Scientometrics Research Journal

Scientific Bi-Quarterly of Shahed University
Vol. 8, No. 1, Spring & Summer 2022 (Serial 15)

Research Article

DOI: 10.22070/RSCI.2020.5871.1440

Small-Data and Its Application among Various Scientific Areas: A Scientometric Study

Razieh Farshid [1]

Yousef Abedi [2]

Somayeh Jafari [3*]

1. PhD student in Information and Knowledge Science, Kharazmi University. Email: Razieh.farshid@gmail.com

2. PhD student in Information and Knowledge Management, University of Tehran. Email: y.abedi@gmail.com

3. PhD student in Information and Knowledge Management, University of Tehran. (Corresponding Author)

Email:jafari.somayyeh@gmail.com

Date of Reception: 19/09/2020

Date of Acceptation: 09/03/2021

Abstract

Purpose: The purpose of this study is to identify the characteristics of scientific products in the field of small-data indexed in the Web of Science database and to explain its application based on identifying the words of scientific products related to this subject separately by scientific fields.

Methodology: This research is a descriptive study based on the scientometric approach and content analysis method, which has been done by using the common techniques of co-word analysis and social network analysis. Data analysis was performed by HistCite, Bibexecl, Gephi, and SPSS software; and the data mapping is done by VOSviewer.

Findings: Over the past decades, the rate of publications in the field of small data has had an increasing trend with an average annual growth rate of 15.59%. The main language of these works is in English. Although the National Cheng Kung University (Taiwan) ranked the first of organizations in this field, the United States, China and Germany recognized the top countries in this field, overall. More than 90% of these products are in the fields of Computer Science (8 clusters), Engineering (6 clusters), Mathematics (7 clusters), Telecommunications (5 clusters), and Physics (3 clusters). The greatest degree of centrality belongs to Machine Learning, the Internet of Things, and Universal existence; the most closeness centrality belongs to Adaptation, Bipartite Graph, and Machine Learning; and the most betweenness centrality belongs to Machine Learning, Long-Term Evolution Technology, and Global Existence.

Conclusion: The pattern of dissemination of scientific products in the field of small data indicates a continuous growth situation. Theoretical discussions of microdata have further evolved in mathematics and physics, and its applications in computer science and other fields are expanding.

Keywords: Small data, Scientometrics, Big data, Co-word, Intellectual Structure of Knowledge, Social Network Analysis.

پژوهش‌نامه علم‌سنجی
دوفصلنامه علمی دانشگاه شاهد/ دوره ۸/ شماره ۱/ بهار و تابستان ۱۴۰۱ (پیاپی ۱۵)

مقاله پژوهشی

DOI: 10.22070/RSCI.2020.5871.1440

ریزداده‌ها و کاربرد آن در حوزه‌های علمی مختلف: مطالعهٔ علم‌سنجی

راضیه فرشید [۱]

یوسف عابدی [۲]

سمیه جعفری [۳*]

۱. دانشجوی دکتری بازیابی اطلاعات و دانش، دانشگاه خوارزمی.
Email: Razieh.farshid@gmail.com

۲. دانشجوی دکتری مدیریت اطلاعات و دانش، دانشگاه تهران.
Email: y.abedi@gmail.com

۳. دانشجوی دکتری مدیریت اطلاعات و دانش، دانشگاه تهران. (نویسنده مسئول)
Email: jafari.somayyeh@gmail.com

صفحه ۲۸۱-۲۵۵
دریافت: ۱۳۹۹/۰۶/۲۹
پذیرش: ۱۳۹۹/۱۲/۱۹

CC BY

چکیده

هدف: هدف از پژوهش، شناسایی ویژگی‌های تولیدات علمی حوزهٔ ریزداده نمایه‌شده در پایگاه وب آو ساینس و تبیین کاربرد آن بر مبنای شناسایی واژگان تولیدات علمی مرتبط با این موضوع به تفکیک حوزه‌های علمی است.

روش‌شناسی: پژوهش حاضر از نوع توصیفی با رویکرد علم‌سنجی و روش تحلیل محتوا و بهره‌مندی از فنون تحلیل هم‌واژگانی و تحلیل شبکه اجتماعی انجام شده است. تحلیل داده‌ها با نرم‌افزارهای هیست‌سایت، بایب‌اکسل، گفی و اس.پی.اس.اس؛ و ترسیم داده‌ها با نرم‌افزار وی او اس ویوئر انجام شده است.

یافته‌ها: طی دهه‌های گذشته میزان انتشار تولیدات علمی حوزهٔ ریزداده روندی افزایشی با میانگین نرخ رشد سالانهٔ ۱۵.۵۹ درصد داشته است. هرچند دانشگاه چنگ‌کونگ (تایوان) از نظر میزان تولید علم و دریافت استناد جایگاه اول را کسب کرد، درمجموع ایالات متحده، چین و آلمان به‌عنوان کشورهای برتر شناخته شدند. بیش از ۹۰ درصد این تولیدات مربوط به حوزه‌های علوم رایانه، مهندسی، ریاضیات، مخابرات و فیزیک است. خوشه‌بندی هم‌واژگانی در این حوزه‌ها به‌ترتیب منجر به تشکیل ۸، ۶، ۷، ۵ و ۳ خوشه شد. بیشترین مرکزیت درجه مربوط به یادگیری ماشینی، اینترنت اشیاء و وجود جهانی است؛ کلیدواژه‌های انطباق، نمودار دوقطبی و یادگیری ماشینی بیشترین مرکزیت نزدیکی و یادگیری ماشینی، فناوری تکامل بلندمدت و وجود جهانی دارای بیشترین مرکزیت بینابینی در حوزه‌های علمی فوق‌اند.

نتیجه‌گیری: الگوی انتشار تولیدات علمی حوزه ریزداده‌ها نشان‌دهندهٔ وضعیت بالندگی مستمر است. مباحث نظری ریزداده‌ها بیشتر در علوم ریاضیات و فیزیک تکامل یافته و کاربردهای آن در علوم رایانه و حوزه‌های دیگر در حال گسترش است.

واژگان کلیدی: ریزداده، علم‌سنجی، کلان‌داده، هم‌واژگانی، ساختار فکری دانش، تحلیل شبکه اجتماعی.

مقدمه و بیان مسئله

در دنیای امروز، هر لحظه حجم وسیعی از داده‌ها تولید و پردازش می‌شوند. مراکز اداری و دولتی، بخش پزشکی و بهداشت، واحدهای آموزشی و پژوهشی، رسانه‌های اجتماعی و کسب‌وکارهای گوناگون در سراسر جهان مجموعه داده‌های عظیمی را در قالب‌های مختلف تولید می‌کنند که در محمل‌های متفاوت جریان می‌یابند. با چنین حجمی از داده‌ها و رشد سریع آن و همچنین قابلیت دسترسی آسان، می‌توانیم ادعا کنیم که در «عصر داده‌ها» زندگی می‌کنیم. این داده‌ها به یکی از دارایی‌های کلیدی و راهبردی سازمان‌ها تبدیل شده که برای استفاده‌پذیری، مستلزم تحلیل روشمند و مؤثر است.

فرایند کشف دانش از داده‌ها[1] را به استخراج طلا از معادن تشبیه می‌کنند. این فرایند با پیش‌پردازش داده‌ها[2] (پالایش، یکپارچه‌سازی، انتخاب و تبدیل داده‌ها) آغاز می‌شود و درنهایت به کشف الگوهای جالب (واجد اعتبار لازم، بدیع و حاوی سودمندی بالقوه) می‌انجامد تا دانش به‌دست‌آمده در تصمیم‌گیری‌ها استفاده شود و مزیت رقابتی را به ارمغان آورد. داده‌کاوی[3] فارغ از فناوری، ابزار، رویکرد و روش مورد استفاده‌اش، ارزش افزودهٔ چشمگیری برای صاحب آن و ذی‌نفعانش ایجاد می‌کند. داده‌کاوی کاربردهای موفقیت‌آمیز بسیاری در حوزه‌هایی نظیر هوش تجاری، بیوانفورماتیک، مالی، کتابخانه‌های دیجیتال، و دولت الکترونیک به دست آورده است (هان، کمبر و پی[4]، ۱۳۹۴).

با توسعهٔ ابزارهای قدرتمند برای گردآوری و تحلیل داده‌ها، فرصت استفاده از کلان‌داده‌ها[5] فراهم شد. کلان‌داده‌ها عبارت‌اند از داده‌هایی که پردازش آنها خارج از حد توان سیستم‌های رایج است، به‌گونه‌ای که ذخیره، پردازش و تحلیل آنها از طریق فناوری‌های قدیمی پایگاه داده امکان‌پذیر نیست (مینلی[6] و همکاران، ۲۰۱۳). تعاریف زیادی از کلان‌داده وجود دارد که تمام آنها در سه ویژگی اصلی حجم، سرعت و تنوع بسیار زیاد اشتراک دارند (گندمی[7]، ۲۰۱۵؛ سگیروگلو[8]، ۲۰۱۳). پیشرفت‌های مستمر روش‌شناختی و فناورانه در تحلیل کلان‌داده‌ها با تمرکز بر شناسایی الگوهای متغیر استمرار دارد (چن[9]، ۲۰۱۶). در الگوهای کلان‌داده‌ای خوشه‌ها و روابط معنادار به‌صورت پویا از مجموعه دادگان به‌روزشونده عظیم شکل می‌گیرند (هکلر[10]، ۲۰۱۹).

کاستی‌هایی در رویکرد کلان‌داده رخ نمود که زمینهٔ توجه مجدد به ریزداده‌ها[11] را رقم زد. کلان‌داده‌ها با داده‌های زیاد و تحلیل‌شده توسط ماشین سروکار دارند که مستلزم هزینه بالایی است، در مقابل ریزداده حاصل گردآوری نمونه‌های هدفمندی است که در آن تمرکز بر علیت عمیق و تحلیل ملموس واقعیت‌نگر است (لیندستورم[12]، ۲۰۱۶). هدف از رویکرد ریزداده دستیابی به توصیف، پیش‌بینی و درنهایت کنترل واحدهای مرتبط با داده‌هاست، به‌نحوی‌که در بافت خود فهم شوند. در این رویکرد داده‌ها توسط واحدهای انفرادی یا محلی و برای اهداف خاص خود استفاده می‌شوند. این بدان معناست که نیازهای ابرازنشدهٔ افرادی که داده‌ها در مورد ایشان است، بهتر شناسایی می‌شود (هکلر،

1 . Knowledge Discovery from Data (KDD)
2 . Data Preprocessing
3 . Data Mining
4 . Han, Kamber and Pei
5 . Big data
6 . Minelli
7 . Gandomi
8 . Sagiroglu
9 . Chen
10 . Hekler
11 . Small data
12 . Lindstrom

۲۰۱۹). استفاده از ریزداده‌ها در علوم مختلفی مانند پزشکی، رایانه و علوم پایه رو به فزونی است. تعداد فزاینـده‌ای از فناوری‌های متناسب با ریزداده نیز در حال تکامل است. با تحلیل هزینه‌ـفایده مشخص می‌شود که در برخی از شرایط، ریزداده‌ها نسبت به کلان‌داده‌ها ارجحیت دارند. یک نمونهٔ کوچک با کیفیت بالا می‌تواند یافته‌های مفیدتری نسـبت بـه یک نمونهٔ بزرگ با کیفیت پایین داشته باشد. به لحاظ علمی نیز در شرایطی، استنباط آماری روی ریزداده‌ها نسبت بـه کلان‌داده‌ها بهتر عمل می‌کند (فاراوی، ۲۰۱۸).

به‌طور خلاصه، کلان‌داده مجموعه داده‌های گسترده از نظر حجم، تنوع، سرعت تولید، تغییرپذیری و غیره است که نیازمند ابزار و فناوری خاص و معماری مقیاس‌پذیر برای ذخیره و تحلیـل اسـت و توسـط تـیم‌هـای گسـترده انجـام می‌شود؛ درحالی‌که ریزداده مجموعه داده‌هایی که یک تحلیلگر می‌تواند با ابزارهای ساده به‌طور کامل آنها را بفهمـد و تحلیل و ارزیابی کند. کلان‌داده اگرچه به نتایجی فراتر از ذهن فردی منجر می‌شـود، امـا مسـتلزم کاربسـت فنـاوری، تجهیزات و نیروی انسانی گران‌قیمت‌تری است و یافته‌های حاصل نیز از بافت آن جدا هستند. هرچند ریـزداده‌هـا بـه اندازهٔ کلان‌داده جذاب نیستند، اما بسیار کاربردی‌اند و می‌توانند عمق بیشتری داشته باشند. جدول زیر آنها را بر اساس مؤلفه‌های مختلف مقایسه می‌کند (لیندستورم، ۲۰۱۶؛ کیچین و لوریو[1]، ۲۰۱۵).

جدول ۱. مقایسهٔ ریزداده و کلان‌داده

مؤلفه	کلان‌داده	ریزداده
اندازهٔ نمونه	بیشترین حد ممکن	اندازهٔ درست (معمولاً کوچک)
هزینهٔ گردآوری، ذخیره و پردازش داده	زیاد و مستمر (پیوسته)	کم و متناوب (دوره‌ای)
کیفیت داده‌های نهایی	احتمال فروکاهی معنایی	احتمال کشف نیازهای ابرازنشده
قابلیت تحلیل عمیق	متوسط و محدود به سطح	زیاد و قابل پیگیری نامحدود
فناوری، ابزارها و روش‌ها	زیاد، متنوع و به‌روز	متوسط، محدود و در حال تکامل

در کلان‌داده تا حد ممکن از سرشماری استفاده می‌شود، یعنی بیشترین داده‌های ممکن ذخیره و پردازش می‌شوند؛ اما با بزرگ‌ترشدن فزاینده و روزافزون حجم مجموعه داده‌های کلان، مفهوم نمونهٔ به‌اندازه[2] ظهور کرده است. بسیاری از تحلیل‌گران در تلاش‌اند تا به جای اینکه در اقیانوس پرهزینهٔ داده‌ها گم شوند، بر اساس نظریهٔ تیغ اوکام[3] عمل و با پیشگیری از افزایش بدون ضرورت داده‌ها، حد بهینه‌ای از آنها را مبنای کار قرار دهند. این امر با حفظ نسبی کیفیـت، صرفه‌جویی چشمگیری در هزینه‌ها ایجاد می‌کند. این دو رویکرد مبتنی بر فلسفه‌های متفاوت، در ظاهر رقیب به نظـر می‌رسند؛ اما تجارب عملی محدود در استفادهٔ توأم این دو رویکرد توانسته به سطح بالاتری از موفقیـت منجـر شـود. علی‌رغم اینکه استفادهٔ مکمل از این دو رویکرد در ابتدای راه است، ریزداده‌ها توانسته‌اند کلان‌داده‌هـا را سـریع‌تـر، بـا اطمینان بیشتر و با هزینهٔ کمتر به نتیجه‌گیری مفید برسانند (سشی[4]، ۲۰۱۸). یک الگوی ریزداده به‌خودی‌خود ارزشمند است و راه را برای پیشبرد گام‌هـای بعـد روشـن مـی‌سـازد. در مقایسـه بـا رویکـرد کـلان‌داده، رویکـرد ریـزداده از استراتژی‌های منحصربه‌فرد و مکملی برای مدیریت پدیده‌های پیچیـده و پویـا بهـره مـی‌جویـد کـه مـی‌توانـد بـرای کلان‌داده‌ها نیز ارزشمند باشد (هکلر، ۲۰۱۹)، به‌طوری که مطلوبیت هم‌گرایی این دو رویکرد با هدف تحلیـل ایـده‌آل

1 . Kitchin and Lauriault
2 . Right-Sized Data
3 . Occam's Razor
4 . Secchi

داده‌ها تأیید شده است (سرکوئیتلی[1]، ۲۰۱۷).

پژوهشگران به‌منظور دیدن فراسوهای دانش و کسب درک و نمایی کلی از چارچوب‌های یک حوزهٔ علمی، آثار دانشمندان پیش از خود را مرور می‌کنند؛ به‌عبارت‌دیگر، با اتکا به گذشتهٔ علم، آیندهٔ علمی حوزهٔ تخصصی خود را پیش می‌برند. در این راستا طی چند دههٔ گذشته، ترسیم ساختار فکری حوزه‌های علمی به‌عنوان یکی از مهم‌ترین وجوه مطالعات سنجش علم، اهمیت بسیاری در حوزه‌های مختلف علمی کسب کرده است (سهیلی، شعبانی و خاصه، ۱۳۹۵). حوزهٔ علم‌سنجی برای سهولت‌بخشیدن دسترسی به اطلاعات و یاری جستجوگران دانش از طریق پردازش، استخراج و مرتب‌سازی اطلاعات به ترسیم ساختار فکری دانش می‌پردازد و امکان تحلیل، مسیریابی و نمایش دانش را فراهم می‌آورد (نوروزی چاکلی، ۱۳۹۱). ترسیم نقشهٔ علم، ترسیم نقشهٔ دانش و مصورسازی دانش، نام‌های گوناگونی از ترسیم ساختار فکری دانش هستند که بر پایهٔ برون‌دادهای کمی مبتنی است و در سال‌های اخیر کاربرد وسیعی یافته‌اند. ارائهٔ تصویر کلان از وضعیت پژوهش‌های صورت‌گرفته، نحوهٔ ارتباط حوزه‌های مختلف و آگاهی از چگونگی رشد و توسعهٔ این حوزه‌ها در طی زمان از اهداف ترسیم ساختار فکری دانش است.

امروزه مطالعات سنجش علم با هدف بررسی ساختار دانش در حوزه‌های علمی از روش‌ها و فنون مختلفی مانند هم‌استنادی، هم‌واژگانی، و هم‌نویسندگی بهره می‌برد. یکی از روش‌های پرکاربرد برای ترسیم و تحلیل ساختار دانش در حوزه‌های مختلف، هم‌رخدادی واژگان است. این روش که در دههٔ ۱۹۸۰ میلادی مطرح شد بر این فرض بنا شده که استفاده از واژگان کلیدی مشترک در بخش‌های عنوان، چکیده، کلیدواژه‌ها و متن تولیدات علمی نشان‌دهنده نزدیکی مفاهیم آن متن‌ها به یکدیگر است و به‌واسطهٔ آن می‌توان ساختار، مفاهیم و مؤلفه‌های یک حوزه علمی را تعیین کرد. در این شیوه بر مبنای شاخص‌هایی چون نزدیکی و شباهت به ترسیم ساختار حوزه‌ها و زمینه‌های علمی (وایتاکر[2]، ۱۹۸۹)، شناسایی الگوهای پنهان و برجسته، تعیین روابط درونی و بیرونی مفاهیم (عصاره و همکاران، ۱۳۹۴)، آشکارسازی 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 می‌پردازند (احمدی و عصاره، ۱۳۹۶). به بیان دیگر هم‌رخدادی کلیدواژه‌ها از الگوهای هم‌رخدادی در یک مجموعه از متون بهره می‌گیرد تا ارتباط میان اندیشه‌ها در حوزه موضوعات متون را شناسایی کند. ویژگی اصلی تحلیل هم‌واژگانی، دیداری‌سازی ساختار منطقی یک حوزه خاص از طریق ترسیم نقشه مفهومی است؛ بنابراین تحلیل هم‌واژگانی به‌عنوان یکی از روش‌های رایج در مطالعات سنجش علم، خوشه‌های موضوعی ذیل یک حوزه پژوهشی را آشکار ساخته، روابط مفهومی و معنایی آن را در نظر گرفته و ساختار فکری دانش را در حوزه مورد بررسی ترسیم می‌کند.

با شکل‌گیری دو رویکرد کلان‌داده و ریزداده در جریان داده‌کاوی که از مبانی و ابزارهای کاملاً متفاوتی بهره می‌جویند، مطالعهٔ ساختار فکری تولیدات هر یک از این رویکردها می‌تواند به نمایش کاستی‌ها و ظرفیت‌های هر یک بپردازد و از این طریق درک ما را نسبت به هر دو کامل‌تر سازد تا امکان بهره‌مندی به‌جا و متناسب از آنها فراهم شود. در این راستا مسئلهٔ پژوهش حاضر دستیابی به نیم‌رخی از تولیدات علمی این حوزه و همچنین تبیین کاربرد ریزداده در علوم مختلف بر اساس شناسایی موضوعات پژوهشی مطالعات مربوط در وب آو ساینس[3] است. شناسایی ساختار دانش در حوزهٔ ریزداده به تفکیک علوم مختلف، کاربران داده‌کاوی را نسبت به اهمیت ابعاد فرعی ریزداده آگاه ساخته،

1 . Cerquitelli
2 . Whittaker
3 . Web of Science (WOS)

ایشان را به کاربست روش‌شناسی‌های ترکیبی سوق می‌دهد؛ و پژوهشگران را به مطالعات هدفمند در راستای مباحث حوزهٔ علمی خود رهنمون می‌سازد تا در مطالعات آتی با آگاهی بیشتری در این حوزه گام بردارند.

سؤال‌های پژوهش

پژوهش حاضر بر اساس هدف خود یعنی شناسایی ویژگی‌های تولیدات علمی حوزه ریزداده و تبیین کاربرد آن بر اساس واژگان تولیدات علمی نمایه‌شده در وب آو ساینس به تفکیک حوزه‌های علمی به موارد زیر پاسخ خواهد داد:

۱. ویژگی‌های تولیدات علمی حوزه ریزداده از جنبه‌های مختلف مانند تعداد، قالب اثر، زبان، کشورها، مؤسسات و پژوهشگران مشارکت‌کننده، حوزه‌های پژوهشی، نشریات و کلیدواژه‌ها چگونه است؟

۲. تحلیل هم‌واژگانی منجر به شکل‌گیری چه خوشه‌ها و موضوعاتی بر اساس سنجه‌های مرکزیت در حوزه‌های علمی مشارکت‌کننده در موضوع ریزداده شده است؟ میزان شباهت موضوعی حوزه‌های علمی مورد نظر چقدر است؟

۳. آیا تفاوت معناداری میان حوزه‌های علمی مربوط به تولیدات علمی ریزداده به لحاظ تعداد استنادات و تعداد رکوردهای تولیدشده وجود دارد؟

چارچوب نظری

انواع مختلفی از نقشه‌های علمی را می‌توان ترسیم کرد و ساختار یک حوزه علمی را با آن نشان داد. برخی از نقشه‌ها بر پایه روابط هم‌استنادی و برخی بر اساس هم‌رخدادی واژه‌ها شکل می‌گیرند. با مقایسه نقشه‌های حاصل در دوره‌های زمانی مختلف، پویایی علم ردیابی می‌شود (سالمی و کوشا، ۱۳۹۱). فرایند ترسیم نقشه‌های موضوعی علوم مبتنی بر نظر بورنر و همکاران[1] (۲۰۰۳)، شامل شش مرحله است که در ادامه اشاره می‌شود.

نخستین گام در هر فرایند نگاشت یا ترسیم نقشه، استخراج اطلاعات مناسب است. در این مرحله راهبردهای مختلف جستجو کاربرد دارند؛ اما مهم‌ترین نکته اینکه کیفیت نقشه‌هایی که ترسیم می‌شوند، به‌صورت مستقیم وابسته به اطلاعاتی است که مبنای کار قرار می‌گیرند. تعداد مدارکی که می‌توان برای ترسیم نقشه‌ها به کار برد می‌توانند از چند صد مدرک تا چندین هزار مدرک باشند. مرحله دوم انتخاب واحدهای تحلیل؛ بستگی به سؤالی دارد که درصدد پاسخ‌گویی به آن هستیم. رایج‌ترین واحدها برای ترسیم نقشه‌ها، نوشته‌ها هستند که عبارت‌اند از مجله‌ها، مدارک، نویسندگان، واژگان و اصطلاحات توصیف‌گر؛ هرکدام از این واحدها جنبه‌ای متفاوت از حوزه مورد مطالعه را نشان می‌دهد و انواع مختلف تحلیل را فراهم می‌آورد؛ مرحله سوم واژه‌های فنی بسیاری به‌عنوان شاخص‌های شناسایی شباهت بین مقاله‌ها به کار برده می‌شوند، این واژه‌ها از پیشوندهای Co- و Inter- ساخته شده‌اند، واژه‌هایی از این قبیل: Intercitation, Interdocument, Co-classification, Co-citation, Co-word.

شباهت‌های بین مدارک (واحدها) معمولاً با روش‌های مختلفی محاسبه می‌شوند که رایج‌ترین آنها ارتباطات استنادی یا ارجاعی[2]، شباهت‌های هم‌رخدادی[3] و مدل بردار فضایی[4] است. روش‌های دسته‌بندی متنوعی با توجه به کاربرد هر یک در ترسیم نقشه‌ها وجود دارند که مهم‌ترین آنها تجزیه مقدار ویژه، بردار ویژه[5]، تحلیل عاملی[6]،

1 . Borner et all
2 . Citation Linkages
3 . Co-occurrence Similarities
4 . Vector Space Model
5 . Eigen value . Eigenvector Decomposition
6 . Factor Analysis

مقیاس‌بندی چندبعدی[1]، تحلیل معنایی نهفته[2]، تحلیل خوشه‌ای[3] و مثلث‌بندی[4] هستند؛ در آخرین مرحله نوبت به استفاده از فنون نمایش اطلاعات در قالب بصری می‌رسد. نمایش به تمام روش‌های مصورسازی اطلاعات گفته می‌شود که در راستای جستجو و پیمایش اثربخشی فضاهای گسترده اطلاعاتی هستند. از جملهٔ این روش‌ها می‌توان به انواع روش‌های پالایش‌کردن اطلاعات[5]، انواع روش‌های بزرگ‌نمایی[6] و تغییر زاویه دید[7] اشاره کرد (بورنر، چن[8] و چن[8] و بویاک، ۲۰۰۳؛ نقل از رمضانی، علی‌پور حافظی و مؤمنی، ۱۳۹۳).

تحلیل هم‌رخدادی شیوه‌ای از تحلیل محتواست که از طریق هم‌رخدادی واژه‌ها با مفاهیم در متون و مدارک عمل می‌کند و از طریق آن می‌توان مفاهیم اصلی یک زمینه یا حوزه علمی را شناخته و به‌واسطه این شناخت، الگوها و رویدادهای مفهومی حوزه، ترسیم ساختار علمی، شبکه مفهومی، روابط سلسله‌مراتبی مفاهیم و مقولات مفهومی را کشف و در جهت مدیریت حوزه به کار برد. به سخن دیگر، تحلیل هم‌رخدادی واژگان ابزاری در جهت کشف الگوهای پنهان و رویدادهای نوظهور مفهومی است. در تحلیل هم‌رخدادی واژگان محدودیت‌هایی نیز وجود دارد که چنانچه مورد توجه قرار نگیرد تحلیل‌های مورد نظر دچار مشکل می‌گردد. از جمله این محدودیت‌ها می‌توان به کیفیت واژگان منتخب (عدم توجه به محل استخراج واژگان در مدرک، مسائل زبانی واژگان، ترکیب واژگان، ارتباط معنایی واژگان و تأثیر نمایه‌سازی) در حوزه مورد نظر و به‌کارگیری این روش در حوزه‌هایی که از لحاظ واژگان و مفاهیم مستعد نیستند، اشاره کرد (احمدی و عصاره، ۱۳۹۶).

ترسیم نقشه‌های علمی تنها هدف نیست، بلکه پس از ترسیم نقشه‌های مورد نظر باید تحلیل و تفسیر آنها را انجام داد که مرحله‌ای بسیار مهم است. به‌منظور تفسیر نقشه‌های علمی می‌توان از فنون تحلیل شبکه‌های اجتماعی[9] استفاده کرد؛ زیرا نقشه‌های علمی دارای ساختاری مشابه شبکه‌های اجتماعی هستند. تحلیل شبکه‌های اجتماعی به‌عنوان شاخه‌ای از جامعه‌شناسی که به مطالعه شبکه‌ها می‌پردازد شاخص‌های مختلفی را برای تعیین گره‌ها یا بازیگران مهم و مرکزی در شبکه پیشنهاد می‌کند (شکفته و حریری، ۱۳۹۲). درواقع صرف به نمایش‌گذاشتن یک نقشه جذاب علمی، کمکی به پیشبرد حوزه نمی‌کند و تحلیل درست نقشه است که مسیر درست را پیش روی پژوهشگران قرار می‌دهد و راه را جهت پژوهش‌های علمی آینده هموار می‌کند.

پیشینه پژوهش

بررسی پیشینه‌ها در داخل و خارج نشان می‌دهد که هم‌رخدادی واژگان یکی از روش‌هایی است که در بیشتر مطالعات علم‌سنجی برای ترسیم ساختار موضوعی و فکری در حوزه‌های علمی مختلف استفاده شده است؛ ازآنجاکه به‌طور خاص در حوزه ریزداده مطالعاتی با رویکرد علم‌سنجی بازیابی نشد در ادامه به مرور برخی پژوهش‌های دارای رویکرد علم‌سنجی با تأکید بر پژوهش‌های هم‌رخدادی در سال‌های اخیر به‌ویژه در حوزه‌های مرتبط با داده (نظیر کلان‌داده، علم داده، داده‌های پیوندی و غیره) می‌پردازیم.

1 . Multidimensional Scaling
2 . Latent Semantic Analysis
3 . Cluster Analysis
4 . Triangulation
5 . Filtering
6 . Zooming
7 . Distortion
8 . Chen
9 . Social Network Analysis

هرچند این پژوهش‌ها به‌طور مستقیم در حوزه سازمان‌دهی اطلاعات و دانش نبوده‌اند. در ادامه به مرور برخی پژوهش‌های علم‌سنجی با رویکرد هم‌واژگانی یا همکاری علمی در داخل و خارج کشور در سال‌های اخیر می‌پردازیم.

پیشینه پژوهش در داخل

از جمله پژوهش‌های انجام‌شده در حوزه پزشکی برای نمونه می‌توان به مکی‌زاده و همکاران (۱۳۹۵) در حوزه درمان افسردگی، به فیض‌آبادی و وزیری (۱۳۹۶) در حوزه مطالعات دمانس، به رئیس‌زاده و کرمعلی (۱۳۹۷) در حوزه ترومای نظامی و به پژوهش جعفری و همکاران (۱۳۹۹) در حوزه کووید ۱۹ اشاره کرد. در حوزه علم اطلاعات و دانش‌شناسی و موضوعات مرتبط با آن نیز پژوهش‌های فراوانی از جمله مصطفوی و همکاران (۱۳۹۶)، سهیلی و همکاران (۱۳۹۷)، رمضانی و همکاران (۱۳۹۷)، دانیالی و نقشینه (۱۳۹۷)، خزانه‌ها و همکاران (۱۳۹۸) و دانش و نعمت‌الهی (۱۳۹۹) انجام شده است که در ادامه برخی از پژوهش‌های نسبتاً مرتبط‌تر بیشتر توضیح داده شده‌اند.

رمضانی، علی‌پور حافظی و مؤمنی (۱۳۹۷) به مصورسازی و تحلیل شبکه هم‌تألیفی نهادهای پژوهشی حوزه کتابخانه‌های دیجیتالی در ایران پرداخته‌اند. یافته‌ها حاکی از رشد صعودی تولیدات علمی، بالابودن میزان همکاری‌های درون‌سازمانی و پایین‌بودن سهم مشارکت‌های بین‌المللی است. نهادهای پژوهشی استان تهران بیشترین سهم از تولیدات علمی و هم‌تألیفی‌ها را داشته‌اند. همچنین از موانع توسعه علمی یادگیری بیشتر و نوآوری در حوزه کتابخانه‌های دیجیتالی ایران، ارتباطات ضعیف میان گروه‌هاست. دانیالی و نقشینه (۱۳۹۷) به مطالعه روند پژوهش و ترسیم نقشه دانش قلمروهای پژوهشی فعال حوزه بازیابی تصویر در وبگاه علم پرداختند؛ یافته‌ها حاکی از آن بود که حوزه بازیابی تصویر دارای روابط میان‌رشته‌ای نسبتاً گسترده‌ای است، به‌طوری که ۶۸ قلمرو پژوهشی در نوشتن مقالات این حوزه نقش داشته‌اند؛ علوم رایانه حوزه پیشگام بوده و کتابداری جایگاه ششم را کسب نموده است. بالاترین مرکزیت عدد ۰.۰۸ در سال ۲۰۰۴ در رشته سنجش از دور، بالاترین شکوفایی عدد ۱۵.۳۸ در سال ۲۰۰۲ و بالاترین سیگما عدد ۱.۲۶ در سال ۲۰۰۱ در رشته علوم تصویربرداری و تکنولوژی عکس می‌باشد.

دانش و نعمت‌الهی (۱۳۹۹) در پژوهشی به خوشه‌بندی مفاهیم و رویدادهای نوپدید سازمان‌دهی دانش پرداختند. زوج‌های هم‌واژگانی «سیستم اطلاعات جغرافیایی_سیستم اطلاعات جغرافیایی» و «کتابخانه دانشگاهی_سواد اطلاعاتی» بیشترین هم‌رخدادی را به خود اختصاص داده‌اند. خوشه‌بندی موضوعات سازمان‌دهی دانش به روش تحلیل خوشه‌ای حاکی از آن است که در بازه زمانی مورد بررسی درمجموع ۲۷ خوشه وجود دارد که ۱۰ خوشه مربوط به بازه زمانی ۱۹۹۹-۱۹۷۵ و ۱۷ خوشه به بازه زمانی ۲۰۱۸-۲۰۰۰ اختصاص دارد. همچنین نتایج نشان داد که در دوره‌های زمانی، خوشه‌ها از نظر هم‌پوشانی موضوعی تا حد زیادی مشابه‌اند. به‌بیان‌دیگر، شش خوشه و تعدادی از کلیدواژه‌ها از لحاظ محتوا شباهت دارند. بااین‌وجود در بازه زمانی دوم، خوشه‌ها به دلیل تعداد و وسعت گسترده‌تر هستند.

حسینی و همکاران (۱۳۹۹) در پژوهشی با عنوان کتاب‌سنجی و نگاشت هم‌رخدادی واژگان در حوزه داده‌های پیوندی به ترسیم و تحلیل شبکه هم‌رخدادی واژگان، و خوشه‌های موضوعی در حوزه داده‌های پیوندی در بازه زمانی ۲۰۱۸-۱۹۸۶ پرداختند. نتایج به‌دست‌آمده حاکی از آن بود که 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 وب علوم در این حوزه را دارند.

پیشینه پژوهش در خارج

لوپزروبلز[1] و همکاران (۲۰۱۸) در پژوهشی به بررسی ساختار فکری و مضامین حوزه کلان‌داده در وب آو ساینس ساینس در سال‌های ۲۰۱۲ تا ۲۰۱۷ و همچنین تحلیل نویسندگان و کشورهای برتر با استفاده از نرم‌افزار سای‌مت[2] پرداخته است. حوزه‌های علوم رایانه، علم اطلاعات و مهندسی بیشترین مشارکت را در پژوهش‌های حوزه کلان‌داده داشته‌اند. همچنین کشور آمریکا، انگلستان و چین سرآمد بوده‌اند. جین و لی (۲۰۱۸) در پژوهشی به بررسی هات‌اسپات‌ها و ترندهای نوظهور داده‌های بزرگ چندرسانه‌ای‌ها با استفاده از نرم‌افزار سایت‌اسپیس در بازه زمانی ۲۰۱۷–۲۰۰۸ پرداخته‌اند. نتایج به‌دست‌آمده حاکی از آن است که تحقیقات بزرگ داده‌های چندرسانه‌ای به سمت میان‌رشته‌ای در حال پیشرفت است و ریاضیات و سیستم‌ها یک رشته داغ و پزشکی بالینی یک رشته در حال ظهور است. لی[3] و همکاران (۲۰۱۸) در پژوهشی به بررسی داده‌های حوزه پژوهش سلامت الکترونیک در بازه زمانی ۱۹۹۲–۲۰۱۷ با استفاده از تحلیل زمان، مکان و تحلیل هم‌واژگانی و با استفاده از نرم‌افزارهای سایت‌اسپیس و نت‌درا پرداختند. امر استریپ و دمرت (۲۰۱۹) به بررسی ریشه‌های علمی علوم داده با استفاده از استنادهای موجود در گوگل‌اسکالر پرداختند. بعد از تجزیه و تحلیل داده‌های به‌دست‌آمده یک مدل رگرسیون آماری معرفی شده و درنهایت تعریف عینی برای پاسخ به سؤال «علم داده چیست؟» ارائه می‌دهد. پورخانی و همکاران (۲۰۱۹) در پژوهشی به بررسی مقالات منتشر در پایگاه وب آو ساینس در ارتباط با تأثیر رسانه‌های اجتماعی در رشد و عملکرد مشاغل در بازه زمانی ۲۰۰۵ تا پایان ژانویه ۲۰۱۹ پرداخته است. در این پژوهش کشورها و مجلات پیشرو در این حوزه معرفی شده است. ایالات متحده، انگلستان و چین به‌ترتیب بیشترین میزان تولید را در میان کشورها داشته‌اند. رسانه‌های اجتماعی، تأثیر و تبلیغات دهان‌به‌دهان[4] سه کلیدواژه پرتکرار در پژوهش آنها هستند و بیشترین میزان همکاری میان ایالات متحده و چین، کانادا و ایالات متحده و انگلیس و ایالات متحده بوده است. اوزادوگلو و دمر و اوزادوگلو (۲۰۱۹) در پژوهشی به بررسی علم‌سنجی وضعیت هنر در تحقیقات بلاک‌چین با استفاده از مقالات مرتبط در وب آو ساینس و اسکوپوس در بازه زمانی ۲۰۱۸–۲۰۱۳ پرداختند. هدف از انجام این پژوهش روشن‌کردن دیدگاه جامع درباره ادبیات بلاک‌چین و پشتیبانی از پژوهشگران و دست‌اندرکاران این زمینه است. تایبی و نظری و غائلی (۲۰۱۹) در پژوهشی با عنوان رسانه‌های اجتماعی و تجارت الکترونیکی: تجزیه و تحلیل علم‌سنجی، به بررسی مقالات منتشرشده و بررسی همکاری کشورهای مختلف در این حوزه از سال ۲۰۰۹ پرداخته‌اند. ایالات متحده، آلمان و چین بیشترین میزان تولیدات و همکاری‌ها را به خود اختصاص داده‌اند؛ نقشه مفهومی پژوهش آنها شامل ۴ خوشه است که مهم‌ترین کلیدواژه‌ها شامل رسانه اجتماعی، مدیریت اطلاعات، داده‌کاوی و غیره است. جلوه‌گران اصفهانی و همکاران (۲۰۱۹) در پژوهشی به تجزیه و تحلیل داده‌های بزرگ و رسانه‌های اجتماعی در پایگاه اسکوپوس و همچنین به تحلیل موضوعی و بیشترین استنادهای دریافت‌شده توسط کشورها در بازه زمانی ۲۰۱۹–۲۰۱۲ پرداختند. داده‌های به‌دست‌آمده حاکی از آن است که کلان‌داده‌ها، رسانه‌های اجتماعی و شبکه‌های اجتماعی (آنلاین) پرتکرارترین کلیدواژه‌ها در پژوهش بوده‌اند. این حوزه مطالعاتی بیشتر در بین محققان ایالات متحده آمریکا، چین، هند، انگلیس و استرالیا محبوب بوده است. سلیمانی روزبهانی و همکاران (۲۰۱۹) در پژوهشی به بررسی داده‌های حوزه سیستم‌های

1 . López-Robles
2 . SciMAT (Science Mapping Analysis software Tool)
3 . Lee
4 . Word-of-mouth

کلان‌داده‌های مراقبت‌های بهداشتی به بررسی داده‌های دوره زمانی ۲۰۱۸-۲۰۰۸ پرداخته است و در یک پژوهش فراتحلیل با استفاده از تکنیک داده‌کاوی به بررسی مقالات منتشرشده در این حوزه پرداخته است؛ نقشه‌های وی او اس ویوئر منجر به شکل‌گیری ۸ خوشه را نشان می‌دهند؛ کلیدواژه‌های خوشه‌ها شامل بهداشت عمومی، انفورماتیک سلامت، کلان‌داده‌ها، مراقبت‌های بهداشتی تحقیقات، علوم داده، ارتباطات، رمزگذاری سلامت الکترونیکی و موارد دیگر است. موضوع مراقبت‌های بهداشتی با ۷۴ مقاله قابل توجه‌ترین موضوع این حوزه و پس از آن سلامت الکترونیکی با ۵۰ مقاله بیشترین تعداد تولیدات را به خود اختصاص داده‌اند.

جمع‌بندی از مرور پیشینه

مرور پیشینه‌ها نشان می‌دهد استفاده از رویکرد علم‌سنجی و فن هم‌رخدادی واژگان در حوزه‌های مشابه داخلی و خارجی رایج است و در سال‌های اخیر نیز پرطرفدار بوده و دستاوردهای متعددی به همراه داشته، اما تاکنون پژوهش مستقلی درخصوص استفاده از فن هم‌رخدادی واژگان در تولیدات حوزه‌های علمی مختلف در حوزه ریزداده انجام نشده است. در این راستا این پژوهش درصدد است با استفاده از فن فوق، کاربرد ریزداده در حوزه‌های علمی مختلف را تبیین نموده و اساساً به آزمون فرضیهٔ وجود تفاوت معنادار میان حوزه‌های مختلف علمی به لحاظ تعداد تولیدات و استنادات دریافتی در موضوع ریزداده‌ها می‌پردازد.

روش‌شناسی پژوهش

پژوهش حاضر از نوع توصیفی است که با رویکرد علم‌سنجی با استفاده از روش تحلیل محتوا و بهره‌مندی از فنون تحلیل هم‌واژگانی و تحلیل شبکه اجتماعی انجام شده است. فن تحلیل هم‌واژگانی که در این پژوهش مورد استفاده قرار گرفته درواقع یکی از روش‌های تحلیل محتواست. تحلیل محتوا یکی از روش‌های پژوهش است که به‌منظور توصیف منظم و عینی محتوای به‌دست‌آمده از ارتباطات به کار می‌رود (ضیغمی، باقری، حق‌دوست و یادآور، ۱۳۸۷). تحلیل محتوا را از دو دید کمی و کیفی می‌توان تعریف کرد؛ برلسون معتقد است تحلیل محتوای کمی برای تشریح عینی، منظم و کمی محتوای آشکار پیام‌های ارتباطی به کار می‌رود؛ این در حالی است که مسیر حرکت در تحلیل محتوای کیفی عمدتاً از متن به سمت بیرون‌کشیدن مقولات و ایجاد مدل‌ها و نقشه‌های مفهومی است (حسن‌زاده و جعفری باقی‌آبادی، ۱۳۹۶). در این پژوهش منظور از تحلیل محتوا تلفیقی از هر دو روش کمی و کیفی است.

تحلیل هم‌رخدادی واژگان، که امروزه پرکاربردترین روش‌ها برای ترسیم نقشه‌های مفهومی است روشی مناسب برای کشف ارتباطات حوزه‌های پژوهشی علم است و پیوندهای مهمی را نشان می‌دهد که ممکن است کشف آنها به روش‌های دیگر مشکل باشد. در بخش تحلیل شبکه از شاخص‌های مرکزیت که نشان‌دهنده جایگاه یک گره (موجودیت) نسبت به گره‌های دیگر در نقشه‌های علمی است استفاده شده است. جامعه آماری پژوهش حاضر را تمام تولیدات علمی مرتبط با ریزداده در وب آو ساینس تشکیل می‌دهند. در همین راستا ابتدا در فیلد موضوع[1] (عنوان، چکیده، کلمات کلیدی و متن)، عبارت «Small Data» در بخش مجموعهٔ هستهٔ وب آو ساینس[2] مورد جستجو قرار گرفت که ۳۶۶۹ مدرک بازیابی شدند. سپس در مرحله بعد بر اساس نتایج به‌دست‌آمده و با توجه به اینکه در بیش از نود درصد تولیدات علمی ریزداده پنج حوزه علمی علوم رایانه، مهندسی، ریاضیات، مخابرات و فیزیک مشارکت

1 . Topic
2 . Web of Science Core Collection

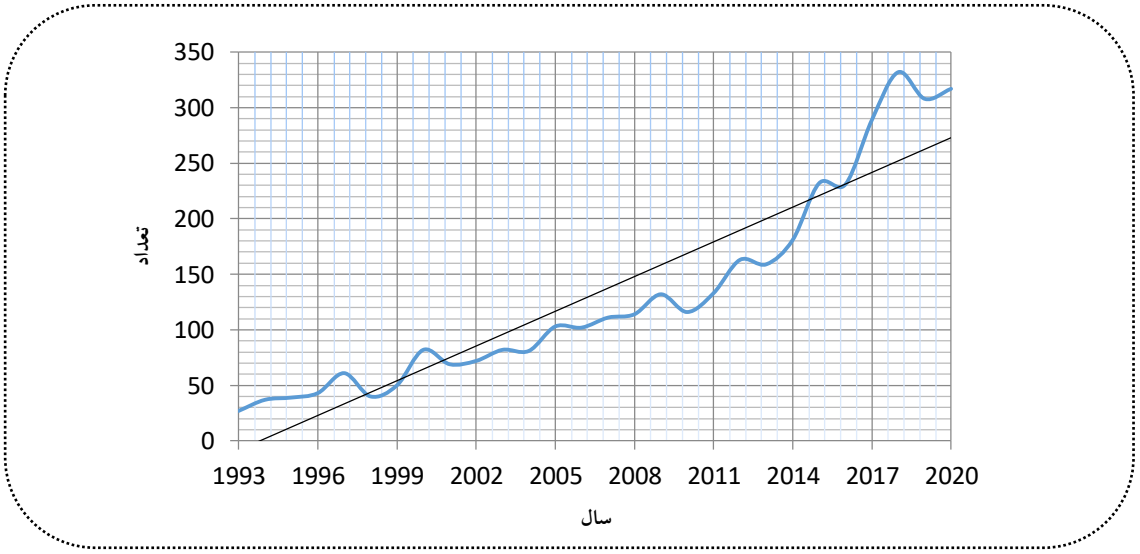
داشتند اقدام به بازیابی تولیدات علمی ریزداده به تفکیک حوزه‌های علوم رایانه با ۱۲۰۹ رکورد، مهندسی با ۹۷۴ رکورد، ریاضیات با ۸۲۵ رکورد، مخابرات با ۲۴۱ رکورد و فیزیک با ۲۲۹ رکورد با فرمت متن ساده[1] گردید؛ بنابراین از سایر حوزه‌های علمی که زمینهٔ مطالعهٔ جدی ریزداده نبوده‌اند صرف نظر شد و تمرکز مطالعه بر ۵ حوزه علمی فوق قرار گرفت. لازم به ذکر است ازآنجایی‌که کلریویت آنالتیکز[2] سقف تعداد رکوردهای دانلودی را ۵۰۰ رکورد تعیین کرده است؛ بنابراین در حوزه‌های پژوهشی که تعداد رکوردهای آنها بیش از ۵۰۰ رکورد بود، استخراج داده‌ها در دو یا سه مرحله بسته به تعداد رکوردهای موجود صورت گرفت.

پس از بازیابی رکوردهای مرتبط و یکپارچه‌سازی داده‌ها، بر اساس اهداف و پرسش‌های پژوهش اقدام به تجزیه‌وتحلیل داده‌ها با استفاده از نرم‌افزارهای هیست‌سایت، اکسل، بیب‌اکسل، گفی و اس.پی.اس.اس شد. برای ترسیم ساختار فکری و شبکه واژگان از نرم‌افزار وی.او.اس‌ویوئر (نسخه ۱۰، ۶، ۱) استفاده شده است. لازم به ذکر است از طریق ساخت اصطلاح‌نامه اقدام به کنترل و یکدست‌سازی کلیدواژه‌ها گردید و کلیدواژه‌های مشابه، یکسان، متشابه و حالت‌های جمع و مفرد ادغام و کلیدواژه‌های غیرتخصصی حذف شدند.

یافته‌های پژوهش

پاسخ به سؤال اول پژوهش. ویژگی‌های آثار حوزه ریزداده از جنبه‌های مختلف مانند تعداد، قالب اثر، زبان، کشورها، مؤسسات و پژوهشگران مشارکت‌کننده، حوزه‌های پژوهشی، نشریات، و کلیدواژه‌ها چگونه است؟

با جستجو در وب آو ساینس به‌واسطه نرم‌افزار هیست‌سایت مشخص شد تعداد ۳۶۶۹ مدرک مرتبط طی سال‌های ۱۹۹۳ تا ۲۰۲۰ میلادی توسط ۹۷۲۳ نویسنده از ۱۳۰ حوزه پژوهشی در این پایگاه نمایه شده‌اند. تصویر زیر توزیع انتشار این آثار را به تفکیک سال نشان می‌دهد.

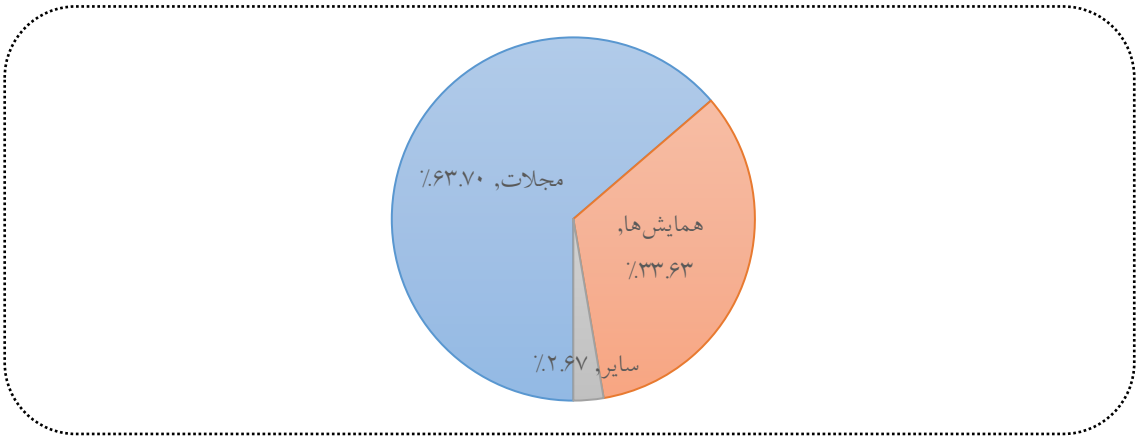


تصویر ۱. روند انتشار تولیدات علمی حوزه ریزداده به تفکیک سال

1 . Plain Text
2 . Clarivate Analytics

* تعداد آثار منتشرشده در سال ۲۰۲۰ م. طی ۸ ماه نخست به دست آمده و بر اساس الگوی زمانی آن برای کل سال برآورد شده است.

روند صعودی در تصویر ۱، استقبال از این رویکرد را نمایش می‌دهد. همچنین سال ۲۰۱۸ با ۳۳۲ اثر (۹ درصد) و سال ۱۹۹۳ با ۲۷ اثر (۱ درصد) به‌ترتیب در میان سال‌های مختلف بیشترین و کمترین تعداد تولیدات علمی را به خود اختصاص داده‌اند؛ میانگین نرخ رشد سالانه[1] انتشار این آثار ۱۵.۵۹ درصد بوده و دامنهٔ نرخ رشد سالانه تولیدات علمی منتشرشده در این حوزه از آغاز تاکنون بین ۸.۵۲ درصد (حداقل) تا ۵۷.۸۱ درصد (حداکثر) نوسان داشته است. همچنین آثار یادشده در قالب‌های متنوعی منتشر شده‌اند (تصویر ۲).



تصویر ۲. قالب‌های اصلی تولیدات علمی حوزه ریزداده

بر اساس تصویر فوق، بیشترین محمل انتشار مربوط به مجلات (مقالات تألیفی، مروری، سخن سردبیر و مقالات زودآیند) به میزان ۶۳.۷۰ درصد بوده است. از سوی دیگر ۲.۶۷ درصد آثار در دسته سایر موارد (فصلی از کتاب، نقد و بررسی، اخبار، اصلاحیه و غیره) ثبت شده‌اند. همچنین تحلیل داده‌ها نشان داد که زبان غالب تولید علم در این حوزه انگلیسی است که ۹۸.۶۰ درصد از آثار را پوشش می‌دهد. علاوه‌بر زبان انگلیسی، آثار این حوزه به ۱۲ زبان دیگر نیز منتشر شده‌اند که مهم‌ترین آنها فرانسوی (۰.۴ درصد)، چینی (۰.۳ درصد)و آلمانی (۰.۲ درصد) است. در ادامه اطلاعات مربوط به سایر ویژگی‌های مختلف پنج رتبه برتر آثار این حوزه در جدول ۲ آمده است. به‌طور متوسط هر اثر مرتبط، ۱۷.۷۶ استناد دریافت کرده است. همچنین شاخص اچ[2] آثار این حوزه در وب آو ساینس ۱۰۱ می‌باشد.

بر اساس جدول ۲ لی دی سی و چنگ سی جی هر دو از دانشگاه ملی چنگ‌کونگ کشور تایوان به‌ترتیب بیشترین تولیدات علمی و استنادهای این حوزه را به خود اختصاص داده‌اند. همچنین دانشگاه چنگ‌کونگ (تایوان) از نظر میزان تولید علم و دریافت استناد در این حوزه دارای جایگاه اول است. همچنین حوزه علوم رایانه با ۱۲۰۹ اثر در میان ۱۳۰ حوزه علمی مشارکت‌کننده، بیشترین سهم را در تولید آثار این حوزه داشته است.

نکته قابل توجه جایگاه ایران در میان کشورهای مشارکت‌کننده است که با ۳۱ تولید علمی رتبه ۲۷ را از آن خود کرده است.

1 . Annual Average Growth Rate
2 . H index

جدول ۲. پنج رتبه برتر تولیدات علمی حوزه ریزداده از نظر ویژگی‌های مختلف آثار علمی

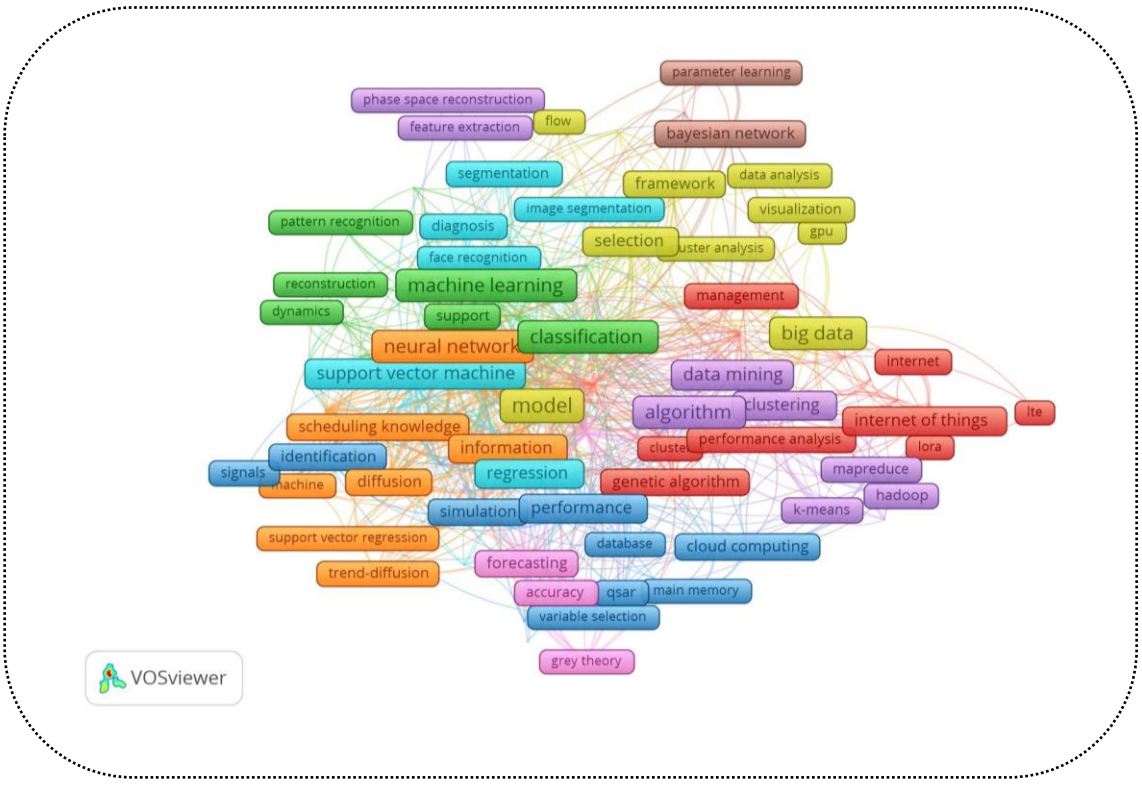
ویژگی آثار علمی	رتبه اول (تعداد رکورد، درصد)	رتبه دوم (تعداد رکورد، درصد)	رتبه سوم (تعداد رکورد، درصد)	رتبه چهارم (تعداد رکورد، درصد)	رتبه پنجم (تعداد رکورد، درصد)
کشور	ایالات متحده (۹۶۳، ۲۶. ۲)	چین (۶۰۵، ۱۶. ۵)	آلمان (۳۰۸، ۸. ۴)	انگلیس (۲۶۸، ۷. ۳)	ژاپن (۲۲۰، ۶)
نویسنده (میزان تولید)	لی دی سی[1] (۶۲، ۱. ۷)	چنگ سی جی[2] (۲۲، ۰. ۶)	چن سی سی[3] (۱۹، ۰. ۵)	ریسیگ ام[4] (۱۸، ۰. ۵)	دی ابیکو ام[5]، اوزاوا تی[6] (۱۷، ۰. ۵)
نویسنده (میزان استناد)	لی دی سی (۲۷۹، ۷. ۴۹)	چنگ سی جی (۸۵، ۲. ۲۸)	تسای تی آی[7] (۸۳، ۲. ۲۳)	چن سی سی (۷۵، ۲. ۰۱)	وو سی اس[8] (۷۲، ۱. ۹۳)
مؤسسه (میزان تولید)	دانشگاه چنگ‌کونگ[9] (۵۸، ۱. ۶)	دانشگاه علوم چین[10] (۳۹، ۱. ۱)	دانشگاه برکلی کالیفرنیا[11] (۲۹، ۰. ۸)	دانشگاه واسدا[12] (۲۷، ۰. ۷)	دانشگاه اوساکا[13] (۲۴، ۰. ۷)
مؤسسه (میزان استناد)	دانشگاه ملی چنگ‌کونگ (۲۴۸، ۱۱)	کالج مدیریت دیوان[14] (۶۴، ۲. ۷)	دانشگاه واسدا (۵۹، ۲. ۵)	دانشگاه سن دیگو کالفرنیا[15] و دانشگاه بوستون[16] (۴۳، ۱. ۸)	دانشگاه ژجیانگ[17] (۳۹، ۱. ۷)
نشریه	نشریه معادلات دیفرانسیل[18] (۴۶، ۱. ۳)	روش‌ها و کاربردهای تحلیل نظری غیرخطی[19] (۳۴، ۰. ۹)	دسترسی IEEE (۲۶، ۰. ۷)	نشریه تحلیل و کاربردهای ریاضی[20] (۲۳، ۰. ۶)	نشریه نظام‌های خبره با کاربردها[21] (۲۲، ۰. ۶)
حوزه پژوهشی مشارکت‌کننده	علوم رایانه[22] (۱۲۰۹، ۳۲. ۹۱۶)	مهندسی[23] (۹۷۴، ۲۶. ۵۱۸)	ریاضیات[24] (۸۲۵، ۲۲. ۴۶۱)	مخابرات[25] (۲۴۱، ۶. ۵۶۱)	فیزیک[26] (۲۲۹، ۶. ۲۳۵)
حوزه پژوهشی استنادکننده	علوم رایانه (۱۲۵۵۰، ۲۱. ۳۷۸)	مهندسی (۱۱۷۹۹، ۲۰. ۰۹۹)	ریاضیات (۷۷۵۹، ۱۳. ۲۱۷)	علوم اکولوژی محیط زیست[27] (۳۵۴۸، ۶. ۰۴۴)	زیست‌شناسی مولکولی بیوشیمی[28] (۳۲۵۲، ۵. ۵۴)
کلیدواژه	ریزداده ۱۷۲	کلان‌داده ۸۱	یادگیری ماشینی[29] ۷۴	وجود جهانی[30] ۵۵	یادگیری عمیق[31] ۴۳

1 . Li DC
2 . Chang CJ
3 . Chen CC
4 . Reissig M
5 . D'Abbicco M
6 . Ozawa T
7 . Tsai TI
8 . Wu CS
9 . Natl Cheng Kung Univ.
10 . Chinese Acad Sci
11 . Univ. Calif Berkeley
12 . Waseda Univ.
13 . Osaka Univ.
14 . Diwan Coll Management
15 . Univ. Calif San Diego
16 . Boston Univ.
17. Zhejiang Univ.
18. Journal of Differential Equations
19. Nonlinear Analysis-theory Methods & Applications
20. Journal of Mathematical Analysis And Applications
21. Expert Systems With Applications
22. Computer Science
23. Engineering
24. Mathematics
25. Telecommunications
26. Physics
27. Environmental Sciences Ecology
28. Biochemistry Molecular Biology
29. Machine Learning
30. Global Existence
31. Deep Learning

پاسخ به سؤال دوم پژوهش. تحلیل هم‌واژگانی منجر به شکل‌گیری چه خوشه‌ها و موضوعاتی بر اساس سنجه‌های مرکزیت در حوزه‌های علمی مشارکت‌کننده در موضوع ریزداده شده است؟ میزان شباهت موضوعی حوزه‌های علمی مورد نظر چقدر است؟

در این مرحله به‌منظور ترسیم و تحلیل ساختار فکری تولیدات علمی حوزه ریزداده در پنج حوزه علمی، تمامی تولیدات علمی استخراج‌شده از وب آو ساینس به تفکیک علوم مختلف وارد نرم‌افزار وی.او.اس ویوئر شد. در ادامه ساختار فکری هر حوزه و کلیدواژه‌های برتر آن بر اساس سنجه‌های مرکزیت آمده است. پس از آن آثار مشترک حوزه‌های علمی و میزان شباهت موضوعی آنها در جدول ۸ گزارش شده است.

الف. حوزه علوم رایانه: به دنبال تحلیل هم‌واژگانی تولیدات علمی این حوزه، ۸ خوشه از واژگان و مفاهیم شناسایی شد. پراستنادترین اثر این حوزه که متعلق به حوزه مهندسی نیز می‌باشد با عنوان « A comparison of methods for multiclass support vector machines» از نویسندگان « Hsu, CW; Lin, CJ» در مجله « IEEE TRANSACTIONS ON NEURAL NETWORKS» در سال ۲۰۰۲ به چاپ رسیده است. تصویر ۳ نقشه مفاهیم آثار ریزداده را در حوزه علوم رایانه نشان می‌دهد. لازم به ذکر است بزرگی مستطیل‌ها نشان‌دهنده کاربرد بیشتر آن مفاهیم در توصیف آثار و رنگ آنها نشان‌دهنده خوشه مفاهیم است. ضمن اینکه در این نقشه دوری و نزدیکی کلیدواژه‌ها نشان می‌دهد که مفاهیم چه مقدار به همدیگر مرتبط هستند.



تصویر ۳. نقشه هم‌رخدادی واژگان تولیدات علمی ریزداده در حوزه علوم رایانه

بر اساس ارتباطات و شبکه اجتماعی به‌دست‌آمده میان کلیدواژه‌ها از طریق نرم‌افزار وی.او.اس ویوئر، در مرحله بعدی و با استفاده از نرم‌افزار گفی به تحلیل کلیدواژه‌های آثار ریزداده در حوزه علوم رایانه بر اساس سنجه‌های

مرکزیت پرداخته شد (جدول ۳). یکی از سنجه‌های مفید جهت تحلیل شبکه‌های اجتماعی، سنجه‌های مرکزیت[1] فریمن شامل مرکزیت‌های درجه، نزدیکی و بینابینی است. مرکزیت، انواع و تعداد روابطی که عضوی از شبکه با سایر اعضای آن شبکه برقرار کرده است را نشان می‌دهد (کیولار[2] و همکاران، ۲۰۱۶).

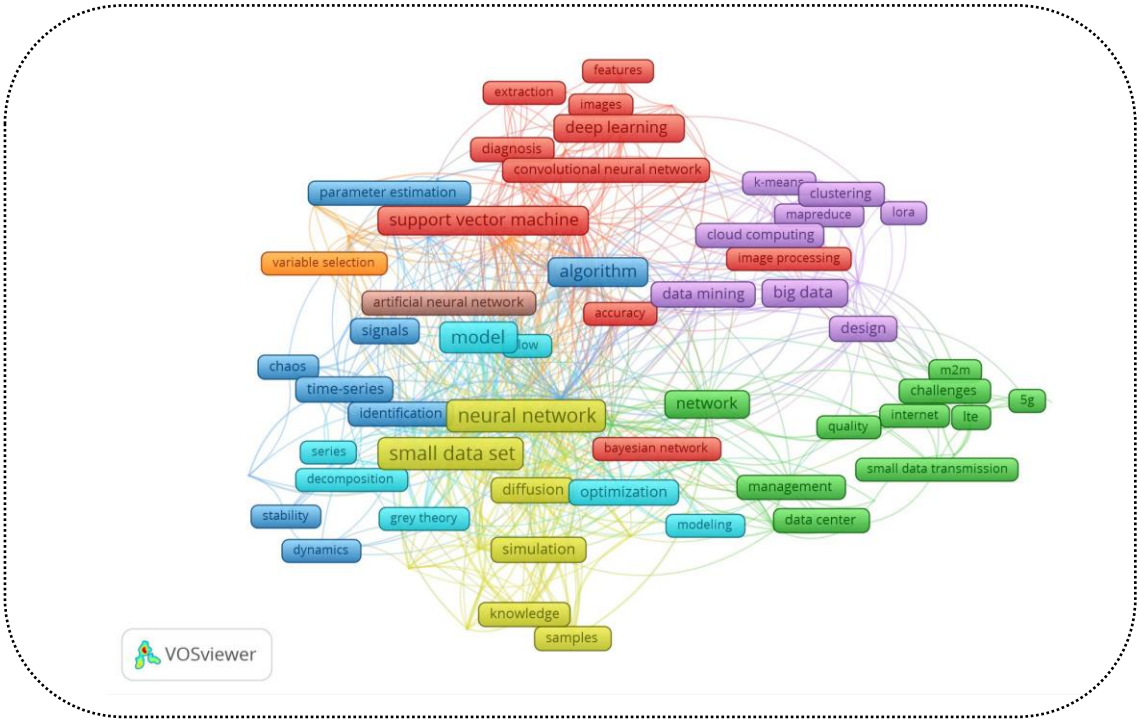
جدول ۳. پنج کلیدواژه برتر تولیدات علمی ریزداده در حوزه علوم رایانه بر اساس سنجه‌های مرکزیت

کلیدواژه	مرکزیت درجه	کلیدواژه	مرکزیت نزدیکی	کلیدواژه	مرکزیت بینابینی
یادگیری ماشینی[3]	۶۸	انطباق[4]	۱	یادگیری ماشینی	۵۵۹۹. ۱۸
کلان‌داده	۶۰	تشخیص گفتار[5]	۱	کلان‌داده	۴۴۱۱. ۷۷
داده‌کاوی[6]	۵۶	یادگیری ماشینی	۰. ۴۸۸	ریزداده	۳۶۶۰. ۶۷
ریزداده	۴۲	کلان‌داده	۰. ۴۶۲	داده‌کاوی	۳۱۷۳
طبقه‌بندی[7]	۳۸	داده‌کاوی	۰. ۴۶۰	طبقه‌بندی	۲۵۲۷. ۵

مرکزیت درجه یک گره نشانگر تعداد پیوندهای آن با سایر گره‌های موجود در شبکه است (عرفان‌منش و ارشدی ۱۳۹۴؛ عباسی، حسین و لیدسدورف، ۲۰۱۲). بر اساس جدول ۲ یادگیری ماشینی، کلان‌داده و داده‌کاوی دارای بیشترین میزان مرکزیت درجه می‌باشند. مرکزیت نزدیکی فاصله یک گره با گره‌های دیگر موجود در شبکه را سنجیده و میانگین طول کوتاه‌ترین مسیر میان آن گره و سایر گره‌های موجود در شبکه را نشان می‌دهد (عباسی، حسین و لیدسدورف، ۲۰۱۲). انطباق، تشخیص گفتار و یادگیری ماشینی به‌ترتیب دارای بیشترین میزان مرکزیت نزدیکی می‌باشند. مرکزیت بینابینی یک گره به تعداد دفعاتی اطلاق می‌شود که آن گره در بین کوتاه‌ترین مسیرهای بین جفت گره‌ها قرار می‌گیرد. گره‌های با مرکزیت بینابینی بالا در یک موقعیت ممتاز نقش یک کارگزار یا دروازه‌بان را برای اتصال گره‌ها و گروه‌ها بازی می‌کنند و به‌عنوان یک شاخص قدرت کنترل‌کننده مستقیم و غیرمستقیم جریان اطلاعات در شبکه شمرده می‌شوند. همچنین مقدار مرکزیت بینابینی همواره عددی بین صفر و یک است؛ در حالت صفر با حذف گره هیچ اتفاق خاصی در شبکه پیش نمی‌آید و همه گره‌ها به هم متصل باقی می‌مانند و حتی فواصل کوتاه میانشان از بین نمی‌رود، اما در حالت یک، گره در موقعیت استراتژیکی قرار دارد که در این صورت این گره خود می‌تواند کاندیدای نقطه عطف بوده و از موقعیت منحصربه‌فردی برخوردار باشد (هانسن، اشنایدرمن و اسمیت، ۲۰۱۰؛ عباسی، حسین و لیدسدورف، ۲۰۱۲). یادگیری ماشینی، کلان‌داده و ریزداده، همان‌طور که در تصویر ۳ مشخص است در مرکز نقشه قرار داشته و دارای بیشترین مرکزیت بینابینی هستند.

ب. حوزه مهندسی: به دنبال تحلیل هم‌واژگانی تولیدات علمی این حوزه، ۶ خوشه از واژگان و مفاهیم شناسایی شدند. تصویر ۴ نیز نقشه مفاهیم تولیدات علمی ریزداده را در حوزه مهندسی نشان می‌دهد.

1 . Betweeness
2 . Cuellar
3 . Machine Learning
4 . Adaptation
5 . Speech Recognition
6 . Data Mining
7 . Classification



تصویر ۴. نقشه هم‌رخدادی واژگان تولیدات علمی ریزداده در حوزه مهندسی

بر اساس ارتباطات و شبکه اجتماعی به‌دست‌آمده میان کلیدواژه‌ها از طریق نرم‌افزار وی.او.اس ویـوئر، در مرحلـه بعدی و با استفاده از نرم‌افزار گفی به تحلیل کلیدواژه‌هـای تولیـدات علمـی ریـزداده در حـوزه مهندسـی بـر اسـاس سنجه‌های مرکزیت پرداخته شد (جدول ۴).

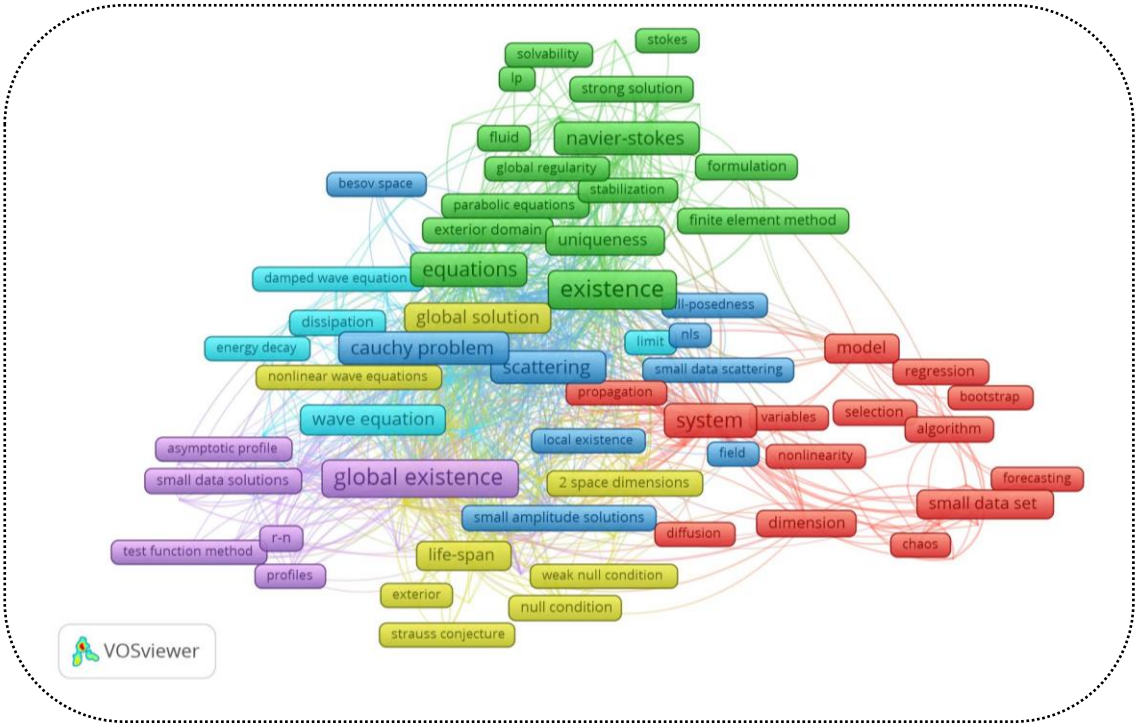
جدول ۴. پنج کلیدواژه برتر تولیدات علمی ریزداده در حوزه مهندسی بر اساس سنجه‌های مرکزیت

کلیدواژه	مرکزیت درجه	کلیدواژه	مرکزیت نزدیکی	کلیدواژه	مرکزیت بینابینی
اینترنت اشیاء[1]	۲۴	نمودار دوقطبی[2]	۱	فناوری تکامل بلندمدت	۱۱۲۰. ۵
ریزداده	۱۶	تنوع مکانی[3]	۱	اینترنت اشیاء	۱۰۸۹. ۲
فناوری تکامل بلندمدت[4]	۱۶	مرکز اطلاعات[5]	۱	تخصیص منابع	۶۲۵. ۲۳
ارتباطات ماشین به ماشین[6]	۱۶	تعادل بار[7]	۱	داده‌کاوی	۴۸۹
ارتباطات نوع ماشینی[8]	۱۶	جمع‌بندی بسته[9]	۱	ریزداده	۴۶۸. ۳

1 . Internet of Things (IoT)
2 . Bipartite graph
3 . Spatial diversity
4 . LTE (Long Term Evolution)
5 . Data center
6 . M2M
7 . Load balancing
8 . MTC
9 . Packet Concatenation

با توجه به جدول شماره ۴ اینترنت اشیاء، ریزداده و فناوری تکامل بلندمدت رتبه اول تا سوم مرکزیت بر اساس درجه را به خود اختصاص داده‌اند. نمودار دوقطبی، تنوع مکانی و مرکز اطلاعات دارای مرکزیت نزدیکی ۱ بوده و فناوری تکامل بلندمدت، اینترنت اشیاء و تخصیص منابع به‌ترتیب دارای بیشترین مرکزیت بینابینی در این حوزه می‌باشند.

ج. حوزه ریاضیات: به دنبال تحلیل هم‌واژگانی تولیدات علمی این حوزه، ۶ خوشه از واژگان و مفاهیم شناسایی و بازنمایی شد. پراستنادترین اثر این حوزه که متعلق به حوزه فیزیک نیز می‌باشد با عنوان « A PRACTICAL METHOD FOR CALCULATING LARGEST LYAPUNOV EXPONENTS FROM SMALL DATA SETS» از نویسندگان «ROSENSTEIN, MT; COLLINS, JJ; DE LUCA, CJ» در مجله «PHYSICA D-NONLINEAR PHENOMENA» در سال ۱۹۹۳ به چاپ رسیده است. تصویر ۵ نیز نقشه مفاهیم تولیدات علمی ریزداده را در حوزه ریاضیات نشان می‌دهد.



تصویر ۵. نقشه هم‌رخدادی واژگان تولیدات علمی ریزداده در حوزه ریاضیات

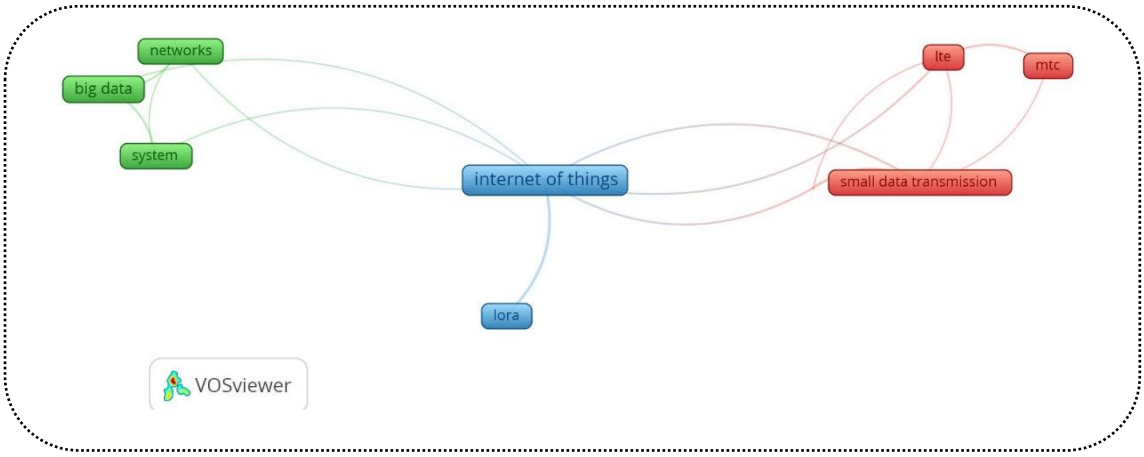
بر اساس ارتباطات و شبکه اجتماعی به‌دست‌آمده میان کلیدواژه‌ها از طریق نرم‌افزار وی.او.اس ویوئر، در مرحله بعدی و با استفاده از نرم‌افزار گفی به تحلیل کلیدواژه‌های تولیدات علمی ریزداده در حوزه ریاضیات بر اساس سنجه‌های مرکزیت پرداخته شد (جدول ۵).

با توجه به جدول ۵ وجود جهانی، معادلات ناویراستوکس و نماینده بحرانی به‌ترتیب دارای بیشترین میزان مرکزیت درجه در حوزه ریاضیات می‌باشند. یادگیری ماشینی، شبکه‌های عصبی و پیش‌بینی به‌ترتیب دارای بیشترین میزان مرکزیت نزدیکی هستند. وجود جهانی، معادلات ناویراستوکس و ریزداده دارای بیشترین میزان مرکزیت بینابینی می‌باشند.

جدول ۵. پنج کلیدواژه برتر تولیدات علمی ریزداده در حوزه ریاضیات بر اساس سنجه‌های مرکزیت

کلیدواژه	مرکزیت درجه	کلیدواژه	مرکزیت نزدیکی	کلیدواژه	مرکزیت بینابینی
وجود جهانی[1]	۵۶	یادگیری ماشینی	۱	وجود جهانی	۵۰۶۷.۷۸
معادلات ناویراستوکس[2]	۳۴	شبکه‌های عصبی[3]	۱	معادلات ناویراستوکس	۲۲۰۷.۵۷
نمای بحرانی[4]	۳۰	پیش‌بینی[5]	۰.۸	ریزداده	۱۵۰۱.۷۱
ریزداده	۲۶	نظریه خاکستری[6]	۰.۸	نماینده بحرانی	۱۴۵۲.۱۳
تخمین استریشارتز[7]	۲۰	خودراه‌اندازی[8]	۰.۵۷	پراکندگی[9]	۱۴۰۷.۷۷

د. حوزه مخابرات: به دنبال تحلیل هم‌واژگانی تولیدات علمی این حوزه، ۳ خوشه از واژگان و مفاهیم شناسایی شد. پراستنادترین اثر این حوزه با عنوان « Comparison of the Device Lifetime in Wireless Networks for the Internet of Things» از نویسندگان «Morin, E; Maman, M; Guizzetti, R; Duda, A» در نشریه «IEEE ACCESS» در سال ۲۰۱۷ به چاپ رسیده است. تصویر ۶ نیز نقشه مفاهیم تولیدات علمی ریزداده را در حوزه مخابرات نشان می‌دهد.



تصویر ۶. نقشه هم‌رخدادی واژگان تولیدات علمی ریزداده در حوزه مخابرات

بر اساس ارتباطات و شبکه اجتماعی به‌دست‌آمده میان کلیدواژه‌ها از طریق نرم‌افزار وی.او.اس ویوئر، در مرحله بعدی و با استفاده از نرم‌افزار گفی به تحلیل کلیدواژه‌های آثار ریزداده در حوزه مخابرات بر اساس سنجه‌های مرکزیت پرداخته شد (جدول ۶).

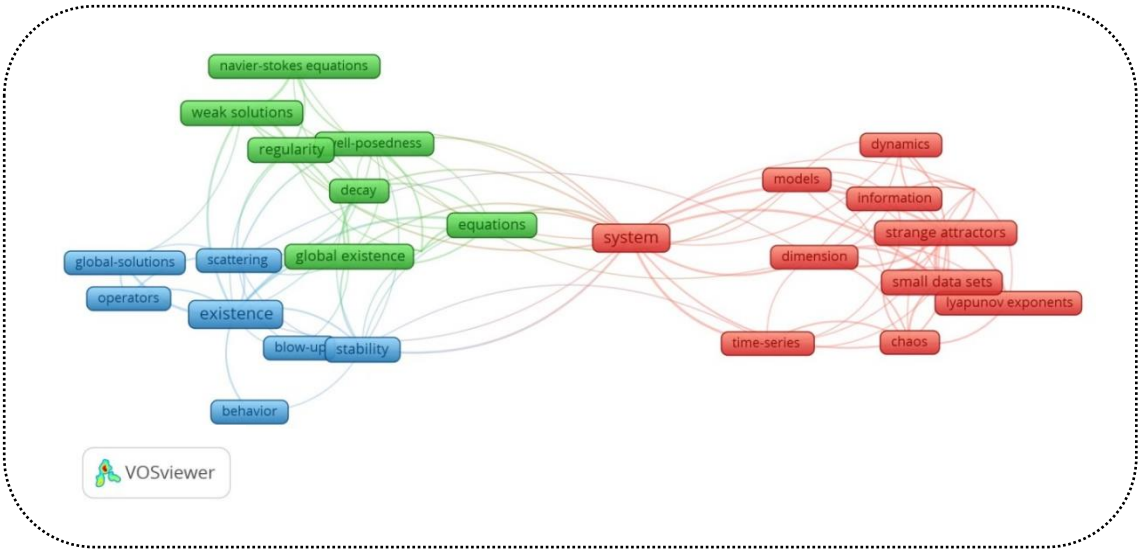
1 . Global existence
2 . Navier-Stokes equations
3 . Neural networks
4 . Critical exponent
5 . Forecasting
6 . Grey theory
7 . Strichartz estimates
8 . Bootstrap
9 . Scattering

جدول ۶. پنج کلیدواژه برتر تولیدات علمی ریزداده در حوزه مخابرات بر اساس سنجه‌های مرکزیت

کلیدواژه	مرکزیت درجه	کلیدواژه	مرکزیت نزدیکی	کلیدواژه	مرکزیت بینابینی
اینترنت اشیاء	۲۴	نمودار دوقطبی	۱	فناوری تکامل بلندمدت	۱۲۲۰. ۵
ریزداده	۱۶	تنوع مکانی	۱	اینترنت اشیاء	۱۰۸۹. ۲
فناوری تکامل بلندمدت	۱۶	مرکز اطلاعات	۱	تخصیص منابع	۶۲۵. ۲۳
ارتباطات ماشین به ماشین	۱۶	تعادل بار	۱	داده‌کاوی	۴۸۹
ارتباطات نوع ماشینی	۱۶	جمع‌بندی بسته	۱	ریزداده	۴۶۸. ۳

با توجه به جدول ۶ بیشترین میزان مرکزیت درجه در حوزه مخابرات مختص به اینترنت اشیاء، ریزداده و فناوری تکامل بلندمدت است. نمودار دوقطبی، تنوع مکانی و مرکز اطلاعات دارای بیشترین مرکزیت نزدیکی هستند و فناوری تکامل بلندمدت، اینترنت اشیاء و تخصیص منابع بیشترین میزان مرکزیت بینابینی را به خود اختصاص داده است.

ه. حوزه فیزیک: به دنبال تحلیل هم‌واژگانی تولیدات علمی این حوزه، ۳ خوشه از واژگان و مفاهیم شناسایی شد. تصویر ۷ نیز نقشه مفاهیم تولیدات علمی ریزداده را در حوزه فیزیک نشان می‌دهد.



تصویر ۷. نقشه هم‌رخدادی واژگان تولیدات علمی ریزداده در حوزه فیزیک

بر اساس ارتباطات و شبکه اجتماعی به‌دست‌آمده میان کلیدواژه‌ها از طریق نرم‌افزار وی.او.اس ویوئر، در مرحله بعدی و با استفاده از نرم‌افزار گفی به تحلیل کلیدواژه‌های تولیدات علمی ریزداده در حوزه فیزیک بر اساس سنجه‌های مرکزیت پرداخته شد (جدول ۷).

با توجه به جدول شماره ۷ نانوذرات، شبکه‌های عصبی مصنوعی و یادگیری عمیق دارای بیشترین مرکزیت درجه در حوزه فیزیک هستند. نانوذرات، رفتار بدون علامت و معادله موج دارای میزان مرکزیت نزدیکی ۱ هستند و نانوذرات، شبکه‌های عصبی مصنوعی و یادگیری عمیق دارای بیشترین مرکزیت بینابینی هستند. مقایسه تولیدات علمی

و خوشه‌های موضوعی پنج حوزه نشان می‌دهد که برخی تولیدات و موضوعات میان پنج حوزه مشترک می‌باشند؛ در این راستا به‌منظور تعیین دقیق شباهت موضوعی پنج حوزه فوق از شاخص دربردارندگی (شمول) استفاده شد که نتایج آن در ادامه آمده است (جدول ۸).

جدول ۷. پنج کلیدواژه برتر تولیدات علمی ریزداده در حوزه فیزیک بر اساس سنجه‌های مرکزیت

کلیدواژه	مرکزیت درجه	کلیدواژه	مرکزیت نزدیکی	کلیدواژه	مرکزیت بینابینی
نانوذرات[1]	۸	نانوذرات	۱	نانوذرات	۶
شبکه‌های عصبی مصنوعی[2]	۶	رفتارمجانبی[3]	۱	شبکه‌های عصبی مصنوعی	۰
یادگیری عمیق[4]	۶	معادله موج	۱	یادگیری عمیق	۰
انتقال یادگیری[5]	۶	معادله موج مرطوب	۱	انتقال یادگیری	۰
معادله موج میرا[6]	۶	برهم‌کنش میان‌سیال[7]	۱	رفتار مجانبی	۰

جدول ۸. هم‌پوشانی حوزه‌های علوم رایانه، مهندسی، ریاضی، مخابرات و فیزیک

حوزه A	حوزه B	تولیدات علمی مشترک (نسبت به کل حوزه A، نسبت به کل حوزه B)	موضوعات مشترک	شباهت موضوعی حوزه A نسبت به حوزه B	شباهت موضوعی حوزه B نسبت به حوزه A
فیزیک	مهندسی	۳۴ (۱۴، ۳)	۵۷	۶.۴ درصد	۱۳.۵ درصد
فیزیک	ریاضی	۳۹ (۱۷، ۵)	۱۰۸	۱۱.۹ درصد	۲۵.۶ درصد
فیزیک	مخابرات	۳۸ (۱۶، ۱۵)	۲۲	۲.۶ درصد	۵.۲ درصد
فیزیک	علوم رایانه	۱۸ (۷، ۱)	۵۰	۵ درصد	۱۱.۸ درصد
مهندسی	ریاضی	۶۶ (۶، ۸)	۷۲	۷.۹ درصد	۸ درصد
مهندسی	مخابرات	۳۵ (۱۴، ۳)	۲۷۴	۳۲.۹ درصد	۳۰.۷ درصد
مهندسی	علوم رایانه	۴۲۹ (۴۴، ۳۵)	۳۷۶	۳۷.۵ درصد	۴۲.۲ درصد
ریاضی	مخابرات	۳۹ (۴، ۱۶)	۲۰	۲.۴ درصد	۲.۲ درصد
ریاضی	علوم رایانه	۷۷ (۹، ۶)	۱۱۱	۱۱ درصد	۱۲.۲ درصد
مخابرات	علوم رایانه	۱۳۵ (۵۶، ۱۱)	۲۱۸	۲۱.۷ درصد	۲۶.۲ درصد

1 . Nanoparticles
2 . Artificial Neural Networks
3 . Asymptotic Behavior
4 . Deep Learning
5 . Transfer Learning
6 . Damped Wave Equation
7 . Fluid-structure Interaction

بر اساس جدول فوق، در میان حوزه‌های علمی مورد بررسی، بیشترین تعداد آثار مشترک (۴۲۹ اثر) و همچنین بالاترین میزان شباهت موضوعی مربوط به حوزه‌های مهندسی و رایانه (۳۷۶ کلیدواژه مشترک) است. همان‌طور که جدول فوق نشان می‌دهد حوزه‌های مهندسی و مخابرات با داشتن ۳۵ اثر مشترک، در مقایسه با سایر حوزه‌ها، هم‌پوشانی موضوعی نسبتاً بیشتری دارند. نکته جالب در مورد جدول فوق آن است که هرچند حوزه‌های فیزیک و علوم رایانه کمترین آثار مشترک (۱۸ اثر مشترک) را داشته‌اند اما کمترین شباهت موضوعی مربوط به حوزه‌های ریاضی و مخابرات (۲۰ کلیدواژه مشترک) با ۳۹ اثر مشترک است.

پاسخ به سؤال سوم پژوهش و آزمون فرضیه. آیا میان حوزه‌های علمی مشارکت‌کننده در تولیدات علمی ریزداده به لحاظ تعداد استنادات و تعداد رکوردهای تولیدشده تفاوت معناداری وجود دارد؟

به‌منظور بررسی تفاوت میان حوزه‌های علمی مشارکت‌کننده در تولیدات علمی ریزداده به لحاظ تعداد استنادات دریافتی و تعداد تولیدات علمی هر حوزه از آمار توصیفی و آمار استنباطی استفاده شد. آمار توصیفی مربوط به تعداد استنادات دریافتی هر یک از حوزه‌ها در جدول ۹ گزارش شده است.

جدول ۹. آمار توصیفی استنادات دریافتی حوزه‌های علمی مشارکت‌کننده در حوزه ریزداده

حوزه/آمار	میانگین	انحراف معیار	کمترین	بیشترین
علوم رایانه	۶.۰۶	۷.۷۲۹	۰	۳۳
مهندسی	۵.۶۵	۷.۲۸۷	۰	۲۷
ریاضیات	۱۸.۰۶	۱۶.۶۰۸	۱	۵۸
مخابرات	۰.۶۸	۱.۲۱۱	۰	۴
فیزیک	۱۸.۰۶	۱۶.۶۰۸	۱	۵۸
مجموع	۱۰.۶۳	۱۳.۸۸۹	۰	۵۸

همان‌طور که جدول بالا نشان می‌دهد میانگین دریافت استنادات در حوزه فیزیک و ریاضیات در طی سال‌های مختلف از سایر حوزه‌های علمی بیشتر بوده است. از سوی دیگر بیشترین میزان استناد دریافتی نیز مربوط به حوزه‌های فیزیک و ریاضیات با ۵۸ استناد است. در بخش آمار استنباطی مربوط به بررسی تفاوت معناداری در دریافت استنادهای حوزه‌های مختلف از آزمون آنوای یک‌سویه[1] یا تحلیل واریانس استفاده شد. آزمون تحلیل واریانس که یکی از آزمون‌های پارامتریک است دارای پیش‌فرض‌هایی است و تفاوت‌های میانگین در بیش از دو جامعه (نمونه) را بررسی می‌کند. در جدول ۱۰ نتیجهٔ آزمون آنوا گزارش شده است.

جدول ۱۰. آمار استنباطی استنادات دریافتی حوزه‌های علمی مشارکت‌کننده در حوزه ریزداده

نوع/آمار	مجموع مربعات	درجه آزادی	مربع میانگین	آماره F	معناداری
میان گروه‌ها	۷۴۷۶.۵۲۳	۴	۱۸۶۹.۱۳۱	۱۲.۶۱۱	۰.۰۰۰
درون گروه‌ها	۲۲۳۱.۵۱۶	۱۵۰	۱۴۸.۲۱۰		
مجموع	۲۹۷۰۸.۰۳۹	۱۵۴			

1 . ANOVA (one-way)

بر اساس جدول ۱۰، نتایج تحلیل واریانس یک‌سویه نشان داد که میان پنج حوزه مورد بررسی تفاوت معناداری به لحاظ تعداد استنادات دریافتی وجود داشته است (P-value= ۰.۰۰۰ و F (150,4)=12 .611). در ادامه و به‌منظور اینکه مشخص شود بین کدام‌یک از گروه‌ها در متغیر مورد نظر تفاوت وجود دارد باید از آزمون‌های تعقیبی[1] استفاده کرد. آزمون‌های تعقیبی به مقایسه جفتی و دوبه‌دوی گروه‌ها می‌پردازد. یکی از این آزمون‌ها آزمون شفه[2] است که در پژوهش حاضر از آن استفاده شد. درنهایت، نتیجه آزمون‌های آنوا و شفه می‌تواند به‌صورت زیر گزارش شود.

تفاوت معناداری بین «حوزه‌های علوم رایانه، ریاضیات و فیزیک»، بین «حوزه‌های مهندسی، ریاضیات و فیزیک» و بین «حوزه‌های مخابرات، ریاضیات و فیزیک» به لحاظ تعداد استنادات دریافتی وجود دارد؛ بنابراین در میان حوزه‌های علمی مشارکت‌کننده در تولیدات ریزداده، دو حوزه ریاضیات و فیزیک در مقایسه با سایر حوزه‌ها اثربخش‌تر بوده‌اند. در ادامه آمار توصیفی مربوط به تعداد تولیدات علمی هر یک از حوزه‌ها در جدول ۱۱ گزارش شده است.

جدول ۱۱. آمار توصیفی تعداد تولیدات حوزه‌های علمی مشارکت‌کننده در حوزه ریزداده

حوزه /آمار	میانگین	انحراف معیار	کمترین	بیشترین
علوم رایانه	۳۷. ۴۴	۳۳. ۴۰۷	۱	۱۱۹
مهندسی	۳۱. ۲۹	۲۷. ۰۱۳	۱	۹۷
ریاضیات	۲۳. ۲۶	۱۸. ۲۹۴	۱	۶۹
مخابرات	۱۰. ۸۶	۱۰. ۳۴۸	۱	۳۷
فیزیک	۶. ۷۴	۵. ۳۵۶	۱	۲۶
مجموع	۲۲. ۴۰	۲۴. ۵۲۸	۱	۱۱۹

همان‌طور که جدول بالا نشان می‌دهد میانگین تولیدات در حوزه‌های علوم رایانه و مهندسی طی سال‌های مختلف از سایر حوزه‌ها بیشتر بوده است. از سوی دیگر بیشترین تعداد تولیدات مربوط به دو حوزه فوق است. در جدول ۱۲، نتیجه آزمون آنوا مربوط به بررسی تفاوت میان حوزه‌ها به لحاظ تعداد تولیدات گزارش شده است.

جدول ۱۲. آمار استنباطی تعداد تولیدات حوزه‌های علمی مشارکت‌کننده در حوزه ریزداده

نوع / آمار	مجموع مربعات	درجه آزادی	مربع میانگین	آماره F	معناداری
میان گروه‌ها	۲۰۹۸۲. ۸۸۳	۴	۵۲۴۵. ۷۲۱	۱۰. ۹۹۹	۰. ۰۰۰
درون گروه‌ها	۷۱۰۶۲. ۱۵۶	۱۴۹	۴۷۶. ۹۲۷		
مجموع	۹۲۰۴۵. ۰۳۹	۱۵۳			

بر اساس جدول ۱۲، نتایج تحلیل واریانس یک‌سویه نشان داد که میان پنج حوزه مورد بررسی تفاوت معناداری به لحاظ تعداد تولیدات علمی وجود داشته است (P-value= ۰.۰۰۰ و F (149,4)=10 .999). نتیجه آزمون‌های آنوا و شفه درنهایت می‌تواند به‌صورت زیر گزارش شود.

1 . Post Hoc
2 . Scheffe

تفاوت معناداری بین حوزه‌های علوم رایانه و مخابرات و فیزیک، بین حوزه‌های مهندسی و مخابرات و فیزیک، بین حوزه‌های ریاضیات و فیزیک به لحاظ تعداد تولیدات علمی وجود دارد؛ بنابراین در میان حوزه‌های علمی مشارکت‌کننده در تولیدات علمی ریزداده، دو حوزه علوم رایانه و فیزیک در مقایسه با سایر حوزه‌ها دارای بهره‌وری بیشتری بوده‌اند.

بحث و نتیجه‌گیری

پژوهش حاضر با هدف شناسایی ویژگی‌های تولیدات علمی حوزهٔ ریزدادهٔ نمایه‌شده در پایگاه وب آو ساینس و تبیین کاربرد آن بر مبنای شناسایی واژگان تولیدات علمی این حوزه در حوزه‌های علمی مختلف با استفاده از فن هم‌رخدادی واژگان انجام شد. یافته‌ها نشان‌دهنده روند صعودی و بالندگی مستمر تولیدات علمی این حوزه است، به‌گونه‌ای که در سال ۲۰۱۸ بیشترین تعداد تولیدات علمی این حوزه (۳۳۲ رکورد) ثبت شده است. سه کشور اول مشارکت‌کننده در تولیدات این حوزه ایالات متحده (۹۶۳ رکورد)، چین (۶۰۵ رکورد) و آلمان (۳۰۸ رکورد) هستند، هرچند دانشگاه چنگ‌کونگ (تایوان) از نظر میزان تولید علم و دریافت استناد جایگاه اول را کسب کرده است. کشور ایران با ۳۱ اثر رتبه ۲۷ را در میان ۹۱ کشور مشارکت‌کننده این حوزه از آن خود کرده است. در پژوهش‌های لوپزروبلز (۲۰۱۸)، جلوه‌گران و همکاران (۲۰۱۹) و پورخانی و همکاران (۲۰۱۹) ایالات متحده و چین نیز در بین کشورهای مشارکت‌کننده برتر هستند. لی دی سی، چنگ سی جی و چن سی سی با ۶۲، ۲۲ و ۱۹ رکورد فعال‌ترین نویسندگان در این حوزه بودند. همچنین لی دی سی، چنگ سی جی و تسای تی آی بیشترین استناد را در این حوزه دریافت کرده‌اند. بیش از ۹۰ درصد تولیدات علمی ریزداده مربوط به حوزه‌های علوم رایانه با ۱۲۰۹، مهندسی با ۹۷۴، ریاضیات با ۸۲۵، مخابرات با ۲۴۱ و فیزیک با ۲۲۹ رکورد است. در پژوهش لوپزروبلز (۲۰۱۸) حوزه‌های علوم رایانه، علم اطلاعات و مهندسی بیشترین مشارکت را در پژوهش‌های حوزه کلان‌داده داشته‌اند. واژگان ریزداده، کلان‌داده و یادگیری ماشین به‌ترتیب با فراوانی ۹۳، ۸۱ و ۷۴ هسته اصلی پژوهش‌های این حوزه را تشکیل داده‌اند. هم‌رخدادی «ریزداده» و «کلان‌داده» با فراوانی بالا نشان‌دهنده این است که این دو رویکرد مکمل هم بوده و در تعامل با هم به تقویت و توسعه یکدیگر کمک می‌کنند و این روش‌شناسی ترکیبی سبب کسب نتایج بهتر در تحلیل داده‌ها می‌شود.

همچنین کلیدواژه‌هایی نظیر «یادگیری عمیق»، «وجود جهانی»، «داده‌کاوی»، «طبقه‌بندی»، «اینترنت اشیاء»، «فناوری تکامل بلندمدت»، «تخصیص منابع» و «پیش‌بینی» نیز در پژوهش‌های این حوزه دارای فراوانی بالایی بوده و سهم زیادی از مباحث این حوزه را به خود اختصاص داده‌اند و جزو زیرشاخه‌های عمده این حوزه محسوب می‌شوند. این نتایج نشان‌دهنده آن است که پژوهشگران این حوزه در تلاش هستند با پژوهش‌های ترکیبی روی انواع ریزداده‌ها و کلان‌داده‌های حاصل از اینترنت اشیاء و سایر فناوری‌ها از طریق الگوریتم‌های یادگیری ماشین، داده‌کاوی و طبقه‌بندی، پیش‌بینی و به دنبال آن تخصیص منابع را بهبود دهند. در پژوهش روزبهانی و همکاران (۲۰۱۹) نیز کلان‌داده‌ها و یادگیری ماشینی، متن‌کاوی و داده‌های معامله‌ای موضوعات محوری معرفی شده‌اند. جلوه‌گران و همکاران (۲۰۱۹) کلان داده‌ها، رسانه‌های اجتماعی و شبکه‌های اجتماعی (آنلاین) را پرتکرارترین کلیدواژه‌ها در پژوهش خود بیان کرده‌اند. رسانه‌های اجتماعی، تأثیر و تبلیغات دهان‌به‌دهان سه کلیدواژه پرتکرار در پژوهش پورخانی و همکاران (۲۰۱۹) هستند.

در این پژوهش خوشه‌بندی هم‌واژگانی منجر به تشکیل ۸ خوشه در حوزه علوم رایانه، ۶ خوشه در حوزه

مهندسی، ۷ خوشه در حوزه ریاضیات، ۳ خوشه در حوزه مخابرات و ۳ خوشه در حوزه فیزیک شد. شکل‌گیری تعداد بیشتر خوشه‌های موضوعی در حوزه علوم رایانه از یک‌سو، و وجود تفاوت معنادار بین تولیدات علمی این حوزه با سایر حوزه‌ها حاکی از پیشروبودن این حوزه در مقایسه با سایر حوزه‌هاست؛ در میان هشت خوشهٔ شناسایی‌شده، دو خوشه (زرد و سبز) که مفاهیم و واژگان آنها شامل کلان‌داده، تحلیل داده، چارچوب، مدل، طبقه‌بندی، داده‌کاوی، انتخاب و تحلیل خوشه است، ازآنجایی‌که بیشتر کلیدواژه‌های پرتکرار و رایج در پژوهش‌های ریزداده در این خوشه‌ها هستند؛ بنابراین از جایگاه مرکزی و مهمی برخوردار بوده و مهم‌ترین گرایش‌های پژوهشگران حوزهٔ علوم رایانه هستند؛ همچنین در تولیدات علمی این حوزه کلیدواژه‌های یادگیری ماشینی، کلان‌داده و داده‌کاوی دارای بیشترین میزان مرکزیت درجه، انطباق، تشخیص گفتار و یادگیری ماشینی دارای بیشترین میزان مرکزیت نزدیکی، و یادگیری ماشینی، کلان‌داده و ریزداده دارای بیشترین مرکزیت بینابینی هستند. در میان شش خوشه شناسایی‌شده حوزه مهندسی، چهار خوشه شامل مفاهیم و واژگان دارای هم‌رخدادی بالا نظیر «شبکه عصبی، مجموعه ریزداده، الگوریتم، ماشین‌بردار پشتیبانی، مدل» است، ازآنجایی‌که بیشتر کلیدواژه‌های پرتکرار و رایج در پژوهش‌های ریزداده در این خوشه‌ها هستند؛ بنابراین از جایگاه مرکزی و مهمی برخوردار بوده و مهم‌ترین گرایش‌های پژوهشگران حوزه مهندسی هستند؛ همچنین در تولیدات علمی این حوزه، کلیدواژه‌های اینترنت اشیاء، ریزداده و فناوری تکامل بلندمدت رتبه اول تا سوم در مرکزیت بر اساس درجه، کلیدواژه‌های نمودار دوقطبی، تنوع مکانی و مرکز اطلاعات رتبه‌های برتر در مرکزیت بر اساس نزدیکی، و کلیدواژه‌های فناوری تکامل بلندمدت، اینترنت اشیاء و تخصیص منابع به‌ترتیب رتبه‌های برتر مرکزیت بینابینی می‌باشند. نکته جالب این است که هرچند بیشترین تعداد آثار مشترک و همچنین بالاترین میزان شباهت موضوعی مربوط به حوزه‌های مهندسی و رایانه (۳۷۶ کلیدواژه مشترک) است، اما شباهت موضوعی در میان کلیدواژه‌های برتر دو حوزه مشاهده نمی‌شود (جداول ۳ و ۴). در میان هفت خوشه شناسایی‌شده حوزه ریاضیات، چهار خوشه که مفاهیم و واژگان آنها شامل وجود جهانی، معادلات، سیستم، پراکندگی، مسئله‌کوشی و راه‌حل جهانی است، ازآنجایی‌که بیشتر کلیدواژه‌های پرتکرار و رایج در پژوهش‌های ریزداده در این خوشه‌ها هستند، از جایگاه مرکزی و مهمی برخوردارند و مهم‌ترین گرایش‌های پژوهشگران حوزه ریاضیات هستند. از سوی دیگر در تولیدات علمی این حوزه، کلیدواژه‌های وجود جهانی، معادلات ناوبر–استوکس و نماینده بحرانی به‌ترتیب دارای بیشترین میزان مرکزیت درجه، کلیدواژه‌های یادگیری ماشینی، شبکه‌های عصبی و پیش‌بینی به‌ترتیب دارای بیشترین میزان مرکزیت نزدیکی، و وجود جهانی، معادلات ناوبر–استوکس و ریزداده دارای بیشترین میزان مرکزیت بینابینی می‌باشند. در میان سه خوشه شناسایی‌شده حوزه مخابرات، یک خوشه که مفاهیم و واژگان آن شامل اینترنت اشیاء است، ازآنجایی‌که بیشتر کلیدواژه‌های پرتکرار و رایج در پژوهش‌های ریزداده در این خوشه است، از جایگاه مرکزی و مهمی برخوردار بوده و مهم‌ترین گرایش پژوهشگران حوزه مخابرات است؛ همچنین در حوزه مخابرات بیشترین میزان مرکزیت درجه مختص به اینترنت اشیاء، ریزداده و فناوری تکامل بلندمدت است. نمودار دوقطبی، تنوع مکانی و مرکز اطلاعات دارای بیشترین مرکزیت نزدیکی هستند و فناوری تکامل بلندمدت، اینترنت اشیاء و تخصیص منابع بیشترین میزان مرکزیت بینابینی را به خود اختصاص داده است. نکته جالب این است که شباهت موضوعی در میان کلیدواژه‌های برتر دو حوزه مهندسی و مخابرات مشاهده می‌شود. در میان سه خوشه شناسایی‌شده حوزه فیزیک، مفاهیم و واژگان دارای هم‌رخدادی بالای آنها در دو خوشه شامل وجود، نانوذرات، سیستم، مجموعه ریزداده است، ازآنجایی‌که بیشتر کلیدواژه‌های پرتکرار و رایج در پژوهش‌های ریزداده در این خوشه‌ها هستند، بنابراین از جایگاه مرکزی و مهمی

برخوردار بوده و مهم‌ترین گرایش‌های پژوهشگران حوزه علوم رایانه هستند؛ از سوی دیگر در حوزه فیزیک نانوذرات، شبکه‌های عصبی مصنوعی و یادگیری عمیق دارای بیشترین مرکزیت درجه، همچنین نانوذرات، رفتار بدون علامت و معادله موج دارای میزان بیشترین مرکزیت نزدیکی و نانوذرات، شبکه‌های عصبی مصنوعی و یادگیری عمیق دارای بیشترین مرکزیت بینابینی هستند. نتایج پژوهش حاضر نشان داد در میان حوزه‌های علمی مورد بررسی هرچند بیشترین تعداد آثار مشترک (۴۲۹ اثر) و همچنین بالاترین میزان شباهت موضوعی مربوط به حوزه‌های مهندسی و رایانه (۳۷۶ کلیدواژه مشترک) است اما به‌طور کلی میان تعداد آثار مشترک و میزان شباهت موضوعی رابطه مشخصی وجود ندارد؛ به‌طور مثال حوزه‌های مهندسی و مخابرات با داشتن تنها ۳۵ اثر مشترک، در مقایسه با سایر حوزه‌ها که آثار مشترک بیشتری دارند، هم‌پوشانی موضوعی نسبتاً بیشتری دارند. همچنین حوزه‌های فیزیک و علوم رایانه هرچند کمترین آثار مشترک (۱۸ اثر مشترک) را داشته‌اند اما کمترین شباهت موضوعی مربوط به حوزه‌های ریاضی و مخابرات (۲۰ کلیدواژه مشترک) با ۳۹ اثر مشترک است.

همچنین نتایج پژوهش حاضر مشخص نمود حوزه‌های ریاضیات و فیزیک به لحاظ تعداد استنادات دریافتی و حوزه‌های علوم رایانه و فیزیک به لحاظ تعداد تولیدات از ابتدا تاکنون نسبت به سایر حوزه‌های علمی اثربخش‌تر بوده‌اند.

با توجه به اهمیت داده در صنعت و اقتصاد عصر حاضر، یافته‌ها و نتایج حاصل از این تحلیل می‌تواند سیاست‌گذاران، صنایع و سازمان‌ها، مدیران و پژوهشگران فعال را هدایت نموده تا برنامه‌ریزی مناسبی به‌منظور افزایش کمی و کیفی تولیدات علمی و توسعه متوازن موضوعات این حوزه انجام دهند. همچنین، نمای ارائه‌شده از ساختار فکری دانش ریزداده در حوزه‌های علمی مختلف می‌تواند دیدگاهی علمی از شکاف‌های موضوعی و موضوعات در حال رشد این حوزه‌ها ارائه نموده و از پژوهش‌های کم‌کاربرد و تکراری جلوگیری کند.

پیشنهادهای اجرایی پژوهش

با توجه به اهمیت ریزداده و روند تولیدات علمی این حوزه می‌توان نسبت به برداشتن گام‌های عملی برای بهره‌برداری بیشتر از آن هم‌سو با مرزهای جهانی دانش اقداماتی را مناسب اجرا دانست که اهم آنها عبارت‌اند از:

۱. گسترش همکاری‌های علمی پژوهشگران پنج حوزه مورد بررسی و سایر حوزه‌های مشارکت‌کننده به‌ویژه دو حوزه علوم رایانه و مهندسی به دلیل شباهت موضوعی نسبتاً بالای این حوزه‌ها به‌منظور تقویت دیدگاه میان‌رشته‌ای و نهایتاً اثربخشی بیشتر تولیدات علمی حوزه ریزداده؛

۲. شناسایی فرایندها، ابزارها و فناوری‌های مورد نیاز در حوزه ریزداده بر اساس واژگان و مفاهیم شناسایی‌شده در علوم مختلف و سپس انجام برنامه‌ریزی، آماده‌سازی و استفاده از آنها در پروژه‌های اجرایی و پژوهشی مرتبط؛

۳. توجه به رویکرد ریزداده به‌عنوان مکمل طرح‌های پژوهشی کلان‌داده‌ای بر اساس فراوانی پربسامد واژه «کلان‌داده» در تولیدات علمی.

پیشنهاد برای پژوهش‌های آتی

بر اساس پژوهش حاضر، پیشنهادهای زیر برای انجام پژوهش‌های آتی ارائه می‌شوند:

۱. تحلیل ساختار و مفاهیم تولیدات علمی این حوزه در دیگر پایگاه‌های استنادی نظیر اسکوپوس و گوگل‌اسکالر و همچنین وب اجتماعی به‌منظور ارزیابی جامع‌تری از وضعیت پژوهش این حوزه در جهان؛

۲. بررسی تطبیقی تولیدات علمی ریزداده با تولیدات علمی کلان‌داده برای مشخص‌شدن نسبت میان دو رویکرد؛
۳. تحلیل محتوا و ساختار مفاهیم و واژگان مدارک علمی زیرحوزه‌ها و خوشه‌های حاصل از مطالعه حاضر به‌طور مجزا، با هدف ترمیم کاستی‌های ابزار و فناوری‌های مورد نیاز مراکز پژوهشی در ایران؛
۴. مطالعه همکاری‌های علمی پژوهشگران علوم مختلف در حوزه ریزداده بر اساس شاخص‌های هم‌نویسندگی.

فهرست منابع

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

دانش، فرشید؛ نعمت‌اللهی، زهرا. (۱۳۹۹). خوشه‌بندی: مفاهیم و رویدادهای نوپدید سازمان‌دهی دانش. *کتابداری و اطلاع‌رسانی*، ۲۳ (۲): ۵۳-۸۵.

جعفری، سمیه؛ فرشید، راضیه؛ و جباری، لیلا. (۱۳۹۹). تحلیل موضوعی مطالعات کووید ۱۹ در پنج قاره بزرگ. *پژوهش‌نامه علم‌سنجی* ۶ (۱۱): ۲۷۷-۲۹۷.

حسینی، الهه؛ غائبی، امیر؛ و برادر، رؤیا. (۱۳۹۹). کتاب‌سنجی و نگاشت هم‌رخدادی واژگان در حوزه داده‌های پیوندی. *پژوهش‌نامه علم‌سنجی*.

رمضانی، هادی؛ علی‌پور حافظی، مهدی؛ و مؤمنی، عصمت. (۱۳۹۳). نقشه‌های علمی: فنون و روش‌ها. *فصلنامه ترویج علم*، ۵ (۶)، ۵۳-۸۴.

رئیس‌زاده، محمد؛ کرمعلی، مازیار. (۱۳۹۷). ترسیم نقشه علمی مقالات حوزه ترومای نظامی با استفاده از تحلیل هم‌واژگانی در مدلاین. *مجله طب نظامی*، ۲۰ (۵)، ۴۷۶-۴۸۷.

سالمی، نجمه؛ کوشا، کیوان. (۱۳۹۱). مقایسه تحلیل هم‌استنادی و تحلیل هم‌واژگانی در ترسیم نقشه کتابشناختی مطالعه موردی: دانشگاه تهران. *پژوهشنامه پردازش مدیریت اطلاعات*، ۲۹ (۱): ۲۵۳-۲۶۶.

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

سهیلی، فرامرز؛ عصاره، فریده. (۱۳۹۲). مفاهیم مرکزیت و تراکم در شبکه‌های علمی و اجتماعی. *فصلنامه مطالعات ملی کتابداری و سازمان‌دهی اطلاعات*، ۲۴ (۳)، ۹۲-۱۰۸.

شکفته، مریم؛ حریری، نجلا. (۱۳۹۲). ترسیم و تحلیل نقشه علمی پزشکی ایران با استفاده از روش هم‌استنادی موضوعی و معیارهای تحلیل شبکه اجتماعی. *مدیریت سلامت*، ۱۶ (۱۵): ۴۳-۵۹.

ضیغمی، رضا؛ باقری نسامی، معصومه؛ حق‌دوست، فاطمه؛ و یادآور، منصوره. (۱۳۸۷). تحلیل محتوا. *فصلنامه پرستاری ایران*، ۲۱ (۵۳): ۴۱-۵۲.

فیض‌آبادی، منصوره؛ وزیری، اسماعیل. (۱۳۹۶). ترسیم ساختار حوزه‌های علمی مطالعات دمانس با استفاده از روش

تحلیل هم‌رخدادی واژگان. *مجله دانشگاه علوم پزشکی سبزوار*، ۲۵ (۲)، ۲۳-۳۰.

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

نوروزی چاکلی، عبدالرضا. (۱۳۹۱). نقش و جایگاه مطالعات علم‌سنجی در توسعه. *پژوهش‌نامه پردازش و مدیریت اطلاعات*، ۲۷ (۳)، ۷۲۳-۷۳۶.

هان، ژیاوی؛ کمبر، میشلین؛ و پی، ژان (۱۳۹۴). داده‌کاوی: مفاهیم و تکنیک‌ها. ترجمهٔ مهدی اسماعیلی. تهران: نیاز دانش.

Augustin, N. H., & Faraway, J. J. (2018). When small data beats big data. Statistics and Probability Letters, Special Issue on The role of Statistics in the era of big data, to appear.

Cerquitelli, T., Quercia, D., & Pasquale, F. (Eds.). (2017). Transparent data mining for Big and small data. Springer International Publishing

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation research part C: emerging technologies*, 68, 285-299.

Emmert-Streib, F., & Dehmer, M. (2019). Defining data science by a data-driven quantification of the community. *Machine Learning and Knowledge Extraction*, 1 (1), 235-251.

Estrin, D. (2014). Small data, where n= me. *Communications of the ACM*, 57 (4), 32-34.

Fang, Z., Costas, R., Tian, W., Wang, X., & Wouters, P. (2020). An extensive analysis of the presence of altmetric data for Web of Science publications across subject fields and research topics. *Scientometrics*, 124 (3), 2519-2549.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35 (2), 137-144.

Hekler, E. B., Klasnja, P., Chevance, G., Golaszewski, N. M., Lewis, D., & Sim, I. (2019). Why we need a small data paradigm. *BMC medicine*, 17 (1), 1-9.

Javid, E., Nazari, M & Ghaeli, M. (2019). Social media and e-commerce: A scientometrics analysis.*International Journal of Data and Network Science*, 3 (3), 269-290.

Jin, Y., & Li, X. (2019). Visualizing the hotspots and emerging trends of multimedia big data through scientometrics. *Multimedia Tools and Applications*, 78 (2), 1289-1313.

Karimi,Ramin. (1394). *Easy statistical analysis guide with spss*.Tehran: Hengam.

Kitchin, Rob and Lauriault, Tracey P. (2015) Small data in the era of big data. *GeoJournal*, 80. pp. 463-475.

Li, T., Gu, D., Wang, X., & Liang, C. (2018, July). Visualizing the Intellectual Structure of Electronic Health Research: A Bibliometric Analysis. In International Conference on Smart Health (pp. 315-324). Springer, Cham.

Lindstrom, M. (2016). Small data: the tiny clues that uncover huge trends. St. Martin's Press.

López-Robles, J. R., Otegi-Olaso, J. R., Gomez, I. P., Gamboa-Rosales, N. K., Gamboa-Rosales, H., & Robles-Berumen, H. (2018, November). Bibliometric network analysis to identify the intellectual structure and evolution of the big data research field. In International Conference on Intelligent Data Engineering and Automated Learning (pp. 113-120). Springer, Cham.

Minelli, M., Chambers, M., & Dhiraj, A. (2013). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses (Vol. 578). John Wiley & Sons.

Ozdagoglu G., Damar M., Ozdagoglu A. (2020) The State of the Art in Blockchain Research (2013–2018): Scientometrics of the Related Papers in Web of Science and Scopus. In: Hacioglu U. (eds) Digital Business Strategies in Blockchain Ecosystems. Contributions to Management Science. Springer, Cham. https: . .doi.org .10.1007 .978-3-030-29739-8_27

Papić, A., & Eskić, E. (2018). Big Data and Data Science: A Scientometrics Approach. *Management*, 16, 18.

Pourkhani, A., Abdipour, K., Baher, B., & Moslehpour, M. (2019). The impact of social media in business growth and performance: A scientometrics analysis. *International Journal of Data and Network Science*, 3 (3), 223-244.

Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In 2013 international conference on collaboration technologies and systems (CTS) (pp. 42-47). IEEE.

Secchi, P. (2018). On the role of statistics in the era of big data: A call for a debate. *Statistics & Probability Letters*, 136, 10-14.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* .William R. Shedish, Thomas D. Cook, Donald T. Campbell. Boston: Houghton Mifflin,.

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic Books.

Soleimani-Roozbahani, F., Ghatari, A. R., & Radfar, R. (2019). Knowledge discovery from a more than a decade studies on healthcare Big Data systems: a scientometrics study. *Journal of Big Data*, 6 (1), 1-15.