Scientometrics Research Journal

Scientific Bi-Quarterly of Shahed University
Vol. 7, No. 1, Spring & Summer 2021 (Serial 13)

Research Article

DOI: 10.22070/rsci.2020.4904.1333

Bibliometrics and Mapping of Co-words in the Field of Linked Data

Elaheh Hosseini [1*]

Amir Ghaebi [2]

Roya Baradar [3]

1. *Ph.D. student, Department of Information Science and Knowledge Studies, Faculty of Psychology & Educational Sciences, Alzahra University, Tehran, Iran. (Corresponding Author)*
2. *Associate Professor, Department of Information Science and Knowledge Studies, Faculty of Psychology & Educational Sciences, Alzahra University, Tehran, Iran. Email: Ghaebi@alzahra.ac.ir*
3. *Associate Professor, Department of Information Science and Knowledge Studies, Faculty of Psychology & Educational Sciences, Alzahra University, Tehran, Iran. Email: rbaradar@alzahra.ac.ir*

Email: E.hosseini@alzahra.ac.ir

Date of Reception: 03/07/2019

Date of Acceptation: 01/12/2019

Abstract

Purpose: The research aims to visualize and analyze co-word network and thematic clusters in the field of linked data during 1986-2018.

Methodology: The study is an applied research in terms of the purpose, which conducted by using co-word analysis as a methodology and descriptive approach. Clusters determined by three methods. VOSViewer, SPSS, and UCINet were used for data analysis and network visualization.

Findings: The keywords *linked data* and *semantic web* in terms of co-word pairs had the highest frequencies. Co-word clustering generated five clusters, while hierarchical clustering produced two clusters. The USA was the most productive country and the highest share of documents published in various sub-categories of the *Computer Sciences*. Studies mostly published in *health* and *cultural heritage* contexts. The cluster *core concepts of the semantic web* was the most mature and central cluster, while *linked data usage in the context of cultural heritage* was a well-developed but isolated cluster.

Conclusion: The results can identify underlying trends and core themes by highlighting thematic gaps to avoid duplicate studies. Policymakers, researchers, and designers of the semantic technologies can plan predictably to develop themes in balance for the future and increase the quality and quantity of scientific outputs.

Keywords: Linked data, Intellectual Structure, Co-word analysis, thematic clusters, Hierarchical Clustering, Strategic Diagram.

پژوهش‌نامه علم‌سنجی
دوفصلنامه علمی دانشگاه شاهد/ دوره ۷/ شماره ۱/ بهار و تابستان ۱۴۰۰ (پیاپی ۱۳)

مقاله پژوهشی

DOI:10.22070/rsci.2020.4904.1333

کتاب‌سنجی و نگاشت هم‌رخدادی واژگان در حوزه داده‌های پیوندی

الهه حسینی [*1]

امیر غائبی [2]

رؤیا برادر [3]

۱. دانشجوی دکتری، علم اطلاعات و دانش‌شناسی، دانشگاه الزهرا (س)، تهران، ایران. (نویسنده مسئول)

۲. دانشیار، گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه الزهرا (س)، تهران، ایران. Email: Ghaebi@alzahra.ac.ir

۳. دانشیار، گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه الزهرا (س)، تهران، ایران. Email:rbaradar@alzahra.ac.ir

Email:E.hosseini@alzahra.ac.ir

صفحه ۱۱۶-۹۱
دریافت: ۱۳۹۸/۴/۱۲
پذیرش: ۱۳۹۸/۹/۱۰

چکیده

هدف: هدف از پژوهش، ترسیم و تحلیل شبکه هم‌رخدادی واژگان، و خوشه‌های موضوعی در حوزه داده‌های پیوندی در بازه زمانی ۲۰۱۸-۱۹۸۶ است.

روش‌شناسی: پژوهش از نظر هدف، نوعی مطالعه کاربردی است که با روش تحلیل هم‌رخدادی واژگانی با رویکرد توصیفی انجام شده است. خوشه‌بندی با استفاده از سه شیوه تعیین شده‌اند. تحلیل و ترسیم شبکه‌ها با استفاده از نرم‌افزارهای«وی.او.اس.ویوئر»، «اس.پی.اس.اس.» و «یو.سی.آی.نت.» انجام شد.

یافته‌ها: از نظر هم‌رخدادی واژگان، «داده‌های پیوندی» و «وب معنایی» بیشترین فراوانی را داشته‌اند. خوشه‌بندی هم‌واژگانی منجر به تشکیل ۵ خوشه و خوشه‌بندی سلسله‌مراتبی منجر به تشکیل ۲ خوشه شد. کشور «آمریکا» و حوزه‌های مختلف «علوم کامپیوتر» بیشترین فراوانی در دسته‌بندی موضوعی وب‌علوم در این حوزه را دارند. عمدتاً، مطالعات منتشرشده در دو بافت «سلامت» و «میراث فرهنگی» بودند. خوشه «مفاهیم هسته در داده‌های پیوندی» بالغ‌ترین و مرکزی‌ترین خوشه و خوشه «کاربرد داده‌های پیوندی در بافت میراث فرهنگی » خوشه توسعه‌یافته اما مجزا می‌باشد.

نتیجه‌گیری: نتایج حاصل می‌تواند با پررنگ‌کردن شکاف‌های موضوعی و جلوگیری از پژوهش‌های تکراری، روندهای اساسی، و موضوعات هسته و محبوب را شناسایی کند. سیاست‌گذاران، محققان، و طراحان فناوری‌های معنایی با آگاهی از این نتایج، می‌توانند برنامه‌ریزی پیش‌بینی‌کننده‌ای به‌منظور توسعه متوازن موضوعات و افزایش کمی و کیفی تولیدات علمی داشته باشند.

واژگان کلیدی: داده‌های پیوندی، ساختار فکری دانش، هم‌رخدادی واژگان، خوشه‌های موضوعی، خوشه‌بندی سلسله‌مراتبی، نمودار راهبردی.

مقدمه و بیان مسئله

«تحلیل هم‌واژگانی[1]» مانند «تحلیل زوج کتابشناختی» و «هم‌نویسندگی» و «هم‌استنادی» از متداول‌ترین تکنیک‌های کتاب‌سنجی و علم‌سنجی است. کلمات کلیدی یا اصطلاحات استخراج‌شده از عنوان، چکیده یا مجموعه مدارک، واحد تحلیل مورد استفاده در آن می‌باشد که نوع رابطه مورد بررسی در آن هم‌رخدادی اصطلاحات یا واژگان است (کوبو و همکاران[2]، ۲۰۱۱). «تحلیل هم‌واژگانی» در دهه ۱۹۸۰ مطرح شد و بر این فرض استوار است که استفاده از واژه‌های مشترک در دو یا چند مدرک، دلالت‌کننده نزدیکی مفاهیم آن متن‌ها به یکدیگر است که به‌واسطه آن می‌توان ساختار، مفاهیم، و مؤلفه‌های یک حوزه علمی را تعیین کرد. این شیوه به ترسیم ساختار حوزه‌ها و زمینه‌های علمی (وایتاکر[3]،۱۹۸۹)، شناسایی الگوهای پنهان و برجسته، روابط درونی و بیرونی مفاهیم (عصاره و همکاران، ۱۳۹۴)، رویدادهای در حال ظهور، تعیین روابط سلسله‌مراتبی مفاهیم در هستان‌شناسی‌های حوزه‌های علمی و زمینه‌های دانش تخصصی، خوشه‌بندی مفاهیم حوزه‌های علمی، و سیاست‌گذاری علم و دانش کمک می‌کند (احمدی و عصاره، ۱۳۹۶). «تحلیل هم‌رخدادی واژگان» نمونه‌ای از روش مدل‌سازی گرافیکی است که در آن از اندیشه‌های مربوط به تحلیل رابطه استفاده می‌شود (نف و کرلی[4]، ۲۰۰۹). به‌عبارت‌دیگر، «تحلیل هم‌واژگانی» یکی از انواع تحلیل‌های هم‌رخدادی به شمار می‌رود که برای نگاشت رابطه میان مفاهیم، اندیشه‌ها، و مشکلات در علوم پایه و علوم اجتماعی به کار می‌رود (لیو و دیگران[5]، ۲۰۱۲). «تحلیل هم‌رخدادی واژگان» می‌تواند به‌عنوان ابزاری قدرتمند امکان پیگیری تحولات ساختاری و تکامل شبکه ادراکی و اجتماعی را میسر سازد (مکی‌زاده و دیگران، ۱۳۹۵).

«داده‌های پیوندی» جدیدترین دستاورد در تکامل طبیعی «وب معنایی» است (آلمنگ و هندلر[6]، ۲۰۱۱). علاقه‌مندی کتابخانه‌ها، موزه‌ها و آرشیوها به بحث «داده‌های پیوندی» در حال گسترش است؛ چراکه «داده‌های پیوندی» کمک می‌کنند تا داده‌ها به شیوه‌ای نوآورانه و در فرمتی استاندارد قابلیت استفاده مجدد[7] یابند، و به منابع دیگر پیوند بخورند تا قابل کشف[8] شوند (گوئرینی[9]، ۲۰۱۳). «داده پیوندی» مفهومی است که برای اولین‌بار توسط «تیم برنرزلی[10]» در سال ۲۰۰۶ معرفی گردید. داده‌های پیوندی مطابق با اصولی است که پیوند میان پایگاه‌های داده، عناصر و واژگان را تسهیل می‌کند (مولازون و میلین[11]، ۲۰۱۴). پتانسیلی قوی در داده‌های پیوندی وجود دارد که راهکارهایی را برای بسیاری از مسائل کتابخانه‌ای مانند افزایش جستجوی وبی، کنترل مستند[12]، رده‌بندی و انعطاف‌پذیریِ داده ارائه می‌کند (زنگنه[13]،۲۰۱۳). روش داده‌های پیوندی، به مجموعه‌ای از بهترین روش‌ها برای ساختارمندسازی و پیوند میان داده‌های دسترس‌پذیر در محیط وب اشاره دارد (بایزر، هیث و برنزلی[14]، ۲۰۰۹). منظور از «داده پیوندی»، داده‌های

1 . Co-word Analysis/ Co-occurrence Analysis
2 . Cobo et al.
3 . Whittaker
4 . Neff & Corley
5 . Liu et al.
6 . Allemang & Hendler
7 . Reusability
8 . Discoverable
9 . Guerrini
10 . Tim Berners-Lee
11 . Moulaison & Million
12 . Authority control
13 . Zengenene
14 . Bizer, Heath & Bernes-Lee

منتشرشده در وب به‌صورت ماشین‌خوان است که معنای آنها به‌صورت دقیق تعریف شده، به سایر دادگان‌های موجود در وب پیوند داده شده و می‌تواند توسط سایر دادگان‌ها نیز مورد پیوند قرار گیرد. داده‌های پیوندی در سال‌های اخیر در ادبیات علوم کتابداری و اطلاع‌رسانی محل تمرکز و مطالعه بوده است (سوث ویک[1]، ۲۰۱۵). درواقع از مهم‌ترین دلایل تلاش برای بهره‌مندی از داده‌های پیوندی، ارزش افزوده‌ای است که به‌واسطه کشف و ثبت روابط میان موجودیت‌ها ایجاد می‌شود. این ارزش افزوده به معنای دانشی است که از این طریق تولید می‌شود، چون تعامل میان موجودیت‌های اطلاعاتی، کاربر نهایی را به سطح بالاتری از شناخت سوق می‌دهد که دانش نامیده می‌شود. از طرفی دیگر، برقراری روابط میان موجودیت‌ها باعث افزایش دسترس‌پذیری می‌شود. بدین معنا که اگر لایه‌های روابط بر مبنای رویکرد هستان‌شناختی گسترش یابند، تعداد نقاط دسترسی شیء نیز توسعه خواهد یافت و در ادامه زمینه برای انتخاب مرتبط‌ترین موجودیت فراهم می‌شود (طاهری، ۱۳۹۴). «جستجوی یکپارچه هوشمند» قوی‌تر می‌شود و بسیاری از مشکلات مربوط به موتورهای جستجوی یکپارچه مانند عدم توانایی پشتیبانی از پرس‌وجوهای پیچیده و رتبه‌بندی ضعیف و کند رفع می‌گردد و باعث افزایش دقت جستجو شود. پیوند داده‌های مرتبط به هم به کاربران اجازه می‌دهد که به‌طور یکپارچه اطلاعات مرتبط را مورد جستجو قرار دهند. به لطف وجود ارتباطات معنادار و مبتنی بر استانداردها، در مقیاس وسیع، یکپارچگی رخ می‌دهد، قابلیت میانکنش‌پذیری و استدلال روی داده‌ها ممکن می‌شود. همچنین، جستجوها دقیق‌تر می‌شوند زیرا به‌واسطه شناساهای یکتای منبع (یو.آر.آی.) ابهام از بین می‌رود (سوث ویک، ۲۰۱۵).

با توجه به اهمیت بحث داده‌های پیوندی و نقش پررنگی که در تحقق وب معنایی دارد، شناسایی ساختار فکری دانش در این حوزه ضروری است تا به محققان آن کمک کند تا تصویر واضح‌تری از تکامل این حوزه به دست آورند و مسیر مطالعات آتی را مشخص کنند؛ بنابراین، هدف از پژوهش حاضر، ترسیم و تحلیل نقشه موضوعی و ساختار فکری دانش حوزه «داده‌های پیوندی» در بازه زمانی ۲۰۱۸-۱۹۸۶ با استفاده از تکنیک هم‌رخدادی واژگان است؛ بنابراین این مطالعه تلاش می‌کند تا به این سؤال کلی پاسخ دهد که خوشه‌های موضوعی در جهت بازنمون ساختار فکری دانش در حوزه «داده‌های پیوندی» با استفاده از تکنیک هم‌رخدادی واژگان شامل چه مفاهیمی بوده، و از نظر رشد، بلوغ، و توسعه دارای چه مختصاتی هستند؟

سؤال‌های پژوهش

این پژوهش تلاش می‌کند تا به پرسش‌های زیر پاسخ دهد:

۱. پنج رتبه برتر تولیدات علمی این حوزه از نظر دسته‌بندی موضوعی وب‌علوم، سال انتشار، سازمان فعال، نویسنده، کشور و حوزه پژوهشی چیست؟
۲. ساختار شبکه و خوشه‌های موضوعی تشکیل‌دهنده این حوزه بر اساس هم‌واژگانی از چه مفاهیمی تشکیل شده‌اند و وضعیت آنها از نظر توزیع فراوانی، تعداد لینک‌ها (پیوند) و قدرت کلی پیوند چگونه است؟
۳. نتایج مربوط به تحلیل هم‌واژگانی منجر به شکل‌گیری چه خوشه‌هایی و با چه موضوعاتی شده است؟
۴. ترسیم نقشه مقیاس‌بندی چندبعدی[2] برای پژوهش‌های حوزه «داده‌های پیوندی» باعث شناسایی چه خوشه‌های کلی می‌شود؟

1 . Southwick
2 . Multidimensional Scaling

خوشه‌های حاصل از تحلیل هم‌واژگانی از نظر میزان بلوغ و توسعه‌یافتگی در نمودار راهبردی[1] در حوزه «داده‌های پیوندی» در چه وضعیتی قرار دارند؟

چارچوب نظری

شبکه مفهومی حاصل از «تحلیل هم‌واژگانی» بر این دلالت می‌کند که اگر دو اصطلاح با هم در یک مدرک به کار روند و بسامد تکرار آنها بالا باشد، این دو واژه ارتباط معنایی بیشتری با هم دارند که به کشف و درک پیوند و رابطه میان دو موضوع در یک حوزه پژوهشی کمک می‌کند (احمدی و عصاره، ۱۳۹۴). به‌عبارت‌دیگر، هرچه واژه‌های مشترک دو مقاله بیشتر باشد، این مقالات ارتباط محتوایی و معنایی نزدیک‌تری دارند و دارای مشابهت موضوعی می‌باشند (نویون و ون‌ران[2]، ۱۹۹۸) و نیز میزان ارتباط شناختی میان یک مجموعه مدارک را نشان می‌دهد (سالمی و کوشا، ۱۳۹۲). این تکنیک همچنین حرکت و پویایی علم را مشخص می‌کند (کارون، کورتیال و ترنر[3]، ۱۹۸۶) بدین معنی که حوزه مورد بررسی چه سیر تحولی را پشت سر گذاشته و احتمالاً در آینده چه زمینه‌هایی در کانون توجه قرار خواهند گرفت و اینکه چه رابطه مفهومی بین حوزه‌ها و زیرحوزه‌های مفهومی وجود دارد (احمدی و عصاره، ۱۳۹۶). علاوه‌براین، این تکنیک این امکان را فراهم می‌کند تا خوشه‌های موضوعی در حال ظهور و همچنین خوشه‌های توسعه‌یافته در راستای پیش‌بینی مسیر پژوهش‌های آتی آشکار شوند (سهیلی، شعبانی و خاصه، ۱۳۹۴؛ لی و سو[4]، ۲۰۱۰).

پیشینه پژوهش

بررسی پیشینه‌ها نشان می‌دهد که همایندی واژگان یکی از روش‌هایی است که در بیشتر مطالعات علم‌سنجی جهت ترسیم ساختار موضوعی و فکری شکل‌گرفته در حوزه‌های علمی مختلف استفاده شده است.

پیشینه پژوهش در داخل

ترسیم و تحلیل شبکه هم‌رخدادی و بررسی ساختار دانش در ایران در حوزه «داده‌های پیوندی» و حوزه‌های وابسته به آن انجام نشده است.

نتایج مطالعه هم‌رخدادی در حوزه رفتار اطلاعاتی توسط سهیلی، شعبانی و خاصه (۱۳۹۴) نشان داد که از نظر هم‌رخدادی دو کلیدواژه «نیازهای اطلاعاتی» و «رفتار اطلاعاتی» بیشترین فراوانی را داشته‌اند. شبکه دارای ۱۱ خوشه است که خوشه‌های «رفتار اطلاعات سلامت»، «مطالعات کاربران»، «شبکه‌های اجتماعی» و «ربط و بازیابی اطلاعات» نقشی محوری دارند و خوشه‌های مرکزی و بالغ به حساب می‌آیند. نتایج همچنین نشان داد که مطالعات رفتار اطلاعاتی، ارتباط تنگاتنگی با حوزه پزشکی و سلامت دارند.

نتایج هم‌رخدادی حوزه مدیریت اطلاعات توسط خادمی و حیدری (۱۳۹۵) حاوی ۹ خوشه موضوعی شامل «سیستم‌های اطلاعاتی»، «مدیریت عملکرد»، «مدیریت اطلاعات»، «داده»، «سلامت الکترونیک»، «بیوانفورماتیک»، «اطلاعات»، «سازمان‌دهی دانش»، و «مدیریت اطلاعات سلامت» است. نتایج مطالعه نشان داد که ساختار موضوعی

1 . Strategic diagram (SD)
2 . Noyons & Van Raan
3 . Collan, Courtial, &Turner
4 . Lee & Su

شناسایی‌شده می‌تواند به سیاست‌گذاری علمی این حوزه کمک کند.

البته در حوزه‌های دیگر مانند افسردگی (مکی‌زاده و دیگران، ۱۳۹۵)، آر.اف.آی.دی.[1] (منصوری و دیگران، ۱۳۹۵)، علوم قرآنی و حدیث (قاضی‌زاده، سهیلی و خاصه،۱۳۹۷)، علم اطلاعات و دانش‌شناسی (سهیلی، خاصه و کرانیان، ۱۳۹۷) مطالعاتی انجام شده است که در ادامه برخی از تحقیقات جدیدتر بیشتر توضیح داده شده‌اند.

امامی، ریاحی‌نیا، و سهیلی (زودآیند) در مطالعه‌ای حوزه تجهیزات پزشکی و آزمایشگاهی را با استفاده از تحلیل هم‌رخدادی تحلیل کردند. نتایج مطالعه نشان داد که از لحاظ زوج هم‌رخدادی، دو کلیدواژه «تجهیزات پزشکی» و «ایمنی بیمار» بیشترین فراوانی را داشته‌اند. خوشه «روش‌های درمانی» و «تجهیزات قلب و عروق»، خوشه‌های بالغ و مرکزی هستند و موضوعات محوری و نوظهور را در مطالعات این حوزه هدایت می‌کنند. علاوه‌براین خوشه «تجهیزات پزشکی» و «بیماری‌ها» حالت حاشیه‌ای دارند، چون از نظر مرکزیت و تراکم در سطح پایینی قرار دارند و در مطالعات این حوزه کمتر مورد توجه بوده‌اند.

پروانه‌های ثبت اختراع در حوزه کشف دانش توسط رستمی، سهیلی، خاصه (زودآیند) نیز تحلیل شده‌اند. نتایج نشان دادند که از نظر هم‌رخدادی واژگانی، زوج‌های «رایانه و ماشین‌ها» و «مدیریت پردازش آنلاین و پردازش آنلاین» بیشترین فراوانی را داشته‌اند. ترکیب مفاهیم این مطالعه نیز نشان داد که حوزه کشف دانش به‌طور قوی به ابزارهای فناورانه، حوزه‌های دانش فناوری اطلاعات، ارتباطات و انسان وابسته است.

سهرابی و غفاری (زودآیند) موضوعات پرکاربرد حوزه ارتباطات علمی با استفاده از روش هم‌رخدادی واژگان را بررسی کرده‌اند. یافته‌های پژوهش بر این دلالت داشت که دو کلیدواژه دسترسی آزاد و ارتباطات علمی، از نظر هم‌رخدادی بیشترین فراوانی را در این حوزه دارند. خوشه «دسترس‌پذیری علم» خوشه بالغ و مرکزی است و خوشه «سنجش علمی»، خوشه‌ای مرکزی ولی توسعه‌نیافته است.

پیشینه پژوهش در خارج

ترسیم و تحلیل شبکه هم‌رخدادی در حوزه‌هایی مانند حوزه امنیت اطلاعات (لی[2]، ۲۰۰۸)، علوم کتابداری و اطلاع‌رسانی (ونگ، ژانگ و وی[3]، ۲۰۱۱)، بازیابی اطلاعات (دینگ، چودهوری و فو[4]، ۲۰۱۱)، کتابخانه دیجیتال در چین (لیو، هو و ونگ[5]، ۲۰۱۲) انجام شده است. علاوه‌براین، ترسیم و تحلیل ساختارفکری دانش با استفاده هم‌زمان از تکنیک‌های شبکه هم‌رخدادی، تحلیل شبکه اجتماعی، خوشه‌بندی سلسله‌مراتبی (دندروگرام)، و تحلیل مقیاس‌بندی چندبعدی در حوزه‌های مختلف مانند مدیریت دانش (صدیقی و جلالی‌منش، ۲۰۱۴)، مخازن مؤسسه‌ای[6] (چو[7]، ۲۰۱۴)، مقالات کنفرانس سی.اچ.آی در طول دو دهه ۲۰۱۳-۱۹۹۴ (لیو و دیگران[8]، ۲۰۱۴)، مقالات مجله ساینتومتریکس در طول ۲۰۰۵-۲۰۱۰ (راویکومار و دیگران[9]، ۲۰۱۴)، خلاقیت (ژانگ و دیگران، ۲۰۱۵)، اینترنت اشیاء

1 . RFID
2 . Lee
3 . Wang, Zhang, & Wei
4 . Ding, Chowdhury, & Foo
5 . Liu, Hu, Wang
6 . institutional repository
7 . Cho
8 . Lie & et al.
9 . Ravikumar & et al.

اشیاء (یانگ، لی، و لی[1]، ۲۰۱۵)، و آی‌متریکس (خاصه و دیگران، ۲۰۱۷) انجام شده است. نتایج این پژوهش‌ها، روندها و حوزه‌های موضوعی اصلی را تعیین کرده، و موضوعات توسعه‌یافته، در حال توسعه، و در حال تکامل را نیز مشخص کرده‌اند.

در حوزه «داده‌های پیوندی» تنها دو مطالعه یافت شد که در ادامه بیان شده‌اند:

نیک‌نیا و میرطاهری (۲۰۱۵) در پژوهشی با استفاده از تولیدات علمی پایگاه اسکوپوس و تحلیل شبکه هم‌رخدادی این حوزه را در طول یک دهه تجزیه و تحلیل کردند. نتایج پژوهش آنها نشان داد که حوزه‌های موضوعی بسیاری متمرکز بر اصطلاحات مرتبط با حوزه کامپیوتر مانند کلان داده، رایانش ابری، داده معنایی، فناوری‌های معنایی، وب معنایی، هوش مصنوعی، برنامه‌نویسی کامپیوتری، و جستجوی معنایی هستند.

کیاو و وانگ[2] (۲۰۱۸) در مطالعه‌ای با استفاده از تکنیک‌های مذکور حوزه داده‌های پیوندی را با ۹۶۴ مقاله از پایگاه وب‌علوم تحلیل کردند. نتایج پژوهش نشان داد که ۹ خوشه موضوعی، حاوی مفاهیم اینترنت اشیاء، پیوند موجودیت، آموزش، وب معنایی، داده‌های پیوندی، وب داده، دی.بی.پدیا، یکپارچه‌سازی داده و آنتولوژی هستند.

جمع‌بندی از مرور پیشینه

بررسی پیشینه‌های پژوهش نشان می‌دهد که شبکه هم‌رخدادی واژگان در حوزه‌های مختلف علمی، ترسیم و تحلیل شده است. تنها دو مطالعه درباره تحلیل شبکه هم‌رخدادی واژگان در حوزه داده‌های پیوندی یافت شد که البته از نظر تعداد مقاله، پایگاه علمی مورد بررسی، و تکنیک‌های به‌کاررفته، تفاوت‌هایی با پژوهش حاضر دارند که این تفاوت‌ها در بخش روش پژوهش، به‌طور مشروح توضیح داده شده و برجسته شده است.

روش‌شناسی پژوهش

این پژوهش از نظر هدف، نوعی مطالعه کاربردی است که با روش تحلیل هم‌رخدادی واژگان، با رویکرد توصیفی و کتاب‌سنجی انجام شده است. «تحلیل هم‌واژگانی» که بر اساس «هم‌رخدادی واژگان» عمل می‌کند به‌عنوان یک روش تحلیل محتوا، یکی از روش‌های علم‌سنجی است (کینگ[3]، ۱۹۸۷). خوشه‌بندی با سه شیوه مختلف تحلیل هم‌واژگانی با استفاده از نرم‌افزار «وی.او.اس. ویوئر»، خوشه‌بندی سلسله‌مراتبی، و خوشه‌بندی با مقیاس‌بندی چندبعدی انجام شده است. برای تحلیل و ترسیم شبکه‌ها از نرم‌افزارهای«وی.او.اس.ویوئر (نسخه ۱.۶.۹)»، «اس.پی.اس.اس. (نسخه ۲۲)» و «یو.سی.آی.نت.(نسخه ۶)» استفاده شد. تفاوت عمده پژوهش حاضر با پژوهش‌های مشابه در این است که استراتژی پژوهشی به‌کاررفته در این مطالعه شامل چند تکنیک مختلف (خوشه‌بندی با استفاده از تحلیل هم‌واژگانی با نرم‌افزار وی.او.اس.ویوئر، مقیاس‌بندی چندبعدی، خوشه‌بندی سلسله‌مراتبی (دندوگرام)) است که به‌طور هم‌زمان برای مصورسازی و تحلیل شبکه موضوعی این حوزه به کار رفته است تا تصویر چندجانبه‌ای از این حوزه را برجسته کند. به‌عبارت‌دیگر، خوشه‌بندی حاصل از تحلیل هم‌واژگانی شامل پنج خوشه اصلی حاوی ۲۸۴ کلیدواژه انجام شد، ولی چون دارای کلیدواژه‌هایی بودند که از لحاظ بافت داده‌های پیوندی، ارتباط عمیقی نداشتند، از تکنیک‌های دیگر برای خوشه‌بندی استفاده شد. خوشه‌بندی سلسله‌مراتبی، برای ۵۰ واژه پرتکرار و پربسامد و مرتبط از

1 . Yang, Lee, Lee
2 . Kyaw, Wang
3 . King

نظر بافت موضوعی انجام شد تا تصویر دقیق‌تر و عمیق‌تری از ساختار دانش این حوزه ارائه کند. خوشه‌بندی با استفاده از مقیاس‌بندی دوبعدی نیز به ادغام و تقلیل خوشه‌ها کمک می‌کند تا خوشه‌های دارای هم‌پوشانی واژگانی در یک خوشه قرار گیرند. نهایتاً، با استفاده از نمودار راهبردی وضعیت پنج خوشه اصلی از نظر و بلوغ توسعه آنها نیز بررسی شد. این تکنیک‌ها به‌عنوان، تکنیک‌های مکمل هم کمک کردند تا ساختار دانش این حوزه از چند جهت توصیف و بررسی شود. بنابراین، خوشه‌ها در سه مرحله نهایی شده‌اند. تحلیل اولیه خوشه‌ها با استفاده از تکنیک‌های خوشه‌بندی در روش تحلیل هم‌واژگانی در نرم‌افزار «وی.او.اس.ویوئر» تعیین شده‌اند. تحلیل دوم، متمرکز بر کلمات پرتکرار این خوشه‌ها در خوشه‌بندی سلسله‌مراتبی است. تحلیل سوم، تقلیل خوشه‌های اولیه به خوشه‌های دارای همبستگی درونی است. این تکنیک‌ها به نهایی‌کردن مفاهیم خوشه‌های هسته به‌طور عمیق و تخصصی کمک می‌کند. تفاوت عمده دیگر در رکوردهای این مقاله است که سعی می‌کند ساختار فکری این حوزه را با تولیدات علمی بیشتر (۴۰۰۱ مقاله) در پایگاه وب علوم و با دیدی جامع‌تر و به‌روزتر تحلیل کند. از سوی دیگر، نویسندگان این مقاله، از دیدگاه متخصصان حوزه «داده‌های پیوندی» نیز در تأیید نتایج، و نام‌گذاری خوشه‌ها بهره برده‌اند؛ بنابراین، تحلیل‌ها و نتایج دارای دیدگاه میان‌رشته‌ای هستند. موارد مذکور بر نوآوری این مطالعه اشاره دارد.

مراحل مختلف روش‌شناسی پژوهش عبارت‌اند از:

گردآوری داده‌ها و پیش‌پردازش

جامعه این پژوهش را کلیه مدارکی تشکیل می‌دهد که در بازه زمانی ۲۰۱۸-۱۹۸۶ با موضوع «داده‌های پیوندی» در پایگاه «وب علوم» نمایه شده‌اند و در مهر ۱۳۹۷ برای این پژوهش مورد جستجو قرار گرفتند. در همین راستا در بخش موضوع[1] یا عنوان[2]، عبارت Linked data در بخش مجموعه «هسته وب علوم» مورد جستجو قرار گرفت که ۴۰۰۱ مدرک بازیابی شدند. به‌عبارت‌دیگر، تمرکز این مطالعه بر روی تولیداتی است که با موضوع یا عنوان «داده‌های پیوندی» در فیلد عنوان یا موضوع مجموعه هسته وب علوم نمایه شده‌اند. همان‌گونه که کلریویت آنالتیکز[3] ادعا می‌کند که نتایج جستجو در فیلد موضوع حاوی عنوان، چکیده، کلمات کلیدی نویسندگان مقاله، و کلمات کلیدی بیشتری مبتنی بر یک الگوریتم خاص[4] است. کلریویت ادعا می‌کند که این شیوه جستجو احتمال اینکه اسناد و کلمات کلیدی مهمی نادیده گرفته شوند، را کاهش می‌دهد (کلریویت آنالتیکز ۲۰۱۹).

پس از جستجو رکوردها با فرمت متن ساده[5] دانلود شدند، و به فایل اکسل منتقل شدند. پس از بررسی ستون کلیدواژه‌ها، ۳۸۲ مدرک دارای کلیدواژه نبودند که از کل مدارک حذف شدند و ۳۶۱۹ مدرک برای بررسی نهایی باقی ماندند. سپس روی ستون کلیدواژه‌ها یکدست‌سازی انجام شد. نام کشورها و ایالت‌هایی مانند United-States، New-South-Wales، Western-Australia، Australia، Norway، England، UK حذف شدند و با استفاده از نظر متخصص موضوعی در این حوزه، موارد مفرد و جمع واژه‌ها به یک حالت تبدیل شدند. تغییرات انجام شده روی کلیدواژه‌ها در جدول شماره ۱ ذکر شده است.

1 . Topic
2 . Title
3 . Clarivate Analytics
4 . Algorithm-based keyword plus
5 . Plain text

جدول ۱. تغییرات کلیدواژه‌ها به منظور یکدست‌سازی

ردیف	کلیدواژه موجود در مدرک	واژه منتخب پس از یکدست‌سازی	ردیف	کلیدواژه موجود در مدرک	واژه منتخب پس از یکدست‌سازی
۱	Resources	Resource	۱۲	RDF data	RDF
۲	Languages	Language	۱۳	Resource description framework	RDF
۳	Breast-cancer	Breast cancer	۱۴	Risk factors	Risk factor
۴	Record-linkage	Record linkage	۱۵	Costs	Cost
۵	socioeconomic-status	socioeconomic status	۱۶	Models	Model
۶	Information-retrieval	information retrieval	۱۷	Systems	System
۷	Open linked data	Linked data	۱۸	Databases	Database
۸	LOD	Linked data	۱۹	Thesauri	Thesaurus
۹	Linked-data	Linked data	۲۰	Ontologies	Ontology
۱۰	Linked (open) data	Linked data	۲۱	Algorithms	Algorithm
۱۱	Semantic-web	semantic web	۲۲	Vocabularies	Vocabulary

داده‌های حاصل از این بخش به پاسخ‌گویی پرسش اول پژوهش انجامید.

تحلیل هم‌واژگانی

با بهره‌گیری از نرم‌افزار «وی.او.اس.ویوئر» پیوند کلی بین مفاهیم در خوشه‌ها مصورسازی شدند. با آزمون و خطا آستانه تکرار کلیدواژه‌ها در نرم‌افزار «وی.او.اس.ویوئر» ۱۰ در نظر گرفته شد. برای تعیین آستانه تکرار، به تعداد کلیدواژه‌ها نیز توجه شد. تحلیل این نرم‌افزار در آستانه تکرار ۱۰، شامل پنج خوشه که حاوی ۲۸۴ کلیدواژه هستند را نشان می‌دهد. به‌عبارت‌دیگر، ۲۸۴ کلیدواژه موجود در پنج خوشه شناسایی‌شده، دارای ۱۰۲۵۵ هم‌رخدادی بوده‌اند. این اعداد، حدود ۳۰ درصد کل فراوانی کلیدواژه‌های مورد مطالعه را تشکیل می‌دهد که به قانون یک‌سوم برادفورد نیز نزدیک است. ازاین‌رو، می‌توان گفت که هم‌رخدادی‌های این ۲۸۴ کلیدواژه قادرند محتوای اصلی پژوهشی (ساختار فکری دانش) در حوزه «داده‌های پیوندی» را نمایان سازد. درمجموع، مفاهیم و واژگان نهایی مجدداً به تأیید متخصص موضوعی این حوزه رسید. داده‌های حاصل از این بخش منجر به پاسخ‌گویی به پرسش دوم پژوهش شد.

خوشه‌بندی سلسله‌مراتبی و مقیاس‌بندی چندبعدی

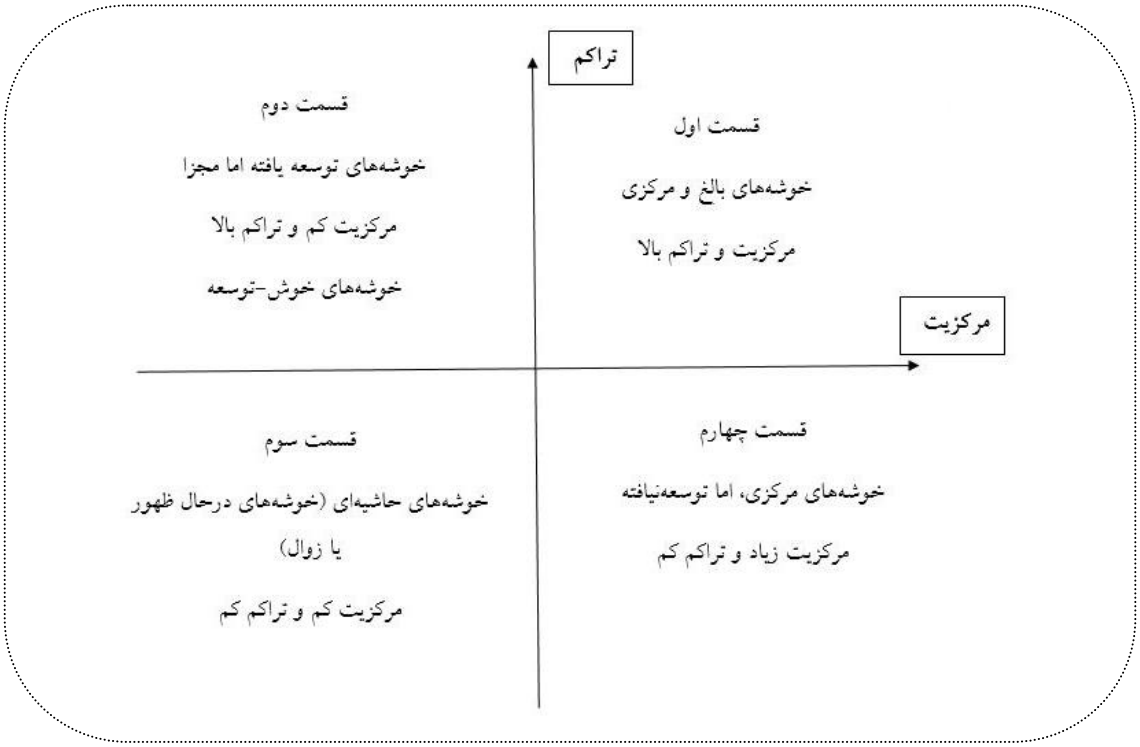
از «خوشه‌بندی سلسله‌مراتبی (دندوگرام)» و «مقیاس‌بندی چندبعدی» نیز می‌توان برای «تحلیل هم‌واژگانی» استفاده کرد. با استفاده از نرم‌افزار «اس.پی.اس.اس.» می‌توان روابط سلسله‌مراتبی بین مفاهیم در خوشه‌ها را نشان داد که برای پاسخ‌گویی به پرسش سوم کمک‌کننده است. ماتریس همبستگی که طبق ماتریس فراوانی هم‌واژگانی تهیه شد، به نرم‌افزار «اس.پی.اس.اس.» منتقل شد و سپس با استفاده از خوشه‌بندی سلسله‌مراتبی[1]، روش وارد[2]، و مربع فاصله

1 . Hierarchical Clustering
2 . Ward's Method

اقلیدسی[1]، خوشه‌ها و نمودار دندروگرامی[2] برای این ماتریس ترسیم شد که در پاسخ به پرسش سوم توضیح داده شده است.

علاوه‌براین، نقشه مقیاس‌بندی چندبعدی، می‌تواند خوشه‌های مهم و جایگاه آنها را در بین سایر خوشه‌ها بر اساس درجه همبستگی آشکار سازد (سهیلی، شعبانی و خاصه، ۱۳۹۴). جایگاه و فاصله کلیدواژه‌ها در مقیاس‌بندی چندبعدی به ادغام و تقلیل خوشه‌ها کمک می‌کند. نقشه‌های دوبعدی ابزار مناسبی برای دیدن میزان نزدیکی خوشه‌ها هستند. محور افقی (بعد اول) در نقشه دوبعدی بیانگر درجه همبستگی درونی هر یک از خوشه‌های موضوعی است و محور عمودی (بعد دوم) بیانگر تأکید خوشه‌های موضوعی می‌باشد (خاصه و دیگران، ۲۰۱۷) که برای پاسخ‌گویی به پرسش چهارم از این تکنیک استفاده شده است. بدین منظور، با استفاده از ماتریس هم‌رخدادی ۵۰ کلیدواژه پرتکرار در نرم‌افزار «اس.پی.اس.اس.» نمودار چندبعدی PROXSCAL ترسیم شد.

نمودار راهبردی



شکل ۱. بخش‌های چهارگانه یک نمودار راهبردی (هو و دیگران، ۲۰۱۳)

در یک نمودار راهبردی، اگر مرکزیت یک خوشه موضوعی بالاتر باشد براین دلالت می‌کند که آن خوشه از جایگاه مهم‌تری در حوزه پژوهشی برخوردار است. از سوی دیگر، اگر تراکم خوشه‌ای بیشتر باشد نشان‌دهنده بلوغ و قابلیت بیشتر آن حوزه است. درنتیجه می‌توان در یک نمودار راهبردی، محور X را به‌عنوان مرکزیت رتبه و محور Y را دلالت‌کننده بر تراکم دانست. یعنی اگر نمودار راهبردی مطابق شکل ۱ به چهار قسمت تقسیم شود، خوشه‌های قرارگرفته در هر قسمت بر اساس میزان مرکزیت و تراکم وضعیت متفاوتی دارند. قسمت اول بیانگر خوشه‌های بالغ و مرکزی است که در مرکز مطالعات حوزه مورد نظر قرار دارد که حاوی مرکزیت و تراکم بالایی هستند. قسمت دوم

1 . Squared Euclidean Distance
2 . Dendogram

معروف به خوشه‌های توسعه‌یافته اما مجزا، بیانگر مرکزیت کم و تراکم بالا و حاوی خوشه‌هایی است که خوشه مرکزی نیستند ولی مجزا و خوش توسعه می‌باشند. خوشه‌های قسمت سوم حاوی مرکزیت کم و تراکم کم می‌باشند و بیانگر خوشه‌های حاشیه‌ای (خوشه‌های در حال ظهور یا زوال) هستند که توجه اندکی را به خود جلب نموده‌اند. قسمت چهارم حاوی خوشه‌های مرکزی، اما توسعه‌نیافته است که اگرچه مرکزی هستند اما نابالغ‌اند؛ یعنی مرکزیت زیاد و تراکم کم دارند (هو و دیگران[1]، ۲۰۱۳). درنهایت، برای هریک از خوشه‌ها بر اساس تعداد کلیدواژه‌هایی که دارند یک ماتریس مربعی و سپس همبستگی ایجاد گردید و سپس برای هریک از ماتریس‌های همبستگی یک مرکزیت و یک تراکم محاسبه شد. در پایان، یک نمودار راهبردی ترسیم گردید که منجر به پاسخ‌گویی به پرسش پنجم شد.

یافته‌های پژوهش

پاسخ به سؤال اول پژوهش. پنج رتبه برتر تولیدات علمی این حوزه از نظر دسته‌بندی موضوعی وب علوم، سال انتشار، سازمان فعال، نویسنده، کشور و حوزه پژوهشی چیست؟

پنج رتبه اول در دسته‌بندی موضوعی وب‌علوم مربوط به حوزه‌های مختلف «علوم کامپیوتر» است که در جدول ۲ ذکر شده‌اند. در بین رتبه‌های بعدی حوزه علم کتابداری و علم اطلاعات با ۳۰۰ تولید علمی در رتبه هفتم وب‌علوم و در رتبه چهارم حوزه‌های پژوهشی قرار دارد. کشور «آمریکا» بیشترین تولیدات علمی در این حوزه را داراست.

جدول ۲. پنج رتبه برتر تولیدات علمی حوزه داده‌های پیوندی بر اساس اطلاعات توصیفی وب علوم

رتبه	دسته‌بندی موضوعی وب‌علوم	حوزه‌های پژوهشی	سال انتشار	سازمان	نویسندگان	کشورها
۱	Computer Science Information Systems	Computer Science	۲۰۱۵	National University of Ireland NUI Galway	Auer, S.	USA
	تعداد ۱۳۹۰	تعداد ۲۱۷۸	تعداد ۶۳۶	تعداد ۱۳۱	تعداد ۵۰	تعداد ۱۰۳۹
۲	Computer Science Theory Method	Engineering	۲۰۱۶	University of Western Australia	Decker, S.	Germany
	تعداد ۱۲۲۶	تعداد ۴۳۱	تعداد ۶۰۶	تعداد ۱۳۱	تعداد ۳۴	تعداد ۴۷۸
۳	Computer Science Artificial Intelligence	Public Environmental Occupational health	۲۰۱۴	University of California System	Verborgh, R.	England
	تعداد ۹۱۶	تعداد ۳۰۲	تعداد ۵۴۳	تعداد ۱۱۸	تعداد ۳۴	تعداد ۴۱۹
۴	Computer Science Software Engineering	Information Science Library Science	۲۰۱۷	Harvard university	Mannens, E.	Australia
	تعداد ۴۰۶	تعداد ۳۰۰	تعداد ۵۰۸	تعداد ۹۲	تعداد ۳۳	تعداد ۳۲۶
۵	Computer Science Interdisciplinary Applications	General Internal Medicine	۲۰۱۳	University of Leipzig	Lehmann, J.	Spain
	تعداد ۳۳۸	تعداد ۱۵۳	تعداد ۳۶۵	تعداد ۸۰	تعداد ۲۸	تعداد ۲۸۷

1 . Hu & et al.

پاسخ به سؤال دوم پژوهش. ساختار شبکه و خوشه‌های موضوعی تشکیل‌دهنده این حوزه بر اساس هم‌واژگانی از چه مفاهیمی تشکیل شده‌اند و وضعیت آنها از نظر توزیع فراوانی، تعداد لینک‌ها (پیوند) و قدرت کلی پیوند چگونه است؟

به‌طور کلی، با استفاده از الگوریتم‌ها و تحلیل‌های نرم‌افزار «وس.ویوئر» پنج خوشه حاوی ۲۸۴ کلیدواژه، ۱۰۲۲۵ هم‌رخدادی، ۳۴۰۳۴ قدرت کلی پیوند، و ۱۴۱۶۳ لینک می‌باشد. در جدول‌های ۳-۷، اطلاعات هشت کلیدواژه پرتکرار هر خوشه ذکر شده است. اعداد مربوط به تعداد لینک‌ها و قدرت کلی پیوند برگرفته از تحلیل نرم‌افزار است.

جدول ۳. کلیدواژه‌های پربسامد حاصل از تحلیل هم‌واژگانی در خوشه اول

خوشه اول ۱۴۰ کلیدواژه/ مجموع رخداد ۱۵۸۶۵/۳۹۰۰ قدرت کلی پیوند/۷۸۸۰ تعداد لینک‌ها				
رتبه	کلیدواژه	فراوانی	پیوند	قدرت کلی پیوند
۱	mortality	۲۰۶	۱۶۱	۸۸۷
۲	outcomes	۱۳۹	۱۴۴	۵۶۴
۳	Population	۱۰۷	۱۳۶	۴۷۵
۴	Risk	۱۰۷	۱۲۹	۳۷۴
۵	Health	۱۰۶	۱۳۳	۳۸۸
۶	Care	۹۶	۱۲۷	۳۷۴
۷	epidemiology	۸۲	۱۲۲	۳۵۲
۸	survival	۸۲	۱۱۳	۴۱۱

جدول ۴. کلیدواژه‌های پربسامد حاصل از تحلیل هم‌واژگانی در خوشه دوم

خوشه دوم ۶۰ کلیدواژه/ مجموع رخداد ٤۳۷۳/ قدرت کلی پیوند ۱۱۵۹۳/ تعداد لینک‌ها ۳۰۵۸				
۱	Linked data	۱۵۶۴	۲۴۹	۳۳۸۰
۲	Semantic Web	۶۱۶	۱۵۴	۱۶۵۶
۳	Ontology	۳۴۶	۱۴۲	۱۰۶۰
۴	Web	۲۸۰	۱۲۴	۷۲۸
۵	RDF	۲۲۹	۱۱۵	۶۹۰
۶	SPARQL	۱۴۰	۸۸	۴۳۴
۷	DBpedia	۸۳	۵۵	۲۳۲
۸	Big Data	۶۷	۷۳	۲۰۷
۹	Open Data	۶۷	۶۳	۱۹۲

جدول ۵. کلیدواژه‌های پربسامد حاصل از تحلیل هم‌واژگانی در خوشه سوم

خوشه سوم ٤۲ کلیدواژه/ مجموع رخداد ۱۱۱۰ / قدرت کلی پیوند ٤۱۲۵/ تعداد لینک‌ها ۱۹۵۲				
۱	System	۱۲۶	۱۲۶	۴۴۰
۲	Management	۷۰	۱۳۶	۳۰۶
۳	Data Integration	۶۹	۷۲	۲۰۹

ادامه جدول ۵. کلیدواژه‌های پربسامد حاصل از تحلیل هم‌واژگانی در خوشه سوم

خوشه سوم ٤٢ کلیدواژه/ مجموع رخداد ۱۱۱۰ / قدرت کلی پیوند ٤۱۲۵/ تعداد لینک‌ها ۱۹۵۲				
۴	interoperability	۶۷	۷۲	۲۱۳
۵	Information	۶۵	۱۲۰	۲۶۲
۶	database	۵۳	۹۸	۲۰۰
۷	Services	۵۰	۱۰۴	۲۱۸
۸	semantics	۴۰	۵۳	۱۳۴

جدول ۶. کلیدواژه‌های پربسامد حاصل از تحلیل هم‌واژگانی در خوشه چهارم

خوشه چهارم ۲۵ کلیدواژه/ مجموع رخداد ۵۳۰/ قدرت کلی پیوند ۱۵۷۲/ تعداد لینک‌ها ۸۵۵				
۱	Model	۷۹	۱۱۸	۲۶۴
۲	provenance	۴۴	۴۲	۱۱۶
۳	data quality	۳۸	۳۱	۸۷
۴	performance	۳۵	۶۳	۱۰۰
۵	design	۳۴	۵۴	۱۲۰
۶	language	۳۰	۵۵	۹۹
۷	algorithm	۲۸	۴۴	۸۷
۸	education	۲۵	۳۱	۵۵

جدول ۷. کلیدواژه‌های پربسامد حاصل از تحلیل هم‌واژگانی در خوشه پنجم

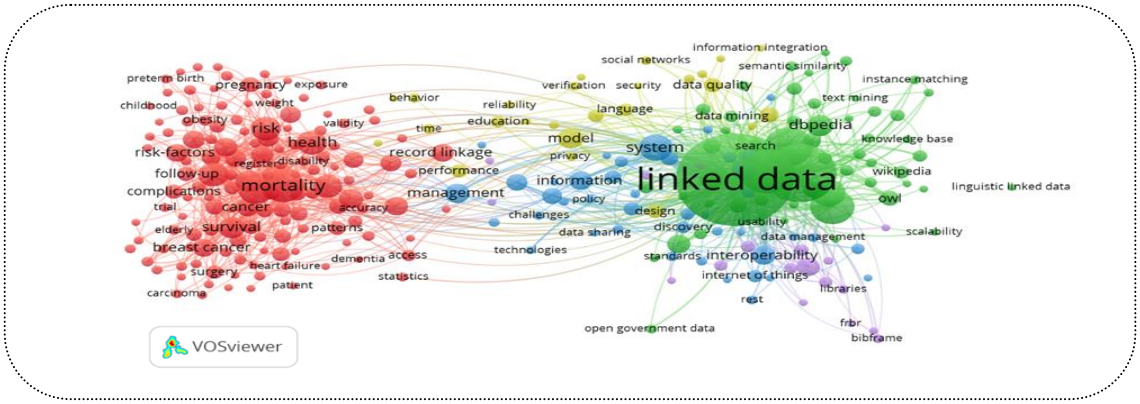
خوشه پنجم ۱۷ کلیدواژه/ مجموع رخداد ۳۱۲/ قدرت کلی پیوند ۸۷۹/ تعداد لینک‌ها ٤۱۸				
۱	metadata	۶۶	۶۰	۱۷۲
۲	thesaurus	۲۶	۳۶	۸۲
۳	digital libraries	۲۴	۲۷	۶۱
۴	skos	۲۳	۳۲	۷۷
۵	cultural heritage	۲۲	۲۲	۴۷
۶	vocabulary	۲۱	۳۳	۷۱
۷	libraries	۱۶	۲۴	۵۳
۸	bibframe	۱۴	۱۱	۳۱

با توجه به مفاهیم موجود در خوشه‌ها و با عنایت به نظرات متخصصان موضوعی این حوزه، پنج خوشه اصلی زیر نام‌گذاری شدند:

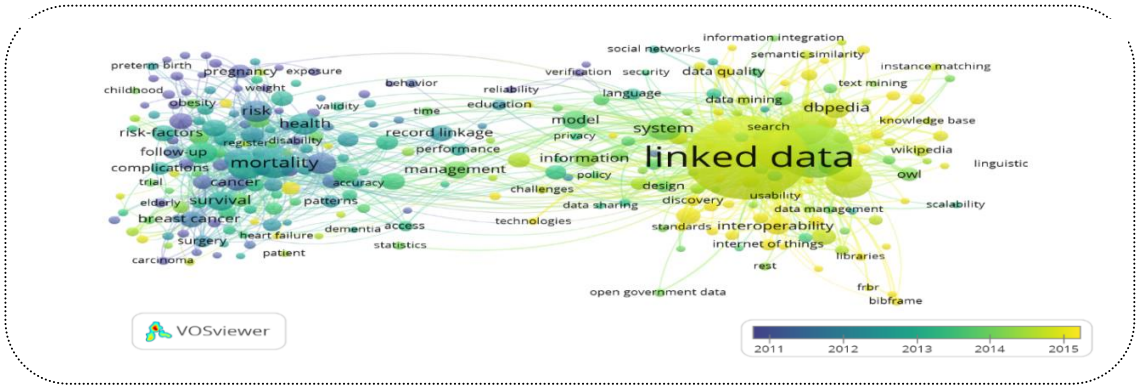
خوشه اول «کاربرد داده‌های پیوندی در حوزه سلامت»؛ خوشه دوم «مفاهیم هسته داده‌های پیوندی»؛ خوشه سوم «یکپارچگی و میانکنش‌پذیری در داده‌های پیوندی»؛ خوشه چهارم «طراحی الگوریتم برای بهبود کیفیت داده در داده‌های پیوندی» و خوشه پنجم «کاربرد داده‌های پیوندی در بافت میراث فرهنگی با بهره‌گیری از اصطلاح‌نامه‌ها،

الگوهای مفهومی و استانداردهای فراداده‌ای».

شکل ۲ نشان‌دهنده ساختار شبکه ۲۸۴ کلیدواژه پربسامد در حوزه داده‌های پیوندی است که با استفاده از نرم‌افزار «وس. ویوئر (نسخه ۱.۶.۹) ترسیم شده است. همان‌طور که شکل یک نشان می‌دهد ساختار شبکه این حوزه از پنج خوشه که با رنگ‌های مختلف نشان داده شده، تشکیل شده است.



شکل ۲. ساختار شبکه مفاهیم حوزه داده‌های پیوندی برگرفته از خوشه‌بندی نرم‌افزار «وی.او.اس.ویوئر»

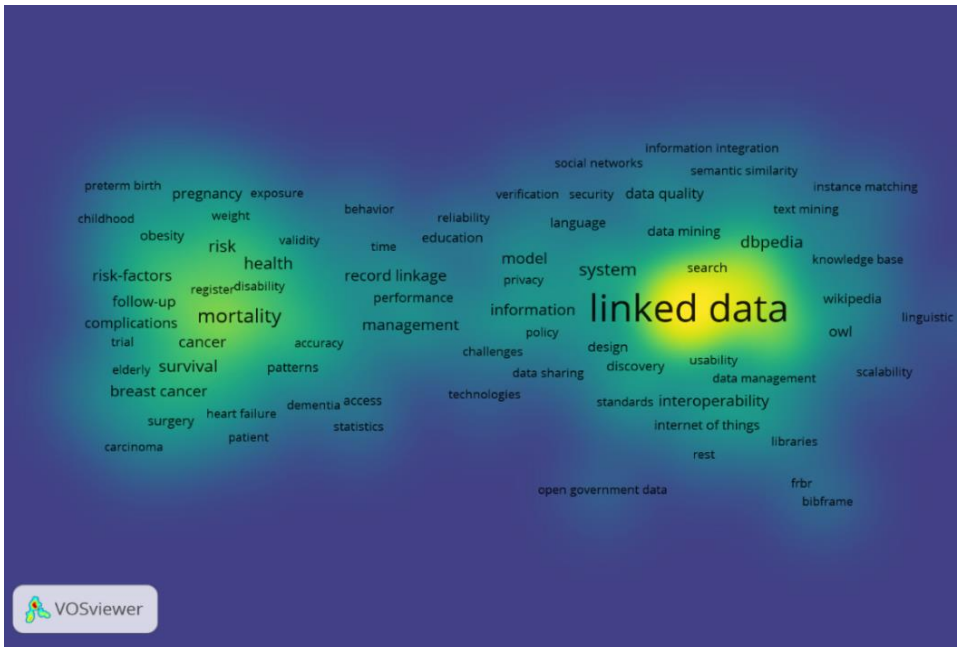


شکل ۳. ساختار لایه‌ای شبکه مفاهیم حوزه داده‌های پیوندی برگرفته از نرم‌افزار «وی.او.اس.ویوئر»

شکل ۳، «مصورسازی لایه‌ای[1]» شبکه در این حوزه را نشان می‌دهد. رنگ‌های این نقشه توسط وزن و امتیاز آنها در شبکه تعیین می‌شود. رنگ آبی دارای کمترین امتیاز و رنگ سبز امتیاز متوسط و رنگ زرد دارای بیشترین امتیاز است. یعنی هرچه از رنگ آبی به رنگ زرد پیش برود، امتیاز بیشتر و اهمیت کلیدواژه در شبکه نیز بیشتر می‌شود (ون اک و ولتمن[2]، ۲۰۱۸، ص. ۸). علاوه‌براین، بر اساس رنگ‌های نوار موجود در زیر نقشه که به سال‌های انتشار مقالات نیز دلالت می‌کند، می‌توان گفت که از سال ۲۰۱۱ به ۲۰۱۵ از وزن کمتر به بیشتر می‌رسد، یعنی سمت راست نقشه، وزن و اهمیت بیشتری در شبکه دارد و کلیدواژه‌های مهم‌تر و کاربردی‌تر را شامل می‌شود.

شکل ۴ «مصورسازی چگالی خوشه‌ای[3]» را نشان می‌دهد. هرچه که رنگ خوشه شبکه به زرد نزدیک‌تر باشد، یعنی چگالی آن خوشه بیشتر و در میان سایر خوشه‌های شبکه بااهمیت‌تر است (ون اک و ولتمن، ۲۰۱۸، ص. ۱۰).

1 . Overlay visualization
2 . Van Eck & Waltman
3 . Cluster Density visualization



شکل ۴. ساختار چگالی خوشه‌ای شبکه مفاهیم حوزه داده‌های پیوندی برگرفته از نرم‌افزار «وی.او.اس.ویوئر»

پاسخ به سؤال سوم پژوهش. نتایج مربوط به تحلیل هم‌واژگانی منجر به شکل‌گیری چه خوشه‌هایی و با چه موضوعاتی شده است؟

پس از ترسیم یک ماتریس متقارن ۲۸۴*۲۸۴ در یک فایل اکسل، هم‌رخدادی کلیدواژه‌های تولیدات علمی (۳۶۱۹ تولیدات علمی حاوی کلیدواژه) در این ماتریس ثبت شد که در جدول زیر زوج‌های برتر این هم‌واژگانی همراه با تحلیل آنها در ادامه ذکر شده‌اند.

جدول ۸. هم‌رخدادی کلیدواژه‌ها در حوزه «داده‌های پیوندی»

ردیف	کلیدواژه	فراوانی	ردیف	کلیدواژه	فراوانی
۱	semantic web*linked data	۵۲۱	۲۱	IOT*linked data	۲۹
۲	ontology*linked data	۱۷۰	۲۲	Linked data*provenance	۲۹
۳	RDF*linked data	۱۶۳	۲۳	RDF*ontology	۲۶
۴	datamining*linked data	۱۶۲	۲۴	Linked data*visualization	۲۶
۵	Open data*big data	۱۵۹	۲۵	NLP*linked data	۲۲
۶	RDF*semantic web	۹۵	۲۶	semantics*linked data	۲۲
۷	SPARQL*linked data	۹۱	۲۷	IR*linked data	۲۱
۸	ontology*semantic web	۶۹	۲۸	Recommender system*linked data	۲۱
۹	Open data*linked data	۵۰	۲۹	SKOS*linked data	۲۰
۱۰	RDF*SPARQL	۴۶	۳۰	semantic search*linked data	۱۸
۱۱	metadata*linked data	۴۵	۳۱	Web of data*semantic web	۱۸

ادامه جدول ۸. هم‌رخدادی کلیدواژه‌ها در حوزه «داده‌های پیوندی»

ردیف	کلیدواژه	فراوانی	ردیف	کلیدواژه	فراوانی
۱۲	Web of data*linked data	۴۲	۳۲	SPARQL*ontology	۱۸
۱۳	data integration*linked data	۴۲	۳۳	Semantic web*web of data	۱۸
۱۴	Linked data*DBpedia	۴۲	۳۴	Linked data*data quality	۱۸
۱۵	Linked data*data integration	۴۲	۳۵	KM*semantic web	۱۷
۱۶	Semantic web*SPARQL	۳۹	۳۶	Data integration*semantic web	۱۷
۱۷	Linked data*big data	۳۹	۳۷	OWL*RDF	۱۷
۱۸	SPARQL*semantic web	۳۸	۳۸	Semantic web*DBpedia	۱۷
۱۹	interoperability*linked data	۳۵	۳۹	Linked data*data mining	۱۷
۲۰	OWL*linked data	۲۹	۴۰	Linked data*semantic search	۱۷

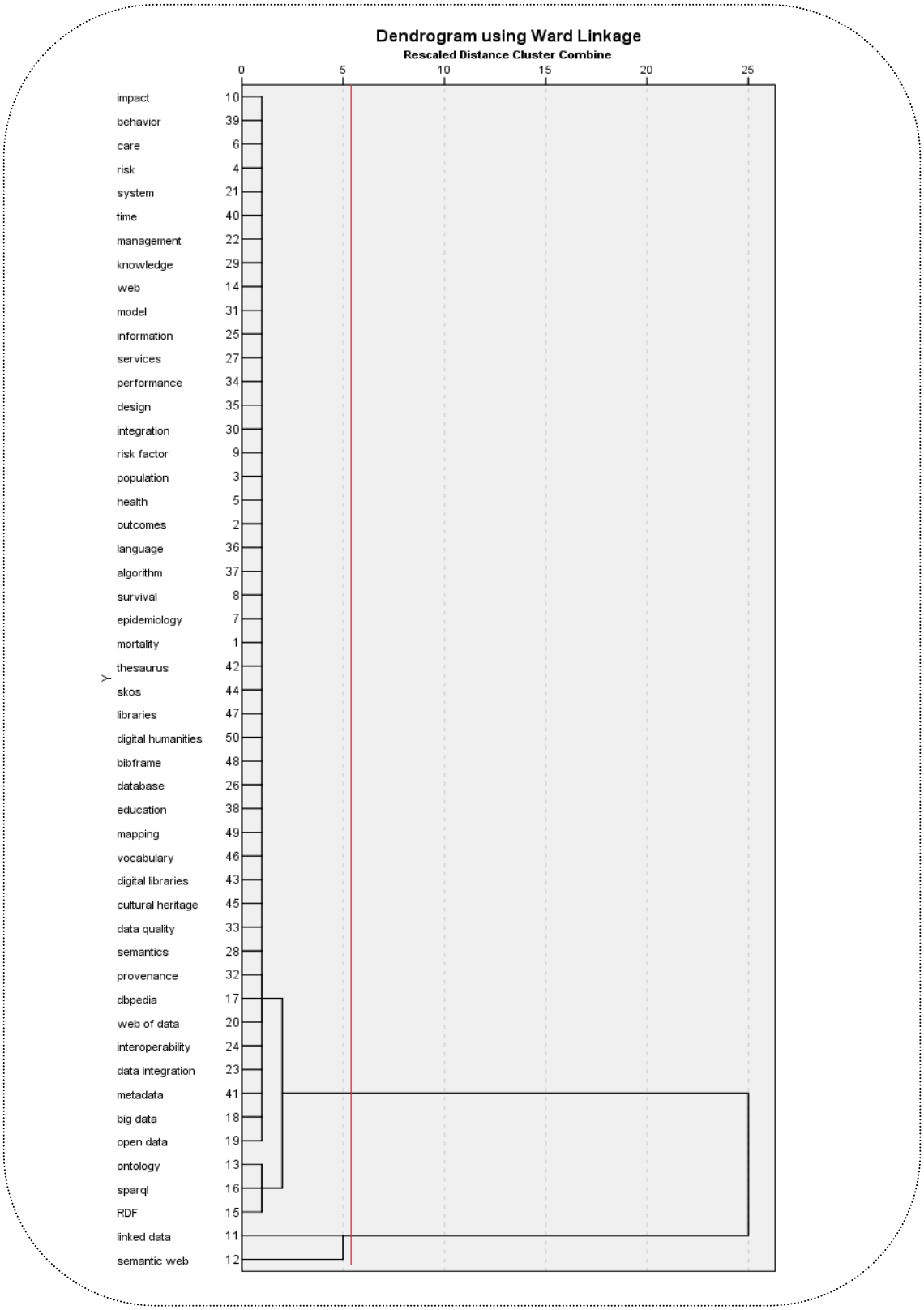
در ادامه، از هر پنج خوشه مذکور در سؤال دوم، ۱۰ کلیدواژه که دارای پرتکرارترین هم‌رخدادی بودند انتخاب شدند و یک ماتریس ۵۰*۵۰ به وجود آمد. ماتریس همبستگی حاصل از این ماتریس هم‌رخدادی با استفاده از نرم‌افزار اس.پی.اس.اس. (نسخه ۲۲) ایجاد شد و سپس با استفاده از خوشه‌بندی سلسله‌مراتبی، روش وارد، و مربع فاصله اقلیدسی، خوشه‌ها و نمودار دندروگرامی برای این ماتریس ترسیم شد. دندروگرام حاصل از خوشه‌بندی سلسله‌مراتبی در شکل ۵ نشان داده شده است.

همان‌طور که در دندروگرام (شکل ۵) مشخص است، تجزیه و تحلیل یافته‌های مربوط به هم‌واژگانی واژه‌های پررخداد در پنج خوشه قبلی (۵۰ کلیدواژه) منجر به شکل‌گیری دو خوشه موضوعی نهایی شده است. هدف از به‌کارگیری واژه‌های پرتکرار این بود که در دندوگرام نهایی کلیدواژه‌هایی قرار بگیرند که بیشترین و مستقیم‌ترین ارتباط معنایی و موضوعی را با موضوع اصلی تحلیل و خوشه‌ها داشته باشند که از حضور کلیدواژه‌های غیرضروری خودداری شود؛ بنابراین خوشه‌های تشکیل‌شده عبارت‌اند از:

خوشه اول: این خوشه بزرگ‌ترین خوشه موجود است که حاوی ۴۸ کلیدواژه است. این خوشه را می‌توان «کاربرد داده‌های پیوندی در بافت‌های سلامت، کتابخانه‌ها، و میراث فرهنگی» نام نهاد. بسیاری از کلیدواژه‌های پرتکرار در جدول بالا در این خوشه قرار دارند. از مهم‌ترین کلیدواژه‌های موجود در این خوشه می‌توان به «یکپارچگی»، «میانکنش‌پذیری»، «معنا»، «وب داده»، «اسپارکل»، «آر.دی.اف.»، «هستان‌شناسی»، «کلان داده»، «سلامت»، «کتابخانه‌ها»، و «میراث فرهنگی» اشاره کرد که کلیدواژه‌های بسیار مهم در پژوهش‌ها و مطالعات مربوط به حوزه داده‌های پیوندی می‌باشند که شکل ۱ نیز مؤید این مطلب است.

خوشه دوم: این خوشه از دو کلیدواژه اساسی «داده‌های پیوندی» و «وب معنایی» تشکیل شده است. می‌توان موضوع این خوشه را «وب معنایی و داده‌های پیوندی» نام نهاد. به‌عبارت‌دیگر، داده‌های پیوندی به‌عنوان چهارچوبی برای وب معنایی و یکی از کلیدی‌ترین مباحث حوزه «وب معنایی» است. ایده «داده‌های پیوندی» با ایده «وب معنایی» هم‌راستاست. این روش ساختار وب را از یک فضای مدارک پیوندخورده به فضای داده‌های به هم پیوندخورده تغییر داده و نام «وب داده» را به خود اختصاص داده است (بیزر، هیث و برنرزلی، ۲۰۰۹). مفهوم داده‌های پیوندی به معنای

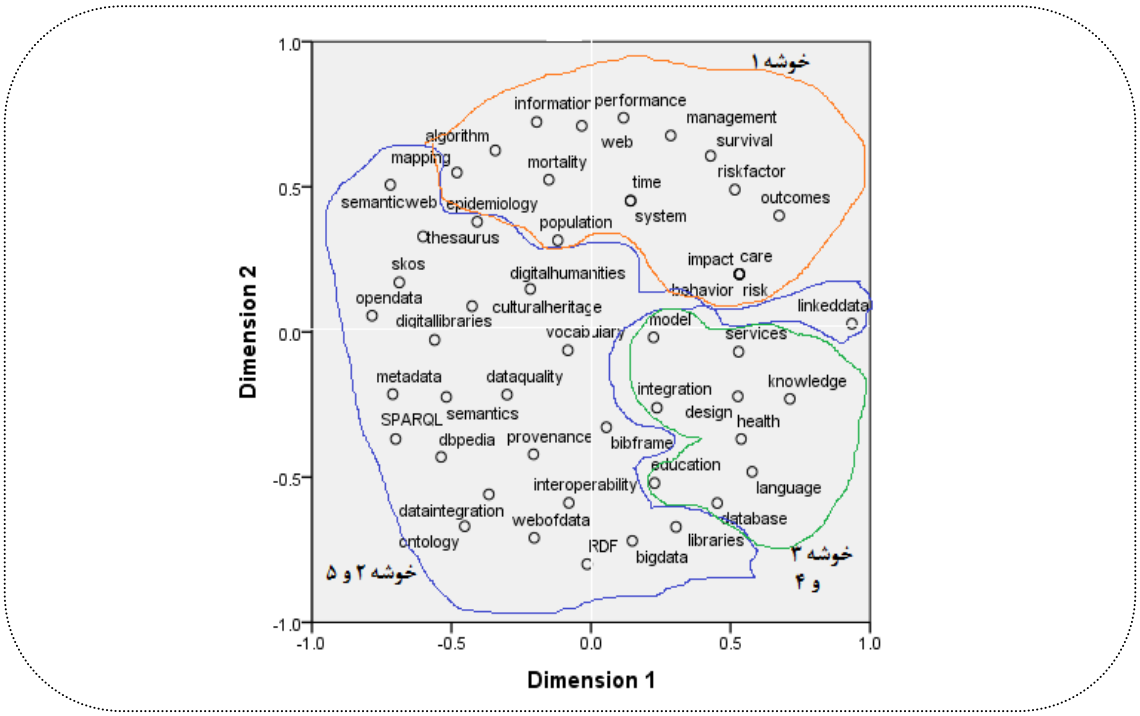
پیوند داده‌ها در سطح وب و انتشار آن در فرمتی ماشین‌خوان با استفاده از استانداردهایی مانند «آر.دی.اف.» و «یو.آر.آی.» می‌باشد (هیث و بیزر، ۲۰۱۱).



شکل ۵. دندروگرام حاصل از خوشه‌بندی سلسله‌مراتبی به روش هم‌واژگانی

پاسخ به سؤال چهارم پژوهش. ترسیم نقشه مقیاس‌بندی چندبعدی (دوبعدی) برای پژوهش‌های حوزه «داده‌های پیوندی» باعث شناسایی چه خوشه‌های کلی می‌شود؟

برای پاسخ‌گویی به این سؤال به‌منظور کسب دیدگاهی جامع‌تر، درکی عمیق‌تر و بهتر درباره مفاهیم این حوزه از روش مقیاس‌بندی چندبعدی استفاده گردید. به‌منظور ترسیم نقشه چندبعدی از کلیه کلیدواژه‌های استفاده‌شده در ماتریس ترسیم دندروگرام (۵۰ کلیدواژه برتر یعنی ۱۰ کلیدواژه از پنج خوشه اصلی پرسش دوم) استفاده شد. با استفاده از نرم‌افزار «اس.پی.اس.اس.»، و ترسیم نمودار چندبعدی PROXSCAL طبق شکل ۶ این نمودار ترسیم شد. استفاده از مقیاس‌بندی چندبعدی کمک می‌کند که بر اساس جایگاه و فاصله کلیدواژه‌ها برخی خوشه‌ها در هم ادغام شوند. به‌گونه‌ای که پنج خوشه اصلی و اولیه را بتوان به خوشه‌های کلی‌تری تقلیل داد. بنابراین، این سه خوشهٔ نهاییِ درهم‌ادغام‌شده خوشه‌هایی هستند که هم‌پوشانی واژگانی دارند که باعث شده است در یک خوشه قرار گیرند. همان‌طور که ملاحظه می‌شود در برخی قسمت‌های این نقشه دوبعدی، تراکم موضوع‌ها بیشتر به چشم می‌خورد که اغلب همان موضوعات پرتکرار و در بافت «داده‌های پیوندی» جزو مفاهیم اصلی هستند. نتایج این تحلیل نشان داد که خوشه‌های «۳ و ۴» و «۲ و ۵» به علت ارتباطی که بین مفاهیم موجود در آنها بوده است در یک خوشه قرار گرفته‌اند.



شکل ۶. نقشه مقیاس‌بندی دوبعدی حاصل از تحلیل هم‌واژگانی حوزه «داده‌های پیوندی»

پاسخ به سؤال پنجم پژوهش. خوشه‌های حاصل از تحلیل هم‌واژگانی از نظر میزان بلوغ و توسعه‌یافتگی در حوزه «داده‌های پیوندی» در چه وضعیتی قرار دارند؟

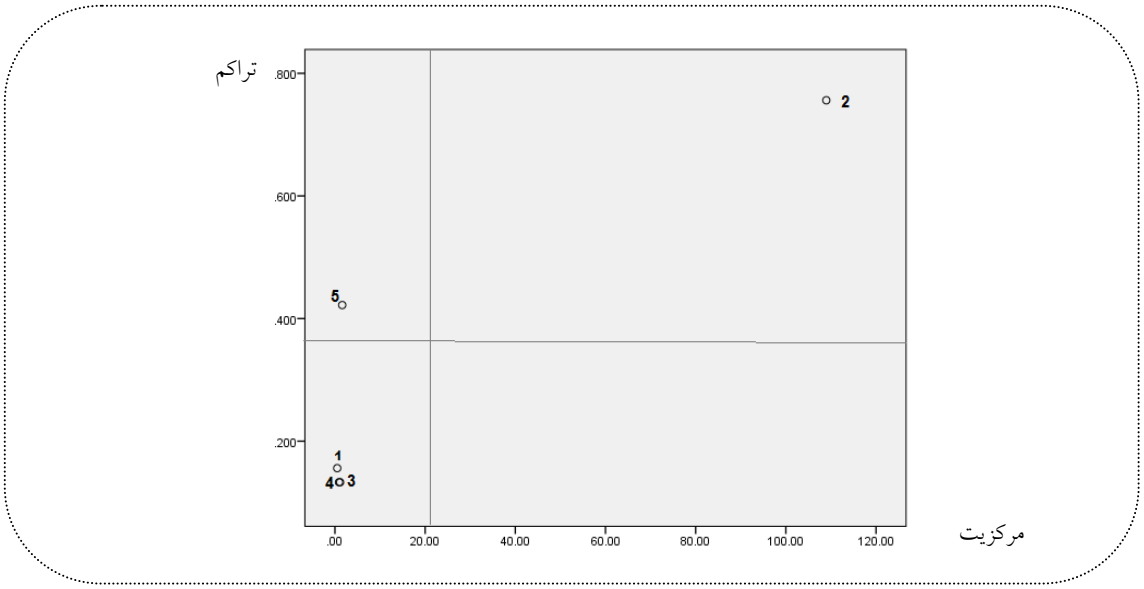
با استفاده از مفاهیم تحلیل شبکه‌های اجتماعی مانند مرکزیت و تراکم می‌توان نمودار راهبردی طراحی کرد و وضعیت بلوغ و توسعه‌یافتگی هر خوشه را بررسی و تحلیل کرد. ازاین‌رو، برای هر پنج خوشه اصلی، ماتریس فراوانی و سپس ماتریس همبستگی ایجاد شد و با استفاده از نرم‌افزار «یو.سی.آی.نت» رتبه و تراکم هر خوشه و میانگین آن به دست آمد.

جدول ۹. تراکم و مرکزیت ۵ خوشه اصلی

خوشه	مرکزیت	تراکم
خوشه اول: «کاربرد داده‌های پیوندی در حوزه سلامت»	۵.۰	۱۵۶.۰
خوشه دوم: «مفاهیم هسته در داده‌های پیوندی»	۹۷.۱۰۸	۷۵۶.۰
خوشه سوم: «یکپارچگی و میانکنش‌پذیری در داده‌های پیوندی»	۹۱.۰	۱۳۳.۰
خوشه چهارم: «طراحی الگوریتم برای بهبود کیفیت داده در داده‌های پیوندی»	۱.۱۳	۱۳۳.۰
خوشه پنجم: «کاربرد داده‌های پیوندی در بافت میراث فرهنگی با بهره‌گیری از اصطلاح‌نامه‌ها، الگوهای مفهومی و استانداردهای فرادادهای»	۶۱.۱	۴۲۲.۰

بنابراین طبق جدول ۹، خوشه دوم با بالاترین مرکزیت در میان خوشه‌ها از جایگاه مهم و مرکزی در این حوزه برخوردار است و همچنین از طرف دیگر با دارابودن بالاترین تراکم، قابلیت بیشتری برای حفظ و توسعه خود در آینده دارد. با استفاده از داده‌های جدول ۹، نمودار راهبردی ترسیم شد و مبدأ نمودار با توجه به میانگین مرکزیت و تراکم به‌ترتیب روی ۲۲/۶ و ۰/۳۲ تنظیم شد.

نمودار راهبردی اغلب برای توصیف روابط داخلی یک خوشه و تعاملات میان حوزه‌های مختلف به کار می‌رود. نمودار راهبردی همچنین با تمرکز بر شاخص‌های مرکزیت و تراکم، پویایی تم‌های پژوهشی را توصیف می‌کند (هو و ژانگ[1]، ۲۰۱۵؛ خاصه و دیگران، ۲۰۱۷). محور افقی در نمودار راهبردی، بر مرکزیت دلالت می‌کند و بیانگر قدرت تعامل هریک از خوشه‌ها در حوزه مورد نظر می‌باشد، هرچه مرکزیت یک خوشه بیشتر باشد، آن خوشه از جایگاهی مهم و مرکزی برخوردار است. علاوه‌براین، محور عمودی در نمودار راهبردی بیانگر تراکم است و بر رابطه درونی در یک حوزه خاص دلالت می‌کند. هرچه تراکم یک خوشه بالاتر باشد، یعنی آن خوشه قابلیت بیشتری برای حفظ و توسعه خود دارد (لیو و دیگران، ۲۰۱۲؛ سهیلی، شعبانی و خاصه، ۱۳۹۴).



شکل ۷. نمودار راهبردی ۵ خوشه اصلی

1 . Hu & Zhang

همان‌طور که شکل ۷ نشان می‌دهد، خوشه پنجم که در قسمت دوم نمودار قرار گرفته است، محوری نیست اما خوش‌توسعه بوده و آن مختصات نمودار معروف به «محلی آرام برای تفکر[1]» می‌باشد. مفاهیم خوشه دوم که به دلیل مرکزیت بالا در این حوزه، جایگاه مرکزی دارد در قسمت اول نمودار راهبردی یعنی به‌عنوان خوشه بالغ و مرکزی تعیین شده است. خوشه‌های اول، سوم و چهارم در قسمت سوم نمودار راهبردی قرار دارند و بر این دلالت می‌کنند که جزو خوشه‌های در حال ظهور هستند و به خوشه‌های بی‌نظم و ساختارنیافته معروف‌اند. این خوشه‌ها حالت حاشیه‌ای دارند و مغفول مانده‌اند. در پایان، هیچ خوشه‌ای در قسمت چهارم نمودار راهبردی قرار نگرفته است یعنی هیچ خوشه‌ای مرکزی و توسعه‌نیافته نیست.

بحث و نتیجه‌گیری

در این پژوهش تلاش شد تا با استفاده از تکنیک‌های مختلف مانند تحلیل‌های هم‌واژگانی، مصورسازی ساختار شبکه، خوشه‌بندی سلسله‌مراتبی، و نقشه دوبعدی، نمای مناسبی از ساختار فکری حاکم بر تولیدات علمی حوزه داده‌های پیوندی ارائه شود. هرچند پژوهش‌های مرتبط با مصورسازی موضوعات و نقشه هم‌رخدادی واژگان به‌خودی‌خود سیاست خاصی را ارائه نمی‌دهد، اما می‌تواند در فهم وضعیت دانش موجود و هدایت سیاست‌های علمی راهگشا باشد (صدیقی، ۱۳۹۳).

یافته‌های پژوهش نشان دادند که تولیدات علمی دارای بیشترین فراوانی در دسته‌بندی وب علوم مربوط به حوزه‌های مختلف «علوم کامپیوتر» هستند که با نتایج پژوهش نیک‌نیا و میرطاهری (۲۰۱۵) مطابقت دارند.

عبارت «داده‌های پیوندی» بیشترین فراوانی را داراست، یعنی هسته اصلی پژوهش‌های این حوزه با همین کلیدواژه، رایج، کاربردی و شناخته شده است. شایان ذکر است که کلیدواژه‌هایی نظیر «وب معنایی»، «سلامت»، «هستان‌شناسی»، «آر.دی.اف.»، «میانکنش‌پذیری»، «یکپارچگی داده»، «میراث فرهنگی»، و «فراداده» نیز دارای فراوانی بالایی هستند و سهم زیادی از مباحث این حوزه را به خود اختصاص داده‌اند و جزو زیرشاخه‌های عمده این حوزه محسوب می‌شوند.

بیشترین قرابت مفهومی در مقایسه با مطالعات پیشین با تحقیق کیا و وانگ (۲۰۱۸) می‌باشد. کلیدواژه‌های موجود در خوشه دوم پژوهش حاضر (وب معنایی، داده‌های پیوندی، دی.بی.پدیا، آنتولوژی)، و واژه یکپارچه‌سازی داده از خوشه سوم، و واژه آموزش از خوشه چهارم با کلیدواژه‌های نهایی گزارش‌شده در مطالعه کیا و وانگ (۲۰۱۸) هم‌راستاست. هرچند که مطالعه آنها تعداد کمی از منابع را بررسی کرده‌اند، اما این شباهت نشان می‌دهد که این کلیدواژه‌ها، عمده‌ترین بحث‌های مربوط به این حوزه را نشان می‌دهند. در ادامه، برخی بحث‌های اصلی با این مفاهیم به نگارش درآمده‌اند که با تمرکز بر داده‌های پیوندی چرایی کاربرد این مفاهیم در این حوزه توضیح داده می‌شوند.

موضوعاتی که نقش کمتری در شبکه دارند نباید فراموش شوند و بررسی دلایل حضور و عدم حضور آنها در شبکه حائز اهمیت است (شکفته و حریری، ۱۳۹۲). به‌عنوان مثال کلیدواژه‌های «بازیابی اطلاعات»، «جستجوی معنایی»، «چهارچوب»، «فناوری‌های معنایی»، «بازنمون دانش»، «اف.آر.بی.آر.»، «آر.دی.ای.»، «اسکاس»، «داده‌کاوی»، «پردازش زبان طبیعی»، «یادگیری ماشینی»، «متن‌کاوی»، «استخراج اطلاعات»، «تطابق نمونه»، «استدلال»، «ترازبندی آنتولوژی»، «مهندسی آنتولوژی»، در بین کلیدواژه‌های ماتریس ۲۸۴* ۲۸۴ مشاهده شده‌اند که به دلیل فراوانی کمتر در بررسی نهایی ذکر نشده‌اند. از دیدگاه متخصصین موضوعی، کلمات کلیدی مذکور، ارتباط موضوعی دقیقی با این

1 . Ivory tower

حوزه دارند که در حال حاضر و نیز برای پژوهش‌های آینده حوزه «داده‌های پیوندی» محل تمرکز می‌باشند. مثلاً «ترازبندی آنتولوژی»، «مهندسی آنتولوژی»، «تطابق نمونه»، در مباحثی مانند توسعه و نگاشت هستان‌شناسی‌ها در بهبود و گسترش داده‌های پیوندی و «داده‌کاوی»، «پردازش زبان طبیعی»، «یادگیری ماشینی»، «متن‌کاوی»، «استخراج اطلاعات» در مباحثی مانند کشف الگوها در بهبود شناسایی مفاهیم، سیستم‌های پرسش و پاسخ و الگوریتم‌های بازیابی معنایی با استفاده از اصول «داده‌های پیوندی» محل مطالعه و پژوهش محققان این حوزه هستند.

کلیدواژه‌هایی مانند «بازیابی اطلاعات»، «جستجوی معنایی»، «چهارچوب»، «فناوری‌های معنایی»، «بازنمون دانش» در مباحثی مانند تمرکز بر چهارچوب‌های معنایی در بافت‌های مختلفی مانند سلامت و میراث فرهنگی وابسته به الگوهای مفهومی و طرح‌های استاندارد فراداده‌ای مانند «اف.آر.بی.آر.»، «آر.دی.ای.»، «اسکاس» به‌منظور بهبود بازیابی معنایی و بازنمون دقیق‌تر دانش و بهبود جستجوی معنایی در وب داده به کار می‌روند. به همین دلیل هم خوشه پنجم «کاربرد داده‌های پیوندی در کتابخانه‌ها و میراث فرهنگی با بهره‌گیری از اصطلاح‌نامه‌ها، الگوهای مفهومی و طرح‌های فراداده‌ای استاندارد». نام گرفته است. درواقع پژوهشگران و محققان حوزه «وب معنایی» در تلاش هستند از طریق بهبود ابزارها و تکنیک‌ها به تحقق بازیابی معنایی کمک کنند تا وب داده را از این پراکندگی نجات دهند، دانش نهفته میان داده‌ها را کشف کنند، و دانش را قابل بازیابی کنند. همان‌گونه که این واژه‌ها در جدول ۸ (زوج‌های هم‌واژگانی برتر) نیز ذکر شده‌اند که نشان‌دهنده قرابت زیاد این مباحث با یکدیگر است.

خوشه اول از پنج خوشه اصلی تحت عنوان «کاربرد داده‌های پیوندی در حوزه سلامت» در برخی مطالعات مانند گارسیا و دیگران[1] (۲۰۱۸)، جوانویک و تراجانو[2] (۲۰۱۷)، کوزاکی و دیگران[3] (۲۰۱۷)، ژاو[4] (۲۰۱۰)، و ناواس دلگادو و دیگران[5] (۲۰۱۵)، محل تمرکز بوده است که از نظر اهمیت با خوشه اول نمودار دوبعدی نیز هم‌راستاست و نشان‌دهنده ارتباط تنگاتنگ این دو حوزه است.

خوشه دوم «مفاهیم هسته در داده‌های پیوندی» که با دو خوشه حاصل از دندروگرام و نمودار دوبعدی نیز هم‌راستاست و یکی از مهم‌ترین مباحث این حوزه تلقی می‌شود و در زوج‌های برتر هم‌واژگانی نیز بیشترین فراوانی را به خود اختصاص داده است. درواقع، «داده‌های پیوندی» اصولی برای انتشار داده‌ها در سطح وب ارائه می‌کند که معنا و ارتباط با مجموعه داده‌های خارجی را غنی می‌کند (بیزر، هیث و برنرزلی، ۲۰۰۹).

خوشه سوم «یکپارچگی و میانکنش‌پذیری در داده‌های پیوندی» نام دارد. این دو کلیدواژه نیز در زوج‌های هم‌رخداد جدول ۸ دارای فراوانی بالا می‌باشند. درواقع، داده‌های پیوندی فرمتی رایج برای همه انواع داده‌هاست که می‌تواند به‌عنوان مزیتی بزرگ برای میانکنش‌پذیری و یکپارچگی همه انواع سیستم‌ها در نظر گرفته شود (بیرن و گودارد[6]، ۲۰۱۰).

خوشه چهارم «طراحی الگوریتم برای بهبود کیفیت داده در داده‌های پیوندی» نامیده شده است. به‌طور کلی، در وب داده‌های پیوندی، کیفیت داده یک چالش و نگرانی مهم است. کارایی برنامه‌های کاربردی مانند برنامه‌ریزی، توسعه، تصمیم‌گیری، و سیاست‌گذاری به کیفیت داده بستگی دارد. تمرکز اصلی بر این است که یک هستان‌شناسی

1 . Garcia & et al
2 . Jovanovik & Trajanov
3 . Kozaki et al
4 . Zhao
5 . Navas-Delgado & et al
6 . Byrne & Goddard

می‌تواند در ارزیابی و بهبود کیفیت داده‌های پیوندی به کار رود. همچنین می‌تواند با استفاده از اگزیوم‌های (اصول متعارف) خود در کشف بی‌ثباتی داده در داده‌های پیوندی کمک نماید (دوتا[1]، ۲۰۱۴).

در تحلیل هم‌رخدادی زوج هم‌واژگان‌های برتر مانند هم‌رخدادی «داده‌های پیوندی» و «آر.دی.اف.» با فراوانی بالا در متون بررسی‌شده می‌توان گفت که ستون پایه زیرساخت داده‌های پیوندی عبارت‌های ساده‌ای هستند که «سه‌گانه‌های آر.دی.اف.» نامیده می‌شوند (پاتوئلی، پروو و ثورسن[2]، ۲۰۱۵). هم‌رخدادی «داده‌های پیوندی» و «هستان‌شناسی» با فراوانی بالا مؤید این مطلب است که این دو مکمل هم هستند و در تعامل با هم به تقویت و توسعه یکدیگر کمک می‌کنند و این رویکرد باعث قدرتمندشدن وب می‌شود. بنابراین، هستان‌شناسی‌ها در مقابل داده‌های پیوندی از نظر قابلیت معنایی مجموعه داده‌ها، پشتیبانی از استدلال و استنباط دانش جدید، بهبود کیفیت داده، و کسب اعتماد اجتماعی سهم عمده‌ای دارند. درحالی‌که، در مقابل، داده‌های پیوندی می‌توانند شیوه‌ای که هستان‌شناسی‌ها توسعه می‌یابند را تغییرات اساسی دهند و بهبود ببخشند. توسعه هستان‌شناسی‌های کاربرمحور را ارتقا دهند و مدل‌سازی هستان‌شناسی داده‌محور را حمایت کنند (دوتا، ۲۰۱۷) که به‌عنوان مؤلفه مرکزی زیرساخت مورد نیاز برای درک وب معنایی شناخته شده است (برنرزلی و فیشتی[3]، ۱۹۹۹).

حضور چشمگیر کلیدواژه «اسپارکل» در زوج‌های هم‌رخداد مؤید این نکته است که اسپارکل از فناوری‌های کلیدی وب معنایی به شمار می‌آید. چراکه حجم عمده داده‌ها در وب معنایی، در محیط پایگاه‌های داده‌ای و در بستر «آر.دی.اف.» ذخیره شده‌اند که دسترسی به آنها نیاز به ابزار خاصی دارد. اسپارکل زبان درخواست اشیای پیاده‌سازی‌شده در قالب «آر.دی.اف.» است که بازیابی و دسترسی به ارزش‌های ذخیره‌شده در عناصر «آر.دی.اف.» را در محیط پایگاه داده‌ای ممکن می‌سازد (تومارلو و دیگران[4]، ۲۰۰۸؛ لو و چو[5]،۲۰۱۴). برای بازیابی داده‌های «آر.دی.اف.» چندین استاندارد و پروتکل موجود است. استانداردی که بیشتر به این منظور مورد استفاده قرار می‌گیرد اسپارکل است (سوث‌ویک، ۲۰۱۵) که در سال ۲۰۰۸ منتشر شد (بیکر وساتون[6]، ۲۰۱۵).

هم‌رخدادی واژه «اُدبلیو.ال.[7]/ زبان هستان‌شناسی وب» در زوج‌های هم‌رخدادی بیانگر این است که این استاندارد در محیط داده‌های پیوندی و در نحو «ایکس.ام.ال.» به‌عنوان یک زبان بین‌المللی در وب معنایی برای کدگذاری و تبادل در هستی‌شناسی‌ها به کار می‌رود تا بر اساس مدل شیءگرا، دانش و روابط بین اشیای محتوایی را از طریق استفاده از هستان‌شناسی‌ها بیان کند (هیتزلر و دیگران[8]، ۲۰۱۲). درواقع، «اُدبلیو.ال.» و «آر.دی.اف.» مشابه هم هستند؛ اما «اُدبلیو.ال.» زبانی قوی‌تر با تفسیرپذیری بیشتری، واژگانی بیشتر، و نحوی قوی‌تر نسبت به «آر.دی.اف.» است (کنسرسیوم وب جهانی[9]، ۲۰۰۸).

حضور کلیدواژه‌های دیگر مانند «دی.بی.پدیا[10]» و «اسکاس» در زوج‌های هم‌رخداد را می‌توان این‌گونه تحلیل کرد

1 . Dutta
2 . Pattuelli, Provo, & Thorsen
3 . Fischetti
4 . Tummarello & et al
5 . Lou & Qiu
6 . Baker, & Sutton,
7 . OWL: Web Ontology Language
8 . Hitzler & et al
9 . World Wide Web Consortium
10 . DBpedia

که مخازن داده‌های پیوندی[1] یا دادگان‌ها[2] مانند «دی.بی.پدیا[3]» مجموعه‌ای از سه‌گانه‌های «آر.دی.اف.» هستند که به توصیف و انتشار انواع گوناگون داده می‌پردازند و دارای پیوندها و روابط بسیاری با یکدیگر هستند (طاهری، ۱۳۹۴). یکی دیگر از مخازن که بیشتر در محیط میراث فرهنگی (کتابخانه‌ها، موزه‌ها و آرشیوها) به کار می‌رود، سیستم سازمان‌دهی ساده دانش (اسکاس[4]) است (مایلز و بکهوفر[5]، ۲۰۰۹) است. واژه‌های مذکور با کلیدواژه‌های برتر پژوهش کیا و وانگ (۲۰۱۸) مطابقت دارند.

خوشه پنجم که در قسمت دوم نمودار راهبردی قرار گرفته است دارای روابط داخلی قوی و سطح مناسبی از بلوغ در این حوزه می‌باشد. سه خوشه‌ای که در قسمت سوم نمودار راهبردی قرار گرفته‌اند به دلیل نداشتن روابط داخلی و خارجی در این حوزه توسعه‌نیافته باقی مانده‌اند. خوشه دوم که خوشه «مفاهیم هسته در داده‌های پیوندی» نام دارد با چگالی و مرکزیت بالا، قوی‌ترین و بالغ‌ترین خوشه است که از جایگاه مرکزی در این حوزه برخوردار می‌باشد. استقلال و ظهور مستقل این خوشه در تکنیک‌های دیگر مانند دندروگرام و نمودار دوبعدی نیز بر قدرت و بلوغ این خوشه نیز دلالت می‌کند. این خوشه دارای جامع‌ترین حوزه‌های موضوعی در این زمینه است و بیشتر از موضوعات دیگر در این حوزه توسعه یافته و مفاهیم این خوشه در مرکز موضوعات این حوزه قرار دارند.

نتایج حاصل از این تحلیل می‌تواند به سیاست‌گذاران و سازمان‌های فعال این حوزه مانند «کنسرسیوم وب جهانی»، توسعه‌دهندگان معنایی، طراحان فناوری‌های معنایی، و تمامی متخصصان در بافت میراث فرهنگی کمک کند تا برنامه‌ریزی مناسبی به‌منظور افزایش کمی و کیفی تولیدات علمی و توسعه متوازن موضوعات این حوزه داشته باشند. همچنین، نمای ارائه‌شده از ساختار فکری دانش این حوزه می‌تواند دیدگاهی علمی از شکاف‌های موضوعی و موضوعات در حال رشد ایجاد کند تا با شناسایی موضوعات هسته و روندهای اساسی از پژوهش‌های کم‌کاربرد و تکراری جلوگیری شود.

پیشنهادهای اجرایی پژوهش

- با توجه به اینکه خوشه پنجم، خوشه خوش‌توسعه می‌باشد، محققان علم اطلاعات و دانش‌شناسی به‌عنوان فعالان بافت میراث فرهنگی می‌توانند با بهره‌مندی از توانایی‌های خود در حوزه سازمان‌دهی و بازیابی دانش با تمرکز بر الگوهای مفهومی و هستان‌شناسی‌ها، نقش مؤثر و قابل توجهی در تولیدات علمی این بخش داشته باشند؛
- با توجه به دو بافت سلامت و میراث فرهنگی در حوزه داده‌های پیوندی، این حوزه باید به‌عنوان یک حوزه دارای پتانسیل میان‌رشته‌ای در سرفصل‌های وزارت علوم و وزارت بهداشت، درمان، و آموزش پزشکی مورد توجه بیشتر و جدی‌تری قرار گیرد؛
- با توجه به فراوانی پربسامد واژه «أ.دبلیو.ال.» در جدول هم‌رخدادی، هستان‌شناسان و توسعه‌دهندگان معنایی باید این نحو را به دلیل تفسیرپذیری بیشتر، علاوه بر نحو «آر.دی.اف.»، در توسعه هستان‌شناسی‌ها بیشتر مورد توجه قرار دهند.

1 . LD repositories
2 . Datasets
3 . DBpedia
4 . Simple Knowledge Organization System (SKOS)
5 . Miles & Bechhofer

پیشنهاد برای پژوهش‌های آتی

- تحلیل و بررسی ساختار فکری دانش در موضوعات مرتبط مانند «وب معنایی»، «هستان‌شناسی»، «میانکنش‌پذیری معنایی»، و «بازنمون دانش» تا مفاهیم و خوشه‌های مشترک و نیز خلأهای پژوهشی این حوزه‌های وابسته و مرتبط به هم شناسایی شوند.

فهرست منابع

احمدی، حمید، و عصاره، فریده (۱۳۹۶). مروری بر کارکردهای تحلیل هم‌واژگانی. *مطالعات ملی کتابداری و سازمان‌دهی اطلاعات*، ۲۸ (۱)، ۱۲۵-۱۴۵.

امامی، مریم، ریاحی‌نیا، نصرت، و سهیلی، فرامرز (زودآیند). ترسیم ساختار علمی حوزه تجهیزات پزشکی و آزمایشگاهی با استفاده از تحلیل هم‌رخدادی واژگان. *پژوهش‌نامه علم‌سنجی*.

خادمی، روح‌اله، و حیدری، غلامرضا (۱۳۹۵). ترسیم ساختار موضوعی مدیریت اطلاعات با استفاده از روش همایندی واژگان طی سال‌های ۲۰۱۲-۱۹۸۶. *فصلنامه علوم و فنون مدیریت اطلاعات*،۲ (۲)، ۱۱۱-۷۱.

رستمی، مصطفی، سهیلی، فرامرز، و خاصه، علی‌اکبر (زودآیند). ساختار دانش در پروانه‌های ثبت اختراع حوزه کشف دانش: مصورسازی با استفاده از تحلیل هم‌رخدادی واژگان. *پژوهش‌نامه علم‌سنجی*.

سالمی، نجمه، و کوشا، کیوان (۱۳۹۲). مقایسه تحلیل هم‌استنادی و تحلیل هم‌واژگانی در ترسیم نقشه کتاب‌شناختی (مطالعه موردی: دانشگاه تهران). *پردازش و مدیریت اطلاعات*، ۲۹ (۱)، ۲۵۳-۲۶۶.

سهرابی، طیبه، و غفاری، سعید (زودآیند). شناسایی موضوعات پرکاربرد تولیدات علمی حوزه ارتباطات علمی با استفاده از روش تحلیل هم‌رخدادی واژگان. *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 (۱۳۹۴). ساختار فکری دانش در حوزهٔ رفتار اطلاعاتی: مطالعهٔ هم‌واژگانی. *تعامل انسان و اطلاعات*. ۲ (۴)، ۳۶-۲۱.

شکفته، مریم، و حریری، نجلا. (۱۳۹۲). ترسیم و تحلیل نقشه علمی پزشکی ایران با استفاده از روش هم استنادی موضوعی و معیارهای تحلیل شبکه اجتماعی. *فصلنامه مدیریت سلامت*. ۱۶ (۵۱)، ۵۹-۴۳.

صدیقی، مهری (۱۳۹۳). بررسی کاربرد روش تحلیل هم‌رخدادی واژگان در ترسیم ساختار حوزه‌های علمی (مطالعه موردی: حوزه اطلاع‌سنجی). *پردازش و مدیریت اطلاعات*، ۳۰ (۲)، ۳۹۶-۳۷۳.

طاهری، مهدی (۱۳۹۴). *ذخیره و بازیابی اطلاعات و دانش با تأکید بر رویکردهای نوین*. تهران: کتابدار: کنسرسیوم محتوای ملی، ۱۳۹۴.

عصاره، فریده، سهیلی، فرامرز، و منصوری، علی (۱۳۹۴). *علم‌سنجی و دیداری‌سازی اطلاعات*. اصفهان: انتشارات دانشگاه اصفهان.

قاضی‌زاده، حمید، سهیلی، فرامرز، و خاصه، علی‌اکبر (۱۳۹۷).*ترسیم ساختار دانش در پژوهش‌های علوم قرآن و حدیث ایران با استفاده از تحلیل هم‌واژگانی*، ۴ (۸)، ۱۲۲-۱۰۱.

مکی‌زاده، فاطمه؛ حاضری، افسانه، حسینی‌نسب، سید حسین، و سهیلی، فرامرز (۱۳۹۵). تحلیل موضوعی و ترسیم نقشه علمی مقالات مرتبط با حوزه درمان افسردگی در پاب‌مد. *مدیریت سلامت*، (۱۹)۶۵، ۶۳-۵۱.

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

Allemang, D., & Hendler, J. A. (2011). Semantic web for the working ontologist: Effective modeling in RDFS and OWL. Waltham, MA: Morgan Kaufmann/Elsevier.

Baker, T., & Sutton, S. A. (2015). Linked data and the charm of weak semantics: Introduction: The strengths of weak semantics. *Bulletin of the Association for Information Science and Technology*, 41(4), 10-12.

Berners-Lee, T., & Fischetti, M. (1999). Weaving the Web: *The original design and ultimate destiny of the World Wide Web by its inventor*. San Francisco, CA: Harper San Francisco.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5 (3), 1-22.

Byrne, G., & Goddard, L. (2010). The strongest link: Libraries and linked data. D-Lib magazine, 16(11/12). Retrieved from: http://www.dlib.org/dlib/november10/byrne/11byrne.print.HTML

Callon, M., Courtial, J. P., & Turner, W. A. (1986). Future developments. In M. Callon, J. Law, & A. Rip(Eds).Mapping the dynamics of science and technology: Sociology of science in the real world, 211–217. London: Macmillan.

Cho, J. (2014). Intellectual structure of the institutional repository field: A co-word analysis. *Journal of Information Science*, 40(3), 386-397.

Clarivate Analytics (2019). Web of Science Core Collection: Search Tips. Retrieved November 2019. https://clarivate.libguides.com/woscc/searchtips

Cobo, M. J., & Lopez-Herrera, A. G., & Herrera-Viedma, E. (2011). An approach fordetecting, quantifying, and visualizing the evolution of a research field: A Practical application to the fuzzy sets theory field. Journal of Informetrics, 5 (1), 146-166.

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.

Dutta, B. (2014). Symbiosis between Ontology and Linked Data. *Librarian*, 21(2), 15-24.

Dutta, B. (2017). Examining the interrelatedness between ontologies and Linked Data. *Library Hi Tech*, 35(2), 312-331.

Garcia, A., Lopez, F., Garcia, L., Giraldo, O., Bucheli, V., & Dumontier, M. (2018). Biotea: semantics for Pubmed Central. *PeerJ*, 6, e4201.

Guerrini, M., & Possemato, T. (2013). Linked data: a new alphabet for the semantic web. *JLIS.it.* 4(1), 67-90.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1), 1-136.

Hitzler, P., Krotzsch, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (eds.) (2012). *OWL 2 web ontology language primer* (2nd ed.). Retrieved from http://www.w3.org/TR/owl2-primer/

Hu, J., & Zhang, Y. (2015). Research patterns and trends of Recommendation System in China using co-word analysis. *Information processing & management*, 51(4), 329-339.

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.

Jovanovik, M., & Trajanov, D. (2017). Consolidating drug data on a global scale using Linked Data. *Journal of biomedical semantics*, 8(1), 3.

Khasseh, A. A., Soheili, F., Moghaddam, H. S., & Chelak, A. M. (2017). Intellectual structure of knowledge in iMetrics: A co-word analysis. *Information Processing & Management*, 53(3), 705-720.

King, J. (1987), A review of bibliometric and other science indicators and their role in research evaluation, *Journal of Information Science*, Vol. 13 No. 5, pp.261–276.

Kozaki, K., Yamagata, Y., Mizoguchi, R., Imai, T., & Ohe, K. (2017). Disease Compass–a navigation system for disease knowledge based on ontology and linked data techniques. *Journal of biomedical semantics*, 8(1), 22.

Kyaw, Wang (2018). Mapping the Intellectual Structure of the Linked Data Field: A Co-Word Analysis and Social Network Analysis. *International Journal of AdvancedResearch in Science, Engineering and Technology*, 5(8), 6632-6647.

Lee, W. H. (2008). How to identify emerging research fields using scientometrics: An example in the field of information security. *Scientometrics*, 76(3), 503–525.

Lee, P.C., & Su, H.N. (2010). Investigating the structure of regional innovation system research Information Science, 19(1), 71-85.

Liu, Y., Goncalves, J., Ferreira, D., Xiao, B., Hosio, S., & Kostakos, V. (2014, April). CHI 1994-2013: mapping two decades of intellectual progress through co-word analysis. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3553-3562). ACM.

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.

Lou, W. & Qiu, J. (2014), Semantic information retrieval research based on co-occurrence analysis, *Online Information Review*, 38(1),4-23.

Miles, A., & Bechhofer, S. (2009). SKOS Simple Knowledge Organization System reference. W3C. Retrieved from http://www.w3.org/TR/skos-reference/

Moulaison, H. L., & Million, A. J. (2014). The disruptive qualities of linked data in the library environment: analysis and recommendations. *Cataloging & classification quarterly*, 52(4), 367-387.

Navas-Delgado, I., García-Godoy, M. J., López-Camacho, E., Rybinski, M., Reyes-Palomares, A., Medina, M. Á., & Aldana-Montes, J. F. (2015). kpath: integration of metabolic pathway linked data. *Database*, 2015.

Neff, M. W., & Corley, E. A. (2009). 35 years and 160,000 articles: a bibliometric exploration of the evolution of ecology. *Scientometrics*, 80 (3), 657–682.

Niknia, Massoomeh, & Mirtaheri, Seyedeh Leili (2015). Mapping a decade of linked data progress through co-word analysis. *Webology*, 12(2), Article 141.

Noyons, E., & van Raan, A. (1998). Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. *Journal of the American Society for Information Science*, 49(1), 68-81.

Pattuelli, M. C., Provo, A., & Thorsen, H. (2015). Ontology building for linked open data: a pragmatic perspective. *Journal of Library Metadata*, 15(3-4), 265-294.

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(1), 929-955.

Sedighi, M., & Jalalimanesh, A. (2014). Mapping research trends in the field of knowledge management. *Malaysian Journal of Library and Information Science*, 19(1), 71-85.

Southwick, S. B. (2015). A guide for transforming digital collections metadata into linked data using open source technologies. *Journal of Library Metadata*, 15(1), 1-35.

Tummarello, G., Morbidoni, C., Puliti, P. & Piazza, F. (2008), A proposal for textual encoding based on semantic web tools, *Online Information Review*, 32(4), 467-477.

van Eck, N. J., & Waltman, L. (2018). VOSviewer manual. *Erişim adresi: http://www. vosviewer. com/download/f-z2w2. pdf*.

Wang, L. Y., Zhang, Z. Q., & Wei, J. Z. (2011). A study on foreign research subjects of library and information science based on the co-word analysis during the last ten years. *Journal of intelligence (in China)*, 30(3), 50–58.

Whittaker, J. (1989). Creativity and conformity in science: Titles, keywords and co-word analysis. Social Studies of Science, 19 (3), 473-496.

World Wide Web Consortium (2008). XML RDF. Retrieved from: https://www.w3schools.com/xml/xml_rdf.asp

Yan, B. N., Lee, T. S., & Lee, T. P. (2015). Mapping the intellectual structure of the Internet of Things (IoT) field (2000–2014): a co-word analysis. *Scientometrics*, 105(2), 1285-1300.

Zengenene, D. (2013). Global interoperability and linked data in libraries. *New Library World*, 114(1/2), 84-87.

Zhao, J. (2010). Publishing Chinese medicine knowledge as Linked Data on the Web. *Chinese medicine*, 5(1), 27-49..