۹۱). ترسیم نقشه دانش مدیریت فناوری در ایران با هدف کمک به سیاست‌گذاری دانش در این حوزه. سیاست علم و فناوری، ۱(۵)، ۴۵-۷۲.

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology Behind Search* (2nd ed.). Addison Wesley.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37(1), 179-255.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management*, 37(6), 817-842.

Glenisson, P., Glänzel, W., Janssens, F., & De Moor, B. (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing and Management*, 41(6), 1548-1572.

Glenisson, P., Glänzel, W., & Persson, O. (2005). Combining full-text analysis and bibliometric indicators. A pilot study. *Scientometrics*, 63(1), 163-180.

Grefenstette, G. (1994). Corpus-Derived First, Second and Third-Order Word Affinities. In *In Proceedings of Euralex* (pp. 279-290).

Hu, C.-P., Hu, J.-M., Deng, S.-L., & Liu, Y. (2013). A co-word analysis of library and information science in China. *Scientometrics*, 97(2), 369-382.

Janssens, F., Glänzel, W., & Moor, B. (2008). A hybrid mapping of information science. *Scientometrics*, 75(3), 607-631.

Janssens, F., Leta, J., Glänzel, W., & De Moor, B. (2006). Towards mapping library and information science. *Information Processing & Management*, 42(6), 1614-1642.

Lee, B., & Jeong, Y. Il. (2008). Mapping Korea's national R&D domain of robot technology by using the co-word analysis. *Scientometrics*, 77(1), 3-19.

Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2), 159-165.

Milojević, S., Sugimoto, C. R., Yan, E., & Ding, Y. (2011). The cognitive structure of Library and Information Science: Analysis of article title words. *Journal of the American Society for Information Science and Technology*, 62(10), 1933-1953.

Mohammadi, E. (2012). Knowledge mapping of the Iranian nanoscience and technology: A text mining approach. *Scientometrics*, 92(3), 593-608.

Noyons, E. C. M., Moed, H. F., & Luwel, M. (1999). Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science*, 50(2), 115–131.

Provalis Research. (2015). WordStat 7 (Content analysis module for SIMSTAT and QDA miner): User's Guide.

Rencher, A. C. (2003). *Methods of multivariate analysis* (2nd ed.). John Wiley & Sons.

Sasaki, M., & Suzuki, T. (1989). New directions in the study of general social attitudes: trend and cross-national perspectives. *Behaviormetrika*, 16(26), 9–30.

Xu, H., Guo, T., Yue, Z., Ru, L., & Fang, S. (2016). Interdisciplinary topics of information science: a study based on the terms interdisciplinarity index series. *Scientometrics*, 106(2), 583–601.

Zong, Q.-J., Shen, H.-Z., Yuan, Q.-J., Hu, X.-W., Hou, Z.-P., & Deng, S.-G. (2013). Doctoral dissertations of Library and Information Science in China: A co-word analysis. *Scientometrics*, 94(2), 781–799.