Scientometrics Research Journal

Scientific Bi-Quarterly of Shahed University
Vol. 6, No. 2, Autumn & Winter 2020-21 (Serial 12)

Research Article

DOI:10.22070/rsci.2019.3841.1240

Knowledge Structure in Knowledge Discovery Patents: Visualization based on Co-word Analysis

Mostafa Rostami [1*]
Faramarz Soheili [2]
AliAkbar Khasseh [3]

1. M.S. in Scientometrics, The Academic Institute for Ethics and Education. (Corresponding Author)
2. Associate Professor, Department of library and information science. Payame Noor University, Tehran, Iran.
Email: fsohieli@gmail.com
3. Assistant Professor of Knowledge and Information Science, Payame Noor University, Tehran, Iran.
Email: khasseh.lisclass@gmail.com

Email: mostafarostami.mostafa@gmail.com

Date of Reception: 10/01/2019

Date of Acceptation: 08/05/2019

Abstract

Purpose: The purpose of this study is to explain the structure of knowledge in the field of knowledge discovery using Co-word Analysis.

Methodology: Bibliometric methodology and social network analysis are used as the research methods. The research population consists of all keywords in 304 patents in the field of Knowledge Discovery (KD) (Reporting, knowledge discovery (KD) systems) during 1995-2014 extracted from USPTO.

Findings: Findings indicated that "System" and "Computer" are the most frequent keywords in Knowledge Discovery patents. Moreover, findings showed that "Computer***Apparatus" and "OLAP***OLAP Management" are highly frequent co-word pairs, respectively. The results of hierarchical clustering uncovered that during the 1995-2014 period 17 thematic clusters are formed. Drawing the strategic diagram showed that clusters "input devices", "business analysis", and "responsive modules" are among mature clusters in Knowledge Discovery.

Conclusion: The combination of concepts in this study showed that the domain of knowledge discovery is heavily dependent on computer tools, information technologies, communication technologies, and human being as the last loop in the period of systematic data analysis.

Keywords: Content analysis, Co-word analysis, Knowledge discovery, Patents.

پژوهش‌نامه علم‌سنجی
دوفصلنامه علمی دانشگاه شاهد/ دوره ۶/ شماره ۲/ پاییز و زمستان ۱۳۹۹ (پیاپی ۱۲)

مقاله پژوهشی

DOI:10.22070/rsci.2019.3841.1240

ساختار دانش در پروانه‌های ثبت اختـراع حـوزه کشف دانش: مصورسازی با اسـتفاده از تحلیـل هم‌رخدادی واژگان

مصطفی رستمی [1*]
فرامرز سهیلی [2]
علی‌اکبر خاصه [3]

۱. کارشناس‌ارشد علم سنجی، موسسه آموزش عالی اخلاق و تربیت، قم، ایران. (نویسنده مسئول)
۲. دانشیار گروه علم اطلاعات و دانش شناسی، دانشگاه پیام نور، تهران، ایران.
Email: fsohieli@gmail.com
۳. استادیار گروه علم اطلاعات ودانش شناسی، دانشگاه پیام نور، تهران، ایران.
Email: khasseh.lisclass@gmail.com

Email:mostafarostami.mostafa@gmail.com

صفحه ۶۰-۴۱
دریافت: ۱۳۹۷/۱۰/۲۰
پذیرش: ۱۳۹۸/۲/۱۸

چکیده

هدف: هدف از این پژوهش تبیین ساختار دانش در حوزهٔ کشف دانـش بـا اسـتفاده از تحلیـل هم‌رخدادی واژگان است.

روش‌شناسی: ایـن پـژوهش از نـوع مطالعـات کتـاب‌سـنجی اسـت و بـا اسـتفاده از تحلیـل هم‌رخدادی واژگان و تحلیل شبکه‌های اجتماعی انجام شده اسـت. داده‌هـای ایـن پـژوهش، از مفاهیم موجود در عناوین ۳۰۴ عنوان پروانه ثبت اختراع حـوزه کشـف دانـش بـین سـال‌هـای ۲۰۱۴-۱۹۹۵ است که از پایگاه ثبت اختراعات و علائم تجاری آمریکا استخراج شده است.

یافته‌ها: یافته‌های پژوهش نشان دادند، کلیدواژه‌های «نظام» و «رایانه» بیشـترین فراوانـی را در عناوین پروانه‌ها را داشتند و از نظر هم‌رخدادی واژگانی، زوج‌هـای «رایانـه و ماشـین‌هـا» و «مدیریت پردازش آنلاین و پردازش آنلاین» بیشترین فراوانـی را داشـتند. یافتـه‌هـای تحلیـل خوشه‌ای نشان دادند که در بازهٔ زمانی پژوهش حوزه کشف دانش از ۱۷ خوشهٔ دانشی تشـکیل شده است. ترسیم نمودار راهبردی نشان داد که خوشه‌های «دستگاه‌هـای ورودی»، «تجزیـه و تحلیل کسب و کار» و «ماژول پاسخ‌گو» در این حوزه به بلوغ دانشی رسیده‌اند.

نتیجه‌گیری: ترکیب مفاهیم در این پژوهش نشان داد، حوزه کشف دانش به‌شدت به ابزارهـای رایانه‌ای، حوزه‌های دانشی فناوری اطلاعات، فناوری ارتباطات و انسان به‌عنوان آخرین حلقه در زمان بررسی داده‌های منظم‌شده وابسته می‌باشد.

واژگان کلیدی: تحلیل محتوا، تحلیل هم‌رخدادی واژگان، حوزه کشف دانش، پروانه‌هـای ثبـت اختراع.

مقدمه و بیان مسئله

در عصر حاضر انسان در تمامی نقاط جهان اقدام به تولید محتوا و اطلاعات می‌کند. این اطلاعات با توجه به زبان و مکان جغرافیایی خود معنا و مفهوم متفاوتی پیدا می‌کند. این امکان نیز وجود دارد که این اطلاعات در همان زمان مورد استفاده قرار نگیرد و در اثر ترکیب با دیگر اطلاعات، داده‌های جدیدی را برای آیندگان تولید نماید. برای اینکه این داده‌های جدید تولیدشده ساختار و مفهوم درستی را برای جستجوگر القا نماید، نیازمند ذخیره، نگهداری و بازشناسایی اطلاعات می‌باشد که از آن به‌عنوان کشف دانش[1] نام می‌برند. کشف دانش شامل ثبت[2]، مرتب‌کردن[3]، ترکیب[4]، محاسبه[5]، جمع‌بندی[6]، ذخیره[7]، بازیافت[8]، تولید مجدد[9] و نمایش[10] (بهشتیان و ابوالحسنی، ۱۳۸۴، ص، ۱۷۳) بوده و توسط انسان با استفاده از سیستم‌های رایانه‌ای و هوش مصنوعی گزاره‌ها، نظم و مفاهیم حقیقی جهان اشیاء (غضنفری، علی‌زاده و تیمورپور، ۱۳۸۷) را توصیف می‌کند. حوزه کشف دانش زیرمجموعه پردازش داده می‌باشد که می‌تواند داده‌ها را از شکل غیرساختارمند به شکل ساختارمند تبدیل کند. موفقیت نظام کشف دانش به عوامل زیادی از جمله سخت‌افزار برای ورود و نگهداری برای نمایش و نرم‌افزار برای شناسایی و برقراری ارتباط بین داده‌ها و درنهایت نیروی انسانی متخصص برای مفهوم‌پردازی گذشته و آینده داده بستگی دارد. ماتینز[11] (۲۰۱۱) معتقد است کیفیت ارتباط بین داده‌ها به مدیریت داده‌ها، تهیه داده‌ها و تمیزکردن داده‌ها بستگی دارد (آریستودمو[12] و دیگران، ۲۰۱۷). درواقع اگر تمامی داده‌های محیطی و انسانی بدون توجه به منبع تولید با ارزش قلمداد شوند و به‌درستی و در جایگاه مناسبی نگهداری و ذخیره شوند، زمانی نخواهد گذشت که مورد استفاده قرار گیرند.

با توجه به ماهیت حوزه کشف دانش که از زیرمجموعه علوم رایانه می‌باشد و این علوم با نام نظام فناوری در حوزه‌های علمی شناخته می‌شوند، بیشتر اطلاعات این حوزه در پروانه‌های ثبت اختراع ضبط می‌شوند. حوزه کشف دانش نیز برای اینکه بتواند دانش خود را به جامعه علمی، کسب و کار، تجارت و صنعت عرضه کند؛ باید اقدام به ثبت اطلاعات خود با استفاده از پروانه‌های ثبت اختراع نماید. پروانه‌های ثبت اختراع به دلیل ماهیت اطلاعاتی که دارند به‌عنوان منابع دست اول در حوزه فناوری شناخته می‌شوند. بنابراین در این پژوهش از این اسناد برای تحلیل ساختار دانش حوزه کشف دانش استفاده شده است. این پژوهش می‌تواند برای سیاست‌گذاران و برنامه‌ریزان کلان کشوری که به دنبال طراحی و سیاست‌گذاری علمی کشور (لی[13] و دیگران، ۲۰۱۴) هستند، نقشه راهبردی بسیار مناسبی را برای رفع عدم اجرای موفق برنامه‌های اجراشده در حوزه تولید دانش و تجاری‌سازی آن داشته باشد. این پژوهش با استفاده از تحلیل هم‌رخدادی واژگان عناوین پروانه‌های ثبت اختراع حوزه کشف دانش انجام شده است.

روش تحلیل هم‌رخدادی واژگان زمینه‌های علمی و میزان ارتباط اصطلاحات در بین اسناد مختلف را نشان

1 . Knowledge discovery (KD)
2 . Recording
3 . Sorting
4 . Merging
5 . Calculation
6 . Accumulating
7 . Storing
8 . Retrieving
9 . Reproducing
10 . Displaying
11 . Martinez
12 . Aristodemou
13 . Li

می‌دهد. این روش با استفاده از میزان سنجه‌های تراکم و مرکزیت رتبه در یک حوزه پژوهشی، روابط مفهومی، معنایی (وینگ، جین و کوان[1]، ۲۰۱۸) و تکامل زمینه‌های علمی را از طریق ساختن خوشه‌ها و نمودارهای راهبردی[2] (دلروکس و اپاستین، ۲۰۰۴) اندازه‌گیری و ترسیم می‌کند. تحلیل هم‌رخدادی واژگان رابطه همبستگی بین دو اصطلاحی را که با هم در یک یا چند مدرک به کار رفته‌اند را به جای استنادهای مشترک بررسی (هی[3]، ۱۹۹۹؛ دلکرویکس و اپاستین، ۲۰۰۴) می‌کند. این پژوهش با استفاده از تحلیل هم‌رخدادی واژگان به بررسی ساختار فکری دانش در حوزه کشف دانش پرداخته و نشان خواهد داد حوزه کشف دانش از چند خوشهٔ دانشی تشکیل شده، به چه حوزه‌های دانشی وابسته بوده و در چه مرحله‌ای (بُعدی) از رشد دانشی قرار دارد؟

سؤال‌های پژوهش

۱. توزیع فراوانی کلیدواژه‌های حوزه کشف دانش بر اساس میزان هم‌رخدادی واژگان چگونه است؟
۲. نتایج مربوط به تحلیل خوشه‌ای هم‌رخدادی واژگان، منجر به شکل‌گیری چه خوشه‌هایی و با چه موضوعاتی در حوزه کشف دانش شده است؟
۳. پراکندگی هم‌رخدادی واژگانی حوزه کشف دانش بر اساس نقشه ترسیمی دوبعدی چگونه می‌باشد؟
۴. خوشه‌های حاصل از تحلیل هم‌رخدادی واژگانی در حوزه کشف دانش از نظر میزان بلوغ و توسعه‌یافتگی هر یک در چه وضعیتی می‌باشند؟

چارچوب نظری

امروزه تحلیل داده و کشف ارتباط میان آنها توسط رایانه و در ابعاد مختلف به‌صورت مشترک در روش اجرا و تخصصی در نوع واکنش محیطی داده‌ها انجام می‌شود و همان‌گونه که پیش‌تر بیان شد حوزه کشف دانش وابستگی مستقیمی به حوزه فناوری دارد و حال این سؤال ایجاد می‌شود که با توجه به اینکه اکثر کشورها و پژوهشگران به ابزارهای مورد نیاز حوزه کشف دانش در خصوص ذخیره، نگهداری و نمایش داده‌ها دسترسی دارند، چرا نمی‌توانند همانند کشورهای توسعه‌یافته دانش نهفته در میان داده‌های ذخیره‌شده را کشف کنند و همچنان این کشورها به لحاظ علمی وابسته به کشورهای مبدأ در تولید دانش می‌باشند. حال اگر نظام‌ها و ماشین‌های کشف دانش به‌درستی شناخته شوند و در کنار علوم مختلف توسعه یابند می‌توانند در کنار پژوهش‌های علمی و تخصصی مورد نیاز کشورهای درحال‌توسعه رشد کرده و از هزینه‌های موازی پژوهش‌های علمی جلوگیری کرده و باعث صرفه‌جویی در عنصر زمان در نتایج علمی گردند.

تحلیل هم‌رخدادی واژگان روشی است که امکان رهگیری این نوع مطالعات را فراهم می‌سازد و این امکان را ایجاد می‌کند که با بهره‌گیری از مفاهیم و با واژگان اسناد در یک مجموعه، ارتباط متون و اسناد را مفهوم‌سازی کرده و بر اساس هستی‌شناسی درونی و بیرونی واژگان، آنها را طبقه‌بندی و ساختار و ارتباط پنهان دانش درون آنها را ترسیم کرد. این روش اولین‌بار توسط کالون، کورتیال و ترنر[4] در سال ۱۹۹۱ استفاده شد (کالون و دیگران، ۱۹۹۱؛ نقل در

1 . Wanying, Jin, Kun
2 . strategic diagram
3 . He
4 . Callon, Courtial & Turner

دلکرویکس و اپاستین[1]، ۲۰۰۴).

پیشینه پژوهش

پیشینه پژوهش در داخل

در پژوهش منصوری، توکلی‌زاده راوری، مکی‌زاده و طوسی (۱۳۹۵) بررسی ساختار دانشی حوزه فناوری RFID نشان داد این حوزه با ۳۶ حوزه موضوعی مرتبط است و محور اصلی این حوزه‌ها، "ارتباطات راه دور و ارتباطات الکتریکی" است. همچنین"ارتباطات الکتریکی" بیشترین نقش را در انتقال RFID و تغییر موضوع آن در طول زمان بوده است.

سهیلی، شعبانی و خاصه (۱۳۹۵) در بررسی ساختار دانش حوزه رفتار اطلاعاتی نشان دادند خوشه‌های «رفتار اطلاعات سلامت»، «مطالعه کاربران»، «شبکه‌های اجتماعی» و «ربط و بازاریابی اطلاعات» جزء خوشه‌های بالغ و خوشه‌های «منابع اطلاعاتی»، «جستجوی وبی»، «بازاریابی اطلاعات و مدیریت اطلاعات» جزء خوشه‌های در حال ظهور یا زوال و خوشه‌های «رابط کاربری» و «فناوری اطلاعات» به‌عنوان توسعه‌نیافته قرار دارند.

سهیلی، خاصه و قاضی‌زاده (۱۳۹۷) در بررسی ساختار دانش در پژوهش‌های علوم قرآنی و حدیث در ایران نشان دادند، این حوزه دارای ۱۱ خوشه مطالعاتی می‌باشد. خوشه‌های «اختلاف قرائات و نقش آن در تفسیر»، «بررسی تطبیقی روش‌های تفسیر قرآن در فریقین» و «سبک‌شناسی مدیریت سیاسی جامعه از دیدگاه امام علی (ع)» بلوغ یافته و خوشه‌های «اعجاز علمی در قرآن» و «جایگاه امامت و عصمت در هدایت‌گری جامعه» به‌عنوان خوشه‌های نابالغ قرار گرفته‌اند.

سهیلی، خاصه و کرانیان (۱۳۹۷) با روش تحلیل هم‌رخدادی واژگان، همه مجله‌های فارسی‌زبان حوزه علم اطلاعات و دانش‌شناسی ایران که رتبه علمی-پژوهشی دارند و در پایگاه استنادی علوم جهان اسلام نمایه شده‌اند را مورد بررسی قرار دادند. نتایج مربوط به نمودار راهبردی نشان دادند مباحث علم‌سنجی بهترین جایگاه را در پژوهش‌های علم اطلاعات و دانش‌شناسی ایران دارند و مباحثی نظیر رابط کاربر، معماری اطلاعات، موتورهای جستجو، کتابخانه دیجیتالی، ابرداده، جستجوی اطلاعات، حفاظت اطلاعات، مدیریت دانش، هستی‌شناسی، مصوّرسازی، و شبکه‌های اجتماعی جزء موضوعات نوظهور در مطالعات علم اطلاعات و دانش‌شناسی ایران هستند.

پیشینه پژوهش در خارج

کالون، کورتیال و لاویل[2] (۱۹۹۱) مصورسازی می‌تواند تغییرات منطقه‌ای دانش در طول زمان را نشان دهد و با استفاده از ارتباط بین کلمات کلیدی و زیرگروه‌های آن زمینه‌های موضوعی و زمینه‌های مورد علاقه پژوهشگران را ترسیم نماید.

دینگ، چوودوری و فو[3] (۲۰۰۱) ساختار دانش حوزه بازیابی اطلاعات را مورد بررسی قرار دادند. این پژوهش نشان داد این حوزه دارای ۵ خوشه موضوعی اصلی بوده و به‌سرعت در حال تکامل می‌باشد.

چن و هوانگ[4] (۲۰۰۸) یک مدل جدید کشف دانش برای مدل‌های متوالی چندسطحی فازی در پایگاه داده‌های

1 . Delecroix & Eppstein
2 . Callon, Courtial & Laville
3 . Ding, Chowdhury & Foo
4 . Chen & Huang

متوالی نشان دادند در داده‌کاوی‌های ضمنی که در یک پایگاه اطلاعاتی متوالی قرار دارند، استفاده از روش سلسله‌مراتبی می‌تواند باعث نمایان‌شدن روابط بین داده‌هایی شود که به لحاظ ظاهری هیچ‌گونه ارتباطی با هم ندارند و این روش می‌تواند بخش‌های پنهان یک حوزه علمی را نشان دهد. این روش را می‌توان در تجزیه و تحلیل کسب و کار، پشتیبانی از تصمیم‌گیری و نمایان‌کردن عدم صداقت در داده‌ها به کار برد.

رانکان، پیسادو و گارسیامارتینز[1] (۲۰۱۰) در فرایند کشف دانش بر اساس موضوعات پایه بیان می‌کنند که پایگاه دانش نماینده‌ای از داده‌هایی است که در آن گنجانده شده است و با توجه به انبوه دانش ذخیره‌شده در آن امکان ازهم‌پاشیدگی اطلاعات و تغییر دانش خروجی وجود دارد. بنابراین نظام کشف دانش با فرایند خوشه‌بندی و نقشه‌های ترسیمی می‌تواند مرز بین داده‌ها را جداسازی و دانش نهان را نمایان سازد.

لیو، هو و وانگ[2] (۲۰۱۲) با روش هم‌رخدادی واژگان، حوزه کتابخانه‌های دیجیتال را در چین مورد مطالعه قرار دادند. تجزیه و تحلیل یافته‌های آنان منجر به شکل‌گیری ۷ خوشه که همبستگی کمی با همدیگر دارند شد. پژوهشگران به این نتیجه رسیدند که موضوعات پژوهشی در حوزه کتابخانه‌های دیجیتال در چین از تمرکز خوبی برخوردار هستند.

پژوهش دیگری نیز به روش هم‌رخدادی واژگان در کشور چین روی رساله‌های علوم کتابداری و اطلاع‌رسانی توسط زانگ[3] و دیگران (۲۰۱۳) انجام شده است. تحلیل آنها که مبتنی بر ۵۶ کلیدواژه پرتکرار در رساله‌های مورد بررسی بود منجر به شناسایی ۱۵ خوشه از قبیل منابع اطلاعاتی، هستی‌شناسی‌ها، دولت الکترونیکی، مدیریت دانش، کتابخانه‌های رقومی، بازیابی اطلاعات، شبکه اجتماعی، ارزیابی علوم انسانی و اجتماعی، ارزیابی عملکرد، مجله‌های علمی، هوش رقابتی، مدیریت کتابخانه‌ها، کتاب‌سنجی، علم‌سنجی و مدیریت اطلاعات شد که بسیاری از آنها نابالغ بودند و معدود حوزه‌های موضوعی ازجمله منابع اطلاعاتی، دولت الکترونیک، کتابخانه‌های رقومی و مدیریت دانش خوش-توسعه و بالغ بودند.

نتایج پژوهش وانگ، لیو و شنگ[4] (۲۰۱۴) در رابطه با تحلیل هم‌واژگانی مقالات حوزهٔ کشف دانش منجر به شناسایی شش موضوع پژوهشی در این حوزه گردید که عبارت بودند از کشف دانش بر اساس پژوهش‌های داده‌ای، پژوهش‌های بهینه‌سازی الگوریتم‌های کشف دانش، مدل کشف دانش و مطالعات پیشینه‌ای، مدیریت دانش بر مبنای هستی‌شناسی، پژوهش‌های ساخت نظام‌های خبره و پژوهش‌های کاربردی کشف دانش.

راوی‌کومار، آگراهاری و سینگ[5] (۲۰۱۵) در پژوهش تحلیل هم‌رخدادی واژگان را روی مجلهٔ علم‌سنجی انجام داده‌اند. نتایج پژوهش نشان دادند که مقاله‌های منتشرشده در مجلهٔ علم‌سنجی به موضوعاتی تعلق دارند که برخی از این موضوعات ریشه‌دار بوده و برخی به‌تدریج تغییر یافته و مباحث جدیدی را ایجاد می‌کنند. به‌عبارت‌دیگر، یافته‌های آنان نشان دادند که برخی مباحث نظیر تحلیل استنادی، بهره‌وری، تحلیل کتاب‌سنجی و مانند آن دارای ریشه‌های استواری در مجلهٔ علم‌سنجی می‌باشند، درحالی‌که برخی مباحث دیگر نظیر نگاشت دانش، تحلیل بیزی[6] و مانند آن قدمت چندانی نداشته و در مدت‌زمان کوتاهی ظهور یافته‌اند.

1 . Rancan, Pesado & Garcia-Martinez
2 . Liu, Hu & Wang
3 . Zong
4 . Wang, Liu & Sheng
5 . Ravikumar, Agrahari & Singh
6 . Bayesian analysis

چن، چن، وو، شی و لی[1] (۲۰۱۶) ساختار دانش در علوم مدیریت و مهندسی کشور چین را ترسیم کردند و نشان دادند این حوزه از ۸ خوشهٔ اصلی تشکیل شده است. همچنین این پژوهش نشان داد خوشه‌هایی نظریه بازی، مدیریت زنجیره تأمین و داده‌کاوی موضوعات داغ و توسعه‌یافته این حوزه در چین می‌باشند. اما بااین‌حال علوم مدیریت و مهندسی در چین هنوز نابالغ است.

در پژوهش دیگری خاصه، سهیلی، شریف‌مقدم و موسوی چلک (۲۰۱۷) به ترسیم ساختار دانش در حوزه سنجه‌های اطلاعاتی پرداختند. نتایج خوشه‌بندی سلسله‌مراتبی آنها نشان دادند که ساختار این حوزه از ۱۱ خوشه تشکیل شده است. این خوشه‌ها عبارت‌اند از شاخص‌ها و پایگاه‌های علم‌سنجی، تحلیل استنادی و مبانی نظری، جامعه‌شناسی علم، موضوعات مربوط به رتبه‌بندی دانشگاه‌ها، مجلات و ...، مصورسازی و بازیابی اطلاعات، ترسیم ساختار فکری علم، وب‌سنجی، ارتباطات صنعت- دانشگاه- دولت، فناوری‌سنجی (نوآوری و پروانه ثبت اختراع)، تحلیل شبکه و همکاری‌های علمی در دانشگاه‌ها.

جمع‌بندی از مرور پیشینه

همان‌طور که در پیشینه‌ها بیان شد بررسی ساختار دانش در حوزه‌های مختلف علمی با استفاده از تجزیه و تحلیل مفاهیم و اصطلاحات علمی برای نشان‌دادن سیر تولید تا مرگ آن حوزه مورد استفاده قرار می‌گیرد. اما اینکه این مفاهیم از چه پایگاه‌های اطلاعاتی استخراج می‌شوند به نظر پژوهشگر و پوشش مفاهیم حوزه مورد نظر که می‌توانند علمی یا فنی باشند بستگی دارد. همچنین از روش‌های مختلف تحلیل محتوا، تحلیل خوشه‌ای با استفاده از هم‌رخدادی واژگان روش بسیار مناسبی در پژوهش‌ها بیان شده است.

روش‌شناسی پژوهش

این پژوهش از نوع مطالعات کتاب‌سنجی است و با استفاده از تحلیل هم‌رخدادی واژگان و تحلیل شبکه‌های اجتماعی انجام شده است. برای به‌دست‌آوردن داده‌ها ابتدا حوزه پردازش داده مورد بررسی قرار گرفت که منجر به شناسایی رده موضوع حوزه کشف دانش به شماره ۶۰۳ شد. در پایگاه USPTO روش‌های مختلفی برای جستجوی پروانه‌های ثبت اختراع لحاظ شده است؛ بنابراین برای این پژوهش از روش جستجوی رده‌بندی جاری آمریکا (ccl[2]) که همانند رده‌بندی دیویی برای هر موضوع شماره‌ای اختصاص داده شده است که بازیابی و شناسایی پروانه‌ها را به‌صورت موضوعی امکان‌پذیر ساخته است. در این پژوهش با استفاده از روش دستوری CCL/707/603، ۳۰۴ عنوان پروانه ثبت اختراع حوزه کشف دانش بدون اعمال محدودیت زمانی از سال‌های ۲۰۱۴-۱۹۹۵ شناسایی شدند. اطلاعات پروانه‌ها با استفاده از نرم‌افزار Uspto2.exe استخراج و با استفاده از نرم‌افزار Aostsoft All Document به فایل متنی ساده تبدیل شدند. برای تحلیل هم‌رخدادی واژگان، عناوین پروانه‌های ثبت اختراع به واژگان مفهومی تجزیه و برای نمایه‌سازی و یکسان‌سازی مفهوم واژگان از وب‌سایت‌های پی.سی.دات.نت[3]، گوگل و یاهو استفاده شد. در ادامه برای تحلیل هم‌واژگانی، فراوانی واژگان، ماتریس مربع و ماتریس همبستگی از نرم‌افزارهای اکسل، راورپریمپ، یو.سی.آی.نت استفاده شد. برای محاسبه زوج‌های هم‌واژگانی از نرم‌افزار بایب‌اکسل و برای تحلیل خوشه‌ای از

1 . Chen, Chen, Wu, Xie, Li
2 . Current US Classification
3 . PC.NET

نرم‌افزار اس.پی.اس.اس.۲۲ و مقیاس‌بندی چندبعدی هم از نرم‌افزار یو.سی.آی.نت و برای ترسیم تصاویر گرافیکی و دیداری‌سازی روابط بین واژگان از نرم‌افزار ووس–ویور[1] استفاده شد.

یافته‌های پژوهش

پاسخ به سؤال اول پژوهش. توزیع فراوانی کلیدواژه‌های حوزه کشف دانش بر اساس میزان هم‌رخدادی واژگان چگونه است؟

جدول ۱. رتبه‌بندی کلیدواژه‌های حوزه کشف دانش بر اساس فراوانی

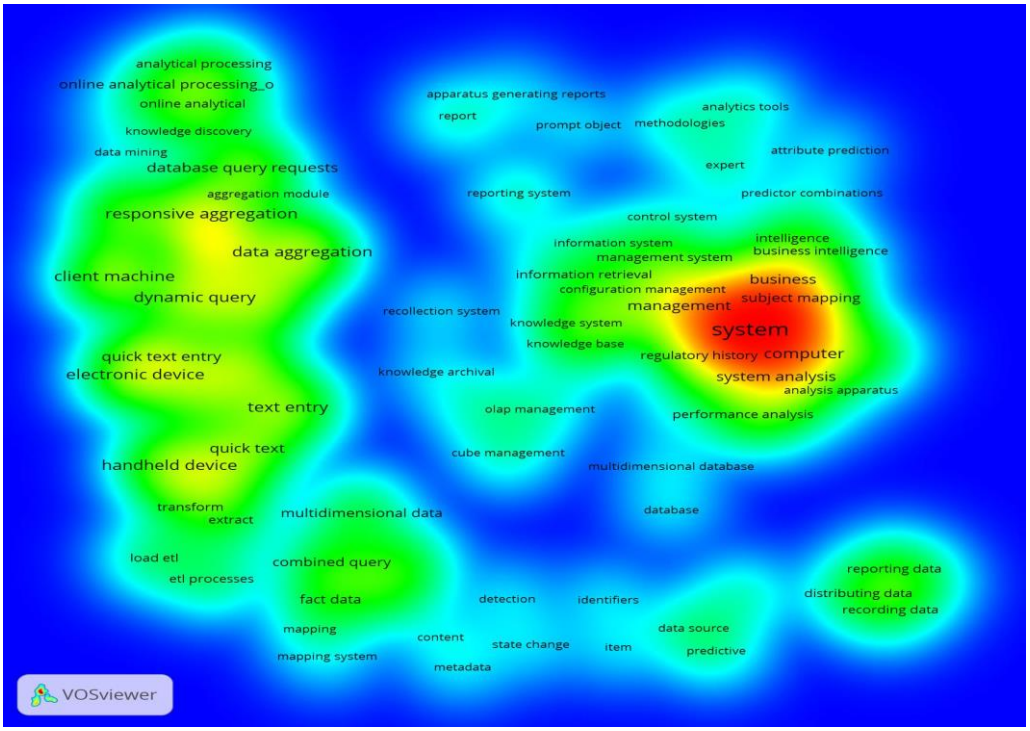
رتبه	کلیدواژه	فراوانی	رتبه	کلیدواژه	فراوانی
1	System	43	16	Information	7
2	Computer	19	17	Relationship	6
3	Apparatus	15	18	Metadata	6
4	Report	15	19	Analysis	5
5	Database	14	20	Application	5
6	Management	13	21	computer program	5
7	Knowledge	10	22	data extraction	5
8	system analysis	10	23	Document	5
9	olap management	9	24	Environment	5
10	Business	9	25	identifiers	5
11	data management	9	26	matching system	5
12	online analytical processing(olap)	8	27	multidimensional data	5
13	Content	8	28	reporting system	5
14	management system	8	29	analyzing data	4
15	knowledge management	7	30	data aggregation	4

با استفاده از نرم‌افزار راورپریمپ، تعداد و فراوانی کلیدواژگان عناوین پروانه‌ها مشخص و نشان داد در این حوزه از ۱۲۴۴ کلیدواژه اختصاصی استفاده شده بود. این واژگان درمجموع ۱۷۰۳ بار تکرار شده بودند. به‌طور میانگین هر پروانه ثبت اختراع حاوی ۵.۶ کلیدواژه بود. کلیدواژگان حوزه کشف دانش که فراوانی بالاتری از ۳ داشتند در جدول ۱ نشان داده شده است. همان‌طور که مشاهده می‌شود، کلیدواژه «نظام» با ۴۳ بار تکرار بیشترین فراوانی به خود اختصاص داده است. بعدازآن کلیدواژه‌های «رایانه»، «ماشین‌ها»، «گزارش»، «پایگاه‌های اطلاعاتی»، «مدیریت»، «دانش»، «تجزیه و تحلیل سیستم»، «مدیریت پروسه‌های تحلیلی آنلاین»، «تجارت»، «مدیریت داده» هریک به‌ترتیب ۱۹، ۱۵، ۱۵، ۱۴، ۱۳، ۱۰، ۱۰، ۹، ۹ و ۹ بیشترین فراوانی را دریافت کرده‌اند.

برای نشان‌دادن تراکم و ساختار فراوانی کلیدواژگان و میزان تجمع این واژگان با استفاده از نرم‌افزار ووس- ویور

1 . Vosviewer

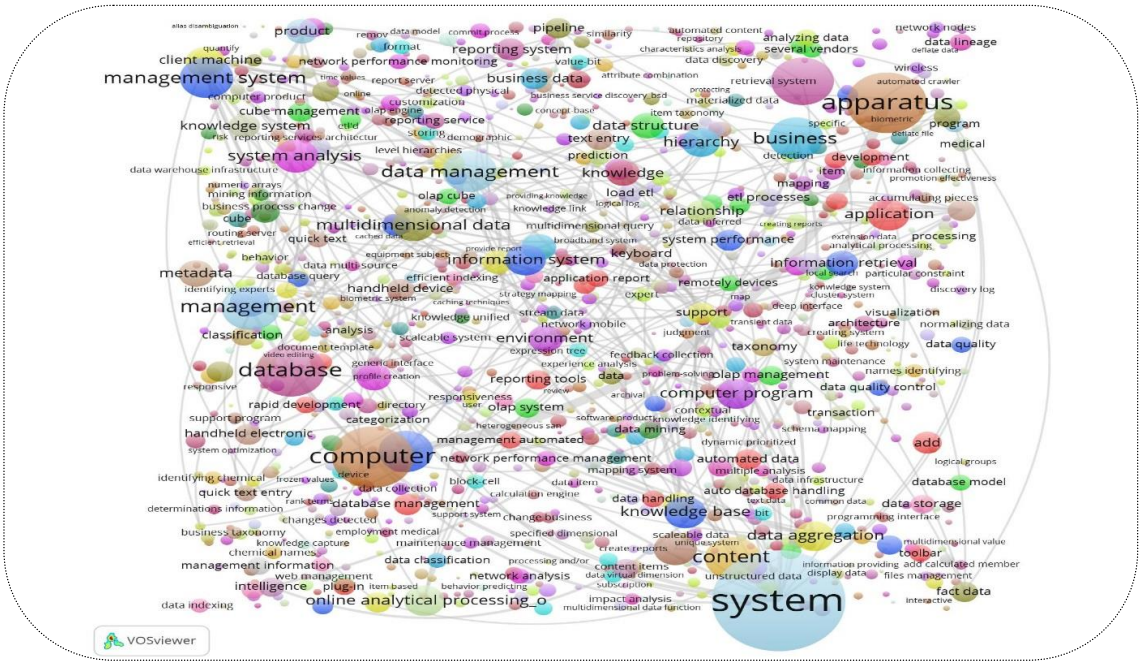
نقاط داغ این حوزه ترسیم شد که همان‌طور که در تصویر ۱ ملاحظه می‌شود نقاط تراکمی واژگان حوزه کشف دانش و میزان به‌کارگیری واژگان تخصصی در پروانه‌ها را به‌صورت نقاط حرارتی نشان داده شده است. واژگانی که در محدوده رنگ قرمز قرار دارند درواقع بیشترین استعمال را در حوزه کشف دانش دارند. به‌عبارتی بیشتر اختراعات حول این واژگان قرار دارند. هرچه از این بخش دور می‌شویم رنگ نقشه از قرمز به نارنجی، زرد، سبز و آبی تغییر می‌کند و این نشان می‌دهد که از تراکم واژگان در یک منطقه کاسته شده و به سمت واژگانی که در مجاورت و کنار خوشه‌ها قرار دارند حرکت می‌کند.



تصویر ۱. تراکم موضوعی واژگان حوزه کشف دانش

در ادامه برای ترسیم حداکثر ارتباط کلیدواژگان حوزه کشف دانش از نرم‌افزار ووس-ویور استفاده شد که در تصویر ۲ نشان داده شده است. همان‌طور که در تصویر مشخص شده است، دایره‌های بزرگ‌تر دارای بیشترین فراوانی هستند که توانسته اند بیشترین ارتباط را با دیگر واژگان حوزه کشف دانش برقرار کنند.

حضور هم‌رخدادی واژگان در یک مدرک، نشان‌دهنده محتوای آن مدرک می‌باشد. بعد از تعیین آستانه شمول واژگان و با استفاده از نرم‌افزار بایب‌اکسل، میزان هم‌رخدادی واژگان حوزه کشف دانش به دست آمد. در این مرحله، میزان هم‌رخدادی واژه‌های پربسامد با واژگان موجود در عناوین پروانه‌های ثبت اختراع محاسبه شد. نتایج نشان دادند که زوج «رایانه*ماشین‌ها» بیشترین فراوانی را به خود اختصاص داده است. این زوج وابستگی این حوزه به ماشین‌های رایانه‌ای را نشان می‌دهد. همچنین زوج «مدیریت پردازش تحلیل آنلاین*پردازش تحلیل آنلاین» با ۷ بار تکرار در ردیف دوم قرار دارد. یافته‌های مربوط به توزیع فراوانی ۱۰ زوج برتر هم‌واژگانی در جدول ۲ نشان داده شده است.



تصویر ۲. نقشه ارتباط موضوعی تمامی واژگان حوزه کشف دانش

جدول ۲. توزیع فراوانی ۳۷ زوج برتر هم‌واژگانی

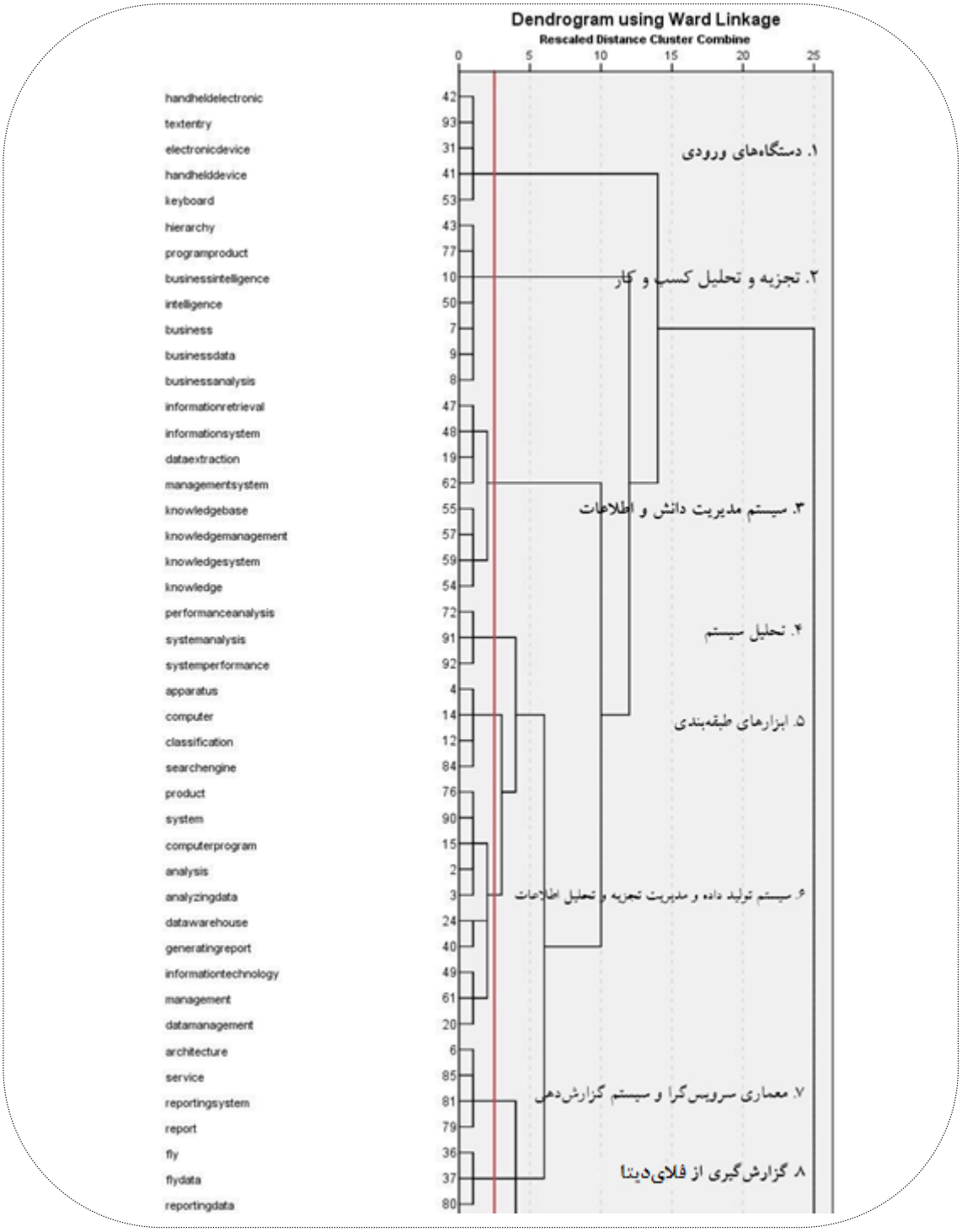
رتبه	زوج‌های هم‌واژگانی	فراوانی
1	Apparatus ** computer	10
2	olap management ** online analytical processing(olap)	7
3	data aggregation ** module	4
4	performance analysis ** system analysis	4
5	data aggregation ** servicing	3
6	business intelligence ** business	3
7	business intelligence ** intelligence	3
8	data management ** management	3
9	handheld device ** electronic device	3
10	Fly ** fly data	3

پاسخ به سؤال دوم پژوهش. نتایج مربوط به تحلیل خوشه‌ای هم‌رخدادیِ واژگان منجر به شکل‌گیری چه خوشه‌هایی و با چه موضوعاتی در حوزه کشف دانش شده است؟

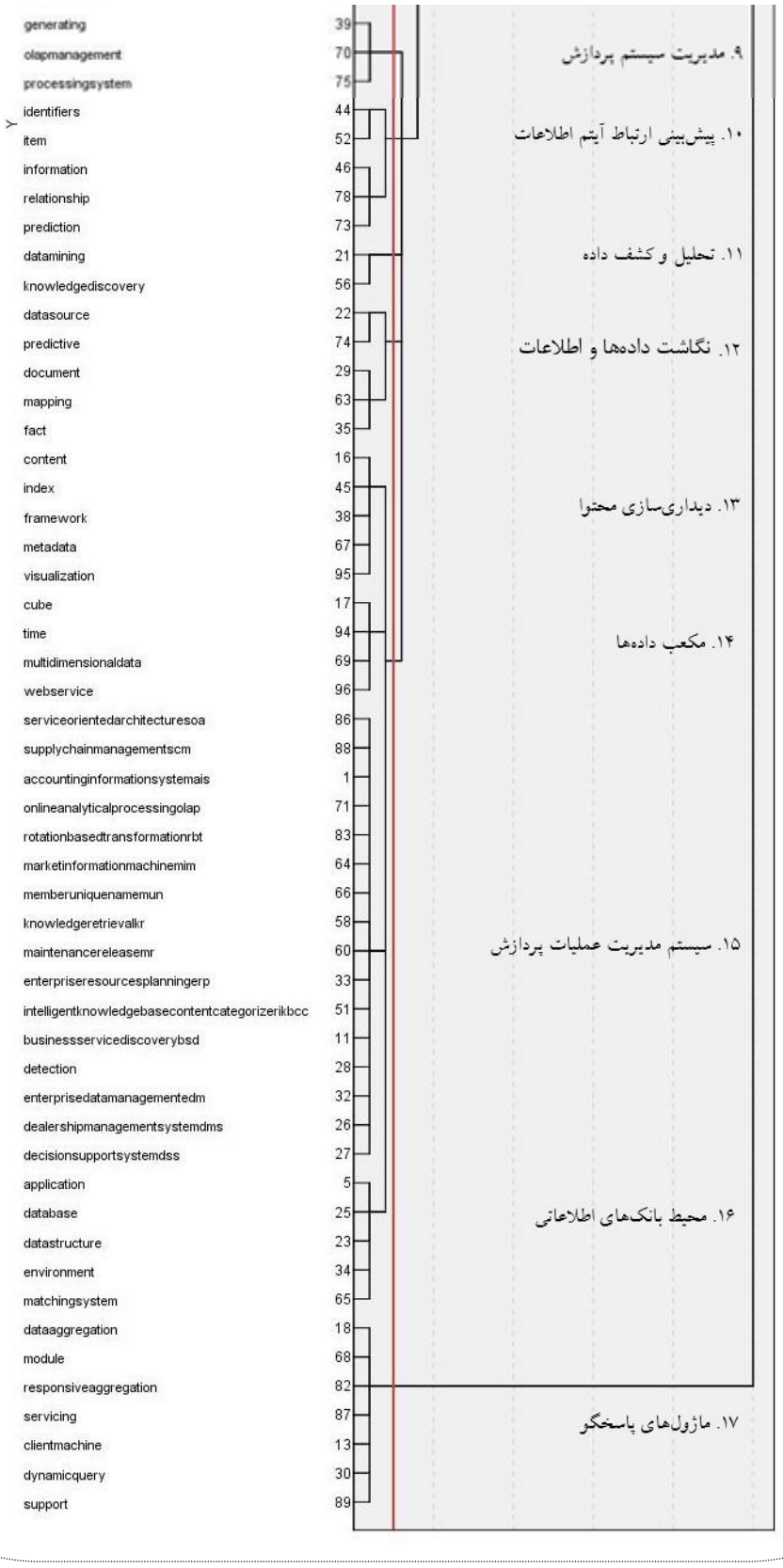
دندروگرام حاصل از خوشه‌بندی سلسله‌مراتبی در شکل ۳ نشان داده شده است. ازآنجایی‌که تعداد کلیدواژه‌های مورد بررسی نسبتاً زیاد بودند، نمودار دندروگرام ایجادشده به دو صفحه گسترش پیدا کرد. همان‌طور که در دندروگرام مشخص شده است، نتایج تحلیل خوشه‌ای منجر به شکل‌گیری ۱۷ خوشه در حوزه کشف دانش شده است. در مواقعی امکان دارد در برخی از خوشه‌ها علاوه بر واژگان اصلی و مهم واژگانی وجود داشته باشند که به نظر ارتباط مستقیم معنایی با موضوع خوشه مورد نظر نداشته باشد که چنین موردی در تحلیل‌های هم‌واژگانی تازگی ندارد و این به دلیل کم‌توجهی پژوهشگران به این واژگان است (کالون، کورتیال و لاویل[1]، ۱۹۹۱؛ سهیلی، شعبانی و خاصه،

1 . Callon, Courtial & Laville

۱۳۹۵). برای نام‌گذاری خوشه‌ها از راهنمایی اساتید و متخصصان موضوعی بخش فناوری، و علم اطلاعات و دانش‌شناسی (کتابداری) استفاده شد. بنابراین بعد از بررسی واژگان توسط ایشان، نزدیک‌ترین و پذیرفته‌شده‌ترین نام انتخاب و پس از بررسی این عناوین در پایگاه‌های اطلاعاتی اقدام شد. با توجه به اینکه بعضی از خوشه‌ها کمتر از سه واژه داشتند تصمیم گرفته شد همچون رشته‌های مهندسی، اصل محتوای واژگان حفظ و ترجمه شوند، تا ابهام یا داده‌ای غیرمتعارفی در این حوزه توسط پژوهشگر ایجاد نشود. در ادامه بررسی نتایج خوشه‌ها بیان می‌شود.



تصویر ۳. دندوگرام حاصل از خوشه‌بندی سلسله‌مراتبی به روش هم‌واژگانی



ادامه تصویر ۳. دندوگرام حاصل از خوشه‌بندی سلسله‌مراتبی به روش هم‌واژگانی

خوشه اول دستگاه‌های ورودی. این خوشه از ۵ واژه تشکیل شده است که با توجه به معنای واژگان مثل «الکترونیکی دستی»، «ورودی متن»، «دستگاه الکترونیکی»، «دستگاه دستی» و «صفحه‌کلید» می‌توانند به‌عنوان ورود داده‌ها توسط انسان به کار گرفته شود.

خوشه دوم: تجزیه و تحلیل کسب و کار. این خوشه از ۷ واژه تشکیل شده است. این خوشه به تجزیه و تحلیل کسب و کار پرداخته است؛ به‌عبارتی به تجارت و تجاری‌سازی داده‌ها با استفاده از کشف دانش می‌پردازد. از مهم‌ترین واژگانی که در این خوشه قرار دارند می‌توان به «سلسله‌مراتب»، «محصول برنامه»، «کسب و کار»، «هوش کسب و کار»، «هوش مصنوعی»، «داده‌های کسب و کار» و «تجزیه و تحلیل کسب و کار» اشاره کرد. با توجه به معنای واژه‌هایی که در این خوشه قرار دارند، تجارت و تجاری‌سازی داده با استفاده از کشف دانش‌ها امکان‌پذیر است.

خوشه سوم: سیستم مدیریت دانش و اطلاعات. این خوشه از ۸ واژه تشکیل شده است. این خوشه به‌عنوان یکی از اصلی‌ترین خوشه‌ها در حوزه کشف دانش که به مدیریت دانش پرداخته است شناخته می‌شود. واژگانی که در این خوشه قرار دارند شامل «بازیابی اطلاعات»، «سیستم اطلاعاتی»، «استخراج داده»، «نظام‌های مدیریت»، واژگان «پایگاه دانش»، «مدیریت دانش»، «سیستم دانش» و «دانش» می‌باشد.

خوشه چهارم: تحلیل سیستم. این خوشه از ۳ واژه ساخته شده است. این خوشه به بررسی سیستم‌هایی که در حوزه کشف دانش به کار می‌روند مربوط است. واژگان این خوشه شامل «تجزیه و تحلیل عملکرد»، «تحلیل سیستم» و «عملکرد سیستم» می‌باشد.

خوشه پنجم: ابزارهای طبقه‌بندی. این خوشه از ۴ واژه تشکیل شده است. واژگان این خوشه شامل «دستگاه»، «کامپیوتر»، «طبقه‌بندی» و «موتور جستجو» می‌باشد که برای طبقه‌بندی و ذخیره‌سازی داده‌ها به کار می‌روند.

خوشه ششم: سیستم تولید داده و مدیریت تجزیه و تحلیل اطلاعات. این خوشه از ۱۰ واژه تشکیل شده است. واژگان این خوشه شامل «محصول»، «سیستم»، «برنامه کامپیوتری»، «تحلیل و بررسی» و «تجزیه و تحلیل داده‌ها»، «انبار داده»، «گزارش تولید»، «فناوری اطلاعات»، «مدیریت» و «مدیریت اطلاعات» می‌باشد. درواقع برای تحلیل داده‌هایی با حجم بزرگ و برخط از موتورهای جستجو که بر روی کامپیوتر و ماشین‌های بزرگ قرار دارند استفاده می‌شود و این ماشین‌ها می‌توانند بعد از ورود داده‌ها به سیستم، آن را تحلیل کرده و گزارش جدید را تولید کنند.

خوشه هفتم: معماری سرویس‌گرا و سیستم گزارش‌دهی. این خوشه از ۴ واژه ساخته شده است. این خوشه شامل واژگان «معماری»، «سرویس»، «سیستم گزارش‌دهی» و «گزارش» می‌باشد.

خوشه هشتم: گزارش‌گیری از فلای‌دیتا. این خوشه از ۳ واژه ساخته شده است. این خوشه شامل واژگان «پرواز»، «پرواز داده‌ها» و «گزارش داده‌ها» می‌باشد.

خوشه نهم: مدیریت سیستم پردازش. این خوشه از ۳ واژه ساخته شده است. این خوشه شامل واژگان «تولید»، «مدیریت پردازش تحلیلی آنلاین» و «سیستم پردازش» می‌باشد.

خوشه دهم: پیش‌بینی ارتباط آیتم اطلاعات. این خوشه از ۵ واژه ساخته شده است. این خوشه شامل «شناساگر»، «آیتم»، «اطلاعات»، «ارتباط» و «پیش‌بینی» می‌باشد. داده‌ها برای تحلیل در هر یک از مراحل پردازش نیازمند شناسه مخصوص به خود می‌باشند تا بتوانند اطلاعات آینده را پیش‌بینی نمایند.

خوشه یازدهم: تحلیل و کشف داده. این خوشه از ۲ واژه ساخته شده است. تحلیل و کاوش در داده‌ها باعث کشف دانشی جدید خواهد شد. این خوشه از واژه‌های «داده‌کاوی» و «کشف دانش» تشکیل شده است.

خوشه دوازدهم: نگاشت داده‌ها و اطلاعات. این خوشه از ۵ واژه ساخته شده است. واژگان این خوشه شامل «منبع اطلاعات»، «پیش‌بینی‌کننده»، «اسناد»، «نگاشت» و «فاکت» می‌باشد. ساخت نقشه اسناد و مدارک باعث مشخص‌شدن ماهیت داده‌ها و درنتیجه پیش‌بینی آینده داده مورد نظر خواهد شد.

خوشه سیزدهم: دیداری‌سازی محتوا. این خوشه از ۵ واژه ساخته شده است. دیدارسازی محتوا و نمایهٔ اطلاعات باعث تحلیل بهتر اطلاعات می‌شود. واژگان این خوشه شامل «محتوا»، «نمایه»، «چارچوب»، «ابرداده» و «دیداری‌سازی» می‌باشد.

خوشه چهاردهم: مکعب داده‌ها. این خوشه از ۴ واژه ساخته شده است. واژگان این خوشه شامل «مکعب»، «زمان»، «داده‌های چندبعدی» و «سرویس وب» می‌باشد. از مشخصات مکعب طول، عرض، ارتفاع و فضاست. این مشخصات برای ماده می‌باشد؛ بنابراین برای اینکه داده هم درست و به‌موقع مورد استفاده قرار گیرد باید شرایط یک داده را داشته باشد. به‌عبارتی داده باید در زمان (فضا) مناسب در دسترس باشد.

خوشه پانزدهم: سیستم مدیریت عملیات پردازش. این خوشه از ۱۶ واژه ساخته شده است. این خوشه بلندترین خوشه حوزه کشف دانش می‌باشد. این خوشه به بخش‌های مختلف کشف دانش در زمان تحلیل داده تعلق دارد. واژگان این خوشه شامل «معماری سرویس‌گرا»، «مدیریت زنجیره تأمین»، «سیستم اطلاعات حسابداری»، «پردازش تحلیلی آنلاین»، »چرخش بر اساس تحول»، «ماشین‌سازی شکل‌گیری بازار»، «نام منحصربه‌فرد عضو»، «بازیابی دانش»، «نسخه آزمایشی»، «برنامه‌ریزی منابع سازمانی»، «طبقه‌بندی محتوای دانش هوشمند»، «کشف خدمات کسب و کار»، «تشخیص»، «مدیریت داده‌های سازمانی»، «سیستم مدیریت نمایندگی» و «سیستم پشتیبانی تصمیم» می‌باشد.

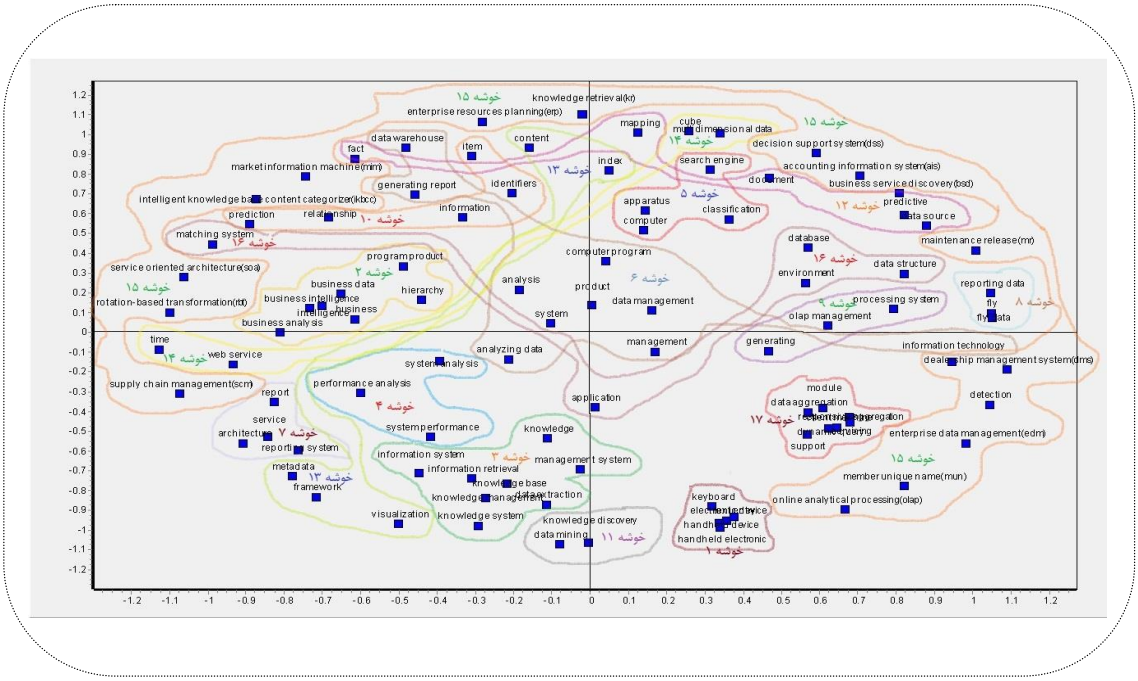
خوشه شانزدهم: محیط بانک‌های اطلاعاتی. این خوشه از ۵ واژه ساخته شده است. واژگان این خوشه شامل «نرم‌افزارهای کاربردی»، «بانک اطلاعاتی»، «ساختار داده‌ها»، «محیط» و «سیستم تطابق» می‌باشد. محیط بانک‌های اطلاعاتی می‌تواند نمایانگر ساختار داده‌ها و به‌کارگیری نرم‌افزارهای مناسب را مشخص نماید.

خوشه هفدهم: ماژول‌های پاسخ‌گو. این خوشه از ۷ واژه ساخته شده است. واژگان این خوشه شامل «جمع‌آوری داده‌ها»، «ماژول»، «پاسخ‌گویی»، «سرویس‌دهی»، «دستگاه مشتری»، «جستجوی پویا» و «حمایت‌کننده» می‌باشد که درواقع بهره‌گیران و جستجوگران را شامل می‌شود. این خوشه آخرین بخش از یک سیستم تحلیل می‌باشد که تمامی بخش‌های حوزه کشف دانش را زیر پوشش قرار داده است و نتیجه پایانی را در اختیار کاربران نهایی قرار می‌دهد.

پاسخ به سؤال سوم پژوهش. پراکندگی هم‌رخدادی واژگانی حوزه کشف دانش بر اساس نقشه ترسیمی دوبعدی چگونه می‌باشد؟

همان‌طور که در تصویر ۴ و نقشه دوبعدی نشان داده شده است، گستردگی تراکم موضوعی خوشه‌ها به لحاظ ارتباطی و نزدیکی که به هم دارند ترسیم شده است. احتمال اینکه این موضوعات پرتکرار باشند وجود دارد (سهیلی، قاضی‌زاده و خاصه، زودآیند). بیشترین تراکم واژگان در بالای محور افقی (X) قرار گرفته و به سمت بخش مثبت محور عمودی (Y) کشیده شده است. همان‌طور که مشاهده می‌شود خوشه ۱۵ در تمام نیمه بالای کشیده شده است؛ همان‌گونه که در تحلیل خوشه‌ای بیان شد، خوشه ۱۵ بیشترین تراکم موضوعات را دارا می‌باشد که نقشه مقیاس‌بندی چندبعدی آن را تأیید می‌کند و نشان‌دهنده منطبق‌بودن تحلیل خوشه‌ای می‌باشد. نتایج این تحلیل نشان می‌دهد که خوشه‌های «ابزارهای طبقه‌بندی» و «سیستم تولید داده و مدیریت تجزیه و تحلیل اطلاعات» به لحاظ مفهومی نزدیکی بیشتری به هم دارند. همچنین در پایین محور افقی و در بخش منفی محور عمودی، خوشه‌های «سیستم مدیریت دانش

و اطلاعات» و «تحلیل سیستم» بیشترین نزدیکی مفهومی را به هم دارند. به‌عبارتی بعد از تحلیل خوشه‌ای می‌توان از ترسیم نقشه دوبعدی برای نمایش مکان فضایی داده‌ها در داخل مخازن داده‌ای و میزان تأثیرگذاری خوشهٔ دانشی استفاده کرد.



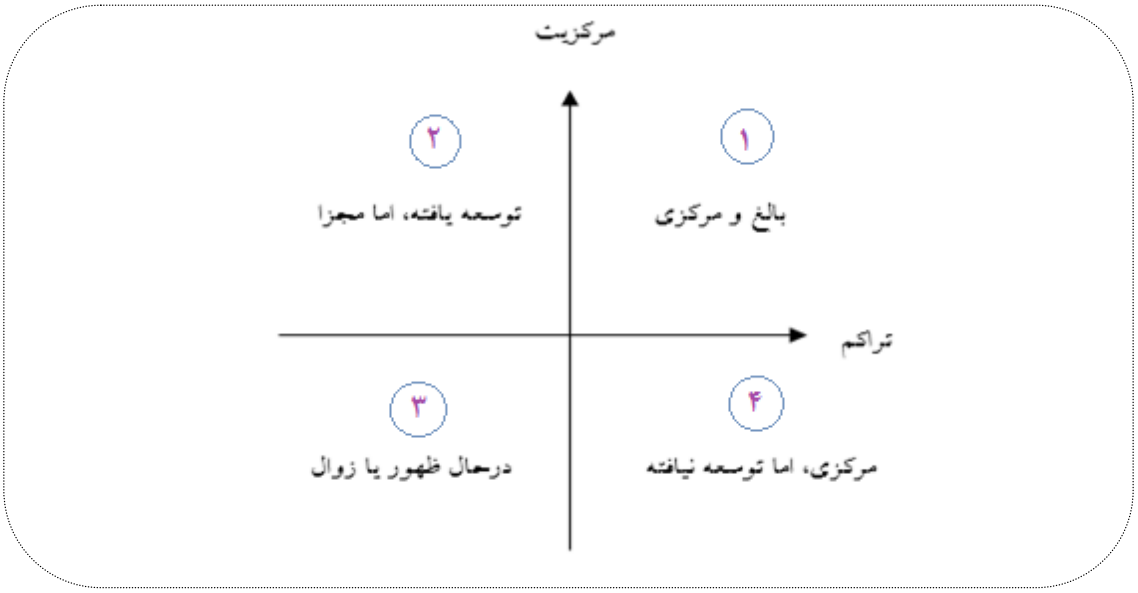
تصویر ۴. نقشه حاصل از مقیاس چندبعدی تحلیل هم‌رخدادی واژگان

پاسخ به سؤال چهارم پژوهش. خوشه‌های حاصل از تحلیل هم‌رخدادی واژگانی در حوزه کشف دانش از نظر میزان بلوغ و توسعه‌یافتگی هر یک در چه وضعیتی می‌باشند؟

کلیدواژه‌های حاضر در یک خوشه که حاصل خوشه‌بندی سلسله‌مراتبی می‌باشند، بیشترین میزان ارتباط را با یکدیگر و کمترین میزان ارتباط را با سایر خوشه‌ها دارند. در یک شبکه، اگر واژه‌ای روابط زیادی با سایر واژه‌ها داشته باشد، از مرکزیت بالاتری برخوردار است. همچنین هرچه تراکم یک خوشه بالاتر باشد، آن خوشه انسجام و همبستگی بیشتری (لاو[1] و دیگران، ۱۹۸۸؛ لیو، هو و وانگ[2]، ۲۰۱۲؛ سهیلی، خاصه و شعبانی، ۱۳۹۵) برای حفظ و توسعه خود خواهد داشت. در این ساختار هیچ فرضی وجود ندارد و تنها بر اساس شباهت و فواصل میان واژگان است که خوشه‌ها دسته‌بندی می‌شوند؛ بنابراین هرچه یک واژه پخته‌تر شده باشد، به‌عبارتی عمومیت بیشتری داشته باشد به همان میزان می‌تواند با واژگان بیشتری ارتباط برقرار کند و اگر واژه‌ای تازه وارد یک حوزه شده باشد به نسبت ارتباطات کمتری خواهد داشت.

همان‌طور که در تصویر ۶ نشان داده شده است با استفاده از میانگین مرکزیت رتبه و تراکم شبکه محاسبه شده در جدول ۳، نمودار راهبردی حوزه کشف دانش ترسیم شد، تا میزان بلوغ و انسجام هریک از خوشه‌های موضوعی حوزه کشف دانش مشخص گردد.

1 . Law
2 . Liu, Hu & Wang

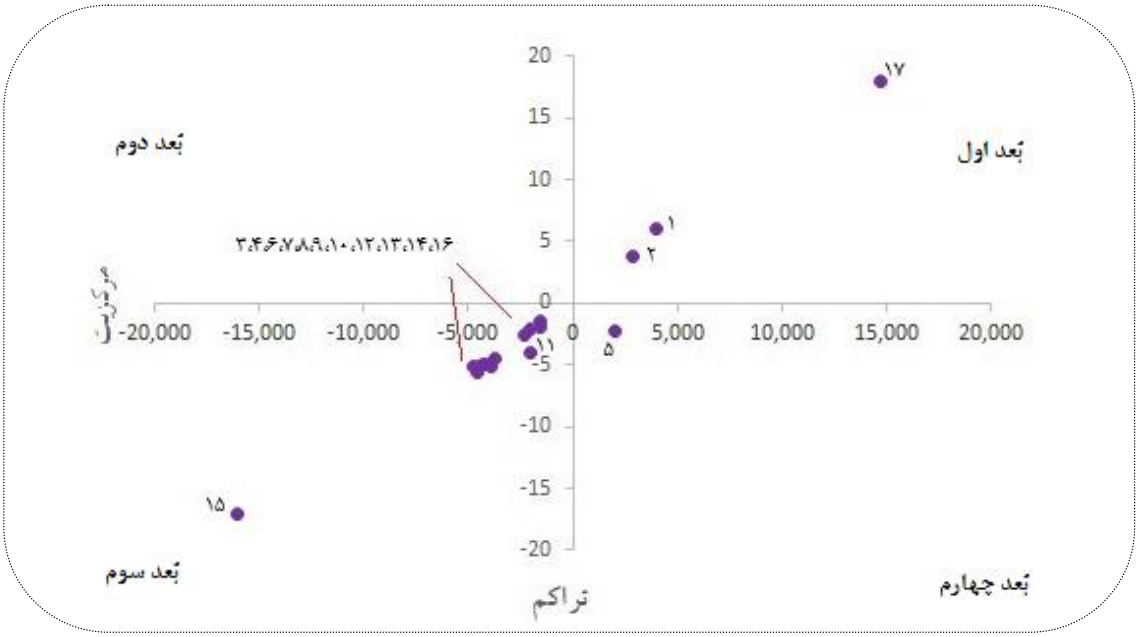


تصویر ۵. ابعاد چهارگانه نمودار راهبردی (اقتباس از ملسر، نگویان، چن، کانوسا، النصر، آیسبیستر[1]، ۲۰۱۵)

جدول ۳. تراکم و مرکزیت خوشه‌های حاصل از تحلیل هم‌واژگانی

ردیف	نام خوشه	مرکزیت	تراکم
۱	دستگاه‌های ورودی	۴.۰۰۰	۵.۹۸
۲	تجزیه و تحلیل کسب و کار	۲.۷۲۷	۳.۷۱۳
۳	سیستم مدیریت دانش و اطلاعات	-۳.۵۴۲	-۵.۱۴۲
۴	تحلیل سیستم	-۱.۶۱۰	-۲.۳۱۴
۵	ابزارهای طبقه‌بندی	-۲.۷۲۹	-۲.۵۹۴
۶	سیستم تولید داده و مدیریت تجزیه و تحلیل اطلاعات	-۱.۱۴۰	-۱.۴۶۹
۷	معماری سرویس‌گرا و سیستم گزارش‌دهی	-۲.۲۸۰	-۵.۰۹۲
۸	گزارش‌گیری از فلای‌دیتا	-۲.۰۰۰	-۲.۰۵۷
۹	مدیریت سیستم پردازش	-۲.۰۰۰	-۱.۸۶۶
۱۰	پیش‌بینی ارتباط آیتم اطلاعات	-۱.۶۳۲	-۵.۱۵۲
۱۱	تحلیل و کشف داده	-۲.۰۰۰	-۴.۰۰۰
۱۲	نگاشت داده‌ها و اطلاعات	-۲.۴۸۰	-۴.۹۱
۱۳	دیداری‌سازی محتوا	-۳.۵۵۳	-۵.۰۲۶
۱۴	مکعب داده‌ها	-۲.۴۲۸	-۴.۵۴۵
۱۵	سیستم مدیریت عملیات پردازش	-۱۶.۰۰۰	-۱۷.۰۶۷
۱۶	محیط بانک‌های اطلاعاتی	-۲.۵۲۷	-۵.۵۴۹
۱۷	ماژول‌های پاسخ‌گو	۱۴.۷۶۵	۱۸.۰۳۴

1 . Melcer, Nguyen, Chen, Canossa, El-Nasr, Isbister



تصویر ۶: نمودار راهبردی خوشه‌های حاصل از تحلیل هم‌رخدادی واژگانی

همان‌طور که در تصویر ۶ مشاهده می‌شود، خوشه‌ها در سه بُعد نمودار راهبردی توزیع شده‌اند که به‌ترتیب اشاره می‌شوند:

بالغ و مرکزی (بُعد اول) شامل خوشه‌های ۱۷ «ماژول‌های پاسخ‌گو»، ۱ «دستگاه‌های ورودی»، ۲ «تجزیه و تحلیل کسب و کار» می‌باشند که این خوشه‌ها از مرکزیت و تراکم بالاتری نسبت به بقیه خوشه‌ها برخوردار بوده و نقش محوری دارند که بسیار خوش توسعه نیز می‌باشند.

توسعه‌یافته، امّا مجزا (بُعد دوم) هیچ خوشه‌ای در آن توزیع نشده است. این بُعد مربوط به خوشه‌هایی است که خوش‌توسعه می‌باشند، اما محوری نیستند؛ بنابراین می‌توان نتیجه گرفت که هیچ‌یک از خوشه‌های هم‌واژگانی استخراج‌شده حوزه کشف دانش در این پژوهش چنین ویژگی را ندارند.

در حال ظهور یا زوال (بُعد سوم) شامل خوشه‌های ۳ «سیستم مدیریت دانش و اطلاعات»، ۴ «تحلیل سیستم»، ۶ «سیستم تولید داده و مدیریت تجزیه و تحلیل اطلاعات»، ۷ «معماری سرویس‌گرا و سیستم گزارش‌دهی»، ۸ «گزارش‌گیری از فلای‌دیتا»، ۹ «مدیریت سیستم پردازش»، ۱۰ «پیش‌بینی ارتباط آیتم اطلاعات»، ۱۱ «تحلیل و کشف داده»، ۱۲ «نگاشت داده‌ها و اطلاعات»، ۱۳ «دیداری‌سازی محتوا»، ۱۴ «مکعب داده‌ها»، ۱۵ «سیستم مدیریت عملیات پردازش» و ۱۶ «محیط بانک‌های اطلاعاتی» می‌باشد. این خوشه‌ها از مرکزیت و تراکم کمتری برخوردار هستند، درواقع در حوزه کشف دانش نادیده گرفته شده یا در حاشیه قرار دارند.

مرکزی، امّا توسعه‌نیافته (بُعد چهارم) شامل خوشهٔ ۵ «ابزارهای طبقه‌بندی» است. این خوشه محوری بوده، امّا هنوز توسعه‌نیافته می‌باشد.

بحث و نتیجه‌گیری

کشف دانش و میزان تأثیرگذاری دانش بر زندگی زیستی انسان سال‌هاست که فکر اندیشمندان و دانشمندان را به خود معطوف کرده است. این درک فلسفی و عملی از ارتباط میان دانش بشری و فعالیت‌های انسانی بسیار حائز

اهمیت می‌باشد. تحلیل هم‌رخدادی واژگان میزان ارتباط مفاهیم بین اصطلاحات به‌کاررفته در اسناد مختلف را اندازه‌گیری و نقشه ارتباطی میان آنها را ترسیم می‌کند. این روش خود زیرمجموعه‌ای از روش تحلیل محتوا می‌باشد که با روش‌های آماری و ترسیم نقشه ارتباط مستقیمی دارد. یکی از این روش‌های آماری تحلیل خوشه‌ای است که می‌تواند با بررسی مفاهیم، دامنه، محیط فضایی و ارتباط معناشناختی یک اصطلاح علمی ارتباط بین داده‌ها و ساختار علمی یک حوزه دانشی را ترسیم کند. تحلیل هم‌رخدادی واژگان با استفاده از اطلاعات تحلیل خوشه‌ای و به‌کارگیری مقیاس‌بندی چندبعدی می‌تواند ابهام‌های تحلیل خوشه‌ای را برطرف و میزان نزدیکی و شباهت داده‌ها را به‌صورت تصویر دوبعدی ترسیم کند. این روش می‌تواند گستردگی مفاهیم علمی در یک خوشه دانشی را نشان دهد. اما این روش‌ها به‌تنهایی قادر به بیان نقشه راهبردی در زمان به‌کارگیری آنها در حوزه‌های پنهان علم نیستند، برای همین باید از روش دیگری که از آن به‌عنوان نمودار راهبردی یاد می‌شود استفاده کرد. نمودار راهبردی می‌تواند با استفاده از داده‌های روش‌های قبلی گراف‌های پیچیده را به ساده‌ترین شکل و اینکه در موقعیت ظهور یا سقوط دانشی قرار دارند را نمایان سازد.

در این پژوهش با استفاده از روش تحلیل هم‌رخدادی واژگان زمینه موضوعی حوزه کشف دانش و ارتباط پنهان این حوزه با حوزه‌های دیگر نمایان شد. همچنین نشان داد حوزه کشف دانش به دلیل ماهیت خود و نیز تغییر روند انجام پردازش و کشف دانش به‌شدت به ابزارهای رایانه‌ای و انسان به‌عنوان آخرین حلقه در زمان بررسی داده‌های منظم‌شده (گارسیاپنالئو، کولوموپالاسیوز و هسو[1]، ۲۰۱۳) وابسته می‌باشد. تحلیل خوشه‌ای نشان داد حوزه کشف دانش از ۱۷ خوشهٔ موضوعی تشکیل شده است. این خوشه‌ها با توجه به میزان مرکزیت و تراکم می‌توانند نقش مؤثری در توسعه یک حوزه علمی داشته باشند. بهره‌گیری از نمودار راهبردی نشان داد خوشه‌های «دستگاه‌های ورودی»، «تجزیه و تحلیل کسب و کار» و «محیط بانک‌های اطلاعاتی» در بُعد اول نمودار به‌عنوان خوشه‌های خوش توسعه و محوری قرار دارند. هیچ خوشه‌ای در بُعد دوم نمودار به‌عنوان توسعه‌یافته، اما مجزا قرار نگرفت. خوشه‌هایی که در این بُعد قرار می‌گیرند علاوه بر اینکه می‌توانند در حوزه میزبان تغییر ایجاد کنند، هم‌زمان می‌توانند حوزه علمی خود را نیز توسعه دهند. خوشه‌های «سیستم مدیریت دانش و اطلاعات»، «تحلیل سیستم»، «سیستم تولید داده و مدیریت تجزیه و تحلیل اطلاعات»، «معماری سرویس‌گرا و سیستم گزارش‌دهی»، «گزارش‌گیری از فلای‌دیتا»، «مدیریت سیستم پردازش»، «پیش‌بینی ارتباط آیتم اطلاعات»، «تحلیل و کشف داده»، «نگاشت داده‌ها و اطلاعات»، «دیداری‌سازی محتوا»، «مکعب داده‌ها»، «سیستم مدیریت عملیات پردازش» و «محیط بانک‌های اطلاعاتی» در بُعد سوم نمودار راهبردی قرار گرفته‌اند که نوظهور یا در حال زوال می‌باشند.

با توجه به مطالب بیان‌شده تحلیل خوشه‌ای حاصل از هم‌رخدادی واژگان روش بسیار مناسبی برای شناسایی حوزه‌های پنهان در یک زمنیه علمی می‌باشد (چانگ[2] و دیگران، ۲۰۱۵). ترکیب مفاهیم در خوشه‌های ایجادشده نشان داد بیشتر مفاهیم مربوط به حوزه‌های فنی می‌باشد که وابستگی حوزه کشف دانش به فناوری اطلاعات، ارتباطات و علوم کامپیوتری را نشان می‌دهد. کشف خودکار آیتم‌های دانش یک روش و موضوع رو به رشدی (رانکان، پیسادو و گارسیامارتینز[3]، ۲۰۱۰) است که توسط ماشین‌های یادگیرنده اجرا می‌شود که می‌تواند در زمان کشف دانش در پایگاه‌های اطلاعاتی مختلف مورد استفاده قرار گیرد. این یافته گویای این واقعیت است بسیاری از پژوهش‌های علمی

1 . García-Peñalvo, Colomo-Palacios, Hsu
2 . Chung
3 . Rancan, Pesado & Garcia-Martinez

با استفاده از رایانه و نظام‌های اطلاعاتی صورت می‌گیرد. به‌طوری که هر کشور و سازمانی که بتواند از این ابزار به‌خوبی بهره‌برداری کند، می‌تواند به نتایج بسیار باارزشی دست پیدا کند. بسیاری از خوشه‌های این حوزه در بُعد سوم نمودار راهبردی قرار دارند. در میان این خوشه‌ها، خوشه «سیستم مدیریت عملیات پردازش» به دور از بقیه خوشه‌ها قرار دارد، امکان اینکه خوشه رو به زوال باشد، وجود دارد؛ زیرا دارای بیشترین واژگان بوده و در تمامی سطح این حوزه پراکنده است. اما دیگر خوشه‌ها به دلیل متمرکزشدن در بخش نزدیک به صفر بُعد سوم تمایل به این دارند که بتوانند بالغ شوند که نشان‌دهنده ظهور این خوشه‌ها می‌باشد. از میان خوشه‌های بُعد سوم، خوشه «تحلیل سیستم» و «سیستم تولید داده و مدیریت تجزیه و تحلیل اطلاعات» خوشه‌هایی هستند که در آینده نزدیک به بلوغ دانشی خواهند رسید. درواقع بیشتر تحلیل‌های حوزه کشف دانش به‌صورت برخط انجام می‌شود و این نوع تحلیل نیازمند مدیریت درست داده‌ها می‌باشد تا نتایج دقیق‌تری حاصل گردد. همچنین ترکیب مفاهیم در خوشه‌های ایجادشده نشان داد بیشتر مفاهیم مربوط به حوزه‌های فنی بوده و وابستگی حوزه پردازش داده به فناوری اطلاعات، ارتباطات و علوم رایانه را نشان می‌دهد. این نتیجه گویای این واقعیت است که حوزه پردازش داده فناوری-بنیان می‌باشد.

پیشنهادهای اجرایی پژوهش

از مفاهیم خوشه دستگاه‌های ورودی در هر چه بهترشدن ذخیره اطلاعات و داده‌ها استفاده شود؛

خوشه‌های بُعد اول نمودار راهبردی می‌توانند در زمان ورود، انباشت و دریافت بسیار مؤثر باشند. از طرفی استفاده بهینه از این خوشه‌ها باعث درک طبقه‌بندی اطلاعات و دانش شده و جامعه هدف را مورد حمایت قرار خواهد داد.

فهرست منابع

بهشتیان، مهدی و ابوالحسنی، حسین. (۱۳۸۴). *سیستم‌های اطلاعات مدیریت نگرشی جامع بر تئوری، کاربردی و طراحی*. تهران: انتشارات شرکت پردیس.

سهیلی، فرامرز، خاصه، علی‌اکبر و قاضی‌زاده، حمید. (زودآیند). ترسیم ساختار دانش در پژوهش‌های علوم قرآن و حدیث ایران با استفاده از تحلیل هم‌واژگانی. *پژوهش‌نامه علم‌سنجی*.

سهیلی، فرامرز، خاصه، علی‌اکبر و کرانیان، پریوش. (۱۳۹۷). روند موضوعی مفاهیم حوزه علم اطلاعات و دانش‌شناسی ایران بر اساس تحلیل هم‌رخدادی واژگان. *مطالعات ملی کتابداری و سازمان‌دهی اطلاعات*، ۲۹ (۲)، ۱۷۱-۱۹۰.

سهیلی، فرامرز، شعبانی، علی و خاصه، علی‌اکبر. (۱۳۹۵). ساختار فکری دانش در حوزه رفتار اطلاعاتی: مطالعه هم‌واژگانی. *تعامل انسان و اطلاعات*، ۲ (۴)،۲۱-۳۶.

غضنفری، مهدی، علیزاده، سمیه و تیمورپور، بابک. (۱۳۸۷). *داده‌کاوی و کشف دانش*. تهران: انتشارات دانشگاه علم و صنعت ایران.

منصوری، علی، توکلی‌زاده راوری، محمد، مکی‌زاده، فاطمه و طوسی، زیبا. (۱۳۹۵). روند تکامل فناوری: مورد مطالعه تحلیل رده‌های موضوعی پروانه‌های ثبت اختراع RFID. *پژوهش‌نامه پردازش و مدیریت اطلاعات*، ۳۲ (۱)، ۷۵-۹۱.

Aristodemou, L., Tietze, F., Athanassopoulou, N. & Minshall, T. (2017). Exploring the Future of Patent Analytics: A Technology Roadmapping Approach. *Centre for Technology Management Working Paper Series*, No 5. This Paper Has Been Accepted to the R&D Management Conference in Leuven, Belgium.

Callon M., Courtial J.P. & Turner W. (1991). La méthode Leximappe: un outil pour l'analyse stratégique du développement scientifique et technique. *Gestion de la recherche: Nouveaux problèmes nouveaux outils*, 207-277.

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 Chemsitry. *Scientometrics*, 22 (1), 155-205.

Chen, X., Chen, J., Wu, D., Xie, Y. & Li, J. (2016). Mapping the Research Trends by Co-Word Analysis Based on Keywords from Funded Project. *Procedia Computer Science*, 91, 547 – 555.

Chen, Y.L. & Huang, T.C.K. (2008). A novel knowledge discovering model for mining fuzzy multi-level sequential patterns in sequence databases. *Data & Knowledge Engineering*, 66 (3), 349–367.

Chung. M.H., Wang, Y., Tang, H., Zou, W., Basinger, J., Xu, X. & Tong, W. (2015). Asymmetric author-topic model for knowledge discovering of big data in toxicogenomics. *Frontiers in pharmacology*, 6, 81

Delecroix, B. & Eppstein, R. (2004). Co-word Analysis for The Non-Scientific Information Example of Reuters Business Briefings. *Data science journal*, 3, 80-87.

Ding, Y., Chowdhury, G.G. & Foo, S. (2001). Bibliometric Cartography of Information Retrieval Research by Using Co-word Analysis. *Information Processing & Management*, 37 (6), 817-842.

García-Peñalvo, F. J., Colomo-Palacios, R. & Hsu, J. Y. (2013). Discovering Knowledge through Highly Interactive Information Based Systems. *Journal of Information Science and Engineering (JISE)*, 29 (1), Foreword to the volume.

He, Q. (1999). Knowledge Discovery through Co-word Analysis. *Library trends*, 48 (1), 133-159.

Khasseh, A., Soheili, F., Sharif moghaddam, H. & Mousavi chelak, A. (2017). Intellectual structure of knowledge of imetrics: A co- word analysis. *Information processing & management*, 53 (3), 705-720.

Law, J., Bauin, S., Courtial, J. & Whittaker, J. (1988). Policy and the Mapping of Scientific Change: A Co-word Analysis of Research into Environmental Acidification. *Scientometrics*, 14 (3), 251-264.

Li, R., Chambers, T., Ding, Y., Zhang, G. & Meng, L. (2014). Patent Citation Analysis: Calculating Science Linkage Based on Citing Motivation. *Journal of the Association for Information Science and Technology*, 65 (5), 1007–1017.

Liu, G.Y., Hu, J.M. & Wang, H.L. (2012). A Co-word Analysis of Digital Library Field in China. *Scientometrics*, 91 (1), 203-217.

Melcer, E., Nguyen, T.H.D., Chen, Z., Canossa, A., El-Nasr, M.S. & Isbister, K. (2015). Games Research Today: Analyzing the Academic Landscape 2000-2014. *In Proceedings of the 10th International Conference on The Foundations of Digital Games*, At Pacific Grove. CA, USA, 9p.

Rancan, C., Pesado, P. & Garcia-Martinez, R. (2010). Issues in Rule Based Knowledge Discovering Process. *Advances and Applications in Statistical Sciences, Proceedings of the IV Meeting on Dynamics of Social and Economic Systems*, 2 (2), 303-314.

Ravikumar, S., Agrahari, A., & Singh, S.N. (2015). Mapping the intellectual structure of scientometrics: A co-word analysis of the journal Scientometrics (2005-2010). *Scientometrics*, 102 (2), 929-955.

Wang, X.D., Liu, J.J. & Sheng, F.S. (2014). Analysis of hotspots in the field of domestic knowledge discovery based on co-word analysis method. *Cybernetics and Information Technologies*, 14 (5), 145-158.

Wanying, Z., Jin, M. & Kun, L. (2018). Ranking Themes on Co-word Networks: Exploring the Relationships. *Information Processing and Management*, 54 (2), 203-218.

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.