

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

محمد حسن زاده<sup>۱</sup>

فاطمه زندیان<sup>۲</sup>

سیده سمیه احمدی<sup>۳</sup>

۱. دانشیار دانشگاه تربیت مدرس (نویسنده مسول) Email: hasanzadeh@modares.ac.ir

۲. استادیار دانشگاه تربیت مدرس

۳. دانشجوی کارشناسی ارشد علم اطلاعات و دانش‌شناسی

## چکیده

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

**روش‌شناسی:** روش به کار رفته تحلیل محتوای مفاهیم استخراج شده براساس فنون متن‌کاوی از ۶۸۳۰ مقاله ۱۰ نشریه هسته براساس گزارش جی‌سی‌آر ۲۰۱۳ (۲۰۱۳-۲۰۰۴) می‌باشد. خوشه‌بندی واژگان و عبارات بعد از استخراج به صورت پارامتریک و انتخاب ۱۵۰ مفهوم براساس وزن tf-idf، با استفاده از ماتریس هم‌رخدادی واژگان براساس شاخص کسینوس تتا و قرابت مرتبه ثانویه و الگوریتم average-linkage، در نرم افزار Provalis Research، انجام شد. همچنین از فراوانی نسبی مفاهیم منتخب جهت تحلیل تناظر، به منظور خوشه‌بندی مجلات و دامنه موضوعی آن‌ها و بررسی چیدمان تاریخی مفاهیم استفاده گردید.

**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:** به طور 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 چشم انداز شناختی در خوشه زمانی نخست مبتنی بر ظهور اینترنت بوده و در خوشه دوم به دلیل استفاده از شاخص‌ها است.

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

دریافت: ۰۰۰/۰۰۰/۰۰

پذیرش: ۰۰۰/۰۰۰/۰۰

## 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، انعطاف و ... مشخص می‌گردد و به عنوان 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 روز به روز افزایش یافته و توجه حوزه‌ها و علوم مختلف را در بهره‌برداری از مزیت آن به ویژه در مباحث شناختی جلب نموده است. علم اطلاعات و دانش‌شناسی به عنوان یکی از حوزه‌هایی که با داده سروکار دارد و در توسعه روش‌های بهره‌برداری به خصوص در حوزه نگاشت علم سهم است، خود نیز به عنوان یک حوزه نیاز به بهره‌برداری از مزیت داده‌ها - به ویژه در شناخت ساختار مفهومی - دارد. آنچه که مساله این تحقیق بوده و در پی کشف الگوی حاکم بر درخت علم اطلاعات و دانش‌شناسی، زیرحوزه‌های تحقیقی بارز و همچنین روابط بین آن‌ها است. در این تحقیق به عنوان 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<sup>۱</sup>، مود<sup>۲</sup> و لوول<sup>۳</sup>، ۱۹۹۹؛ برونر<sup>۴</sup>، چن<sup>۵</sup> و بویاک<sup>۶</sup>، ۲۰۰۳ است، که بر زبان طبیعی استوار است. این تحلیل با تکنیک متن‌کاوی بر روی عنوان، چکیده و متن کامل بروندهای علمی قابل اجرا است، البته اجرای آن بر روی متن کامل با محدودیت‌های سخت‌افزاری و نرم‌افزاری برحسب داده‌های موردتحلیل روبرو است. با توجه به نگاشت‌های مختلف براساس مجلات و متون استخراجی از پایگاه‌ها در این حوزه، این پژوهش به دنبال اجرای این تکنیک بر روی متن کامل مقالات نشریات هسته در طی ده 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 نمود. در کنار دستیابی به این ساختار، بررسی چیدمان تاریخی مفاهیم و خوشه‌بندی مجلات براساس مفاهیم مورد انتظار است و در پی پاسخگویی به سوالات زیر است.

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

## پیشینه پژوهش

<sup>1</sup> Noyons  
<sup>2</sup> Moed  
<sup>3</sup> Luwel  
<sup>4</sup> Börner  
<sup>5</sup> Chen  
<sup>6</sup> Boyack

بررسی پیشینه پژوهش در این حوزه به دلیل ارتباط قوی بین زمینه‌های مختلف تحقیق و فناوری پیچیده است. ارتباط این حوزه با حوزه‌های چون زبان‌شناسی و فرهنگ‌نویسی، متن‌کاوی، مصورسازی، زبان‌شناسی محاسباتی، علوم اجتماعی و... خاستگاه‌های مختلفی را برای این حوزه رقم می‌زند. چنانچه کاستوف<sup>۱</sup> خاستگاه نمود هم‌واژگانی را اولین اثر در (۱) فرهنگ‌نویسی هورنبای<sup>۲</sup> (۱۹۴۲) برای شرح دانش هم‌رخدادی<sup>۳</sup> و (۲) زبان‌شناسی دی سوسور<sup>۴</sup> (۱۹۴۹) با هدف توصیف نحوه ارتباط قرابت دو واحد زبانی با ظهور آن‌ها در زبان، بیان می‌کند (کاستوف، ۱۹۹۳). اما سرآغاز استفاده از روش‌های تحلیل هم‌واژگانی جهت ارزیابی تحقیقات را می‌توان مطالعات میشل کالون و همکارانش در مرکز جامعه‌شناسی نوآوری پاریس<sup>۵</sup> در دهه ۱۹۸۰ برای کمک به ارزیابی وضعیت پژوهش و به عنوان مکمل اطلاعاتی برای هیات کارشناسان، عنوان نمود. وی جهت غلبه بر محدودیت‌های ذاتی تحلیل هم‌استنادی یعنی در نظر گرفتن مدارک رسمی، به عنوان مثال در نظر نگرفتن مدارکی مثل گزارشات غیررسمی، فنی و صنعتی، دسترسی غیرمستقیم به محتوای مدارک، نشان ندادن پویایی علم و در نظر نگرفتن متون استناد نشده این نوع از هم‌رخدادی را پیشنهاد می‌دهد (کالون<sup>۶</sup>، کورشیال<sup>۷</sup>، ترنر<sup>۸</sup> و بوئین<sup>۹</sup>، ۱۹۸۳). این مطالعات توسط (ریپ و کورشیال، ۱۹۸۴؛ کالون ۱۹۸۶) ادامه داشت و با مطرح شدن چالش موجود در این روش توسط (هیلی<sup>۱۰</sup>، راثمن<sup>۱۱</sup> و هوچ<sup>۱۲</sup>، ۱۹۸۶؛ لیدسدروف<sup>۱۳</sup>، ۱۹۸۹) یعنی بررسی غیرمستقیم متن و خطای ناشی از تأثیر نمایه‌ساز و تفسیر متفاوت از نقشه‌ها، در تحقیقات بعدی با در نظر گرفتن این چالش علاوه بر توسعه روش از نمایه‌سازی ماشینی جهت استخراج مفاهیم استفاده گردید (لاو<sup>۱۴</sup>، بوئین، کورشیال و وایتاکر<sup>۱۵</sup>، ۱۹۸۸؛ ترنر، چارترون<sup>۱۶</sup>، لاویله<sup>۱۷</sup> و میشل<sup>۱۸</sup>، ۱۹۸۸؛ وایتاکر، ۱۹۸۹؛ لاو<sup>۱۹</sup> و وایتاکر، ۱۹۹۲). روند توسعه روش با استفاده از مزایای توپولوژیک مقیاس‌بندی چندبعدی<sup>۲۰</sup> (تییجسن<sup>۲۱</sup> و ون‌ران<sup>۲۲</sup>، ۱۹۸۹)، نقشه‌های زمانی از تکامل چگالی و مرکزیت تراکم خوشه (کالون، کورشیال و لاویله، ۱۹۹۱)، ترکیب تحلیل هم‌واژگانی و هم-استنادی (برام<sup>۲۳</sup>، مود و ون‌ران، ۱۹۹۱ الف، ۱۹۹۱ ب)، پتانسیل معرفت‌شناختی نقشه‌های کتابشناختی براساس تحلیل

<sup>1</sup> Kostoff

<sup>2</sup> Hornby

<sup>3</sup> دانش هم‌رخدادی برپایه میزان تعریف یک واحد براساس واحدهای اطراف خود، جدای از ملاک‌های نحوی و یا معنایی است ( Smadja, 1989).

<sup>4</sup> De Saussure

<sup>5</sup> Centre de Sociologie de l'Innovation à l'Ecole des Mines de Paris

<sup>6</sup> Callon

<sup>7</sup> Courtial

<sup>8</sup> Turner

<sup>9</sup> Bauin

<sup>10</sup> Healey

<sup>11</sup> Rothman

<sup>12</sup> Hoch

<sup>13</sup> Leydesdroff

<sup>14</sup> Law

<sup>15</sup> Whittaker

<sup>16</sup> Chartron

<sup>17</sup> Laville

<sup>18</sup> Michelet

<sup>19</sup> Law

<sup>20</sup> Multidimensional Scaling (MDS)

<sup>21</sup> Tijssen

<sup>22</sup> Van Raan

<sup>23</sup> Braam

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

هم‌رخدادی (ون‌ران و تیجسن، ۱۹۹۳)، امکان‌سنجی یک طبقه‌بندی "طبیعی" از پایگاه داده IR&D و شناسایی فرصت‌های تحقیقاتی و تاکسونومی R&D برای پایگاه داده IRD (کاستوف، ۱۹۹۳، ۱۹۹۴) در حوزه‌های مختلف ادامه داشت. تا اینکه لیدسدرف در سال ۱۹۹۷ مطالعاتی با هدف بررسی قابلیت روش هم‌رخدادی برای تحلیل متن با استفاده از لغات و توزیع آن‌ها در سطوح مختلف مثل جملات، بندها و بخش‌ها در سال ۱۹۹۱ بر روی یک مقاله انجام داد. دو یافته مهم این تحقیق باعث گردید تا یک مدل تجزیه و تحلیل را پیشنهاد دهد و برای تست آن پژوهشی بر روی مجموعه محدودی از متن کامل مقالات (۱۷مقاله) از زیرحوزه بیوشیمی تحلیل نماید.

در مورد مطالعات تمام‌متن مقالات همچنین (گلنسون<sup>۱</sup>، گلنزل<sup>۲</sup>، پرسون<sup>۳</sup>، 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 ارائه شده در نهمین کنفرانس بین‌المللی علم‌سنجی و اطلاع‌سنجی<sup>۴</sup> سال ۲۰۰۳ در چین بود. نتایج حاکی از آن بود که پروفایل وزن‌دهی شده متون، توصیفگر خوبی برای موضوع یک مقاله می‌باشد. 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٪ از تمام مقالات در مواقع استفاده از عناوین و چکیده‌ها به درستی اختصاص داده نمی‌شود، 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 منتشر شده 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 به عنوان 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<sup>۵</sup> و همچنین موضوعات فراتر از جریان اصلی را نمایان ساخت (گلنسون<sup>۶</sup>، گلنزل، جانسنز<sup>۷</sup> و دی‌مور<sup>۸</sup>، ۲۰۰۵).

بار دیگر (جانسنز، لتا<sup>۹</sup>، گلنزل و دی‌مور، ۲۰۰۶) نه با هدف ارائه یک روش‌شناسی بلکه با هدف بسط روش ارائه شده بر روی متون ناهمگن علم کتابداری و اطلاع‌رسانی به عنوان یک رشته وسیع مورد مطالعه قرار دادند. لذا ۹۳۸ مقاله و یادداشت تمام‌متن منتشر شده بین سال‌های ۲۰۰۲ تا ۲۰۰۴ در پنج نشریه "مدیریت و پردازش اطلاعات"<sup>۹</sup>، "مجله انجمن علم و فناوری اطلاعات آمریکا"<sup>۱۰</sup>، "مجله دکوماناسیون"<sup>۱۱</sup>، "مجله علوم اطلاعات"<sup>۱۲</sup> و "علم‌سنجی"<sup>۱۳</sup> با محوریت موضوعی روش‌شناختی، مطالعات نظری و رویکردهای کمی رشته انتخاب شدند. نتایج، دو خوشه در

<sup>1</sup> Glenisson

<sup>2</sup> Glänzel

<sup>3</sup> Persson

<sup>4</sup> 9th International Conference on Scientometrics and Informetrics

<sup>5</sup> Technometric

<sup>6</sup> Janssens

<sup>7</sup> De Moor

<sup>8</sup> Leta

<sup>9</sup> Information Processing & Management (IPM)

<sup>10</sup> Journal of the American Society for Information Science and Technology (JASIST)

<sup>11</sup> Journal of Documentation (JDoc)

<sup>12</sup> Journal of Information Science (JIS)

<sup>13</sup> Scientometrics (SciMetr)

کتابسنجی که خوشه بزرگ با کتابسنجی کاربردی یا ارزیابی تحقیقات و کوچکتر با موضوعات روش‌شناختی و نظری، خوشه بزرگ دیگر در ارزیابی اطلاعات و موضوعات عمومی و گوناگون و در نهایت دو خوشه کوچک نوظهور در وب‌سنجی و مطالعات ثبت اختراعات و فناوری را نشان داد. البته در خوشه ارزیابی اطلاعات خوشه کوچک و موقت نیز دیده شده که ناشی از ویژه‌نامه مجله JASIST در این موضوع بود. مطابق انتظارات ارزیابی اطلاعات، موضوعات عمومی و وب‌سنجی موضوعات مورد بحث چهار مجله JASIST، IPM، JIS و JDoc بودند؛ در حالی که کتابسنجی و ثبت اختراعات موضوع اصلی مجله Sci بود. نتایج حاصل از مطالعات نشان داد 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 این‌بار خوشه‌بندی ترکیبی براساس معکوس کای دو فیشر<sup>۱</sup> جهت ترکیب متن کامل با استنادها 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 دی‌مور، ۲۰۰۸).

از دیگر مطالعات می‌توان به مطالعه ساختار حوزه با استفاده از واژگان عناوین ۱۰۳۴۴ مقاله منتشره شده 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 داد. 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 هدایت می‌شود (میلوژویک<sup>۲</sup>، ساگیموتو<sup>۳</sup>، یان<sup>۴</sup> و دینگ<sup>۵</sup>، ۲۰۱۱).

در کنار مطالعات جهانی توجه به مطالعات منطقه‌ای و کشوری در این حوزه دیده می‌شود از جمله این مطالعات به بررسی ۳۱۲۱ رساله دکتری طی سال‌های ۱۹۳۰ و ۲۰۰۹ در برنامه‌های علم کتابداری و اطلاع‌رسانی آمریکای شمالی اشاره نمود. در این مطالعه از تخصیص پنهان دیریکله<sup>۶</sup> جهت شناسایی تحولات زبانی موضوعات پنهان در طول تاریخ و شناسایی پایان‌نامه‌های نماینده آن موضوعات استفاده گردید (ساگیموتو، لی<sup>۷</sup>، راسل<sup>۸</sup>، فاینلی<sup>۹</sup> و دینگ، 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 داد 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 (۳) بیشتر موضوعات حاشیه‌ای و رشدنیافته هستند (هو<sup>۱۰</sup>، هو، دنگ<sup>۱۱</sup> و لیو<sup>۱۲</sup>، ۲۰۱۳). هم‌زمان مطالعه‌ای مشابه بر روی رساله‌های دکتری در این رشته در این

<sup>1</sup> Fisher's inverse chi-square

<sup>2</sup> Milojević

<sup>3</sup> Sugimoto

<sup>4</sup> Yan

<sup>5</sup> Ding

<sup>6</sup> Latent Dirichlet Allocation (LDA)

<sup>7</sup> Li

<sup>8</sup> Russell

<sup>9</sup> Finlay

<sup>10</sup> Hu

<sup>11</sup> Deng

<sup>12</sup> Liu

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

کشور طی ساله‌های ۱۹۹۴ تا ۲۰۱۱ صورت گرفت. نتایج تنوع موضوعی در زمینه‌های تحقیقاتی با تمرکز بر موضوعات منابع اطلاعاتی، هستی‌شناسی، وب‌معنایی، 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<sup>۱</sup> و همکاران، ۲۰۱۳).

استفاده از تحلیل هم‌رخدادی به منظور توسعه روش‌ها این بار با محاسبه مقادیر اصطلاحات موضوعی بین‌رشته‌ای<sup>۲</sup> همراه با مقادیر شرطی، بسامد واژگان و ... 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<sup>۳</sup>، گو<sup>۴</sup>، یو<sup>۵</sup>، رو<sup>۶</sup> و فنگ<sup>۷</sup>، ۲۰۱۶).

در ایران نیز استفاده از تحلیل هم‌رخدادی در کنار مطالعه ساختار حوزه‌های مختلف از جمله حوزه فناوری نانو (محمدی، ۱۳۸۸؛ تیمورپور، سپهری و پزشک، ۱۳۸۸)، روانشناسی ازدواج در ادبیات زیست‌پزشکی (توکلی‌زاده‌راوری و نجابتیان، ۱۳۸۹)، حوزه جراحی (شیرشاهی، عصاره، لونی و حیدری، ۱۳۹۳) و ... در گرایش‌های حوزه علم اطلاعات و دانش‌شناسی نیز مورد استفاده قرار گرفته است. از این مطالعات می‌توان به مطالعه (موسوی زاده، باقری و کربلا آقایی کامران، ۱۳۹۳) در حوزه سازماندهی اطلاعات در ۶ نشریه علوم کتابداری و اطلاع‌رسانی ایران طی سال‌های ۱۳۸۴-۱۳۸۸ نام برد. برای تعیین مقالات حوزه ابتدا سیاهه‌ای با عنوان راهبرد جستجو با بررسی اصطلاح‌نامه‌های "لیستا"<sup>۸</sup>، "یونسکو"<sup>۹</sup>، "ایسیس"<sup>۱۰</sup> و چکیده مقالات همایش "رویکردها و راهکارهای نوین در سازماندهی اطلاعات" ایجاد شد. اما به دلیل نبود پایگاه منسجم جهت انجام جستجو، با مراجعه به نشریات و مطالعه بخش‌های عنوان، چکیده و کلیدواژه، در نهایت ۱۲۵ مقاله دارای این اصطلاحات استخراج گردید. در نهایت تصویر حاصل سه ناحیه (۱) ابر داده، وب و بازیابی اطلاعات (۲) اصطلاح‌نامه‌ها و (۳) فهرست‌نویسی و نمایه‌سازی را برای حوزه سازماندهی اطلاعات نشان داد.

کاتبی در پایان‌نامه کارشناسی ارشد خود، به تحلیل هم‌رخدادی واژگان شکل‌دهنده زیرساخت مدیریت دانش در پایگاه وب آو ساینس در بازه زمانی ۱۹۹۳-۲۰۱۲ به منظور مصورسازی واژگان و تحلیل ارتباط شکل گرفته میان این واژگان پرداخت. نتایج پژوهش نشان داد که زیرساخت مدیریت دانش بر پایه و اساس هفت خوشه واژگان شکل گرفته است که متعلق به هفت حوزه موضوعی قابلیت‌های دانش انسانی، تعاملات متقابل انسانی، عملکرد فکری، مدیریت فکری، تفکر سیستمی، همکاری و پشتیبانی است. مهم‌ترین واژگان تشکیل‌دهنده حوزه مدیریت دانش از لحاظ بسامد تکرار، واژگان مدیریت دانش، اشتراک دانش و هستی‌شناسی بوده و واژگان قابلیت‌های پویایی، دانش آشکار و نظام‌های مبتنی بر دانش از بیشترین هم‌جواری و پیوند با سایر واژگان این حوزه برخوردارند (کاتبی، ۱۳۹۳).

(احمدی و کوکی، ۱۳۹۴) پژوهشی با هدف شناخت پیوند و مرز دو حوزه مدیریت اطلاعات و مدیریت دانش با

<sup>1</sup> Zong

<sup>2</sup> Topic Terms Interdisciplinarity

<sup>3</sup> Xu

<sup>4</sup> Guo

<sup>5</sup> Yue

<sup>6</sup> Ru

<sup>7</sup> Fang

<sup>8</sup> Library, Information Science & Technology Abstract (LISTA)

<sup>9</sup> United Nations Educational, Scientific and Cultural Organization (Unesco)

<sup>10</sup> American Society for Information Science (ASIS)

استفاده از روش تحلیل همبندی واژگان بر روی مقالات نشریات علمی و همایش‌های داخلی انجام دادند. نتایج پژوهش حاکی از این بود که مدیریت دانش حوزه‌ای است که تا حدودی زمینه فعالیت رشته مدیریت با تمام گرایش‌های آن است و مدیریت اطلاعات نیز زمینه مطالعاتی رشته‌هایی، همچون مدارک پزشکی و علم اطلاعات هست و رشته علم اطلاعات مرز مشترک برای هر دو حوزه محسوب می‌شود.

مطالعه (حاضری، توکلی‌زاده راوری و ابراهیمی، ۱۳۹۴) نیز با هدف بررسی ساختار موضوعی حوزه مدیریت دانش با تحلیل کلیدواژه‌های مدارک بازاریابی شده از نتایج جستجوی عبارت "Knowledge Management" در فیلد موضوع و بررسی در فیلدهای عنوان، چکیده، کلیدواژگان نویسنده و نمایه‌ساز در پایگاه وب‌آوساینس صورت گرفت. نتایج نشان داد حوزه‌های مدیریت، علوم کامپیوتر-سامانه‌های اطلاعاتی و علم اطلاعات/ کتابداری به ترتیب بیشترین سهم در تعداد مدارک مربوط به این حوزه را دارند. تحلیل هم‌واژگانی ۹۶ کلیدواژه پرکاربرد از کل ۵۵۷۰ کلیدواژه در سه حوزه مورد بررسی، با روش سلسله‌مراتبی وارد<sup>۱</sup> خوشه‌بندی شد. نتایج حوزه علم اطلاعات/ کتابداری را دارای بیشترین کلیدواژه‌های مشترک با دو حوزه دیگر معرفی کرد.

ترسیم نقشه علم‌سنجی نیز با هدف شناخت سیر تحول پژوهش‌ها و مفاهیم علمی حوزه علم‌سنجی در ایران صورت گرفت. این پژوهش در مجموع ۲۵۸۵ مدرک شامل کلیه بروندهای تولید شده توسط پژوهشگران ایران در قالب کتاب‌ها، مقاله‌ها، پایان‌نامه‌های تحصیلی و طرح‌های پژوهشی حوزه علم‌سنجی در داخل و خارج از ایران اعم از چاپی و الکترونیکی به زبان فارسی و انگلیسی از سال ۱۳۶۱-۱۳۹۳ است. نتایج ۱۸۰۰ مفهوم مطرح را در قالب ۱۸ خوشه موضوعی با مفاهیم غالبی چون "ارزیابی تولیدات علمی، تولید علم و تحلیل استنادی" را نشان داد. همچنین بررسی‌ها حاکی از ارتباط میان‌رشته‌ای حوزه با ۲۲ رشته علمی دیگر همچون فناوری اطلاعات، علوم اجتماعی و مدیریت بود (احمدی و عصاره، ۱۳۹۴).

(خادمی و حیدری، ۱۳۹۵) ساختار موضوعی حوزه مدیریت اطلاعات را به منظور سیاست‌گذاری علمی مورد بررسی قرار دادند. جهت گردآوری متون جستجوی کلیدواژه "Information Management" در فیلد موضوع پایگاه وب آو ساینس در حوزه مدیریت اطلاعات در بازه زمانی ۱۹۸۶ تا ۲۰۱۲ انجام گرفت که در مجموع ۴۵۷۱ مدرک نمایه شده از پایگاه‌های نمایه استنادی علوم، نمایه استنادی علوم اجتماعی و نمایه استنادی علوم انسانی و هنر بازاریابی شد. نتایج ۹ خوشه موضوعی را نشان داد که شامل مدیریت اطلاعات، بیوانفورماتیک، مدیریت عملکرد، اطلاعات، داده، مدیریت اطلاعات سلامت، سیستم‌های اطلاعاتی، سلامت الکترونیک و سازماندهی دانش می‌باشد.

از دیگر پژوهش‌های انجام شده می‌توان به مطالعه جنبه‌های عاطفی در رفتار جستجو و بازاریابی اطلاعات در وب براساس ماتریس همبندی موضوعات در فیلد subject category (یاری‌زنگنه، حریری و باب‌الحوایجی، ۱۳۹۵) و روند تکامل فناوری: مورد مطالعه تحلیل رده‌های موضوعی پروانه‌های ثبت اختراع RFID (منصوری، توکلی‌زاده راوری، مکی‌زاده و طوسی، ۱۳۹۶) نام برد.

اما مرتبط‌ترین مطالعات در این حوزه در ایران، شناسایی ساختار واژگان چکیده مقالات "علم اطلاعات و دانش-شناسی" با استفاده از نگاشت‌های علمی هم‌واژگانی در پایگاه وب‌گاه علم (۲۰۰۹-۲۰۱۳) است. یافته‌ها نشان می‌دهد 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

<sup>1</sup> Ward



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).

مطالعه پیشینه‌ها هر چند اهمیت نگاشت ساختار رشته‌ها براساس 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 براساس متون بیشتر به ویژه متون هسته و مطالعه روند گسترش این مفاهیم در طی سال‌های مختلف و دستیابی به مفاهیم هسته براساس متون هسته در سطح بین‌المللی جهت مقایسه با مفاهیم داخلی 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 محتوا است که در مطالعات علم‌سنجی و برنامه‌های تحقیق و توسعه کاربرد دارد. بدین‌صورت که از الگوی هم‌رخدادی اقلام (مثل: واژگان یا عبارات یعنی مفاهیم) در یک مجموعه از متون جهت شناسایی روابط بین انگاره‌های موجود در حوزه‌های موضوعی ارائه شده به شکل متون استفاده می‌شود. جهت انجام مراحل مختلف تحقیق از بسته آزمایشی Provalis Research<sup>1</sup>، نرم‌افزارهای Word و Excel، Adobe Acrobat XI 11<sup>2</sup>، ABBYY FineReader 12<sup>3</sup> و Internet Download Manager<sup>4</sup> 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 گردید.

مطابق جدول ۱ به طور 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 گردید. در بین منابع مجله Scientometrics با ۲۵،۲۵ درصد و مجله Information and Organization با ۱،۹۱ درصد به ترتیب بیشترین و کمترین تعداد متون را دارا می‌باشند.

ب) پاک‌سازی متون: استخراج واژگان از متن کامل به جای عناوین و چکیده‌ها، مشکلات و پیچیدگی‌های مضاعفی دارد به طوری که مستلزم موارد پیش‌پردازش خاص بوده و به تبع 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

<sup>1</sup> <http://provalisresearch.com/>

<sup>2</sup> <http://www.adobe.com/>

<sup>3</sup> <http://www.abbyy.com/finereader/>

<sup>4</sup> <https://www.internetdownloadmanager.com/>



تصحیح مقالات داری غلطنامه و متون تصویری استخراج شده با اسی آر صورت گرفت.

ج) وارد کردن داده‌های متنی در نرم‌افزار: داده‌های متنی ذخیره شده در فایل‌ها بعد از آماده‌سازی در قالب یک پیکره<sup>۱</sup> وارد نرم افزار کیودی‌ای ماینر گردید و هریک از متون به صورت یک کیس<sup>۲</sup> و فایل‌ها با عناوین مجلات در کنار ۵ دوره (۲۰۰۴-۲۰۰۵، ۲۰۰۶-۲۰۰۷، ۲۰۰۸-۲۰۰۹، ۲۰۱۰-۲۰۱۱ و ۲۰۱۲-۲۰۱۳) به عنوان متغیر تعریف شدند.

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

## ۲) پیش‌پردازش متون

در تحلیل هم‌واژگانی، واژه‌ها و عبارات پایه اصلی تحقیق محسوب می‌شوند، که بسته به هدف و روش تحقیق نوع گردآوری واژه‌ها نیز متفاوت است. همانطور که در بررسی پیشینه روش‌شناسی پژوهش دیدیم جهت استخراج واحدهای تحلیل هم‌واژگانی از مقالات مجلات، کنفرانس‌ها، گزارش‌ها و حتی فصل‌های کتاب از دو روش پارامتریک و ناپارامتریک به عنوان روش‌های اساسی استفاده شده بود. در این پژوهش از روش پارامتریک جهت استخراج واژگان و عبارات با استفاده از نرم‌افزار ورداستیت در مراحل ذیل صورت گرفت.

الف) حذف کلمات بازدارنده<sup>۳</sup>: جهت حذف واژه‌هایی که دارای ارزش معنایی کمتری هستند، از قبیل: ضمایر، حروف ربط و غیره مثل: "about"، "near"، "and" و ... از لیست بازدارنده مختص زبان انگلیسی نرم افزار ورداستیت<sup>۴</sup> با تغییراتی جهت بازیابی عبارات استفاده گردید.

ب) حذف آدرس‌های اینترنتی و برخی اسامی خاص: ویژگی اصلی آدرس‌های اینترنتی و اسنادهای درون متنی، هاپر لینک یا لینک‌دار بودن آنهاست، این ویژگی توسط نرم افزار قابل شناسایی بوده و در متن‌کاوی حذف می‌شود. اما در اغلب متون این ویژگی رعایت نشده، بنابراین از ویژگی این نوع کلمات مثل داشتن "." استفاده گردید.

<sup>۱</sup> Corpus

<sup>۲</sup> Case

<sup>۳</sup> Stop Words

<sup>۴</sup> WordStat

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

بدین جهت با تعریف این محدودیت در استخراج، واژگان دارای نقطه به صورت یکجا از طریق نرم افزار ورداستیت، بازیابی و سپس از طریق مرور آدرس‌هایی دارای **www** و یا برخی ویژگی‌ها 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"** می‌باشد. ورد استیت این امکان را ایجاد کرده تا عبارات متداول در متون استخراج گردد ولی نمی‌توان آن‌ها را به طور معمول با لیست واژگان ترکیب کرد. برای استخراج همزمان آن‌ها با واژگان از **"X"** به جای فاصله و یا **"-"** استفاده گردید بدین صورت که با توجه به وزن آستانه واژگان عباراتی مثل **"LIBRARY AND INFORMATION"** با **"LIBRARYXANDXINFORMATION"** جایگزین و استخراج گردید.

**د) استانداردسازی لغات:** از اساسی‌ترین کارکردهای استاندارسازی لغات، می‌توان به یکدستی واژه‌ها 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 ورداستیت استفاده شد که مبتنی بر روش فرهنگ‌نامه‌ای تعدیل شده است. بدین صورت که قسمتی مبتنی بر الگوریتم جایگزینی پسوند کروتز<sup>۱</sup> یعنی **KSTEM** می‌باشد. همچنین به دلیل استفاده از یک لیست جایگزین به جای برچسب‌زنی اجزای کلامی واژه‌ها، در برخی موارد استثنایی از نظر معنایی نامعتبر است، اما افزایش سرعت پردازش و قابلیت بازبینی و تغییر در لیست جایگزین برای متن‌کاو را فراهم می‌نماید.

### ۳) انتخاب مفاهیم

انتخاب مفاهیم برپایه روش‌های متنوعی صورت می‌گیرد، از جمله این روش‌ها استفاده از ایده لوهن<sup>۲</sup> (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، چو دبری<sup>۳</sup> و فو<sup>۴</sup>، ۲۰۰۱؛ لی<sup>۵</sup> و جونگ<sup>۶</sup>، ۲۰۰۸؛ محمدی، ۲۰۱۲؛ هو و همکاران، ۲۰۱۳؛ ناصری‌جزه، طباطبائیان و فاتح راد، ۱۳۹۱) و... می‌باشد.

در این پژوهش، بنابر تعداد زیاد لغات به ویژه در تحلیل متون تمام‌متن و نبود امکان بازبینی جهت گروه‌بندی ذیل یک واژه، تمایل به انتخاب واژگان خاص در مجموعه مدارک برحسب ویژگی **IDF** از روش وزن‌دهی **TF-IDF** برای ۱۵۰ واژه منتخب استفاده گردید. این روش که عمومیت بیشتری دارد خصوصیت وزن‌دهی فراوانی معکوس متن را با فراوانی واژه ترکیب می‌کند، که توسط سالتون<sup>۷</sup> و یانگ<sup>۸</sup> مطرح شده است. بنابراین هرچه قدر رخداد یک واژه در یک مدرک بیشتر باشد، محتوای متن را بهتر نشان می‌دهد. ولی هرچه تعداد مدارکی که واژه در

<sup>1</sup> Krovetz

<sup>2</sup> Luhn

<sup>3</sup> Chowdhury

<sup>4</sup> Foo

<sup>5</sup> Lee

<sup>6</sup> Jeong

<sup>7</sup> Salton

<sup>8</sup> Yang

آنها رخ می‌دهد بیشتر باشد، از شاخص بودن آن کاسته می‌شود (بایز-یتس<sup>۱</sup> و ریبریو-نتو<sup>۲</sup>، 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 براساس چه نوع هم‌رخدادی تشکیل شوند. در واقع دو نوع خویشاوندی وجود دارد، یکی براساس قرابت مرتبه اولیه<sup>۳</sup> که کلمات را براساس قرار گرفتن کنار هم در یک مدرک و دیگری قرابت مرتبه ثانویه<sup>۴</sup> که مفاهیم براساس 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. به عنوان مثال زمانی که TUMOUR و TUMOR اغلب و یا اصلاً با هم رخ نمی‌دهند این روش آنها را نزدیک به هم دانسته چرا که هر دوی آنها با کلماتی چون BRAIN و CANCER رخ می‌دهند. جالب‌تر گروه‌بندی مترادف‌های نزدیک و کلمات مرتبط از نظر معنایی<sup>۵</sup> است مثل MILK, JUICE و WINE. چرا که آنها گرایش به ترکیب شدن با افعال مشابهی مثل DRINK یا POUR و اسامی مثل GLASS دارند (گرفنسته<sup>۶</sup>، ۱۹۹۴).

ماتریس ۱۵۰×۱۵۰ از هم‌رخدادی مفاهیم در واقع به عنوان داده‌های ورودی برای نگاشت مفاهیم می‌باشد، اما برای تبدیل شدن به نقشه باید با استفاده از سنج‌های مختلف نرمال‌سازی شده و تبدیل به ماتریس شباهت شود. بدین منظور از ضریب کسینوس تتا<sup>۷</sup> 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<sup>۸</sup>، ۲۰۱۵) و به صورت ذیل محاسبه می‌شود.

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

#### ۵) مصورسازی و تحلیل داده‌ها

به منظور رسیدن به نمای کلی از نگاشت مفهومی<sup>۹</sup> همانند مطالعات هم‌واژگانی، از تحلیل آماری چند متغیره<sup>۱۰</sup>، تحلیل شبکه‌ای<sup>۱۱</sup> 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 بکارگیری چندین سنج

<sup>1</sup> Baeza-Yates

<sup>2</sup> Ribeiro-Neto

<sup>3</sup> First Order Clustering

<sup>4</sup> Second Order Clustering

<sup>5</sup> Semantically

<sup>6</sup> Grefenstette

<sup>7</sup> Cosine Theta

<sup>8</sup> Provalis Research

<sup>9</sup> Concept Maps

<sup>10</sup> Multivariate Analysis Methods

<sup>11</sup> Link Analysis

بر روی هر شی یا فرد در یک یا چند نمونه است. سنجه‌ها به عنوان متغیر و اشیاء و افراد به عنوان اجزای تحقیق، نمونه و یا آزمایش هستند (رنچر<sup>۱</sup>، ۲۰۰۳). روش تحلیل چندمتغیره اغلب به عنوان یک معیار یا رویکرد اکتشافی کاربرد دارد تا روشی برای آزمون فرضیات واقعی (ساساکی<sup>۲</sup> و سوزوکی<sup>۳</sup>، 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<sup>۴</sup> و مقیاس‌بندی چندبعدی و تحلیل تناظر<sup>۵</sup> بوده و به صورت دندوگرام‌ها، نقشه‌های مختلف مقیاس‌بندی سه‌بعدی و حرارتی مصورسازی و مورد تحلیل قرار می‌گیرد.

## یافته‌های پژوهش

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

**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** AUCTION DOMAIN STRATEGY SUPPLIER  
SELLER IMPLEMENTATION SOFTWARE INTENTION CHINESE GOVERNMENT 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

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

شکل ۱ نقشه ابری واژگان پایه متون تمام‌متن بین‌الملل

۲. براساس واژگان پایه علم اطلاعات و دانش‌شناسی چه خوشه‌های موضوعی قابل شناسایی است؟

براساس تحلیل خوشه‌ای و قرابت درجه دو مفاهیم، دندوگرام حاصل از مفاهیم انتخابی در متون تمام 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

<sup>1</sup> Rencher

<sup>2</sup> Sasaki

<sup>3</sup> Suzuki

<sup>4</sup> Clustering

<sup>5</sup> Correspondence Analysis

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، هزینه؛ (۲) نظریه‌ها و فرضیه‌ها با مفاهیم رفتار، متغیر، مورد، فرضیه، ارتباطات، اجتماعی، عضو، دانش، اشتراک، فناوری، مدل، ارتباط، نظام، منابع، عملکرد، پییر، شبکه، ساختار، نظریه، یادگیری، شرکت کننده، سایت، 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 می‌شود. 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، موسسه، صنعت، بازار و نوآوری؛ (۷) مذاکره شامل لغات قصد، درک کردن، اعتماد، مذاکره، پاسخگو و امنیت؛ (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) نگاهت 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، موسسات، انتشار، دانشگاه، استناد، مجله، انتشارات، علمی و رشته؛ (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<sup>۱</sup>، 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 بازیابی 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<sup>۲</sup> 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.

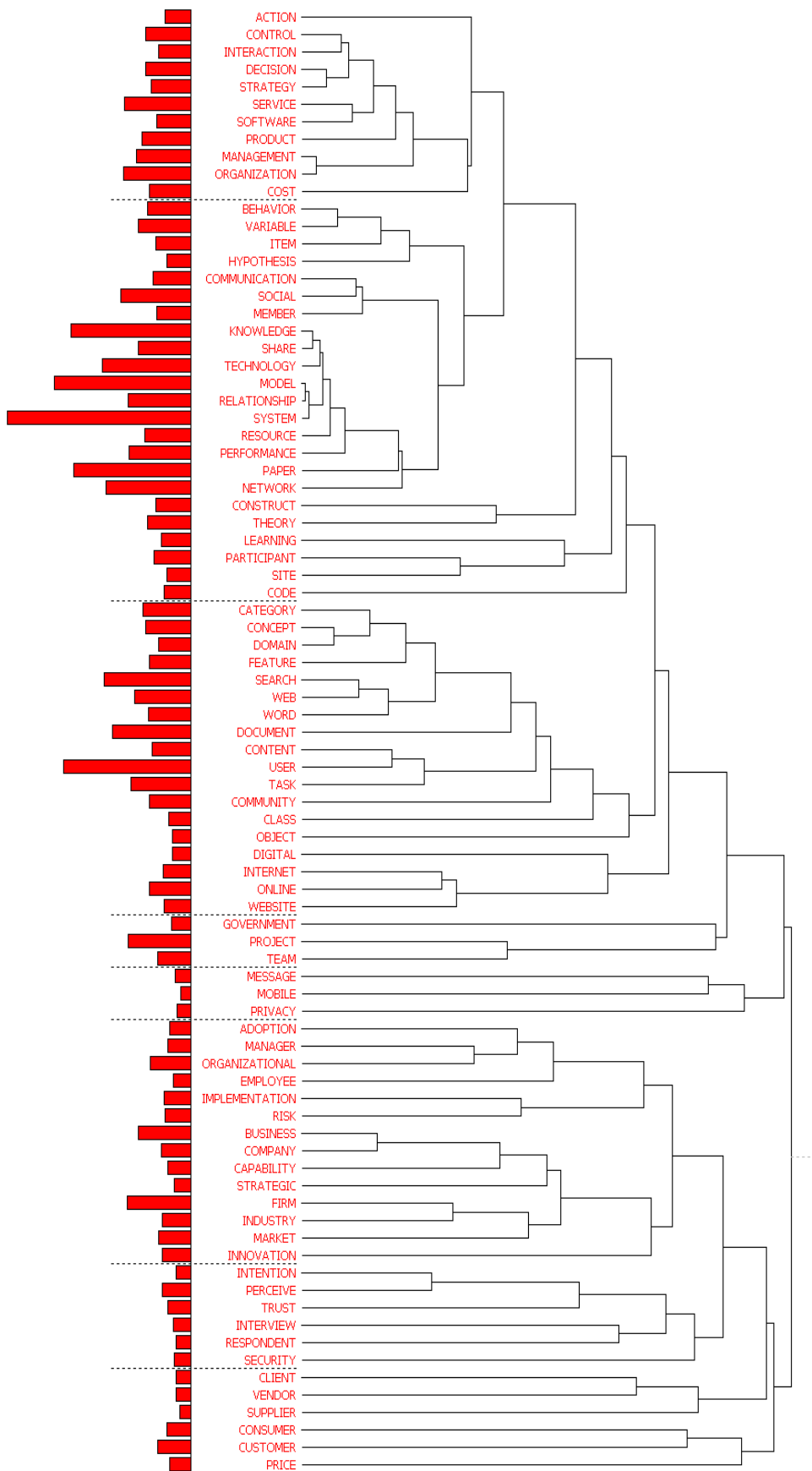
خوشه هفتم متشکل از چهار اصطلاح دانشمند، استناد، ضریب تأثیر و شاخص اچ؛ دیگر خوشه کوچک موجود با واژگان تخصصی حوزه علم‌سنجی است. با توجه به تعداد مفاهیم و ضعیف بودن خوشه و نیز شبکه واژگانی آن با

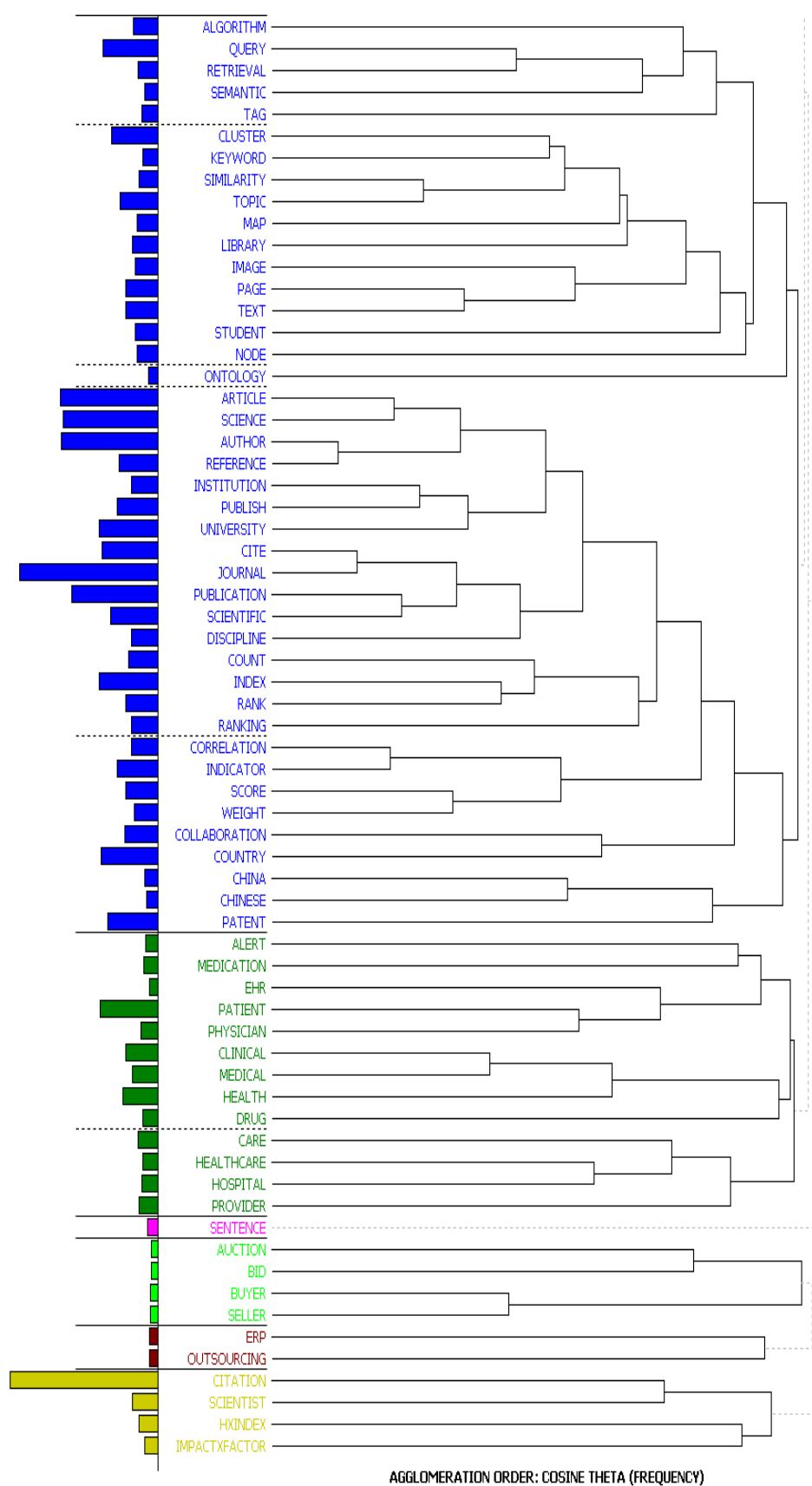
<sup>1</sup> Electronic Health Record

<sup>2</sup> Enterprise Resource Planning (ERP)

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

مباحث علم‌سنجی از خوشه دوم دارای ارتباط بوده و به عبارتی خوشه اشتقاق یافته از این خوشه بزرگ می‌باشد.



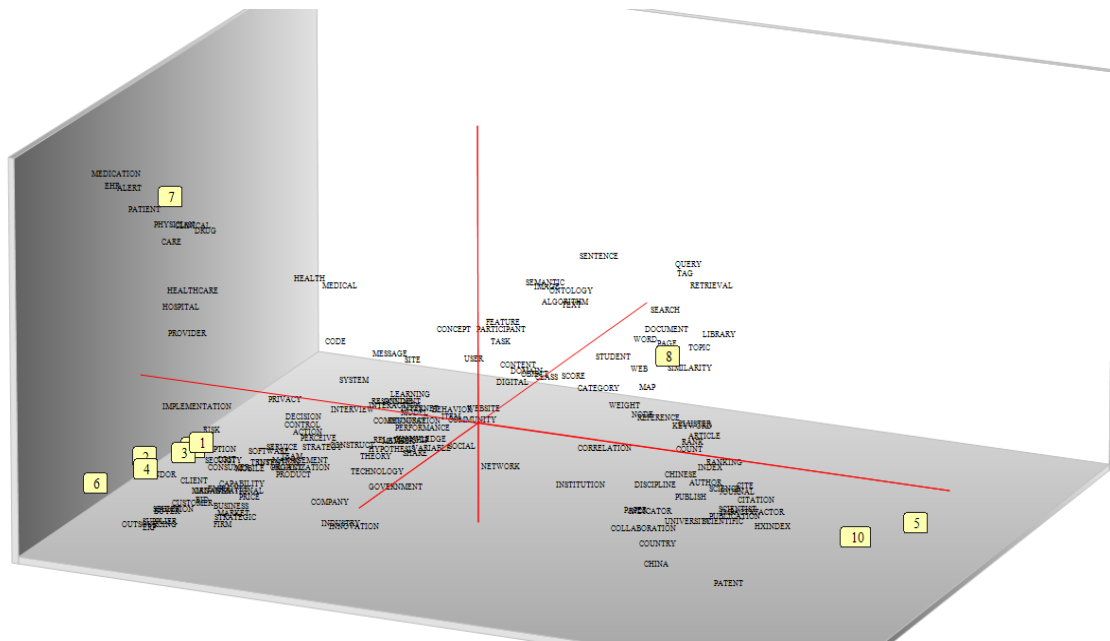


شکل ۲ دندوگرام خوشه‌بندی سطح دو مفاهیم تمام متن بین الملل براساس کسینوس تتا



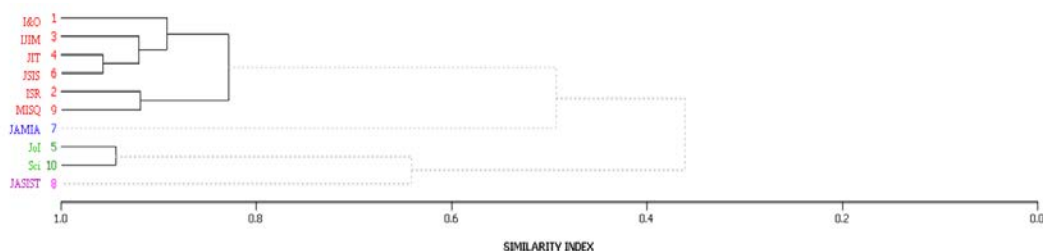
### ۳. دامنه شناختی مجلات هسته براساس مفاهیم منتخب چگونه است؟

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 براساس حضور آن‌ها را نشان می‌دهد. مجلات نزدیک به مرکز نمودار از طیف لغات عام و مشترک تشکیل گردیده 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. به طوری که چهار خوشه موضوعی قابل شناسایی است. اولین خوشه، با بیشترین تعداد مجله شامل MISQ, I&O, JIM, ISR, JIT 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. در نهایت 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 خوشه‌ها است، که خوشه سوم با چهارم همبستگی مثبت و بقیه با هم همبستگی منفی دارند. به عبارتی خوشه‌های دارای همبستگی منفی از نظر دارا بودن واژگان خاص شباهت کمتری با هم دارند.



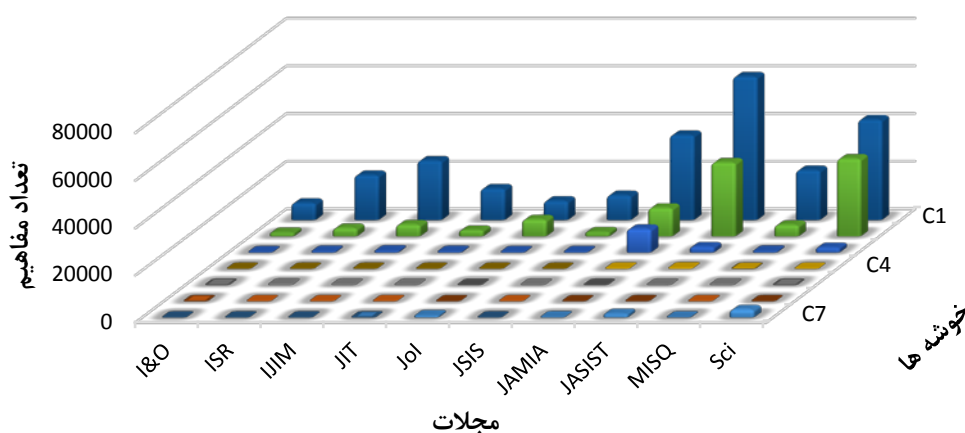
شکل ۳ همبستگی فاصله‌ای مجلات براساس بسامد مفاهیم منتخب متون تمام متن بین‌الملل

برای بررسی دقیق‌تر این شباهت‌ها براساس کاربرد واژه‌ها نمودار حرارتی با خوشه‌بندی مجلات و واژه‌ها ترسیم گردید. این خوشه‌بندی براساس فراوانی نسبی، شباهت توزیع واژه‌ها را در بین گروه‌های مختلف از متغیر مستقل را نشان می‌دهد و قابلیت خوشه‌بندی مجلات و دوره‌های مختلف را دارد. برش در نقطه  $0/7$  شاخص شباهت در دندوگرام مجلات، خوشه‌بندی مجلات به چهار خوشه در تحلیل تناظر را تایید کرد. چنانچه بیشترین شباهت بین



مجله JIT با JSIS و نیز JoI با Sci است.

شکل ۴ دندوگرام مجلات بین‌المللی براساس شاخص شباهت فراوانی نسبی مفاهیم منتخب متن کامل برای بررسی بیشتر، این بار نحوه توزیع مفاهیم هر یک از مجلات، در خوشه‌های حاصل از هم‌رخدادی واژگان در نمودار ۱ ترسیم شد. نحوه توزیع حاکی از آن است که خوشه اول و بزرگ، با بیشترین مفهوم در تمامی مجلات نیز بیشترین مفهوم را به خود اختصاص داده است. بنابراین چنانچه در تحلیل تناظر نیز گفته شده دارای بیشترین لغات عام در حوزه است و بیشترین مفهوم را در مجله JASIST دارد. خوشه دوم و هفتم نیز خوشه مرتبط با حوزه تخصصی علم‌سنجی در مجله Sci، خوشه سوم و مرتبط با حوزه پزشکی در مجله JAMIA و خوشه‌های چهارم باز در مجله JASIST، در نهایت خوشه پنج در ISR و خوشه ششم در MISQ نمودار بیشتری دارند. این نمودار نیز گویای ارتباط و انشقاق خوشه هفت از دو و خوشه چهار از یک و نیز تخصصی بودن مجلات در حوزه می‌باشد.

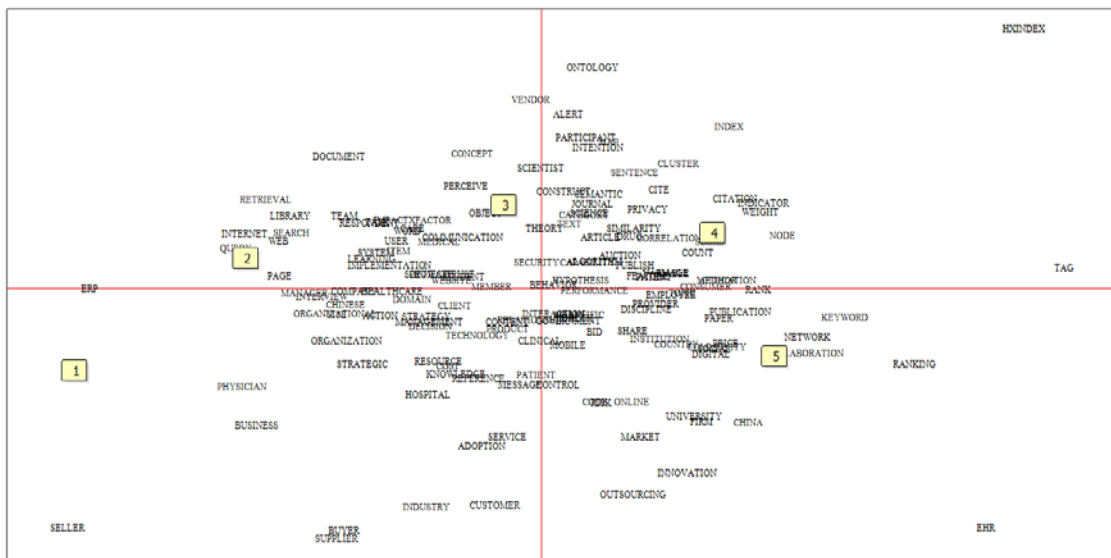


نمودار ۱ توزیع مفاهیم (متن کامل بین‌المللی) هر یک از مجلات در خوشه‌های مختلف

#### ۴. نحوه چیدمان مفاهیم طی سال‌های (۲۰۱۳-۲۰۰۴) چگونه بوده است؟

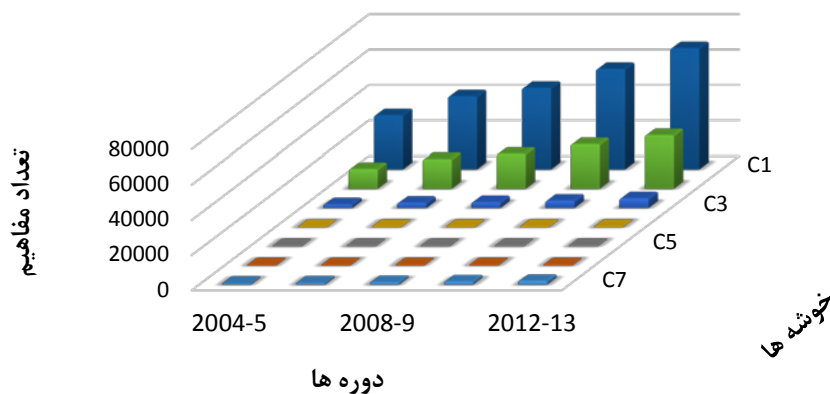
برای درک نحوه تحولات زمانی، بسامد مفاهیم براساس سال نشر به صورت دوره دوساله تحلیل و ارتباط و توزیع مفاهیم در سال‌های مذکور به صورت دوبعدی در شکل ۵ ارائه شد. در این تصویر مفاهیم مشابه از نظر زمانی کنار هم با شماره‌های مختص هر دوره خوشه‌بندی گردید. به عبارتی بازه‌های زمانی مشابه از نظر استفاده مفاهیم نزدیک 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 عقربه‌های ساعت دیده می‌شود. چنانچه فاصله زاویه بین مفاهیم سال‌های نخستین با سال‌های پایانی یک زاویه منفرجه بوده لذا همبستگی بین آن‌ها منفی و سال‌های نزدیک بهم نیز همبستگی

مثبت دارند. یعنی بین سال‌های ۲۰۰۴-۲۰۰۵ با سال‌های 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 می‌آورد.



شکل ۵ تحلیل تناظر بین مفاهیم منتخب تمام متن متون بین‌الملل و دوره‌های دو ساله

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، کوئری، 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، هستی-شناسی 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، برچسب بروز می‌کند. در آخرین دوره نیز مفاهیمی چون تولیدات، قیمت، جامعه، اجتماع، کلیدواژه، شبکه، همکاری، رتبه‌بندی، دانشگاه، چین، پرونده الکترونیک سلامت (EHR) نمود بیشتری می‌یابد. برای بررسی دقیق‌تر میزان رشد و توزیع مفاهیم خوشه‌ها در سال‌های مختلف نمودار ۲ ترسیم شد. بنابر 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 در 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 متن‌کامل مقالات، در ۷ خوشه و ۱۵ زیرخوشه براساس 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 ...؛ وب و اینترنت با مفاهیم دیجیتال، پیوسته، دسته‌بندی ویژگی‌های حوزه و مفاهیم، جستجو، محتوا، کاربران و ...؛ پروژه‌های تیمی دولتی و در نهایت تلفن همراه با مفاهیم امنیت و پیام کوتاه تقسیم می‌شود. شاخه مدیریت کسب و کار نیز به ۳ زیرخوشه به موضوعات راه‌اندازی کسب و کار شامل مدیر، کارمند، ریسک، نوآوری، بازار و ...؛ مذاکره در کنار مفاهیم درک، قصد، امنیت و اعتماد و پاسخ‌دهنده؛ و مقوله فروشنده، مشتری و قیمت و ... 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 برچسب؛ نگاشت با مفاهیم خوشه، کلیدواژه، موضوع، کتابخانه و ...؛ و در نهایت زیرخوشه سوم به هستی‌شناسی می‌پردازد. شاخه دوم حوزه علم‌سنجی نیز شامل زیرخوشه‌های تولیدات علمی شامل مقاله، پدیدآور، ارجاع، موسسات، دانشگاه و ...؛ شاخص‌های رتبه‌بندی با مفاهیم استناد، نشریات، رتبه، همبستگی، وزن و ...؛ و در نهایت همکاری‌های علمی شامل کشور، ثبت اختراع، چین و ... 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.

در واقع مفاهیم علم اطلاعات و دانش‌شناسی در تحلیل هم‌واژگانی براساس 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 گردند- 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 به ویژه مفاهیم دارای نماینده خاصی در مجلات هستند. حضور کم‌رنگ مفاهیم در مرکز نمودار به ویژه در متون تمام متن حاکی اشتراک مفهومی کم در بین مجلات است. یعنی در متون بین‌الملل مجلات به صورت تخصصی عمل می‌کنند و مقالات مرتبط با مفاهیم و موضوعات تشکیل دهنده 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 نشان از تداوم حوزه‌های پژوهشی داشته است. حوزه‌هایی که در تحلیل تناظر زمانی با چیدمان خاص مفاهیم، دو دوره تاریخی را تشکیل می‌دهد. چنانچه اگر دوره اول بر مبنای مفاهیم علم 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 الگوریتم‌ها است و می‌توان این دوره را دوره علم‌سنجی و بازیابی اطلاعات نام نهاد. به طور 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 چشم انداز شناختی در خوشه زمانی نخست مبتنی بر ظهور اینترنت بوده در حالی که در خوشه دوم به دلیل استفاده از شاخص‌ها است.

## منابع

- احمدی، ح. و عصاره، ف. (۱۳۹۴). سیر تحول پژوهش‌های علم‌سنجی در ایران. فصلنامه رهیافت، ۲۵ (۳)، ۶۹-۸۲.
- احمدی، ح. و کوبی، م. (۱۳۹۴). همایندی واژگان: مطالعه‌ای پیرامون پیوند و مرز میان مدیریت اطلاعات و مدیریت دانش بر اساس انتشارات داخلی نویسندگان ایرانی. پژوهشنامه پردازش و مدیریت اطلاعات، ۳۰ (۳)، ۶۴۷-۶۷۶.
- توکلی‌زاده‌راوری، م. و نجابتیان، م. (۱۳۸۹). خوشه بندی مبتنی بر مدرک - اصطلاح: هم جواری موضوعات

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

- روان‌شناسی ازدواج در ادبیات زیست‌پزشکی در دوره‌های زمانی «۱۹۹۹ - ۱۹۹۰» و «۲۰۰۸ - ۲۰۰۰». مدیریت اطلاعات سلامت، ۷(۲)، ۱۷۲-۱۸۶.
- 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.، توکلی زاده 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 همابندی واژگان طی سال‌های ۱۹۸۶ تا ۲۰۱۲. فصلنامه علوم و فنون مدیریت اطلاعات، ۲(۲)، ۵۹-۹۳.
- شیرشاهی، س.، عصاره، ف.، لونی، ن. و حیدری، غ. (۱۳۹۳). ترسیم نقشه علم‌نگاری تولیدات علمی حوزه جراحی، نمایه شده در نمایه استنادی علوم. مدیریت اطلاعات سلامت (ویژه‌نامه مدیریت سلامت و فناوری)، ۱۱(۷)، ۸۳۹-۸۳۰.
- کاتبی، ف. (۱۳۹۳). تحلیل هم‌رخدادی واژگان حوزه مدیریت دانش در پایگاه وب آو ساینس در سال‌های ۱۹۹۳ لغایت ۲۰۱۲ به منظور 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 توکلی زاده 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>
- منصوری، ع.، توکلی زاده راوری، م.، مکی زاده، ف. و طوسی، ز. (۱۳۹۶). روند تکامل فناوری: مورد مطالعه تحلیل رده‌های موضوعی پروانه‌های ثبت اختراع RFID. پژوهشنامه پردازش و مدیریت اطلاعات، ۳۳(۱)، ۲۹۵-۳۱۰.
- موسوی زاده، 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.، طباطبائیان، 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.
- یاری زنگنه، م.، حریری، ن. و باب‌الحوایجی، ف. (۱۳۹۵). مطالعه جنبه‌های عاطفی در رفتار جستجو و بازیابی اطلاعات در وب: متن پژوهی و ترسیم نقشه علمی. پژوهشنامه پردازش و مدیریت اطلاعات، ۳۱(۴)، ۹۵۳-۹۷۹.
- 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.
- Braam, R. R., Moed, H. F., & van Raan, A. F. J. (1991a). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for*

*Information Science*, 42(4), 233–251.

Braam, R. R., Moed, H. F., & van Raan, A. F. J. (1991b). Mapping of science by combined co-citation and word analysis. II: Dynamical aspects. *Journal of the American Society for Information Science*, 42(4), 252–266.

Callon, M. (1986). Pinpointing Industrial Invention: An Exploration of Quantitative Methods for the Analysis of Patents. In M. Callon, J. Law, & A. Rip (Eds.), *Mapping the Dynamics of Science and Technology* (pp. 163–188). London: Palgrave Macmillan UK.

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22(1), 155–205.

Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2), 191–235.

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).

Healey, P., Rothman, H., & Hoch, P. K. (1986). An experiment in science mapping for research planning. *Research Policy*, 15(5), 233–251.

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.

Kostoff, R. N. (1993). Co-Word Analysis. In *Evaluating R&D Impacts: Methods and Practice* (pp. 63–78). Boston, MA: Springer US.

Kostoff, R. N. (1994). Database tomography: Origins and duplications. *Competitive Intelligence Review*, 5(1), 48–55.

Law, J., Bauin, S., Courtial, J.-P., & Whittaker, J. (1988). Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics*, 14(3-4), 251–264.

Law, J., & Whittaker, J. (1992). Mapping acidification research: A test of the co-word method. *Scientometrics*, 23(3), 417–461.

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.

Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. *Journal of the American Society for Information Science*, 48(5), 418–427.

Leydesdorff, L. (1989). Words and co-words as indicators of intellectual organization. *Research Policy*, 18(4), 209–223.

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.

Rip, A., & Courtial, J.-P. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, 6(6), 381-400.

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

Smadja, F. A. (1989). Lexical Co-occurrence: The Missing Link. *Literary and Linguistic Computing*, 4(3), 163-168.

Sugimoto, C. R., Li, D., Russell, T. G., Finlay, S. C., & Ding, Y. (2011). The shifting sands of disciplinary development: Analyzing North American Library and Information Science dissertations using latent Dirichlet allocation. *Journal of the American Society for Information Science and Technology*, 62(1), 185-204.

Tijssen, R. J. W., & Van Raan, A. F. J. (1989). Mapping co-word structures: A comparison of multidimensional scaling and leximappe. *Scientometrics*, 15(3-4), 283-295.

Turner, W. A., Chartron, G., Laville, F., & Michelet, B. (1988). PACKAGING INFORMATION FOR PEER REVIEW ~~WORD ANALYSIS~~ TECHNIQUES. In A. F. J. van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology* (pp. 291-323). Amsterdam: Elsevier.

Van Raan, A. F. J., & Tijssen, R. J. W. (1993). The neural net of neural network research: AN EXERCISE IN BIBLIOMETRIC MAPPING. *Scientometrics*, 26(1), 169-192.

Whittaker, J. (1989). Creativity and Conformity in Science: Titles, Keywords and Co-word Analysis. *Social Studies of Science*, 19(3), 473-496.

Xu, H., Guo, T., Yue, Z., Ru, L., & Fang, S. (2016). Interdisciplinary topics of information science ~~Science studies based on the terms interdisciplinary~~. *Scientometrics*, 94(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 ~~word analysis~~. *Scientometrics*, 94(2), 781-799.

## Mapping the cognitive structure and its evolution in "Knowledge and Information Science": text mining approach (2004-2013)

### Abstract

**Purpose:** The purpose of this study was to map and study the cognitive structure of the field of "Knowledge and Information Science" from the perspective of full text words or phrases, their distribution in journals, and their usage over time.

**Methodology:** Methodology of this study was content analysis of 6830 articles (2004-2013) in 10 core journals based on the JCR 2013. Parametric way of key extraction has been used to select 150 words and phrases according to their TF-IDF weight. Finally, the similarity matrix based on the Cosine theta index and the second-order affinity, is used to construct hierarchical clustering by average-linkage algorithm in "Provalis Research" software. Also, correspondence analysis was used to study concepts historic evaluation and to cluster the journals according to their subject domain.

**Findings:** The results of this study included 3 mature clusters, 1 semi-mature cluster and 3 immature clusters in co-word analysis and the thematic review reveal four clusters of journals and two time clusters in clockwise motion layout.

**Conclusion:** The results show that publication has some well-established topics which



are changing gradually to adopt new themes. These change in the cognitive landscape is based on the advent of Internet in the first time cluster while in the second cluster it is due to the use of indices.

**Keywords:** Concept Clustering, Knowledge and Information Science, Text mining, content analysis