Applied Scientometric Studies
Volume 2, Issue 1, 2025

Visualization of Scientific Production on Big Data in the Citation Database

Fatemeh Dolati Liavali[1], Mohaddeseh Nourmohammadi Amjad[2], and Zahra Hashem[3]

1. B.A. Student in Knowledge and Information Science, University of Qom, Iran (Corresponding Author); Email: dolatifateme9@gmail.com
2. B.A. Student in Knowledge and Information Science, University of Qom, Iran; Email: mn9088853@gmail.com
3. B.A. Student in Knowledge and Information Science, University of Qom, Iran; Email: hh0567729@gmail.com

Article Info

Article type:
Research

Article history:
Received: 2025/02/18
Accepted: 2025/03/30
Available online: 2025/05/31

Keywords:
Visualization, scientific mapping, big data, scientometrics, scientific production, web of science.

ABSTRACT

Propose: The present study aims to visualize the scientific production regarding big data indexed in the Web of Science (WoS) citation database.

Methodology: This study is applied research conducted using a scientometric approach. Data extraction was performed by searching for the query "Ts=Big Data" in the advanced search section of Web of Science, resulting in the retrieval of 1,053 bibliographic records. The statistical population comprises all scientific outputs from Iran related to big data indexed in this database. Following extraction, the data were imported into scientometric software. Specifically, scientific software tools such as VOSviewer, BibExcel, and UCINET were utilized for data analysis and scientific mapping.

Findings: The findings revealed that a total of 1,053 bibliographic records from Iran regarding big data are indexed in this citation database. The first article dates back to 2013. Iran's most significant research collaborations were with the United States, Australia, and the United Kingdom. All scientific productions were published in English, and the Islamic Azad University had the highest volume of scientific output with 240 records. Among the authors, Amir Mosavi from Obuda University ranked first with 48 records, 1,851 citations, and an h-index of 26.

Conclusion: The results indicate a significant growth in Iran's scientific production in the field of big data, particularly within the domains of artificial intelligence and machine learning. The high concentration of researchers on current technological topics and extensive international collaborations demonstrate Iran's substantial potential in this field. However, quantitative fluctuations in production in recent years highlight the necessity for supportive policy-making and more precise direction of research based on the identified thematic clusters. Overall, the data formed four clusters, with the second cluster being the largest, comprising 20 keywords.

Cite this article: Dolati Liavali, Fatemeh; Nourmohammadi Amjad, Mohaddeseh; Hashem, Zahra (2025). Visualization of Scientific Production on Big Data in the Citation Database, *Applied Scientometric Studies*, 2(1), 14 - 30. https://doi.org/10.22091/apss.2025.12399.1032

© Author(s) retain the copyright and full publishing rights.
Publisher: University of Qom.
DOI: http//doi.org/10.22091/apss.2025.12399.1032

مطالعات کاربردی علم‌سنجی
دوره ۲، شماره ۱، ۱۴۰۴

مصوّرسازی تولیدات علمی کلان داده‌ها در پایگاه استنادی

فاطمه دولتی لیاولی[۱]، محدثه نورمحمدی امجد[۲]، و زهرا هاشم[۳]

۱. دانشجوی کارشناسی علم اطلاعات و دانش شناسی دانشگاه قم، ایران(نویسنده مسئول). رایانامه: dolatifateme9@gmail.com
۲. دانشجوی کارشناسی علم اطلاعات و دانش شناسی دانشگاه قم، ایران. رایانامه: mn9088853@gmail.com
۳. دانشجوی کارشناسی علم اطلاعات و دانش شناسی دانشگاه قم، ایران. رایانامه: hh0567729@gmail.com

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی.

تاریخچه مقاله:
تاریخ دریافت: ۱۴۰۴/۱۱/۳۰
تاریخ پذیرش: ۱۴۰۴/۰۱/۱۰
تاریخ انتشار: ۱۴۰۴/۰۳/۱۰

کلیدواژه‌ها:
مصور سازی، ترسیم نقشه علمی، کلان داده، علم‌سنجی، تولیدات علمی، وب‌آوساینس.[۱]

چکیده

هدف: این پژوهش با هدف مصوّرسازی تولیدات علمی کلان داده‌ها در پایگاه استنادی وب‌آوساینس نوشته شده است.

روش: این پژوهش از نوع پژوهش‌های کاربردی بوده و با رویکرد علم‌سنجی انجام شده است. استخراج داده‌ها با جست‌وجوی عبارت"Ts=Big Data " در قسمت جست‌وجوی پیشرفتهٔ وب‌آوساینس، انجام شد که در مجموع ۱۰۵۳ رکورد اطلاعاتی استخراج شد. جامعهٔ آماری این پژوهش کلیهٔ داده‌های مرتبط با تولیدات علمی ایران در این پایگاه استنادی در رابطه با کلان داده است. داده‌ها پس از استخراج، وارد نرم‌افزارهای علم‌سنجی شد. برای تحلیل داده‌ها و ترسیم نقشهٔ علمی از نرم‌افزارهای علمی مثل ووس‌ویوور، بایب‌اکسل، و یوسی‌نت استفاده شد.

یافته‌ها: یافته‌ها نشان داد در مجموع ۱۰۵۳ رکورد اطلاعاتی از ایران در رابطه با کلان‌داده‌ها در این پایگاه استنادی نمایه شده است. اولین مقاله مربوط به سال ۲۰۱۳ بوده است. بیشترین همکاری کشور ایران در نوشتن مقالات با کشورهای آمریکا، استرالیا، و انگلستان بوده است. همهٔ تولیدات علمی به زبان انگلیسی بوده و دانشگاه آزاد با تولید ۲۴۰ رکورد اطلاعاتی، بیشترین میزان تولیدات علمی را داشته است. در بین نویسندگان امیر موسوی از دانشگاه اوبدو[۲] با ۴۸ رکورد اطلاعاتی و با ۱۸۵۱ استناد و با شاخص هرش ۲۶ در رتبهٔ اول قرار دارد.

نتیجه‌گیری: نتایج نشان‌دهندهٔ رشد قابل‌توجه تولیدات علمی ایران در حوزه کلان‌داده، به‌ویژه در محورهای هوش مصنوعی و یادگیری ماشین است. تمرکز بالای پژوهشگران بر موضوعات روز فناوری و همکاری‌های گسترده بین‌المللی، پتانسیل بالای ایران را در این حوزه نشان می‌دهد. با این حال، نوسانات کمی تولیدات در سال‌های اخیر، لزوم سیاست‌گذاری‌های حمایتی و جهت‌دهی دقیق‌تر پژوهش‌ها بر اساس خوشه‌های موضوعی شناسایی‌شده را برجسته می‌سازد. به طور کلی داده‌ها از ۴ خوشه تشکیل شده و خوشه دوم با ۲۰ کلیدواژه، بزرگترین خوشه است.

استناد: دولتی لیاولی، فاطمه؛ نورمحمدی امجد، محدثه؛ هاشم، زهرا (۱۴۰۴). مصوّرسازی تولیدات علمی کلان داده‌ها در پایگاه استنادی. *مطالعات کاربردی علم‌سنجی*، ۲ (۱)، ۳۰-۱۴. https://doi.org/10.22091/apss.2025.12399.1032

ناشر: دانشگاه قم.
© نویسندگان.

1. Web Of Science
2. Obdu Univ

۱. مقدمه

علم‌سنجی به مطالعه و تحلیل علمی از جنبه‌های مختلف علمی و پژوهشی اشاره دارد. این علم به ارزیابی و اندازه‌گیری تولیدات علمی، بهره‌وری پژوهشگران و دانشگاه‌ها، و تأثیرگذاری مقالات و پژوهش‌ها می‌پردازد (اسمیت و دیگران[1]، 2021). علم‌سنجی با استفاده از شاخص‌های متنوعی مانند تعداد مقالات، تعداد استنادات، شاخص هرش[2] ، و سایر شاخص‌های استنادی، عملکرد علمی را ارزیابی می‌کند و به سیا ست‌گذاران و محققان در ت صمیم‌گیری‌های علمی کمک می‌کند. اهمیت علم سنجی به دلیل توانایی آن در شنا سایی روندهای پژوهشی، تحلیل تأثیرات علمی، و بهبود کیفیت تحقیقات علمی ا ست (بران[3]، 2019). با علم سنجی، می‌توان به تشخیص پژوهش‌ها با تأثیر بالا، تخصیص منابع بهینه، و تدوین راهبردهای علمی پرداخت (ویلیامز[4]، 2018)

مصوّرسازی به فرایند تبدیل مفاهیم و اطلاعات به نمایش‌های بصری گفته می‌شود. این فرایند در زمینه‌های مختلف از هنر گرفته تا تبلیغات و آموزش، کاربردهای فراوانی دارد. مصوّرسازی با استفاده از تکنیک‌های گرافیکی، تصاویر و نقاشی‌ها، مفاهیم پیچیده را به شکلی ساده و قابل فهم تبدیل می‌کند. همچنین مصوّرسازی داده‌ها فرایند تبدیل داده‌های خام و پیچیده به تصاویر گرافیکی، نمودارها، و گراف ها است که به درک بهتر و تحلیل سریع‌تر داده ها کمک می‌کند. این روش به‌ویژه در علم داده و تحلیل های بزرگ‌داده‌ها اهمیت ویژه‌ای دارد، چرا که به شناسایی الگوها، روندها، و استثناها در داده‌ها کمک می‌کند. اهمیت مصوّرسازی در این است که به انتقال پیام‌ها و اطلاعات به شیوه‌ای جذاب و مؤثر کمک می‌کند و اطلاعات پیچیده را به شکلی ساده و قابل فهم ارائه می‌دهد و تصمیم‌گیری مبتنی بر داده‌ها را تسهیل می‌کند. همچنین، مصوّرسازی به ارتباط بهتر یافته‌ها با مخاطبان غیرمتخصص کمک می‌کند و امکان مقایسه و تحلیل همزمان چندین مجموعه داده را فراهم می‌آورد(داندیس[5] ، 1973; آرنیم[6]، 1969).

کلان‌داده‌ها به مجموعه‌ای از داده‌ها گفته می‌شود که به دلیل حجم، سرعت و تنوع بالا، پردازش و تحلیل آن‌ها با استفاده از روش‌های سنّتی امکان‌پذیر نیست. این داده‌ها از منابع مختلفی مانند شبکه‌های اجتماعی، سنسورها، تراکنش‌های مالی، و دستگاه‌های دیجیتال جمع‌آوری می‌شوند. اهمیت کلان‌داده‌ها در این است که با تحلیل و بررسی این داده‌ها، می‌توان الگوها، روندها، و ارتباطات پنهان را کشف کرد و به تصمیم‌گیری‌های دقیق‌تر و استراتژیک‌تر دست یافت. کلان‌داده‌ها در حوزه‌های مختلف مانند سلامت، تجارت، علوم اجتماعی، و مهند سی به بهبود کارایی، افزایش نوآوری و ارتقاء کیفیت خدمات و مح صولات کمک می‌کنند. با کلان‌داده‌ها، سازمان‌ها و محققان قادر به پیش‌بینی دقیق‌تر رویدادها و بهینه‌سازی فرایندها می‌شوند که این خود به مزیت‌های رقابتی و رشد پایدار منجر می‌شود. به‌طور خلاصه، تحلیل و مصوّرسازی تولیدات علمی کلان‌داده‌ها در پایگاه استنادی وب‌آوساینس می‌تواند به درک بهتر روندهای پژوهشی، شناسایی مفاهیم کلیدی، و روابط میان آن‌ها کمک کند. با استفاده از روش‌های علم‌سنجی و ابزارهای مصوّرسازی، امکان تحلیل دقیق‌تر و بصری‌سازی اطلاعات فراهم می‌شود. این رویکرد می‌تواند به ارتقاء کیفیت پژوهش‌ها و تسهیل تصمیم‌گیری‌های استراتژیک علمی کمک شایانی کند.

1. Smith et al
2. H-Index
3. Brown
4. Williams
5. Dondis A.Donis
6. Rudolf Arnheim

۲. پیشینه پژوهش

پیشینهٔ این پژوهش شامل بررسی مطالعات وپژوهش‌های قبلی نسبتاً مرتبط با موضوع پژوهش حاضر «مصوّرسازی تولیدات علمی مرتبط با کلان داده‌ها در پایگاه استنادی وب‌آوساینس» است.

۲-۱. پیشینهٔ پژوهش در داخل کشور

کاردگر و سعادت‌دار آرانی(۱۴۰۳) پژوهشی با عنوان « مصور سازی تولیدات علمی ایران و ترکیه با کلان داده در پایگاه استنادی وب آو ساینس در بازه زمانی (۲۰۱۰-۲۰۲۲)» انجام دادند. پژوهش آن‌ها با رویکرد علم‌سنجی بوده و برای گردآوری داده ها از پایگاه استنادی وب‌آوساینس استفاده شده است. نتایج نشان داد که تعداد کل رکوردهای استخراج شده ۷۷۴۰ بوده که مهم‌ترین کلیدواژه‌ها «کلان‌داده»، «یادگیری ماشین»، «یادگیری عمیق» و «اینترنت اشیاء[1]» بوده‌اند. همچنین، بیشترین هم‌واژگانی بین«مهندسی برق و الکترونیک»، «علوم کامپیوتر- سیستم‌های اطلاعاتی» و «علوم کامپیوتر، هوش مصنوعی» مشاهده شده است.

عباداله عموقین و خرم آبادی آرانی(۱۴۰۳) در پژوهشی با عنوان «تحلیل و مصورسازی تولیدات علمی محققان ایرانی در رابطه با یادگیری الکترونیکی در پایگاه استنادی وب آو ساینس» به بررسی دقیق این حوزه پرداخته‌اند. پژوهش آن‌ها بر مبنای رویکرد علم‌سنجی انجام شده و نتایج نشان می‌دهد که ایران با ۵۱۷ رکورد اطلاعاتی در زمینهٔ یادگیری الکترونیکی، در جایگاه نخست جهانی قرار دارد. کشورهای آمریکا و مالزی به ترتیب با ۲۴ و ۱۶ رکورد در رده‌های بعدی قرار گرفته‌اند. در میان دانشگاه‌های ایرانی، دانشگاه آزاد اسلامی با ۱۱۹ رکورد اطلاعاتی در صدر قرار دارد، در حالی که دانشگاه تهران با ۷۲ رکورد و دانشگاه تربیت مدرس با ۳۷ رکورد در رده‌های بعدی جای گرفته‌اند. علاوه‌بر این، نتایج پژوهش آن‌ها نشان می‌دهد که واژه‌های کلیدی در هفت خوشهٔ اصلی طبقه‌بندی شده‌اند که خوشهٔ «یادگیری الکترونیکی» دارای مرکزیت بالاتری است. نقشهٔ چگالی نیز تراکم بالای این کلیدواژه‌ها را نشان می‌دهد.

یوسفی و مهربان(۱۴۰۲) پژوهشی با عنوان «مصورسازی داده‌ها و اطلاعات: ابزاری نوین جهت تسهیل خط مشی‌گذاری و مشارکت عمومی» انجام دادند. در پژوهش آن‌ها، اهمیت استفاده از مصوّرسازی داده‌ها در تسهیل خط مشی‌گذاری و افزایش شفافیت اطلاعاتی بررسی شده است. یافته‌ها نشان می‌دهند که مصوّرسازی داده‌ها باعث تسهیل درک مسائل و چالش‌ها می‌شود، همچنین شناسایی الگوها و روندها را ممکن می سازد. علاوه‌بر این نشان دادند این ابزار نه تنها به بهبود تصمیم‌گیری مبتنی بر شواهد کمک می‌کند، بلکه موجب افزایش مشارکت عمومی و هم‌گرایی ذی‌نفعان می‌شود. بررسی‌ها نشان‌دهندهٔ آن است که دولت‌ها به‌طور گسترده‌ای از این ابزار استفاده می‌کنند و سرمایه‌گذاری قابل‌توجهی در این زمینه انجام داده‌اند.

مکرمی پور و همکاران(۱۴۰۱) در پژوهشی با عنوان «مصورسازی تولیدات علمی پژوهشگران ایرانی حوزه علوم سیاسی در پایگاه اطلاعاتی وب آو ساینس» نشان دادند که از مجموعهٔ ۳۴۹٫۲۴۷ مقاله منتشر شده در حوزهٔ علوم سیاسی در پایگاه وب‌آوساینس بین سال‌های ۱۹۹۰تا ۲۰۲۰، تعداد ۳۶۸ مقاله از سوی پژوهشگران ایرانی ارائه شده است. روند رشد تولیدات علمی پژوهشگران ایرانی از سال ۲۰۱۱ آغاز و در سال‌های ۲۰۱۵ و ۲۰۱۶ به اوج خود رسیده است. دانشگاه آزاد اسلامی، دانشگاه تهران، و دانشگاه شهید بهشتی بیشترین میزان تولید مقالات را به خود اختصاص داده‌اند. همچنین،۳۲مقاله از تولیدات علمی پژوهشگران ایرانی در مجلات برتر علوم سیاسی منتشر شده که۴۲۰ استناد دریافت کرده‌اند.

1. Internet of Things (IOT)

رضایی، میر حسینی و سپهر(۱۴۰۰) پژوهشی با عنوان «آینده پژوهی تأثیر کلان داده بر مدیریت و خدمات کتابخانه های عمومی کشور و ارائه مدل راهبردی» انجام داده‌اند که پژوهش آن‌ها از نوع آینده‌پژوهی بوده و از روش دلفی در سه دور انجام شده ، یافته های این پژوهش نشان داد پژوهش‌ها نشان داده‌اند که در حال حا ضر، کتابخانه‌های عمومی به‌طور بسیار محدودی از کلان‌داده‌ها برای بهبود عملکرد و خدمات خود اسـتفاده می‌کنند. همچنین نتایج آماری نشـان می‌دهد که کاربرد کلان‌داده‌ها در وضـعیت فعلی تنها به میزان ۲۰ در صد موجب افزایش سرعت آماده سازی و پردازش داده‌ها شده ا ست. با این حال، پیش‌بینی‌های آماری نشان‌دهندۀ آن است که در آینده، کاربرد کلان‌داده‌ها می‌تواند تا ۷۵ درصـد کیفیت و هدفمندی اطلاع‌رسـانی را افزایش داده و دسـترسـی سـریع‌تر به داده‌ها را فراهم کند.

میرعرب و نوروزی فیروز(۱۳۹۹) در پژوه شی با عنوان «کلان داده و سیا ست گذاری عمومی» نشان دادند که قرن ۲۰ ام به‌طور چشمگیری در حال تغییر و پیشرفت است و ورود فناوری اطلاعات هر چند با پیچیدگی همراه است اما می‌توان با سیاست‌گذاری‌های درست خدمات و کیفیت را افزایش داد. با اینکه در برخی از نظام‌های حکمرانی استفاده از کلان داده‌ها سخت و با ابهامات روبه‌رو بوده اما واضحا استفاده از کلان داده‌ها در حوزۀ سیاست مفید است و در سیاست کلان داده‌ها مزایایی مانند « کشف الگوهای ناشناخته و بهبود تصمیم گیری های دولتی » نهفته است.

۲-۲. پیشینۀ پژوهش در خارج از کشور

لاواله، ماتِ، و تروجیلو [1](۲۰۲۴) در پژوهشـــی باعنوان «مصـــوّرســـازی‌های مبتنی بر نیاز برای تحلیل داده‌های کلان: یک رویکرد مدل‌محور» به بررســـی رویکردی در چارچوب اســـتاندارد معماری مدل‌محور [2] می‌پردازد که هدف آن کمک به کاربران در اســـتخراج مصوّرسازی‌های مناسب برای نیازهای اطلاعاتی آن‌ها است. یافته‌ها نشان می‌دهد که این رویکرد شامل سه مدل مختلف است: مدل نیازمندی‌ها، مدل پروفایل داده، و مدل مصوّرسازی، که با استفاده از یک سری تبدیلات، فاصلۀ بین نیازهای اطلاعاتی و پیاده‌سازی واقعی را پر می‌کند. نتایج به دست آمده از یک مطالعۀ موردی در سازمان جمع‌آوری مالی نشان‌دهندۀ قابلیت‌پذیری این رویکرد است.

مخنی نی گاهر، آرفاعی، و ساسی حیدری[3](۲۰۲۴) در پژوهشی با عنوان «مسائل پژوهشی باز و ابزارها برای مصوّرسازی و تحلیل داده‌های کلان» به بررســـی اهمیت مصـــوّرســـازی داده‌های کلان در عصـــر دیجیتال می‌پردازد و یافته‌های آن نشـــان می‌دهد که مصوّرسازی اطلاعات به سازمان‌ها کمک می‌کند تا تصمیمات بهتری بگیرند. با استفاده از تکنیک‌های مصوّرسازی، شرکت‌ها قادرند اطلاعات پنهان را شنا سایی و فعالیت‌های خود را به‌طور مؤثرتری مدیریت کنند. نتایج نشان می‌دهد که ابزارهای سنّتی مصوّر سازی نمی‌توانند حجم و تنوع بالای داده‌ها را مدیریت کنند و نیاز به توسـعۀ ابزارهای جدید برای پاسـخ‌گویی به چالش‌های موجود احسـاس می‌شود.

کومار، کومار، و سود [4] (۲۰۲۰) در پژوه شی با عنوان «م صوّر سازی داده بر ا ساس کلان داده‌ها» به برر سی م صوّر سازی کلان داده‌ها پرداخته‌اند. در پژوهش آن‌ها م شخص شد که روش‌های سنّتی م صوّر سازی داده‌ها نمی‌توانند با حجم و سرعت بالای داده‌ها همگام شوند. یافته‌های پژوهش آن‌ها نشان می‌دهد که ا ستفاده از ابزارهای متن‌باز م صوّر سازی داده‌ها می‌تواند پا سخ‌گوی نیازهای کاربران باشـد و ویژگی‌هایی مانند دسـترسـی به داده‌های واقعی و محیط‌های نرم‌افزاری انعطاف‌پذیر را فراهم کند. این ابزارها پس از

1. lavlle, Ana; Mate, Alejandro; Trujillo, Juan
2. Model-driven architecture (MDA)
3. Mkhinini Gahar, Rania; Arfaoui, Olfa; Sassi Hidri
4. Kumar, Nishant; Kumar, Abhijit; Soo

بررسی مقادیر زیادی داده، نتایج قابل اعتماد و کارآمدی ارائه می‌دهند. مصوّرسازی داده‌های بزرگ، فرصتی بی‌نظیر برای تبدیل اطلاعات به ارزش و مدیریت حجم بزرگ داده‌ها ارائه می‌دهد.

آلام[1] (۲۰۱۷) در پژوهشی با عنوان «مطالعهٔ اکتشافی برای مصوّرسازی داده‌های بزرگ در اینترنت اشیاء» نشان داد که از تحلیل پیشینه‌های پژوهش در حوزهٔ اینترنت اشیاء و کلان داده می‌توان نتیجه‌گیری کرد که خدمات اینترنت اشیاء به‌طور چشمگیری تولید مجموعه‌های دادهٔ بزرگ و پیچیده را افزایش داده است. تکنیک‌های کلان داده، به‌ویژه در مصوّرسازی داده‌ها، نقش مهمی در تجزیه و تحلیل این داده‌ها ایفا و به شناسایی روندها و ارائهٔ دیدگاه‌های مفید کمک می‌کنند. یافته‌های پژوهش او نشان می‌دهد که مصوّرسازی داده‌های بزرگ در فرایندهای اینترنت اشیاء، امکان تصمیم‌گیری‌های استراتژیک را فراهم کرده و نقش کلان داده در مصورسازی اینترنت اشیاء و تحلیل بصری این داده‌ها بسیار اهمیت دارد.

اولشانیکوا و همکاران[2](۲۰۱۶) در مقاله‌ای با عنوان «مصورسازی کلان داده‌ها» به بررسی چالش‌های پردازش، تحلیل، و تجزیهٔ کلان داده‌ها پرداخته‌اند. این چالش‌ها شامل حجم بالای داده، تنوع آن‌ها، و تولید سریع اطلاعات است که بر دقت و کارایی تحلیل داده تأثیر دارند. در پژوهش آن‌ها، روش‌های مختلف مصوّرسازی داده‌های کلان طبقه‌بندی شده و تأکید شده که مصوّرسازی مناسب می‌تواند به درک بهتر داده‌ها کمک کند. همچنین، به تمایل روزافزون به استفاده از ابزارهای مبتنی بر تجسم برای پشتیبانی از کسب‌وکار و سایر زمینه‌های مهم اشاره شده است. یافته‌های پژوهش آن‌ها نشان می‌دهد که انتخاب ابزارهای مناسب مصوّرسازی می‌تواند تأثیر قابل توجهی بر کیفیت تحلیل داده‌ها داشته باشد.

نایر، شتی و شتی[3] (۲۰۱۶) در پژوهشی با عنوان «تجزیه و تحلیل تصویری تعاملی روی داده‌های بزرگ: TABLEAU در مقابل D3.JS» به بررسی اثربخشی مصوّرسازی داده‌های کلان با استفاده از دو ابزار محبوب، D3.js و Tableau، پرداخته‌اند. تحلیل بصری مجموعه داده‌های جرایم شیکاگو، نشان داد که هر دو ابزار قادر به آشکارسازی الگوهای مهمی هستند. همچنین، نتایج مقایسهٔ این دو ابزار مشخص کرد Tableau از نظر جذابیت بصری و سهولت استفاده برتری دارد، در حالی‌که D3.js به دلیل قابلیت‌های سفارشی سازی بیشتر نیاز به ارتقاءهای بیشتری دارد تا بتواند به‌طور کامل پاسخ‌گوی چالش‌های مصوّرسازی داده‌های کلان باشد. این نتایج بر اهمیت انتخاب هوشمندانهٔ ابزارهای مصوّرسازی بر اساس نیازهای خاص پروژه تأکید می‌کند.

با توجه به مطالعات انجام شده در داخل و خارج از کشور، پژوهش های فراوانی دربارهٔ کلان داده‌ها به‌مثابه یک خط‌مشی کلیدی برای مقابله با تغییر روزافزون و سریع علم‌سنجی تأکید دارد. این مطالعات به بررسی رویکردهای مختلف در این حوزه پرداخته‌اند. همچنین کلان داده‌ها در کسب‌وکارها تأثیرات قابل‌توجهی گذاشته‌اند. به‌طور کلی این پژوهش‌ها بر ضرورت توجه به کلان داده‌ها در سازمان‌ها تأکید دارند.

۳. روش پژوهش

این پژوهش از نوع کاربردی است و می‌توان مصوّرسازی تولیدات علمی آن را نوعی علم‌سنجی دانست. ارزیابی‌های علم‌سنجی معمولا در سازمان‌های اجتماعی مانند دانشگاه‌ها، مؤسسات علمی، انجمن‌ها، و کشورهای حوزهٔ اصلی علمی و فرعی انجام می‌شود. در این

1. Allam, Sudhir
2. Olshannikova, Ekaterina; Ometov, Aleksandr; Olsson, Thomas; Koucheryavy
3. Lekha R Nair ،Sujala D Shetty Siddhanth D Shetty

پژوهش ا ستخراج داده‌ها از وب‌آو ساینس انجام شده و کلیدواژهٔ کلان داده[1] در آن ج ست‌وجو شده ا ست. ۱۰۵۳ مدرک بازیابی شد. بازیابی منابع در تاریخ ۱۴۰۳/۰۹/۰۵ انجام شـده اسـت. برای این پژوهش ترسـیم خوشـه و نقشـه راه مورد نظر اسـت. داده‌ها پس از یکسان‌سازی وارد نرم افزارهای ووس‌ویور، بایب اکسل، و یوسی‌نت شده و نقشه‌های مورد نظر رسم شد. ووس‌ویور برای ترسیم نقشه، بایب اکسل برای مشخص شدن هم‌رخدادی واژگان و یوسی نت برای ارتباط بین کلیدواژه‌ها استفاده شده است.

۴. هدف اصلی پژوهش

هدف اصـلی این پژوهش تحلیل و مصـوّرسـازی تولیدات علمی مرتبط با کلان داده‌ها با اسـتفاده از داده‌های پایگاه اسـتنادی وب‌آوساینس، است.

اهداف فرعی

معرفی نویسندگان پُرکار در حوزهٔ موضوعی کلان داده‌ها

معرفی مجلات پُرتولید در زمینهٔ کلان داده‌ها

معرفی سال‌ها و موضوعات پُرتولید در زمینهٔ کلان داده‌ها

بررسی همکاری‌های علمی نویسندگان ایرانی با دانشگاه‌ها و کشورهای دیگر

شناسایی و تحلیل پُرتکرارترین کلیدواژه‌ها در تولیدات علمی ایران مرتبط با کلان داده‌ها

خوشه‌بندی کلیدواژه‌ها و موضوعات مرتبط با چهار خوشهٔ اصلی

۵. پرسش‌های پژوهش

هدف از این پژوهش مصوّرسازی تولیدات علمی کلان داده‌ها و شناسایی نویسندگان، مجلات هسته، و کشورهای پُرکار در این حوزه است. در این راستا، پرسش‌های پژوهش عبارتند از:

پرسش ۱: بیشترین تولیدات علمی ایرانیان در حوزهٔ کلان داده، مربوط به کدام نویسندگان، مجلات، چه سالی، و چه موضوعی است؟

پرسش ۲: نویسندگان ایرانی مرتبط با کلان داده، بیشتر با چه دانشگاه‌ها و کشورهایی همکاری علمی داشته‌اند؟

پرسش ۳: پُرتکرارترین کلیدواژهٔ مرتبط با نویسندگان و کلان داده، کدامند؟

پرسش ۴: کلیدواژه‌ها از چند خوشه تشکیل شده و هر خوشه شامل چه موضوعاتی است؟

۶. یافته‌ها

۱ . پرسش اول: بیشترین مقالات نمایه‌شده در حوزهٔ کلان داده‌ها مربوط به کدام نویسندگان، مجلات، سال‌ها و موضوعاتی است؟

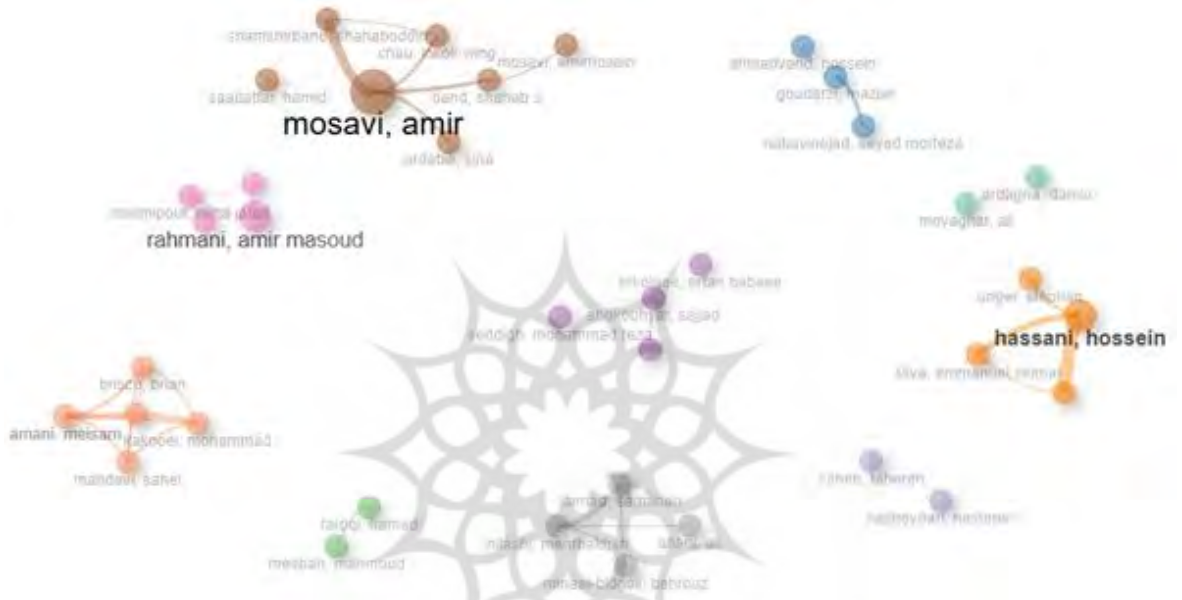
جدول شماره ۱. تولیدات مربوط به نویسندگان

ردیف	نویسنده	تعداد تولیدات	درصد از کل
۱	Mosavi A	48	4.558%
۲	Rahmani Am	31	2.944%
۳	Hassani H	22	2.089%

۱. کلان داده

۴	Shokouhyar S	18	1.709%
۵	Goudarzi M	16	1.519%
۶	Navimipour Nj	15	1.425%
۷	Shamshirband S	15	1.425%
۸	Khosravi Mr	13	1.235%
۹	Huang X	12	1.140%

جدول شماره ۱ نویسندگان حوزهٔ کلان داده را نشان می‌دهد که امیر موسوی با ۴۸ رکورد اطلاعاتی پیشتاز است. همچنین از میان دیگر نویسندگان رحمانی، موسوی، و شکوهیار به ترتیب در جایگاه دوم تا چهارم قرار دارند.



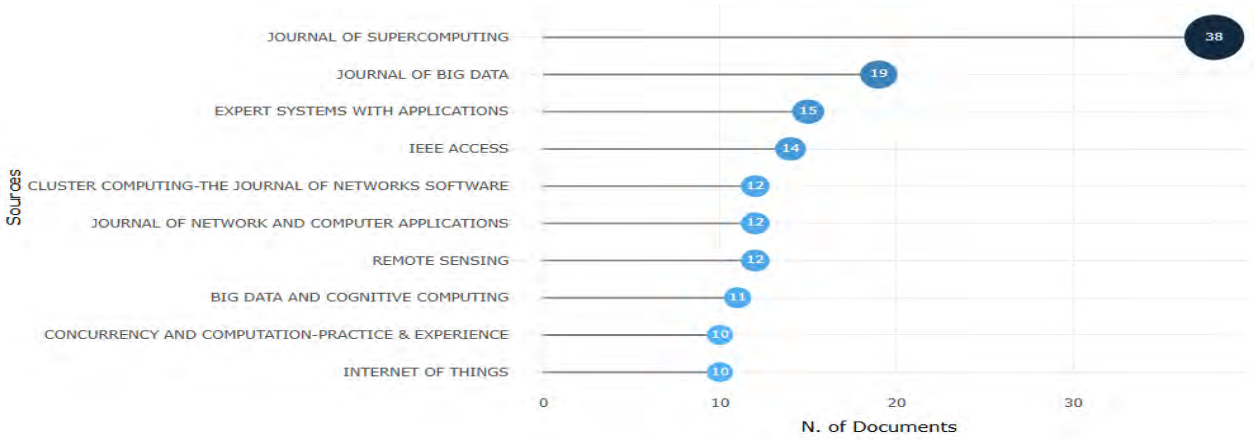
شکل ۱. نمودار هم‌نویسندگی

جدول شماره۲. تولیدات مربوط به مجلات

ردیف	مجلات منتشر شده	تعداد تولیدات	درصد از کل
۱	Journal Of Supercomputing	38	3.609%
۲	Journal Of Big Data	19	1.804%
۳	Expert Systems With Applications	15	1.425%
۴	Ieee Access	14	1.330%
۵	Cluster Computing Of Network Software Tools And Applications	12	1.140%
۶	Journal Of Network And Computer Applications	12	1.140%
۷	Remote Sensing	12	1.140%

جدول شماره ۲ مجلات منتشر شده در این حوزه را نشان می‌دهد. Journal Of Supercomputing با ۳۸ تولید ۳٫۶ درصد در صدر جدول قرار دارد.

همچنین Journal Of Big Data با ۱۹ تولید علمی و کسب ۱٫۸ درصد در جایگاه دوم قرار دارد. Expert Systems With Applications با تولید ۱۵ رکورد اطلاعاتی و ۱٫۴ در جایگاه سوم و Leee Access با ۱۴ تولید علمی و ۱٫۳ در جایگاه چهارم قرار دارد.



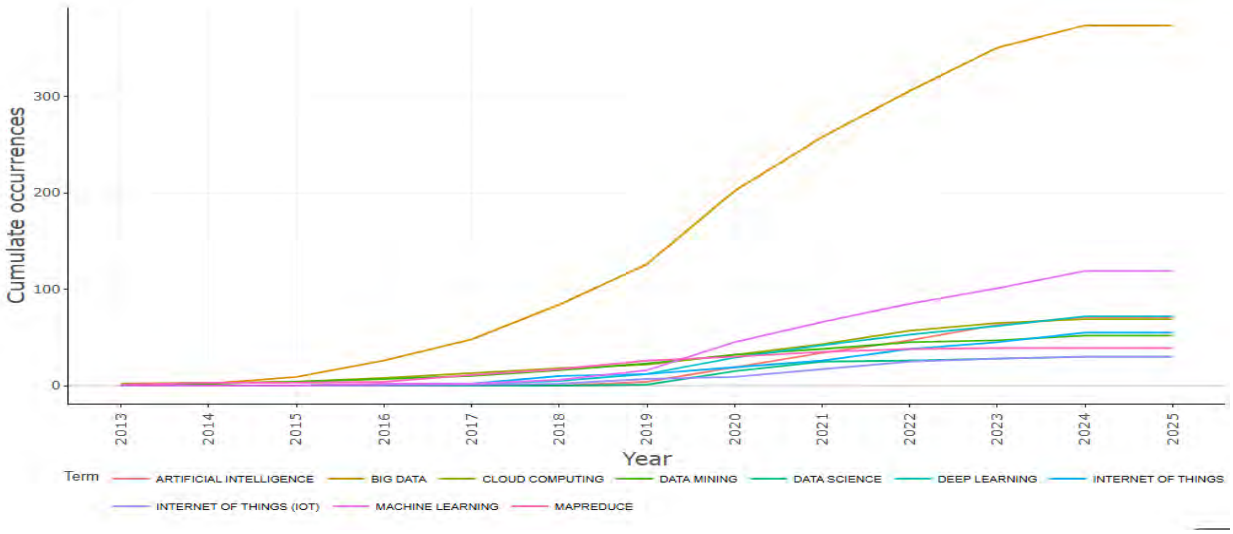
شکل ۲. مجلات با بیشترین فراوانی

۳. بیشترین تولیدات علمی ایران مرتبط با کلان داده، مربوط به کدام سال است؟

جدول شماره ۳. اطلاعات سال نشر تولیدات

ردیف	سال انتشار	تعداد تولیدات	درصد از کل
۱	2024	96	9.117%
۲	2023	125	11.871%
۳	2022	158	15.005%
۴	2021	194	18.424%
۵	2020	179	16.999%
۶	2019	113	10.731%
۷	2018	86	8.167%

سال ۲۰۲۱ در زمینهٔ کلان داده، به تعداد ۱۹۴ رکورد اطلاعاتی تولید شده که ۱۸٫۴% از کل تولیدات است. سال ۲۰۲۰ با ۱۷۹ رکورد اطلاعاتی و ۱۶٫۹ درصد جایگاه دوم، سال ۲۰۲۲با ۱۵۸ رکورد اطلاعاتی و ۱۵٫۰ درصد از کل تولیدات و سال ۲۰۲۳ با ۱۲۵ رکورد و ۱۱٫۸ درصد به ترتیب جایگاه دوم و سوم و چهارم را به خود اختصاص دادند.



شکل۳. روند کلمات کلیدی در زمان

جدول شماره۴. حوزهٔ موضوعی تولیدات

ردیف	حوزه موضوعی	تعداد تولیدات	درصد از کل
۱	Computer Science	۵۲۱	%۴۹٫۴۷۸
۲	Engineering	۳۵۲	%۳۳٫۴۲۸
۳	Telecommunications	۱۰۸	%۱۰٫۲۵۶
۴	Business Economics	۸۴	%۷٫۹۷۷
۵	Environmental Sciences Ecology	۶۷	%۶٫۳۶۳
۶	Science Technology Other Topics	۶۵	%۶٫۱۷۳
۷	Operations Research Management Science	۳۹	%۳٫۷۰۴
۸	Mathematics	۳۸	%۳٫۶۰۹

در تولیدات علمی ایرانیان ۵۲۱ تولید مربوط به حوزهٔ علوم رایانه[1] است. همچنین علوم مهندسی[2] با ۳۵۲ تولید علمی در رتبهٔ دوم، موضوع مخابرات[3] با ۱۰۸ تولید علمی در رتبه سوم و اقتصاد بازرگانی[4] با ۸۴ تولید علمی درجایگاه چهارم قرار دارد.

جدول شماره ۵. تولیدات دانشگاه‌ها

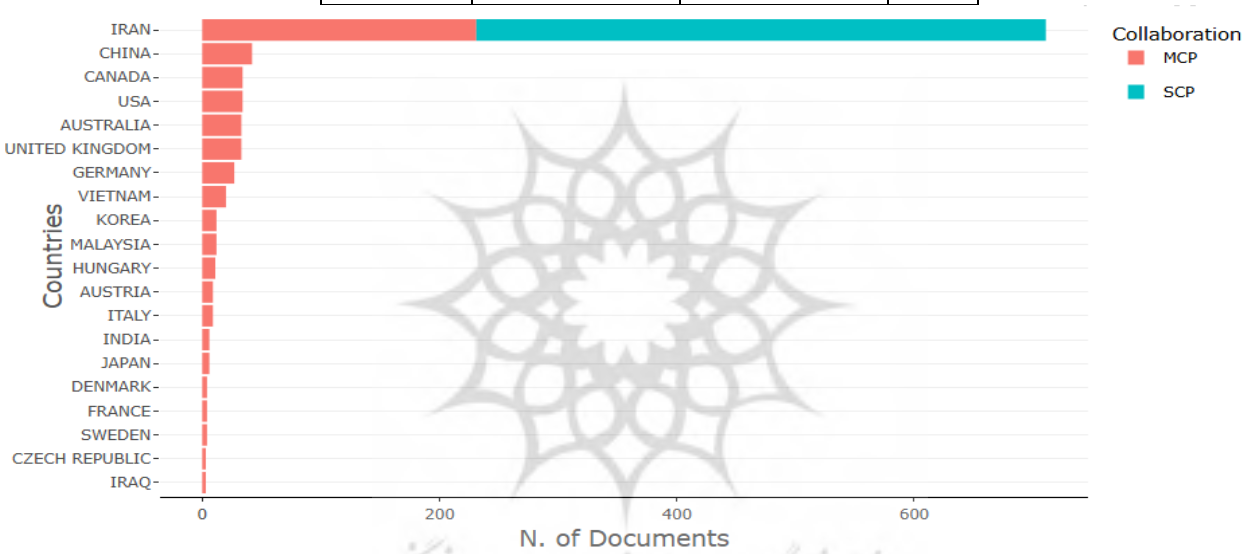
ردیف	دانشگاه	تعداد تولیدات	درصد از کل
1	Islamic Azad University	240	22.792%
2	University Of Tehran	126	11.966%
3	Sharif University Of Technology	66	6.268%
4	Amir kabir University Of Technology	58	5.508%
5	Tarbiat Modares University	57	5.413%
6	Iran University Science Technology	52	4.938%
7	Shahid Beheshti University	49	4.653%

1. Computer Science
2. Engineering
3. Telecommunication
4. Business Economics

در بین دانشگاه‌های ایران دانشگاه آزاد اسلامی با ۲۴۰ رکورد علمی از کل تولیدات اولین دانشگاه ایران است. دانشگاه تهران، شریف، و امیرکبیر به ترتیب با ۱۲۶، ۶۶، ۵۸ و کسب ۱۱/۹، ۶/۲، ۵/۵ رتبه های دوم، سوم و چهارم را در اختیار دارند.

جدول شماره۶. تولیدات کشورها

ردیف	کشور	تعداد تولیدات	درصد از کل
۱.	**ایران**	1053	100.000%
۲.	**آمریکا**	139	13.200%
۳.	**استرالیا**	105	9.972%
۴.	**انگلستان**	94	8.927%
۵.	**کانادا**	78	7.407%
۶.	**آلمان**	71	6.743%
۷.	**چین**	65	6.173%
۸.	**مجارستان**	48	4.558%



شکل ۴. کشورهای با بیشترین تولید علمی

جدول شماره ۶ نشان می‌دهد که بیشترین تولیدات علمی نویسندگان در رابطه با کلان داده‌ها مربوط به کشور ایران با ۱۰۵۳ رکورد و ۱۰۰ درصد تولیدات بوده است.

کشور امریکا با ۱۳۹ رکورد، در رتبهٔ دوم، کشور استرالیا با ۱۰۵ رکورد، در رتبهٔ سوم و کشور انگلیس ۹۴ رکورد و در رتبه چهارم قرار دارد.

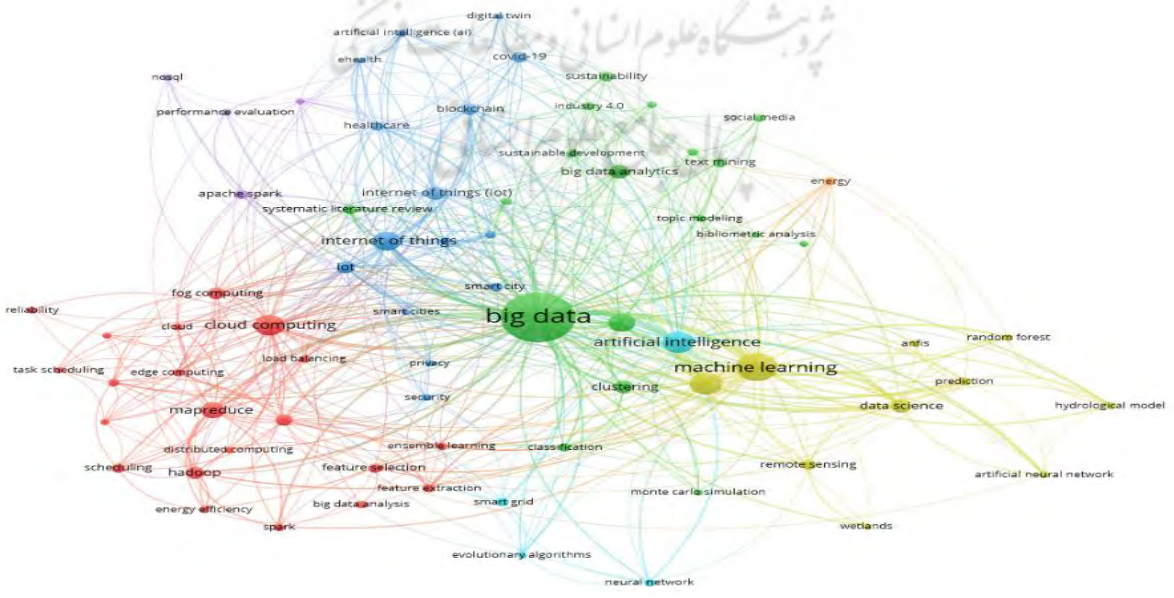
پرسش سوم: پرتکرارترین کلیدواژه‌های حوزهٔ کلان داده کدامند؟

Words	Occurrences
big data	374
machine learning	119
deep learning	72
artificial intelligence	71
cloud computing	69
internet of things	55

data mining	52
Mapreduce	39
data science	30
internet of things (iot)	30
big data analytics	28
Clustering	26

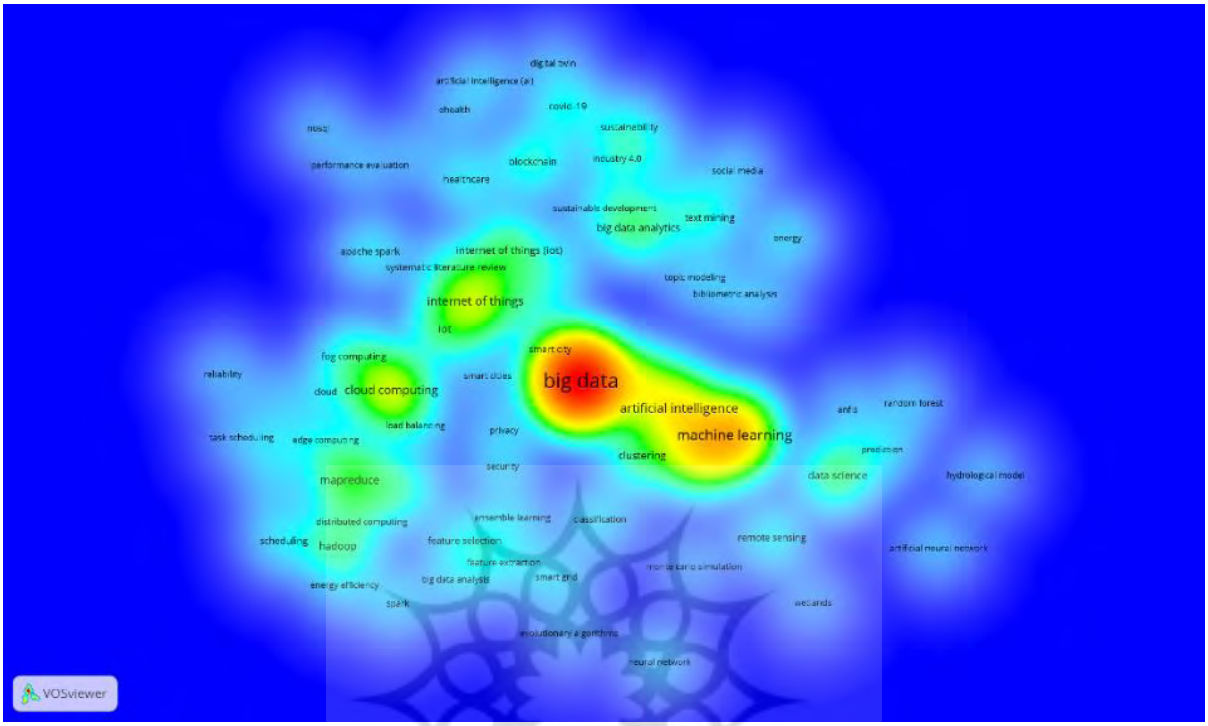
جدول شماره۷. واژگان کلیدی

شکل ۵. ابر واژگان کلیدی



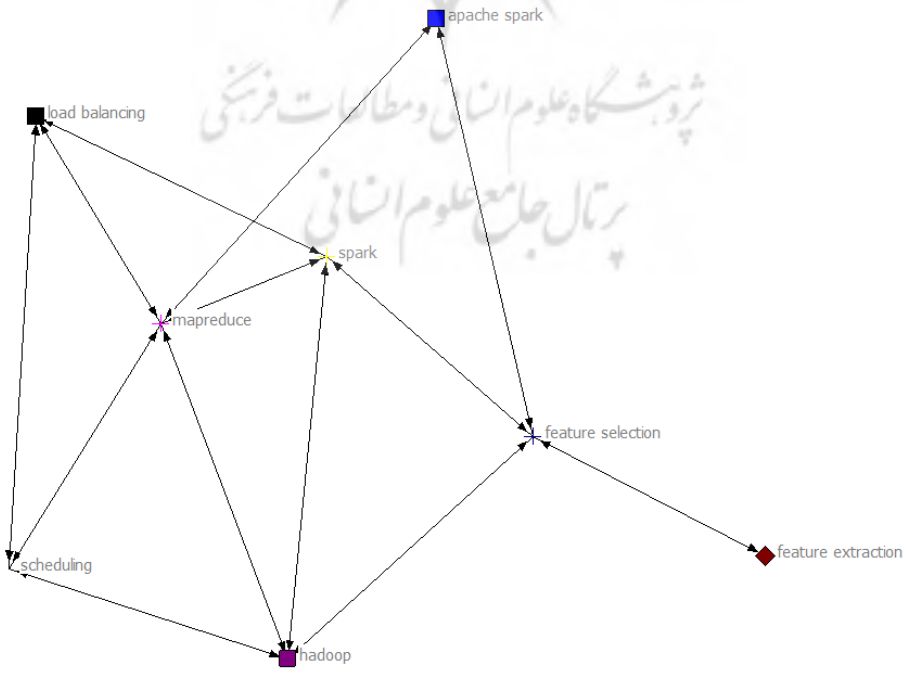
شکل ۶. شبکه هم‌رخداد واژگان

تصویر شماره ۱، نقشهٔ مفاهیم حوزهٔ کلان داده‌ها، تحلیل کلیدواژه‌های مدارک را نشان می‌دهد. بزرگی دایره‌ها نشان‌دهندهٔ کاربرد بیشتر آن مفاهیم یا کلیدواژه‌ها در توصیف مدارک است. رنگ دایره‌ها نیز نشان‌دهندهٔ خوشه‌های مفاهیم است. همان‌طور که در نمودار مشاهده می‌شود، رنگ آبی شامل کلیدواژه‌ای مانند big data است.



شکل ۷. نمودار حرارتی واژگان

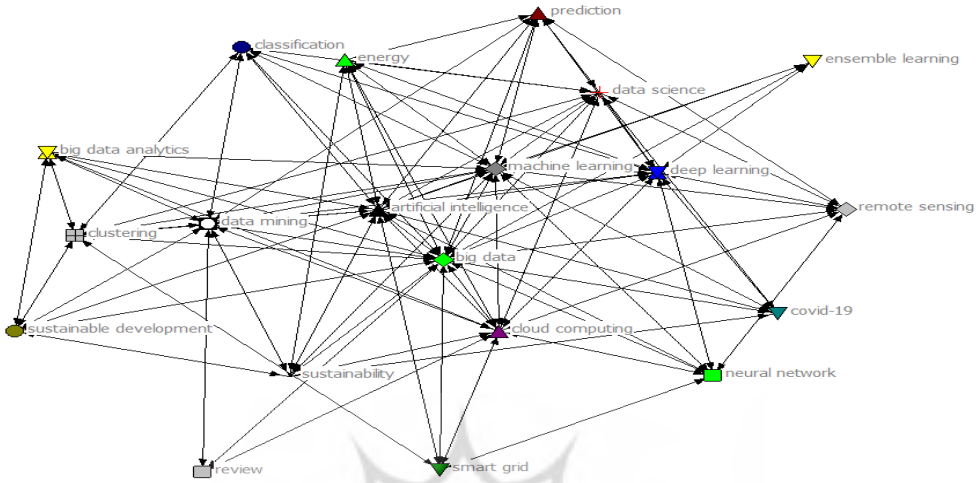
خوشهٔ یک:



شکل۸. خوشه یک

این خوشه به تحلیل و مصوّرسازی تولیدات علمی در زمینه‌های پردازش کلان داده‌ها و یادگیری ماشینی می‌پردازد. مفاهیم کلیدی آن شامل آپاچی اسپارک[1]، مپریدیوس[2]، هادوپ[3] انتخاب و استخراج ویژگی‌ها، تعادل بار، و زمان‌بندی است. این مفاهیم به بهبود پردازش داده‌های بزرگ و بهینه‌سازی کارایی سیستم‌های توزیع‌شده کمک می‌کنند.

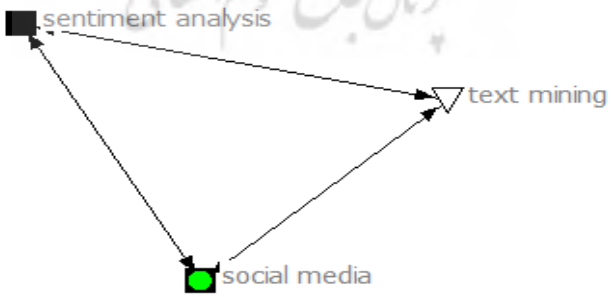
خوشهٔ دو:



شکل ۹. خوشه دو

خوشهٔ شماره دو به تحلیل و مصوّرسازی تولیدات علمی در زمینه‌های کلان‌داده‌ها، شبکه‌های عصبی، و هوش مصنوعی می‌پردازد. این خوشه شامل مفاهیمی چون مصوّرسازی تولیدات علمی در زمینه‌های کلان‌داده‌ها، پردازش موازی، و یادگیری ماشینی تحلیل کلان‌داده‌ها، محاسبات ابری، و تکنیک‌های مصورسازی داده‌ها است. مفاهیم کلیدی آن شامل آپاچی اسپارک برای پردازش داده‌ها، مپریدیوس به‌عنوان مدل برنامه‌نویسی، هادوپ برای ذخیره‌سازی و پردازش داده‌ها، انتخاب ویژگی‌ها و استخراج ویژگی‌ها برای بهینه‌سازی مدل‌های یادگیری ماشینی، و تعادل بار، و زمان‌بندی برای بهینه‌سازی کارایی سیستم‌های توزیع‌شده است. این مفاهیم با هم به بهبود پردازش داده‌های بزرگ و توسع الگوریتم‌های هوش مصنوعی کمک می‌کنند.

خوشهٔ سه:

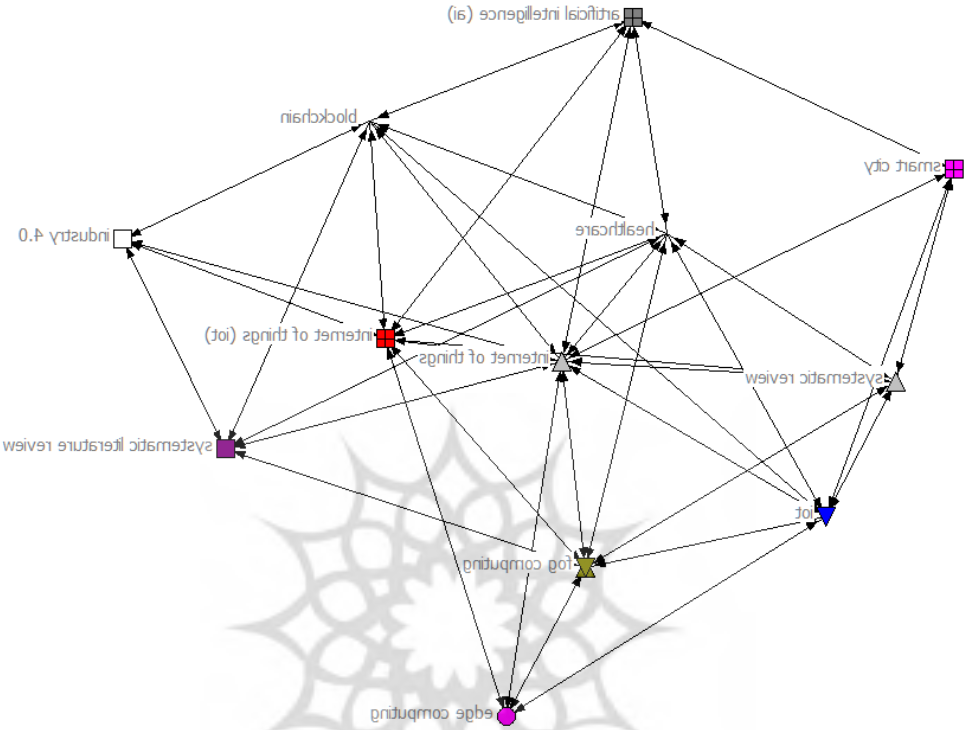


شکل ۱۰. خوشه سه

1. Apache Spark
2. MapReduce
3. Hadoop

خوشهٔ سوم به تحلیل و مصوّرسازی تولیدات علمی در زمینهٔ تحلیل احساسات، متن‌کاوی، و رسانه‌های اجتماعی پرداخته است. مفاهیم کلیدی آن شامل تحلیل احساسات، استخراج متن، و استفاده از داده‌های رسانه‌های اجتماعی است. این مفاهیم با هم به بررسی و بهبود تحلیل داده‌ها و درک بهتر تعاملات اجتماعی کمک می‌کنند.

خوشهٔ چهار:



شکل ۱۱. خوشه چهار

خوشهٔ چهارم به تحلیل و مصوّرسازی تولیدات علمی در زمینهٔ تحلیل احساسات، متن‌کاوی، و رسانه‌های اجتماعی پرداخته است. این خوشه شامل مفاهیمی چون تحلیل احساسات، متن‌کاوی، و استفاده از داده‌های رسانه‌های اجتماعی است. این مفاهیم به بررسی تعاملات اجتماعی و بهبود تحلیل داده‌ها کمک می‌کنند.

۷. نتیجه گیری

نتایج پژوهش علم‌سنجی نشان داد که ۱۰۵۳ مدرک در حوزه موضوعی کلان داده‌ها در پایگاه استنادی وب‌آوساینس نمایه‌سازی شده است. موسوی با ۴۸ مدرک پُرکارترین نویسنده است. یافته‌های پژوهش نشان داد مجله Journal Of Supercomputing در صدر مجلات این حوزه قرار دارد. سال ۲۰۲۱با ۱۹۴ تولید علمی دربارهٔ کلان داده پربارترین سال بوده است. این در حالی است که در سال ۲۰۲۴ تنها ۹۶ مدرک در این باره تولید شده است. حوزهٔ موضوعی Computer Science بیشترین مدرک را در حوزه کلان داده‌ها به خود اختصاص داده است.در بین دانشگاه‌های ایران دانشگاه آزاد اسلامی با ۲۴۰ رکورد اطلاعاتی بیشترین تولیدات را به خود اختصاص داده است. بیشترین تولیدات علمی نویسندگان در رابطه با کلان داده‌ها از کشور ایران است، کشورهای امریکا، استرالیا، و انگلستان در رده‌های بعدی قرار دارند. در این پژوهش واژهٔ کلان‌داده از خوشهٔ ۲ با ۳۷۴ تکرار پربسامدترین کلیدواژه است. یافته‌های

این خوشـــه نشـــان داد که به جز این واژه کلماتی مانند machine learning، deep learning، artificial intelligence و internet of thing نیز واژه‌های پرتکرار بوده‌اند و به آن‌ها توجه شده است. برای یافته‌های این پژوهش ۱۰۵۳ مدرک بازیابی شد.

نتایج پژوهش علم‌سنجی در حوزهٔ کلان‌داده نشان‌دهندهٔ روندهای قابل توجهی در تولید علمی و فعالیت‌های پژوهشی است. این یافته‌ها می‌تواند به محققان و تصمیم‌گیرندگان کمک کند تا روندها و نیازهای پژوهشی را در حوزهٔ کلان‌داده بهتر درک کنند و به توسعهٔ بیشتر این حوزه بپردازند.

سپاس‌گذاری

از همکاری و کمک‌های ارزشمند استاد محمدرضا نصیری در ساخت نقشه‌ها و خوشه‌های این پژوهش، صمیمانه تشکر و قدردانی می‌کنیم.

۸. منابع

رضایی، ز.؛ میرحسینی، ز.؛ سپهر، ف. (۱۴۰۰). آینده پژوهی تأثیر کلان داده بر مدیریت و خدمات کتابخانه های عمومی کشور و ارائه مدل راهبردی. علوم و فنون مدیریت اطلاعات، ۷(۲)، ۸۱-۱۱۰.

عباداله عموقین، ج.؛ خرم آبادی آرانی، م. (۱۴۰۳). تحلیل و مصورسازی تولیدات علمی محققان ایرانی در رابطه با یادگیری الکترونیکی در پایگاه استنادی وب آو ساینس. پژوهش در روش‌های آموزش، ۲(۱)، ۱۱۹-۱۴۶.

کاردگر، د.؛ سعادت دار ارانی، م. (۱۴۰۳). مصور سازی تولیدات علمی ایران و ترکیه مرتبط با *Big Data* در پایگاه استنادی وب‌آوساینس در بازه زمانی (۲۰۱۰-۲۰۲۲). مطالعات کاربردی علم‌سنجی (مقاله آماده انتشار).

مکرمی پور، م.؛ فهیمی فر، س.؛ نقدی نژاد، ا.؛ محمدی، ه. (۱۴۰۱). مصورسازی تولیدات علمی پژوهشگران ایرانی حوزه علوم سیاسی در پایگاه اطلاعاتی وب آو ساینس. جامعه‌شناسی سیاسی ایران، ۵(۱۱)، ۲۹۴۶-۲۹۶۵.

میرعرب، ع.؛ نوروزی فیروز، ر. (۱۴۰۰). کلان داده و سیاست گذاری عمومی. فصلنامه علوم سیاسی، ۲۴(۹۳)، ۱۵۱-۱۷۲.

یوسفی، ع.؛ مهربان، م. (۱۴۰۲). مصور سازی داده ها و اطلاعات: ابزاری نوین جهت خط مشی گذاری و مشارکت عمومی. مرکز پژوهش های مجلس شورای اسلامی (گزارش‌های کارشناسی).

Allam, S. (2017). Exploratory Study for Big Data Visualization in the Internet of Things .*International Journal of Creative Research Thoughts* (IJCRT).(۴)۵ ,

Brown, P. (2019). Measuring Scientific Impact: Scientometric Approaches .*Science Analysis Journal*.

Kumar, N.; Kumar, A.; Soo, I. (2020). Data Visualization Based On Big Data .*IOP Conference Series: Materials Science and Engineering* 993(1), 012068.

Lavalle, A.; Mate, A.; Trujillo, J. (2024). Requirements-Driven Visualizations for Big Data Analytics: a Model-Driven approach .*arXiv preprint arXiv*:2402.07914.

Mkhinini Gahar, R.; Arfaoui, O.; Sassi Hidri, M. (2024). Open Research Issues and Tools for Visualization and Big Data Analytics .*Procedia Computer Science*, 232, 803-812.

Nair, L. R.; Shetty, S. D.; Shetty, S. (2016). Interactive Visual Analytics On Big Data: Tableau Vs D3.JS .*Journal of e-Learning and Knowledge Society*, 12(4), 139-150.

Olshannikova, E.; Ometov, A.; Olsson, T.; Koucheryavy, Y. (2016). Visualizing Big Data. *In Big Data Technologies and Applications* (pp. 101-131). Springer.

Smith, J., et al. (2021). The Role of Scientometrics in Evaluating Research Productivity .*Journal of Research Evaluation*.

Williams, R. (2018). The Importance of Scientometrics in Academic Research .*Journal of Academic Insights*.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی