Arch

JCWR2020

کنفرانس بین المللی وب پژوهی
ایران، تهران، ۲۲ خرداد ۱۳۹۹

Scientific Map of Data Mining Related Articles in CIVILICA Database Based on Co-Word Analysis

Fateme Shahrabi, Babak Teimourpour*, Mina Ghasemi, Meysam Alavi

Information Technology department, Tarbiat Modares University, Tehran, Iran
{fateme.shahrabi,b.teimourpour, minaghasemi, meysamalavi}@modares.ac.ir

Abstract

Today, due to the large volume of data and the high speed of its production, data analysis using traditional methods is infeasible. Data mining, as one of the most popular topics in the present century, has contributed to the advancement of science and technology in a significant number of fields. In recent years, researchers have used data mining extensively for data analysis. One of the important issues for researchers in this field is to identify common areas in the field of data mining and to find areas of active research in this field for future research. On the other hand, the analysis of social networks in recent years as an appropriate tool for examining the present and future relationships between the entities of a network structure has been used by scholars of various sciences to analyze and draw a scientific map of an area of science. In this paper, by using Co-Word Analysis and Social Network Analysis, the map and scientific structure of data mining topics in Iran based on articles indexed during the years 2009 to 2019 in the Civilica database are analyzed and the topic trends in the research field has been discussed. The results of the analysis show that in terms of data mining, concepts such as clustering, decision tree and neural network are the more important topics.

Keywords: Data Mining, Co-Word Analysis, Scientific Map, Social Networks Analysis

www.SID.ir

Arch

JCWR2020

کنفرانس بین المللی وب پژوهی
ایران، تهران، ۲۲ خرداد ۱۳۹۹

ترسیم نقشهٔ علمی مقالات مرتبط با حوزه داده‌کاوی در پایگاه داده سیویلیکا مبتنی بر تحلیل هم‌رخدادی واژگان

فاطمه شهرابی فراهانی[۱]، بابک تیمورپور[۲*]، مینا قاسمی[۳]، سیدمیثم علوی[۳]

[۱] دپارتمان مهندسی فناوری اطلاعات دانشگاه تربیت مدرس
fateme.shahrabi@modares.ac.ir

[۲] استادیاردانشکده مهندسی صنایع، گروه مهندسی فناوری اطلاعات دانشگاه تربیت مدرس
b.teimourpour@modares.ac.ir

[۳] دپارتمان مهندسی فناوری اطلاعات دانشگاه تربیت مدرس
{minaghasemi, meysamalavi}@modares.ac.ir

چکیده

امروزه با توجه به حجم زیاد داده‌ها و سرعت بالای تولید آن، تحلیل داده‌ها با استفاده از روش‌های سنتی عملاً امکان پذیر نیست. در این میان داده‌کاوی به عنوان یکی از موضوعات پر طرفدار در قرن حاضر، در تعداد قابل توجهی از زمینه‌ها به پیشرفت علم و فناوری کمک کرده‌است. در سال‌های اخیرمحققین بطور گسترده از داده‌کاوی برای تحلیل داده ها استفاده کرده‌اند. یکی از مسائل قابل توجه پژوهشگران این حوزه، شناخت زمینه های مشترک در زیرحوزه های مقوله داده‌کاوی و یافتن حوزه‌های پژوهشی فعال در این زمینه جهت پژوهش‌های آتی است. از سوی دیگر تحلیل شبکه های اجتماعی در سال‌های اخیر به عنوان ابزاری مناسب جهت بررسی روابط حال و آینده بین موجودیت‌های یک ساختار شبکه‌ای نظر محققین علوم مختلف را برای تحلیل این روابط و ترسیم نقشهٔ علمی یک حوزه از علم به خود معطوف کرده‌است. در این مقاله با استفاده از روش تحلیل هم‌رخدادی واژگان و تحلیل شبکه‌های اجتماعی، نقشه و ساختار علمی موضوعات حوزه داده‌کاوی در ایران بر اساس مقالات نمایه شده طی سال‌های ۱۳۸۸ تا۱۳۹۸ در پایگاه داده سیویلیکا ترسیم شده و روند موضوعی حاکم بر پژوهش‌های این حوزه بررسی شده است. نتایج تحلیل نشان می‌دهد که در مقولهٔ داده‌کاوی مفاهیمی نظیر خوشه‌بندی، درخت تصمیم و شبکهٔ‌عصبی در بر گیرندهٔ بیشترین حجم از زیربخش‌های حوزهٔ داده کاوی می-باشند.

کلمات کلیدی

داده‌کاوی، تحلیل هم‌رخدادی واژگان، نقشهٔ علم، تحلیل شبکه‌های اجتماعی

۱- مقدمه

امروزه به‌دلیل افزایش حجم و پیچیدگی داده ها، وجود ابزاری مناسب برای تحلیل داده‌های موجود و دستیابی به دانش نهفته درون آنها امری ضروری است. این تمایلات برای حصول دانش نهفته در داده‌ها، باعث رشد چشمگیر داده‌کاوی شده است [1]. داده‌کاوی، فرآیندی پیچیده جهت شناسایی الگوها در حجم وسیعی از از داده است، به‌طوری‌که قابل درک برای انسان باشد [2]. داده‌کاوی به عنوان یکی از موضوعات پر طرفدار در قرن حاضر، در تعداد قابل توجهی از زمینه‌ها و رشته‌ها به پیشرفت علم و فناوری کمک می‌کند [3]. داده‌کاوی یکی از ده دانش در حال توسعه است که دههٔ آینده را متحول خواهد کرد. امروزه داده‌کاوی دارای کاربرد وسیعی در حوزه‌های مختلف است. و به منزلهٔ پل ارتباطی میان علم آمار، کامپیوتر، هوش مصنوعی و بازنمایی بصری داده‌ها است [2]. داده‌کاوی در سال‌های اخیر، تاثیرات شگرفی در محیط های آکادمیک ایجاد کرده و کاربردهای فراوانی در زمینه‌های گوناگون و در علوم مختلف یافته است. به‌عنوان نمونه می‌توان به کاربردهای آن در تجارت، مدیریت و کشف فریب، ورزش، متن‌کاوی[۱] و وب‌کاوی اشاره نمود [19]. فرایند داده‌کاوی از دو مرحلهٔ اصلی تشکیل شده است. مرحله اول پیش پردازش داده‌ها و مرحلهٔ دوم تشخیص الگو است که وظیفهٔ اصلی داده-کاوی می‌باشد [20].

۱-۱-هم‌رخدادی واژگان

با توجه به رشد روزافزون حجم اطلاعات و سرعت فزاینده تولید علم، رصد روندهای علمی و شناسایی دانش پنهان شده در متون کاری دشوار است. لذا امروزه برای کشف لایه‌های پنهان دانش در علوم مختلف از روش‌های

www.SID.ir

مختلفی از جمله متن‌کاوی استفاده می‌شود. محققین توانسته‌اند با بهره‌گیری از این روش‌ها به کشف دانش پنهان موجود در متون علمی دست پیدا کنند. در حالی که دستیابی به چنین دانشی از طریق روش‌های سنتی امکاپذیر نمی‌باشد. یکی از این فنون که طی سالهای اخیر خیلی مورد استقبال و اقبال متخصصان رشته‌های مختلف قرارگرفته تحلیل هم‌رخدادی واژگان [۲] است که به‌عنوان شاخصی نوین به جامعهٔ علمی عرضه شد. این شاخص روشی مناسب برای رصد سیر تکامل در علوم مختلف می‌باشد. هم‌رخدادی واژگان یک فن تحلیل محتوا می باشد که هم فراوانی موضوعات و هم ارتباط بین آنها را بیان می‌کند [21].

تحلیل هم رخدادی‌واژگان، ابزاری برای کشف الگوهای پنهان و رویدادهای نوظهور مفهومی است که از طریق آن می‌توان مفاهیم اصلی یک زمینه یا حوزهٔ علمی را شناخت و به‌واسطه این شناخت، الگوها و رویدادهای مفهومی، ساختار علمی، شبکهٔ مفهومی، روابط سلسله مراتبی مفاهیم، و مقولات مفهومی آن حوزه را کشف، ترسیم و مدیریت کرد [4]. روش تحلیل هم‌رخدادی واژگان با تکیه بر واژه‌های پر بسامد می‌تواند مهم‌ترین موضوع‌های پژوهشی در هر حوزه علمی را مشخص کند؛ یعنی میزان رخداد یک واژه، نشان‌دهندهٔ میزان اهمیت آن در حوزهٔ علمی مورد نظر است. این روش اولین بار در سال ۱۹۸۰ در فرانسه و در مرکز نوآوری جامعه‌شناسی مورد استفاده قرار گرفت. پس از آن این روش برای ترسیم نقشهٔ علمی حوزه‌های تحقیقی به‌کار گرفته شد. کالون [۳] در سال ۱۹۸۳ اولین کسی بود که این روش را توصیف کرد و از آن در پژوهش‌های خود استفاده کرد [22]. ایدهٔ کالون این بود که «باهم آمدن» واژگان در یک سند، نشان دهنده محتوای آن سند می‌باشد. لذا اگر میزان هم رخدادی ها را اندازه‌گیری کنیم، می توانیم شبکه مفاهیم یک زمینه علمی را تعیین کنیم [5].

روش تحلیل هم‌رخدادی واژگان در سالهای اخیر در پژوهش‌های پژوهشگران کشورهای مختلف به‌نحو قابل ملاحظه‌ای مورد توجه قرار گرفته است. بسیاری از پژوهشگران با این روش حوزه‌های خاصی را به‌قصد تحلیل و ترسیم ساختار آن مطالعه کرده‌اند. در این تحلیل، هم‌رخدادی کلیدواژه‌ها در عنوان، چکیده یا متن مقالات بررسی می‌شود. هم‌رخدادی کلیدواژه‌ها میزان ارتباط شناختی میان یک مجموعه مدارک را نشان می‌دهد. با مقایسهٔ نقشه‌های حاصل در دوره‌های زمانی مختلف، پویایی علم ردیابی می‌شود [6]. با به‌کارگیری روش تجزیه و تحلیل هم‌رخدادی واژگان می‌توان موضوعات علمی را استخراج و ارتباط میان آنها را به صورت مستقیم از محتوای موضوعی کشف کرد [23]. تحلیل هم‌رخدادی واژگان یکی از روش‌های علم‌سنجی است که با تحلیل محتوای متون ساختار مفهومی یک حوزهٔ پژوهشی را شناسایی می‌کند. هدف از علم‌سنجی آشکارسازی ویژگی‌های پدیده‌های علمی و فرآیندهای موجود در پژوهش علمی برای مدیریت مؤثرتر علم است؛ بنابراین با روش‌های علم‌سنجی می‌توان علم و در مجموع حوزه‌های مختلف علمی را بهتر مدیریت کرد؛ و مسیر حرکت پژوهش‌هایی که در آینده انجام می‌شوند را هموار ساخت [24].

۲-۱-نقشهٔ علمی

نقشهٔ علمی، نشان‌دهنده ساختارعلمی یا همان ساختار مفهومی یک حوزهٔ علمی است [7]. نقشهٔ‌علمی به ترسیم کردن نتایج برآمده از تجزیه و تحلیل انتشارات یک حوزهٔ علمی از زوایای مختلف و ترسیم یک نگرش کلی حوزه با هدف کشف روابط پنهانی است [8]. این مفهوم در سال ۱۹۷۴ توسط هنری اسمال و در سال ۱۹۸۰ توسط هووارد وایت بیان شد [25]. نقشهٔ علمی، ابزاری است که توانسته است در سیاست‌گذاری علم‌وفناوری کمک شایانی به دولتها کند و در موضوعاتی مانند ارزیابی علم‌وفناوری، ارزیابی و سنجش عملکرد تحقیقاتی و آثار تولیدی نویسندگان، سازمانها و دانشگاهها، ترسیم ساختار موضوعات و حوزه‌های علمی، مطالعه رشد متون در موضوعات خاص و مطالعه کمّی جریان پیشرفت علم، با بررسی متون علمی منتشر شده راه را برای برنامه‌ریزی و سیاست‌گذاری در نظام علمی هموارسازد [9].

۳-۱-مصورسازی حوزه دانش

مصورسازی [۴] حوزه دانش یک فناوری پردازش اطلاعات پشتیبانی شده کامپیوتری است که می‌تواند یک ظاهر بصری از اشیاء داده متون علمی(مانند نویسندگان، کلمات کلیدی) و روابط بین آنها را آشکار کند. برای واقعیت بخشیدن به‌مصورسازی ساختار فکری حوزه‌های دانشی روابط بین اشیاء به صورت دو بعدی یا سه بعدی بیان می‌شوند که نشان‌دهندهٔ دور نمایی از دانش است [26]و[27]. مصورسازی به طور مؤثری می‌تواند شناخت انسان را نسبت به درک مقادیر زیادی از اطلاعات و طرح کلی ساختار و تکامل یک رشته علمی تقویت کند برای مصورسازی ساختار حوزه‌های علمی یکی از روش‌های مورد استفاده، تحلیل هم‌رخدادی واژگان است [28]. تحقیقات نشان داده که مصورسازی دانش از طریق تحلیل هم‌رخدادی واژگان می‌تواند تصویر روشن‌تری از آن حوزه نمایان سازد چرا که تحلیل واژگان موضوعات اساسی آن حوزه را در یک شمای کلی در اختیار پژوهشگر قرار می‌دهد. ارائه تصویر روشنی از وضعیت پژوهش‌های صورت گرفته و چگونگی ارتباط حوزه‌های مختلف یک موضوع از اهداف نقشه علمی است. ازآنجاکه نقشه‌های علمی دارای ساختاری مشابه ساختار شبکه‌های اجتماعی [۵] هستند، برای مصور سازی و تحلیل و تفسیر آنها از فنون تحلیل شبکه‌های اجتماعی استفاده می شود [29].

٤-۱-تحلیل شبکه‌های اجتماعی

شبکه اجتماعی متشکل از گره‌هایی است که توسط روابطی به‌هم متصل شده‌اند. شاخص‌های مختلفی در تحلیل این شبکه‌ها وجود داردکه می تواند در نقشه‌های علمی نیز بکار گرفته شود. اندازه شبکه، چگالی [۶] ، مرکزیت [۷] از شاخص‌های مهم در تحلیل این نوع شبکه‌ها می‌باشد [29]. یک شبکه اجتماعی به‌عنوان گروهی از افراد یا موجودیت‌هایی که با یکدیگر در حال همکاری یا رقابت هستند، تعریف می‌شود [30]. به کلیه همکاران یا رقیبان در اصطلاح، بازیگران [۸] یک شبکه اجتماعی گفته می‌شود. این بازیگران می-

www.SID.ir

توانند موجودیت‌هایی به صورت اشخاص، سازمان‌ها، گروه‌ها، موجودات زنده، بیماری‌ها، صفحات اینترنتی و یا ترکیبی از آنها باشند [31]. در اصل هر موجودیتی که قابلیت اتصال به سایر موجودیت های دیگر را دارد می تواند به‌عنوان یک شبکه اجتماعی در نظر گرفته شود [32]. مدل‌سازی تعاملات شبکه ممکن است با یک گراف ساده یا یک گراف چندگانه بیان شود. شبکه های اجتماعی می‌توانند توسط یک گراف پیچیده، مرکب از گره‌ها و یال‌ها نمایش داده شوند که در آن، افراد یا سازمان ها، همان گره ها و روابط میان آن ها یا همان یال‌های گراف را تشکیل می‌دهند [10].

گراف حاصل از شبکه دارای شکل یا ساختار خاص است که با استفاده از این ساختار می توان معیارهای مختلفی ازجمله انواع مرکزیت را به‌دست آورد و بر اساس آن گره های با اهمیت و تأثیرگذار را شناسایی کرد. همچنین با توجه به روابط میان گره‌های مختلف، می توان اجتماع‌های موجود دربرگیرنده گره‌های مختلف را تعیین نمود [10]. تحلیل شبکه‌های اجتماعی یک موضوع میان رشته‌ای بین رشته‌های مختلف جامعه‌شناسی، ریاضیات و علوم کامپیوتر است که در علوم مختلف مثل جامعه‌شناسی، اقتصاد ، علوم ارتباطی، روانشناسی ، فیزیک و کامپیوتر کاربرد دارد [30]. در سال‌های اخیر، با بررسی ساختار شبکه‌ها و با استفاده از داده‌های واقعی و ابزار های پردازشی قوی، بسیاری از خصوصیات مشابه شبکه‌ها استخراج شده است. مثلاً کمینهٔ کوتاه‌ترین فاصله گره‌ها، بالا بودن ضریب خوشه‌بندی[9]عناصر، تشخیص وجود اجتماع‌ها[10] بین اعضای شبکه ، بعضی از معیارهای طرح شده برای مطالعهٔ خصوصیات مشترک تمامی شبکه‌ها در علوم مختلف هستند [10]. تحلیل شبکه‌های اجتماعی به‌دنبال شناسایی ساختار روابط اجتماعی در یک گروه، با هدف مشخص کردن ارتباطات غیررسمی بین اعضاء است. اعضاء شبکه می‌توانند شخص، گروه، سازمان و یا حتی دولت‌ها باشند. معیارهای سنجش ویژگی‌های شبکه در همه انواع شبکه‌ها از ماتریس همسایگی استخراج می‌شود، که در دو بخش معیارهای مربوط به ویژگی های عامل‌ها و معیارهای کل شبکه دسته‌بندی می‌شوند. دستهٔ اول معیارها را معیارهای مرکزیت نیز می گویندکه اهمیت نسبی و اثرگذار بودن یک عامل را در شبکه نشان می‌دهد. دستهٔ دوم معیارهای مربوط به ویژگی‌های شبکه هستند که وضعیت کلی شبکه را بیان کرده و ساختار آن‌را مشخص می‌کنند [11].

تا کنون محققین بسیاری سعی در ترسیم نقشهٔ علمی حوزه‌های مختلف علوم جهت آشنایی با ابعاد و زاویه‌های پنهان آن داشته‌اند که در بخش مروری بر ادبیات بطور مشروح به آن پرداخته خواهد شد. مرور پیشینه‌ها نشان می‌دهد با توجه به افزایش چشمگیر پژوهش‌های مرتبط با داده‌کاوی و تکنیک‌های آن، تحلیل روند این پژوهش‌ها و تعیین ساختار نقشهٔ علمی آن امری ضروری است. از آنجاکه تاکنون در ایران پژوهش مستقلی در خصوص کاربرد روش هم‌رخدادی واژگان در ترسیم ساختار و نقشهٔ علمی حوزه داده‌کاوی انجام نشده است، مساله اصلی پژوهش حاضر کشف و ترسیم نقشهٔ علمی موضوعات حوزهٔ داده‌کاوی در ایران بر اساس مقالات نمایه شده در پایگاه داده سیویلیکا در بازه ۱۰ ساله ۱۳۸۸ تا ۱۳۹۸ است. اهداف اصلی این پژوهش افزایش شناخت زمینه‌های مشترک در زیرحوزه‌های حوزه داده‌کاوی و یافتن حوزه‌های پژوهشی فعال در این زمینه جهت پژوهش محققین می‌باشد.

ساختار کلی این مقاله به شرح ذیل می باشد. در بخش دوم مروری بر ادبیات پیشین انجام گرفته است. بخش سوم روش‌شناسی تحقیق شرح داده خواهد شد. در بخش چهارم به تحلیل و ارزیابی نتایج پرداخته می‌شود. بخش پنجم مربوط به نتیجه‌گیری و کارهای آینده است.

۲- مروری بر ادبیات

در این بخش به مرور کارهای انجام شده در حوزه ترسیم نقشه علمی حوزه‌های گوناگون در داخل و خارج ایران پرداخته می‌شود.

۲-۱-کارهای انجام گرفته داخلی

خوش‌بختیان و همکاران برای تعیین زیرحوزه‌های اصلی فناوری خودرویی هوشمند از روش هم‌رخدادی واژگان استفاده کرده‌اند. این تحقیق بر اساس بررسی ۱۱۳۴۴سند در بازه سال‌های ۱۹۶۹ تا ۲۰۱۸ میلادی مربوط به پایگاه داده اسکوپوس[11] می‌باشد [12]. در [13] ترسیم ساختار موضوعی مدیریت اطلاعات با استفاده از تحلیل هم‌رخدادی واژگان طی سال‌های۱۹۸۶ تا۲۰۱۲ بر اساس ۴۵۷۱ مدرک موجود که در پایگاه داده‌های ISI نمایه شده، ارائه شده است. محمدی در پژوهشی با عنوان" حوزه‌های تشکلیل دهنده فناوری و علم نانو در ایران" با بکارگیری تحلیل هم‌رخدادی واژگان به ترسیم و بررسی حوزه‌های فناوری و علم نانو پرداخته است که بر اساس آن ۱۵ خوشه در حوزه فناوری و علم نانو در ایران شناخته‌شده است [14]. در تحقیقی دیگر نقی‌زاده و همکاران با استفاده از روش تحلیل هم‌رخدادی واژگان به بررسی و شناسایی جریان‌های غالب در حوزهٔ توسعه نوآوری در مناطق پرداختند. نقشهٔ علمی این حوزه با بررسی۳۰۰ مقاله از پایگاه داده‌های اسکوپوس و سیج در طی سال‌های ۱۹۹۰ تا ۲۰۱۲ ترسیم شده‌است [15]. ناصری جزه و همکاران به‌بررسی و ترسیم نقشه دانش مدیریت فناوری در ایران با استفاده از تحلیل هم‌رخدادی واژگان پرداخته‌اند که در این پژوهش ۱۶۰۰ چکیده مقالات چاپ‌شده درمجلات، مقالات همایش‌ها، پایان‌نامه‌های دانشجویان و کتب فارسی در حوزهٔ مدیریت‌فناوری در بازه سال‌های ۱۳۸۰ تا ۱۳۸۹ بررسی و تحلیل شده‌است [16]. در [17] با استفاده از فن تحلیل هم‌رخدادی واژگان، ساختار فکری دانش در پژوهش‌های رفتار اطلاعاتی با استفاده از رویکردهای تحلیل شبکه و دیداری سازی علم مورد مطالعه قرار گرفته‌است. همچنین در [18] تحقیق دیگری در راستای تحلیل و مصورسازی مجموعه دادهٔ مقالات علمی در حوزهٔ نانو فناوری به منظور تشخیص حوزه‌های عمده پژوهشی و کشف روندهای نوظهور انجام گرفته‌است. در این تحقیق خلاصه ۴۶۸۰۰۰ مقاله در حوزهٔ نانو برگرفته از پایگاه‌های دادهٔ ISI در بازهٔ سال‌های ۱۹۹۰ تا ۲۰۰۸ مبتنی بر تحلیل شبکه‌های هم‌رخدادی واژگان مورد ارزیابی و تحلیل قرار گرفته است.

www.SID.ir

۲-۲-کارهای انجام گرفته خارجی

در [33] تحقیقی با هدف مشخص‌کردن حوزه‌های تحقیقات بازاریابی مالی و ارتباط بین آنها، به‌تحلیل مقالات این حوزه دربازهٔ سال‌های ۱۹۶۱ تا ۲۰۱۰ با استفاده ازهم‌رخدادی کلمات پرداخته شده است. در تحقیق دیگر لی و سو به تحلیل ۲۲۳ مقاله پر استناد در حوزه نانوکامپوزیت‌های هادی الکتریسیته پرداخته‌اند و با ترکیب روش هم‌رخدادی واژگان و تحلیل شبکه‌های اجتماعی نقشه علمی آن را ترسیم و براساس آن زیرحوزه‌های مهم و در حال ظهور در این زمینه را مشخص کرده‌اند [34]. گن و ونگ با استفاده از روش تجزیه وتحلیل هم‌رخدادی واژگان به بررسی ویژگیها و وضعیت رسانه‌های اجتماعی در کشور چین در بین سالهای 2006 تا 2013 پرداخته‌اند [35]. در [36] با استفاده از تجزیه وتحلیل هم‌رخدادی واژگان به‌ترسیم نقشهٔ ساختار اینترنت اشیاء و بررسی توسعه آن در بازه سال‌های 2001 تا 2014 با بررسی ۷۵۸ مقاله از پایگاه داده WoS[12] پرداخته شده‌است. هو و همکاران در تحقیق خود به بررسی علم کتابخانه و اطلاعات در بین سالهای 2008 تا 2012 در کشور چین با استفاده از روش تحلیل هم‌رخدادی واژگان پرداخته‌اند [37]. در تحقیقی دیگر چانگ و همکاران به منظور شناسایی نقاط ضعف تحقیقات در زمینه صلاحیت بالینی دانشجویان پزشکی، از روش هم‌رخدادی واژگان بر روی ۵۸۸مقاله بهره گرفتند [38] .

بررسی ادبیات موضوع نشان می‌دهد که هم‌رخدادی واژگان یکی از روش‌هایی است که در بیشتر مطالعات علم سنجی جهت ترسیم ساختار موضوعی در حوزه‌های علمی مختلف استفاده شده‌است. پژوهش حاضر در صدد است با نگاهی تحلیلی زمینه‌های پژوهشی حوزهٔ داده‌کاوی را به‌کمک روش هم‌رخدادی واژگان شناسایی و ترسیم نماید. این تحقیق به دنبال پاسخگویی به پرسش‌های زیر است:

۱- روند تولیدات علمی ایران در حوزه داده‌کاوی طی یک بازهٔ ۱۰ساله چگونه است؟

۲- واژگان کلیدی و پربسامد در تولیدات علمی ایران در حوزهٔ داده‌کاوی چیست؟

۳- مهم‌ترین و پرکاربرد ترین زیرخوشه‌ها در حوزه داده‌کاوی کدام است؟

۴- حوزه های پژوهشی فعال در زمینه داده‌کاوی کدام حوزه‌ها می باشند؟

۳- روش‌شناسی تحقیق

در این تحقیق از روش تحلیل هم‌رخدادی واژگان که یکی از روش‌های علم-سنجی است، استفاده شده تا از حوزهٔ داده‌کاوی اطلاعات کمّی قابل تحلیل به‌دست آید. جامعه پژوهش ۳۰۵۷ مقاله در حوزه داده‌کاوی است که در پایگاه داده سیویلیکا در بازه سال‌های۱۳۸۸ تا۱۳۹۸ نمایه شده است. بررسی این مقالات در دو دورهٔ ۵ سالهٔ اول(۱۳۹۳- ۱۳۸۸)و دوم(۱۳۹۸ - ۱۳۹۳) صورت گرفته‌است. جهت انجام این تحقیق، کلمهٔ داده‌کاوی به‌عنوان هستهٔ اصلی در نظر گرفته شده‌است. تمام کلمات کلیدی مربوط به ۳۰۵۷ مقاله از سایت سیویلیکا که شامل این عنوان در مجموعه کلید واژه های خود بودند، با استفاده از کتابخانه سلنیوم[13] در پایتون[14] جمع آوری شد. سپس هر یک از کلمات کلیدی به‌عنوان یک گره در شبکه در نظر گرفته شده است. یال‌های شبکه نیز به‌صورت ارتباط بین کلمات تعریف شده‌اند. به‌طوری‌که بین دو گره یال وجود خواهد داشت در صورتی که آن دو گره(کلمه) به صورت مشترک در لیست کلمات کلیدی یک مقاله آمده باشند. پس از جمع آوری تمام کلمات ، ارتباط دو به دو بین تمام کلمات بررسی شد و در صورتی که دو کلمه دارای مقاله مشترک بودند، بین آن ها یک یال تعریف شده است. با توجه به این که محققین برای انتخاب کلمات کلیدی، استاندارد معینی را در نظر نمی‌گیرند و انتخاب این کلمات به صورت سلیقه‌ای انجام می‌شود؛ در مجموعه داده جمع آوری شده، بسیاری از کلمات کلیدی علی‌رغم یکسان بودن از نظر مفهوم و معنا، در مقالات متفاوت در نظر گرفته شده است مانند واژه « شبکهٔعصبی»،« شبکه‌های‌عصبی» ،« شبکه‌های عصبی مصنوعی» و « شبکه‌های عصبی MLP» که همه یک مفهوم را متبادر می‌سازد. این موضوع باعث می‌شود در شبکه حاصل از ارتباط کلمات کلیدی مقالات، به ازای هر کلمه کلیدی یک گره در نظر گرفته شود و در نتیجه شبکه دارای گره‌های تکراری از لحاظ مفهوم باشد. این امر علاوه بر ایجاد گره‌های ایزوله در شبکه و تشکیل اجتماعات بسیار کوچک با حدکثر ۳ یا ۴ گره، باعث می‌شود نتایج تحلیل منعکس‌کننده واقعیت مساله نباشد. برای حل این مشکل، تمام کلمات از نظر ظاهر و مفهومی بررسی شده و کلمات هم‌معنا به‌عنوان یک گره واحد در نظر گرفته شده اند. این‌کار به صورت ترکیب ماشینی و دستی انجام شده است که جدول(۱) بخشی از جدول یکسان‌سازی کلمات هم‌معنا را نشان می‌دهد.

پس از مرحلهٔ یکسان‌سازی کلمات هم‌معنا که بخشی از مرحلهٔ پیش پردازش محسوب می‌شود، شبکه متناظر بر اساس داده‌های پیش‌پردازش شده ترسیم شده و در نهایت تحلیل و تفسیرهای مربوطه صورت گرفته‌است. مراحل انجام پژوهش در شکل (۱) مشاهده می‌شود.

٤- تحلیل و ارزیابی نتایج

همان‌طورکه بیان شد، به منظور بررسی ساختار علمی حوزهٔ داده‌کاوی مقالات در دو بازهٔ ۵ ساله مورد بررسی قرارگرفته‌است که نتایج آن در جدول(۲) مشاهده می‌شود. نتایج نشان می‌دهد که مقالات حوزهٔ داده‌کاوی در ۵ سالهٔ منتهی به سال ۱۳۹۸ از ۱۰۹۸مقاله در دورهٔ اول به ۱۹۵۹ مقاله در دورهٔ دوم افزایش یافته که رشدی ۱.۷۹ درصدی را داشته و به تبع آن فراوانی کلمات کلیدی با رشد۱.۷۷درصدی از۴۱۰۱به ۷۳۲۶ افزایش یافته است.

در شکل (۲) روند رشد مقالات حوزهٔ داده‌کاوی در سال‌های ۱۳۸۸ تا۱۳۹۸ مشاهده می‌شود.

همان‌طورکه در شکل مشاهده می‌شود بیشترین تولید مقالات به لحاظ کمّی در سال ۱۳۹۴ بوده است.

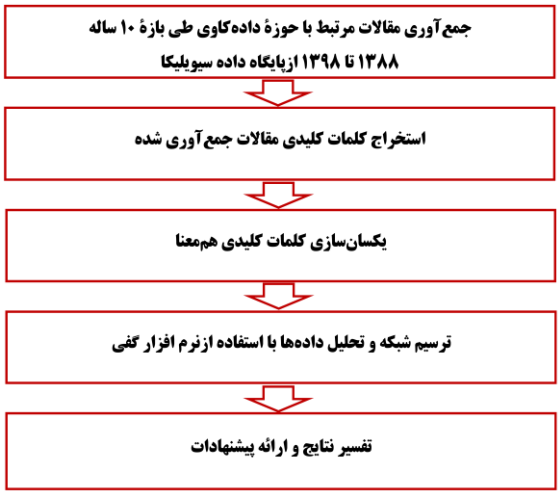
- **تحلیل مرکزیت**

در این مقاله برای شناسایی حوزه‌های داده‌کاوی ، از مفهوم تحلیل شبکه‌های اجتماعی و به طور خاص از تحلیل مرکزیت استفاده شده‌است. در تحلیل

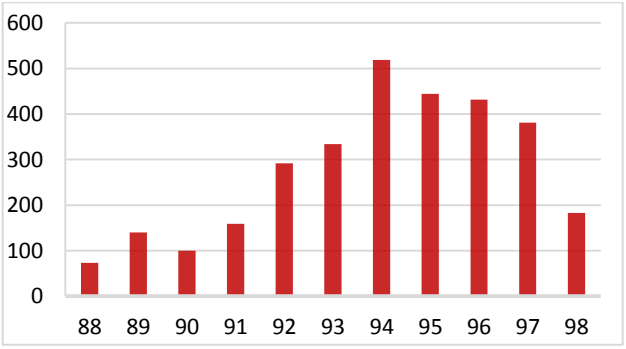
www.SID.ir

مرکزیت، گره‌های مهم و تأثیرگذار از نظر معیارهای مختلف مرکزیت تعیین می شوند. برای کاربردهای ویژه‌ای که در این مقاله مدّنظر است معیارهای مهم مرکزیت عبارت اند از درجه[۱۵]، بینابینی[۱۶]، نزدیکی[۱۷] که شرح هریک از این معیارها درقسمت‌های قبل اشاره شد. مقادیر مرکزیت درجه، بینابینی، نزدیکی برای شبکه مربوطه در دو دورهٔ ۵ سالهٔ اول و دوم محاسبه شده که جداول(۳)تا(۸) مقادیر۶ گره که دارای بیشترین مرکزیت می‌باشند را نشان می‌دهد.

در بین کلمات بررسی شده در شبکه، موضوعات مربوط به «درخت تصمیم» «طبقه‌بندی»، «شبکهٔعصبی» و«خوشه‌بندی» بیشترین مرکزیت-های درجه‌ای، بینابینی و نزدیکی را دارا می‌باشند که نشان‌دهندهٔ تولید مقالات بالا و کار گسترده در این موضوعات است. از منظر مرکزیت بینابینی بیشتر بودن مقادیر مذکور شاهد بر این مدعاست که این کلمات در شبکه نقش واسطه‌گری بیشتری را ایفا می‌نمایند. همچنین از دیدگاه مرکزیت نزدیکی بدین معنی است که این کلمات در شبکه راحتی و سهولت می‌توانند اتصال را با سایر گره‌های دیگر برقرار کنند و ارتباط نزدیک‌تری با بقیه موضوعات دارند.



شکل (۱): مراحل انجام پژوهش



شکل (۲) : روند رشد مقالات حوزهٔ داده‌کاوی

جدول (۱) : یکسان‌سازی کلمات هم‌معنا

کلمات کلیدی هم‌مفهوم	کلمهٔ جایگزین
درخت تصمیم‌گیری، درخت تصمیم بهینه‌شده، درخت تصمیم، درختان تصمیم	درخت تصمیم
شبکهٔمصنوعی، شبکه‌های مصنوعی، شبکه‌های عصبی مصنوعی، شبکه‌های عصبی MLP	شبکهٔ عصبی
یادگیری ماشینی، الگوریتم یادگیری ماشین، یادگیری ماشین	یادگیری ماشین
الگوریتم SVM، الگوریتم طبقه‌بندی SVM، ماشین بردار پشتیبان، طبقه بندSVM	SVM
طبقه‌بند، طبقه‌بندی، کلاس‌بندی، کلاسه‌بندی، الگوریتم طبقه‌بندی، تکنیک طبقه‌بندی	طبقه‌بندی

جدول(۲) : مقادیر مرکزیت بینابینی

دوره	اول(۱۳۸۸-۱۳۹۳)	دوم(۱۳۹۳-۱۳۹۸)
تعداد مقالات	۱۰۹۸	۱۹۵۹
تعداد کلمات کلیدی	٤۱۰۱	۷۳۲٦

جدول (۳) : مقادیر مرکزیت درجه(دورهٔ اول سال‌های ۹۳-۸۸)

نام گره	مقدار
درخت تصمیم	۸۱٦
خوشه‌بندی	۸۰۲
طبقه‌بندی	٦۹٤
شبکهٔعصبی	٥٤۸
پیش‌بینی	٤٦۲

جدو(٤) : مقادیر مرکزیت بینابینی(دورهٔ اول سال‌های ۹۳-۸۸)

نام گره	مقدار
خوشه‌بندی	۱۷۳٦.۹٥
درخت تصمیم	۷۳۰.۷۱
طبقه‌بندی	٦۲۱.۰۹
قوانین انجمنی	۳٦۰.۷۰
پیش‌بینی	۳۰۲.۱۳

جدول(٥) : مقادیر مرکزیت نزدیکی(دورهٔ اول سال‌های ۹۳-۸۸)

نام گره	مقدار
خوشه‌بندی	۰.۸٤
درخت تصمیم	۰.۷٥
طبقه‌بندی	۰.۷٤
شبکهٔعصبی	۰.٦۸
پیش‌بینی	۰.٦٥

www.SID.ir

جدول(٦): مقادیر مرکزیت درجه (دورهٔ دوم سال‌های ۹۸-۹۳)

نام گره	مقدار
درخت تصمیم	١٦٦٨
شبکهٔعصبی	١٤٨٦
طبقه‌بندی	١٢٦٢
داده‌کاوی پزشکی	١١٩٦
خوشه‌بندی	١١٩٦

جدو(۷) : مقادیر مرکزیت بینابینی(دورهٔ دوم سال‌های ۹۸-۹۳)

نام گره	مقدار
خوشه‌بندی	٢٠٤٥.٢٣
درخت تصمیم	١٣٦٢.٩٨
شبکهٔعصبی	١٢٥٤.٥٥
طبقه‌بندی	١٢٣٣.٧٨
پیش‌بینی	٥٤٣.٢٠

جدول۸: مقادیر مرکزیت نزدیکی(دورهٔ دوم سال‌های ۹۸-۹۳)

نام گره	مقدار
خوشه‌بندی	٠.٧٧
شبکهٔعصبی	٠.٧٥
طبقه‌بندی	٠.٧٣
درخت تصمیم	٠.٧٢
پیش‌بینی	٠.٦٨

• اجتماع‌یابی

در مقاله حاضر برای اجتماع‌یابی[۱۸] از الگوریتم گیروان – نیومن[۱۹] استفاده شده است [39] .

روش گیروان نیومن: این روش در ابتدا مرکزیت بینابینی تمام یال‌ها را محاسبه می‌کند. سپس یالی را که بیشترین مقدار مرکزیت را دارد حذف می‌کند. سپس در هر مرحله، قبل از حذف یال بعد، مقدار مرکزیت بینابینی یال‌ها دوباره محاسبه می‌شود. این روند تا زمانی که تمام یال‌ها حذف شوند و یا معیار ماژولاریتی بهینه شود، ادامه پیدا می کند [39].

درشکل‌های (۳) و (۴) خروجی اجتماع‌یابی شبکه بر اساس الگوریتم گیروان نیومن، در دو بازهٔ ۵ ساله مشاهده می‌شود. در خصوص تعداد اجتماعات در این دو بازه نتایج نشان می‌دهد در بازهٔ پنج سالهٔ اول تعداد اجتماعات ۹۰ و در بازهٔ دوم ۱۳۳ اجتماع وجود دارد که با توجه به رشد بیش از ۵۰ درصدی تعداد مقالات و به‌تبع آن کلمات کلیدی امری بدیهی تلقی می‌شود. در این شکل‌ها هر یک از رنگ‌ها نمایندهٔ یک اجتماع می‌باشد که طبعاً به‌دلیل گستردگی تعداد اجتماعات همهٔ آنها در شکل قابل رؤیت نیست.

نمایش شبکهٔ حاصل از تحلیل مقالات حوزهٔ داده‌کاوی در مدت ۱۰ سال در قالب دو بازهٔ پنج ساله در شکل‌های (۵) و (۶) ارائه شده‌است.

همانطور که در این شکل‌ها مشاهده می‌شود فاصلهٔ برخی اجتماعات از یکدیگر کمتر از اجتماعات دیگر است و مفهوم آن این است که پژوهش های مرتبط در بین این اجتماعات نسبت به بقیه بیشتر است. از این منظر می‌توان این‌طور برداشت کرد که خوشه‌ها بر اساس فواصلی که از همدیگر دارند، ارتباط کمتر یا بیشتری با هم دارند؛ اصولاً اجتماع‌های نزدیک‌تر حوزه‌هایی هستند که کار مشترک و ارتباط بیشتری دارند. بطور مثال «خوشه بندی»، «درخت تصمیم»، «داده‌کاوی پزشکی» و «شبکهٔعصبی» جزء موارد موجود در اجتماعات نزدیک و حوزه‌های مرتبط می‌باشند که تعداد پژوهش‌های مشترک در بین آنها نسبت به بقیه موارد افزون‌تر است. همچنین همان‌گونه که بر اساس این شکل‌ها مشاهده می‌شود در بازهٔ ۵ سالهٔ دوم چگالی و تراکم موضوعاتی نظیر خوشه‌بندی، شبکهٔ عصبی، پیش بینی و داده‌کاوی پزشکی به مراتب بیشتر از دورهٔ قبل است که بیانگر تمرکز پژوهش‌ها بر استفاده از این موضوعات است.

در مورد واژگان کلیدی و پربسامد در تولیدات علمی ایران در حوزهٔ داده-کاوی که به عنوان یکی از پرسش‌های تحقیق حاضر می‌باشد، همان‌طور که در جداول(۹) و(۱۰) مشاهده می‌شود «درخت تصمیم»، «خوشه‌بندی»، «شبکهٔ عصبی» و «طبقه‌بندی» بیشترین بسامد تکرار را در مقالات به‌خود اختصاص داده‌اند. همچنین واژگانی مانند داده‌کاوی پزشکی و الگوریتم‌های ابتکاری و فرا ابتکاری در دورهٔ ۵ ساله منتهی به سال ۹۸ به لحاظ بسامد تکرار ، رشد بالاتری را در مقایسه با بقیهٔ واژگان کسب نموده‌اند. این امر نشان می‌دهد که در مقالات و پژوهش‌های اخیر مقولهٔ داده‌کاوی در مسائل پزشکی بیشتر مورد توجه محققین بوده‌است.

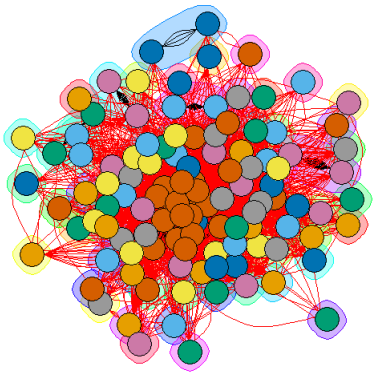
با بررسی نتایج مبتنی بر تحلیل شبکه، مشخص شد واژگان «خوشه-بندی»، «طبقه‌بندی»، «پیش‌بینی»، «شبکهٔعصبی»، «داده‌کاوی‌پزشکی»، «ماشین بردار پشتیبان»، «الگوریتم‌های فرا ابتکاری» و«یادگیری‌ماشین» طی بازهٔ۱۰ سالهٔ مورد بررسی در این مقاله بیشترین بکارگیری در پژوهش‌ها را داشته است. شکل (۷) روند استفاده از کلید واژه‌های مذکور در مقالات این حوزه را نشان می‌دهد.

۵- نتیجه‌گیری

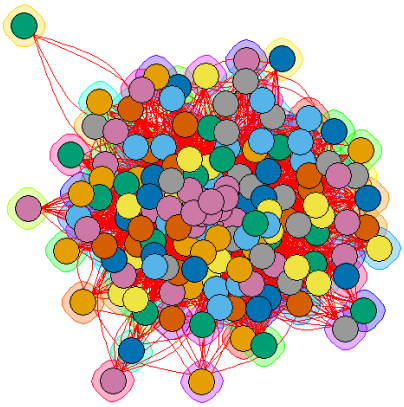
تحلیل متون و مقالات علمی در یک حوزه تخصصی، موضوعات خاص آن رشته را برجسته‌تر و سایر موضوعات مرتبط با موضوع اصلی را آشکار می‌سازد. امروزه داده‌کاوی دارای کاربرد وسیعی در حوزه‌های مختلف است. لزوم آشنایی با حوزه‌های مختلف آن و نیز عناوین مهم و بروز جهت تحقیق برای پژوهشگران بیش از پیش ضروری است. در این مقاله بر اساس تحلیل شبکه-های اجتماعی و نیزهم‌رخدادی واژگان نقشهٔ علمی و شبکه متناظر با مقولهٔ داده‌کاوی و زیرحوزه‌های آن ترسیم شد. نتایج این تحقیق نشان داد داده‌کاوی و زیر حوزه‌های آن همچنان به‌عنوان موضوعات جذاب و مورد علاقه پژوهشگران علوم مختلف به‌شمار می‌آید. همچنین بر اساس این تحقیق مشخص شد در برخی زیر حوزه‌ها و نیز حوزه‌های مرتبط با داده‌کاوی نظیر منطق فازی، سیستم‌های توصیه‌گر، متن‌کاوی، تحلیل احساسات، پتانسیل

www.SID.ir

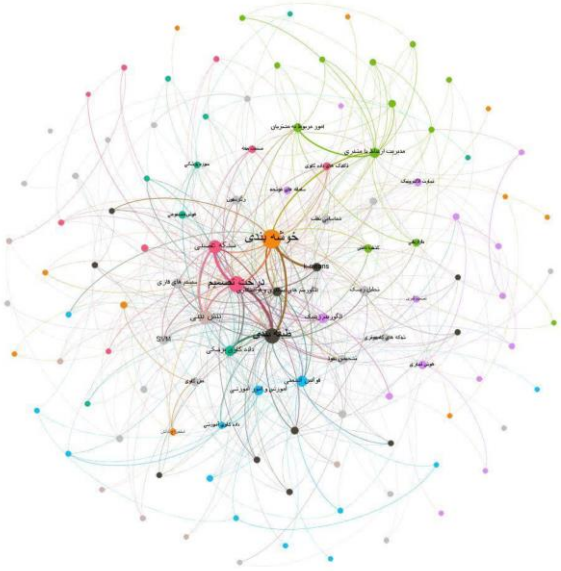
Arch



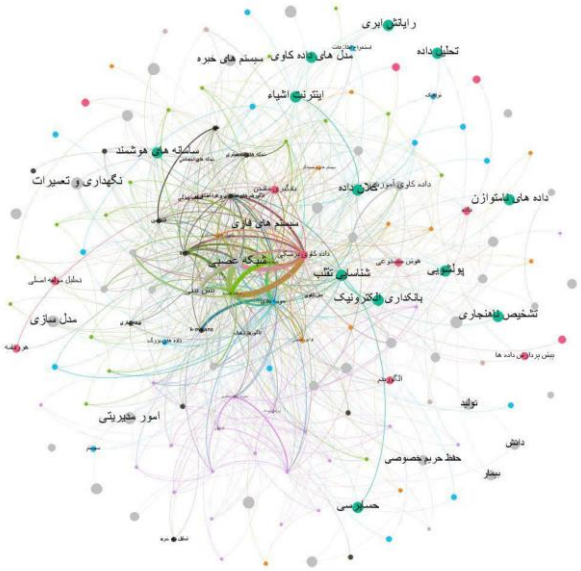
شکل (۳) : اجتماع‌یابی (دوره اول سال‌های ۹۳-۸۸)



شکل (٤) : اجتماع‌یابی (دوره دوم سال‌های ۹۸-۹۳)



شکل (٥): نمایی از شبکه (دوره اول سال‌های ۹۳-۸۸)



شکل (٦) : نمایی از شبکه (دوره دوم سال‌های ۹۸-۹۳)

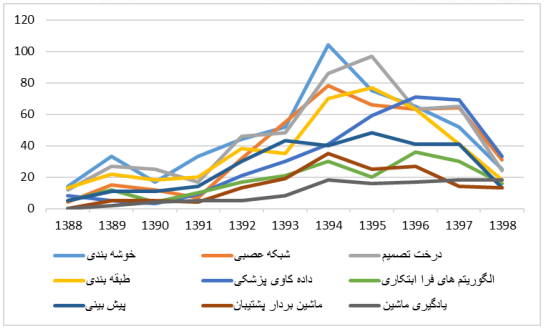
جدول۹ : کلمات کلیدی با بیشترین تکرار ۵ سال منتهی به۹۳

ردیف	نام گره	تعداد
۱	خوشه‌بندی	۲۳۱
۲	درخت تصمیم	۲۰۰
۳	طبقه‌بندی	۱٦۷
٤	شبکهٔ‌عصبی	۱۳۷
٥	پیش‌بینی	۱۲٥
٦	قوانین انجمنی	۹۲
۷	داده‌کاوی پزشکی	۷۷
۸	مدیریت ارتباط با مشتری	۷۷
۹	منطق و سیستم‌های فازی	۷٥
۱۰	الگوریتم‌های ابتکاری و فرا ابتکاری	۷٤

جدول۱۰ : کلمات کلیدی با بیشترین تکرار ۵ سال منتهی به۹۸

ردیف	نام گره	تعداد
۱	درخت تصمیم	۳٤۱
۲	خوشه‌بندی	۳۲٦
۳	شبکهٔ‌عصبی	۳۰۸
٤	طبقه‌بندی	۲۷٥
٥	داده‌کاوی پزشکی	۲۷۳
٦	پیش بینی	۱۸۹
۷	الگوریتم‌های ابتکاری و فرا ابتکاری	۱۳۳
۸	SVM	۱۲٥
۹	مدیریت ارتباط با مشتری	۱۰٥
۱۰	یادگیری ماشین	۸۷

www.SID.ir



شکل۷ : روند استفاده از کلید واژه‌های حوزۀ داده‌کاوی

انجام و ادامه تحقیقات جدید و کاربردی وجود داشته و نسبت بـه موضـوعات مرتبط دیگر کمتر مورد توجه قرار گرفته است که محققین علاقمند می توانند موضوعات پژوهش خود را در این حوزه‌ها جستجو نمایند. در نهایت شناسـایی و پیش‌بینی زیر حوزه‌هایی که احتمـالاًدر آینـده مـی‌تواننـد بیشـتر در کـانون پژوهش محققین قرار گیرد، بررسی روند حوزه‌های مختلف در طی بـازه‌هـای مختلف زمانی و نیز بررسی همین مساله در پایگاه داده‌های معتبر جهانی نظیر اسکوپوس و WoS می تواند بـه عنـوان کارهـای آتـی در ایـن حـوزه مـورد پژوهش و بررسی قرار گیرد.

مراجع

[۱] غضنفری مهدی، علیزاده سمیه، تیمورپور بابک، داده‌کاوی و کشف دانش، چاپ پنجم، تهران، دانشگاه علم و صنعت، ۱۳۹۵.

[۲] مقصودی بهروز، سلیمانی صادق، امیری علی، افشارچی محسن ، ارتقای کیفیت آموزش در سامانه های آموزش الکترونیکی با استفاده از داده کاوی آموزشی، فصلنامه فناوری آموزش، دوره ۶، شماره ۴، ۱۳۹۱.

[۳] رحمانی مهدی، زین‌العابدینی محسن، کاربردهای داده‌کاوی در علم اطلاعات و دانش‌شناسی، فصلنامه مدیریت اطلاعات و دانش‌شناسی، سال دوم، شماره ۳، پاییز ۱۳۹۴.

[۴] احمدی حمید، عصاره فریده، مروری بر کارکردهای تحلیل هم‌واژگانی . فصلنامه مطالعات ملی کتابداری و سازماندهی اطلاعات .دوره28، شماره 1 ،۱۳۹۶.

[۵] ناصری جزه محمود، طباطبائیان سیدحبیب الله، فاتح راد مهدی ، علم سنجی و خوشه بندی دانش مدیریت فناوری در ایران با هدف ارزیابی این دانش و مقایسه آن با وضعیت جهانی، دومین کنفرانس بین المللی و ششمین کنفرانس ملی مدیریت فناوری، تهران، انجمن مدیریت فناوری ایران، ۱۳۹۱.

[۶] سالمی نجمه،کوشا کیوان، مقایسه تحلیل هم‌استنادی و تحلیل هم‌واژگانی در ترسیم نقشه کتاب‌شناختی مطالعه موردی :دانشگاه تهران .پژوهشنامه پردازش ومدیریت اطلاعات، سال بیست و نهم، شماره 1 ،۱۳۹۲.

[۷] حسن‌زاده محمد، خدادوست رضا، ابعاد هم‌نویسندگی بین‌المللی در حوزۀ نانوفناوری. فصلنامۀ سیاست علم و فناوری شماره ۵، ۱۳۹۱.

[۸] محمدی احسان، مفهوم نقشه‌های ساختاری، نشریه رهیافت، شماره ۴۳، ۱۳۸۷.

[۹] عصاره فریده ، حیدری غلام رضا ، زارع فراشبندی فیروزه ، حاجی زین العابدینی محسن، کتاب‌سنجی تا وب‌سنجی: تحلیلی بر مبانی، دیدگاه‌ها، قواعد و شاخص‌ها، انتشارات کتابدار، ۱۳۹۲.

[۱۰] بسطامی اسماعیل، جوادزاده محمدعلی، تحلیل مرکزیت شبکه‌های اجتماعی در فضای سایبری با رویکرد مقابله با تهدیدات نرم، فصلنامه پدافندغیرعامل، سال ششم، شماره ۱، ۱۳۹۴.

[۱۱] محمودزاده مرتضی، البرزی محمود، خلیلی سهراب، تحلیل ساختار شبکۀ نوآوری نانو ایران در حوزۀ سلامت، فصلنامه سیاست علم و فناوری،سال ششم شماره۳، ۱۳۹۳.

[۱۲] خوش بختیان معصومه، افتخاری حسین، محمدی لیلا، مطالعه و ارزیابی وضعیت تولیدات علمی در حوزه فناوری خودرویی هوشمند (خودران و متصل) ، ششمین کنگره بین المللی توسعه و ترویج علوم و فنون بنیادین در جامعه، تهران، انجمن توسعه و ترویج علوم و فنون بنیادین، ۱۳۹۷.

[۱۳] خادمی روح‌الله، حیدری غلامرضا، ترسیم ساختار موضوعی مدیریت اطلاعات با استفاده از روش هم‌آیندی واژگان طی سالهای 1986 تا2012، فصلنامه علوم وفنون اطلاعات، سال دوم، شماره ۲، ۱۳۹۵.

[۱۴] محمدی احسان، حوزه های تشکیل دهنده فناوری و علم نانو در ایران، پنجمین همایش دانشجویی فناوری نانو، تهران، دانشگاه علوم پزشکی تهران، دانشکده فناوریهای نوین پزشکی،۱۳۸۸.

[۱۵] الهی شعبان، نقی‌زاده رضا، قاضی زاده سپهر، منطقی منوچهر، شناسایی جریان‌های غالب در حوزه توسعه نوآوری در مناطق با استفاده از روش تحلیل هم رخدادی کلمات، ۱۳۹۱.

[۱۶] ناصری جزه محمود، طباطبائیان سید حبیب اله، فاتح راد مهدی، ترسیم نقشه دانش مدیریت فناوری در ایران با هدف کمک به سیاست‌گذاری دانش در این حوزه، سیاست علم و فناوری، مجله سیاست علم و فناوری، شماره ۱۷، ۱۳۹۱.

[۱۷] سهیلی فرامرز، شعبانی علی، خاصه علی اکبر، ساختار فکری دانش در حوزه ی رفتار اطلاعاتی: مطالعه ی هم واژگانی،مجله تعامل انسان و اطلاعات،شماره ۸، ۱۳۹۴.

[۱۸] تیمورپور بابک، کشف روندهای نوظهور در حوزه‌های علمی بر پایه خوشه-بندی پویا با رویکرد متن‌کاوی و تحلیل پیوند، رسالۀ دکتری، دانشگاه تربیت مدرس، ۱۳۸۸.

[19] Larose, Daniel T, and Chantal D. Larose. Discovering knowledge in data: an introduction to data mining. Vol. 4. John Wiley & Sons, (2014).

[20] Han, Jiawei, Jian Pei, and Micheline Kamber. Data mining: concepts and techniques. Elsevier, (2011).

[21] Rip, Arie, and J. Courtial. "Co-word maps of biotechnology: An example of cognitive scientometrics." *Scientometrics* 6.6 (1984): 381-400.

[22] He, Qin. "Knowledge discovery through co-word analysis." (1999).

[23] Callon, Michel, Arie Rip, and John Law, eds. Mapping the dynamics of science and technology: Sociology of science in the real world. Springer, (1986).

[24] Vinkler, Peter. The evaluation of research by scientometric indicators. Elsevier, (2010).

[25] Eden, Bradford Lee, and Chaomei Chen. "Holistic sense-making: conflicting opinions, creative ideas, and collective intelligence." Library Hi Tech (2007).

www.SID.ir

Arch

[26] Börner, Katy, Chaomei Chen, and Kevin W. Boyack. "Visualizing knowledge domains." Annual review of information science and technology 37.1 (2003): 179-255
[27] Chen, Yue, and Ze-yuan Liu. "The rise of mapping knowledge domain." Studies In Science of Science 2 (2005).
[28] Ames, L. "Mapping the field of arts and economics." In Paper 16th International Conference on Cultural Economics, ACEI, Copenhagen.2010.
[29] Guns, Raf, Yu Xian Liu, and Dilruba Mahbuba. "Q-measures and betweenness centrality in a collaboration network: a case study of the field of informetrics." Scientometrics 87.1 (2011): 133-147
[30] B. Furht, "Handbook of Social Network Technologies and Applications," Springer, page 3, 2010.
[31] D. Watts and S. Strogatz, "The Small World Problem," Collective Dynamics of Small-World Networks, vol. 393, pp. 440–442, 1998.
[32] A. L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the Social Network of Scientific Collaborations," Physica A: Statistical Mechanics and Its Applications, vol. 311, no. 3, pp. 590–614, 2002.
[33] Muñoz-Leiva, Francisco, et al. "Applying an automatic approach for showing up the hidden themes in financial marketing research (1961–2010)." Expert Systems with Applications 39.12 (2012): 11055-11065.
[34] Lee, Pei-Chun, and Hsin-Ning Su. "Quantitative mapping of scientific research—The case of electrical conducting polymer nanocomposite." Technological forecasting and social change 78.1 (2011): 132-151.
[35] Gan, Chunmei, and Weijun Wang. "Research characteristics and status on social media in China: A bibliometric and co-word analysis." Scientometrics 105.2 (2015): 1167-1182.
[36] Yan, Bei-Ni, Tian-Shyug Lee, and Tsung-Pei Lee. "Mapping the intellectual structure of the Internet of Things (IoT) field (2000–2014): A co-word analysis." Scientometrics 105.2 (2015): 1285-1300.
[37] Hu, Chang-Ping, et al. "A co-word analysis of library and information science in China." Scientometrics 97.2 (2013): 369-382.
[38] Chang, Xing, et al. "Hotspots in research on the measurement of medical students' clinical competence from 2012-2016 based on co-word analysis." BMC medical education 17.1 (2017): 162.
[39] Fortunato, Santo. "Community detection in graphs." *Physics reports* 486.3-5 (2010): 75-174.

زیرنویس‌ها

1 Text mining
2 Co-word analysis
3 Callon
4 Visualization
5 Social networks
6 Density
7 Centrality
8 Actors
9 Clustering coefficient
10 Community
11 Scopus
12 Web of Science
13 Selenium
14 Python
15 Degree centrality
16 Betweenness centrality
17 Closeness centrality
18 Community detection
19 Girvan and Newman

www.SID.ir