J. SLIS, 16(2): 1-20, Summer 2024

ISSN: 2008-5222
slis.scu.ac.ir

Shahid Chamran University of Ahvaz

Journal of Studies in Library and Information Science
(*JSLIS*)

https://doi.org/10.22055/slis.2023.38914.1862

Research Paper

Analysis of information technology research trends using topic modeling

Arman Sajedinejad[1*], Mohammad Rabiei[2]

1. Assistant Professor, Iranian Research Institute for Information Science and Technology (IRANDOC)
2. Assistant Professor, Iranian Research Institute for Information Science and Technology (IRANDOC)

Article Info.

Received: 2023/10/24 **Accepted**: 2021/04/20

Abstract

Background and Objectives: IT's rapid progress and far-reaching impact on other scientific disciplines have not only necessitated significant changes in its own subjects but have also catalyzed extensive changes in the form, amount, and methodology of research in other fields. The objective of the present investigation was to analyze research conducted in the realm of information technology, extract its central themes, and furnish scientometric data pertaining to these themes.

Methodology: This paper explores the topics of the information technology field by extracting and categorizing the relationships between commonly used terms and their temporal evolution. To achieve this, the researchers employed topic modeling, a well-established method for clustering textual data. Topic modeling algorithms utilize statistical methods to analyze and interpret the primary words in documents, allowing for the examination of the presented issues and their interconnections and changes over time. Considering the rapid changes in the field of information technology, this paper drew upon materials spanning the last decade, including 10,000 papers sourced from top-tier journals featured in the Web of Science database.

Findings: The study extracted trends in keyword changes over the past decade and identified important keywords for each paper cluster after grouping them. The papers within the information technology domain were then categorized into eight themes, including hardware, communications, networks, and intelligent applications such as the Internet of Things. The study found that frequently used keywords have been continuously changing over time. The paper highlights that emerging keywords, including the Internet of Things, cloud computing, and Big Data, along with work areas such as Machine Learning and Deep Learning, are shaping the definition of information technology fields in the new era.

Discussion: Given the shift in research emphasis from hardware and communication to analysis and practical applications, it is likely that future scientific fields will focus on creating value through data analysis and human-machine communication in everyday applications, and information technology's

relevance in other sciences will become more apparent. Future research can also concentrate on comparing global trends in information technology with domestic research, enabling the evaluation of the gap between the country's research and that of the world.

Keywords: Information technology, topic modeling, text analysis, scientometrics

***Corresponding author** Email: sajedinejad@irandoc.ac.ir

How to Cite:. Sajedinejad, A., & Rabiei, M. (2024). Analysis of information technology research trends using Topic Modeling. *Journal of Studies in Library and Information Science*, *16*(2), 1-20.

فصلنامه
مطالعات کتابداری و علم اطلاعات. سال ۱۶، شماره ۲، تابستان ۱۴۰۳

مقاله پژوهشی

تحلیل روند پژوهش‌های جهان در حوزه فناوری اطلاعات با استفاده از روش مدل‌سازی موضوعی

آرمان ساجدی‌نژاد[1*]، محمد ربیعی[2]

۱. استادیار، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
۲. استادیار، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

اطلاعات مقاله

تاریخ دریافت: ۱۴۰۲/۰۸/۰۲ **تاریخ پذیرش:** ۱۴۰۳/۰۲/۰۱

چکیده

هدف: پژوهش حاضر با هدف شناسایی و تحلیل پژوهش‌های حوزه فناوری اطلاعات، استخراج موضوعات آن و ارائه اطلاعات علم‌سنجی مربوط به این موضوعات انجام شده است.

روش‌شناسی: در این مقاله، موضوعات حوزه‌ی فناوری اطلاعات استخراج و ارتباطات بین واژگان پرکاربرد و تکامل زمانی آن‌ها تجزیه و تحلیل و در نهایت دسته‌بندی شده است. برای این منظور از مدل‌سازی موضوعی که روشی شناخته شده برای خوشه‌بندی اطلاعات متنی است استفاده شد.

یافته‌ها: روندهای ده‌ساله تغییرات کلیدواژه‌ها در مجموعه مطالعات، استخراج شد و پس از خوشه‌بندی مقالات، کلیدواژه‌های مهم هر خوشه استخراج شد. مقالات استخراج شده در حوزه فناوری اطلاعات به ۸ دسته موضوعی تقسیم شدند که از موضوعات سخت‌افزار، ارتباطات و شبکه تا موضوعات کاربردهای هوشمند مانند اینترنت اشیا را شامل می‌شد. مشخص شد که کلیدواژه‌های پرکاربرد در این منابع همواره در گذر زمان در حال تغییر بوده است.

نتیجه‌گیری: در زمینه فناوری اطلاعات موضوعات در حال ترکیب و در بین دسته‌های مختلف فناوری اطلاعات در حال تغییر است. باتوجه به تغییر سمت و سو پژوهش‌های این حوزه از سخت‌افزار و ارتباطات به سمت کاربردها و تحلیل، به نظر می‌رسد زمینه‌های علمی آتی شامل کاربردهای روزمره و ایجادکننده ارزش‌افزوده با توجه به تحلیل داده و ارتباطات بین انسان و ماشین شکل خواهند گرفت و کاربردهای فناوری اطلاعات در علوم دیگر نمایان‌تر شده است. همچنین تمرکز بر پژوهش‌هایی که کمتر جنبه توسعه‌ای داشته و بیشتر کاربردی هستند و یا ارزش افزوده بیشتری برروی زیرساخت‌های موجود ایجاد نموده‌اند در این تغییرات مشهود است.

کلیدواژه‌ها: *فناوری اطلاعات، مدل‌سازی موضوعی، تحلیل متون، علم‌سنجی*

***نویسنده مسئول** **ایمیل:** sajedinejad@irandoc.ac.ir

استناد به این مقاله: ساجدی‌نژاد، آرمان و ربیعی، محمد. (۱۴۰۳). تحلیل روند پژوهش‌های جهان در حوزه فناوری اطلاعات با استفاده از روش مدل‌سازی موضوعی. *مطالعات کتابداری و علم اطلاعات*، ۱۶(۲)، ۲۰-۱.

مقدمه و بیان مسئله

در دهه‌های اخیر، با رشد و توسعه فناوری اطلاعات شاهد افزایش چشم‌گیر پژوهش‌های این حوزه از دانش هستیم و دسته‌بندی موضوعات مختلف این حوزه همواره محل چالش بوده است (Ghanadinejad et al., 2018). یکی از دلایل این چالش پویایی این حوزه و پتانسیل بالای کاربردی بودن این حوزه از علم است. با وجود انجام پژوهش‌های فراوان با هدف استخراج مدل‌های توسعه فناوری اطلاعات، به نظر می‌رسد کشف مرزهای حوزه فناوری اطلاعات و روندهای آن (در پژوهش‌های پیشرو در جهان) همچنان مورد نیاز است. از سوی دیگر، با وجود اینکه تجمیع داده‌های فعالیت‌های علمی کشور ارزش افزوده فراوانی ایجاد نموده است، اما کمبود درک از این دسته‌بندی‌ها می‌تواند موجب شود تا یک روند تازه ظهور در کشور، کمتر مورد توجه فعالیت‌های علمی، صنعتی و کاربردی واقع شود. اگرچه امکان مشاهده روند جستجوها درباره یک کلیدواژه خاص در نمودارهای پایگاه اطلاعات علمی داخلی و خارجی قابل مشاهده است (Rabiei et al., 2021)، اما کشف روندهای یک حوزه علمی مانند فناوری اطلاعات و مقایسه آن با روندهای جهانی، نیازمند بررسی تخصصی و بکارگیری روش‌های متن‌کاوی است. حوزه فناوری اطلاعات به دلیل پیشرفت روزافزون و تاثیری که در سایر حوزه‌های علمی دارد، علاوه بر تغییرات شگرف در موضوعات و زیرمجموعه‌های خود موجب تغییرات گسترده در شکل، میزان و چگونگی انجام پژوهش در حوزه‌های دیگر نیز شده است. از این رو، تحلیل مطالعات این حوزه مدنظر این پژوهش قرار گرفته است.

هدف از این مقاله، شناسایی و تحلیل روندهای پژوهش در حوزه فناوری اطلاعات در مطالعات جهانی است. برای این منظور، مقالات پژوهشی در این زمینه استخراج و با ابزارهای متن‌کاوی بررسی شده و سپس با بکارگیری روش مدل‌سازی موضوعی، موضوعات این پژوهش‌ها شناسایی و روند پژوهش‌ها در طول زمان ۱۰ ساله کشف شده است. روش مدل‌سازی موضوعی به این دلیل انتخاب شده است که می‌تواند بدون نیاز به دانش پیشین در حوزه موضوعی و بدون وابستگی به نظر خبرگان، موضوعات یک حوزه دانشی را بر مبنای الگوریتم‌های متن‌کاوی شناسایی و استخراج کند. در این گزارش با انتخاب مجموعه‌ای از مجلات مهم و پیشرو در زمینه فناوری اطلاعات در پایگاه «وب‌آف‌ساینس»، ارتباطات بین واژگان پرکاربرد و تکامل زمانی آن‌ها تجزیه و تحلیل شده و در نهایت یک تحلیل سطح بالا از مفاهیم مرتبط با حوزه‌های فناوری اطلاعات ارائه گردیده است. منظور از ارائه سطح بالا از مفاهیم و حوزه‌های فناوری اطلاعات، شناسایی حوزه‌های اصلی روندهای فناوری اطلاعات و نمایش واژگان کلیدی هر یک از حوزه‌ها، در یک دوره زمانی خاص است.

مدل‌سازی موضوعی از روش‌های کارای تحلیل متن است که هدف آن گروه‌بندی مجموعه‌ای از متون مختلف به‌صورت زیرمجموعه‌های معنادار است که موضوع (Topic) نامیده می‌شوند. مدل‌سازی موضوعی بر اساس این ایده تشکیل شده است که اسناد، ترکیبی از موضوعات هستند و موضوعات برگرفته شده از مجموعه کلماتی هستند که با یک توزیع احتمالی بیش از دیگر کلمات مورد استفاده قرار گرفته‌اند، موضوعات (و توزیع آن‌ها) به عنوان ساختار نهفته یا متغیرهای نهفته در نظر گرفته می‌شوند (Rabiei et al., 2017).

الگوریتم‌های مدل‌سازی موضوعی روش‌های آماری هستند که کلمات اصلی اسناد را تجزیه و تحلیل می‌کنند تا مباحثی را که از طریق آن‌ها ارایه می‌شود تحلیل نموده و امکان بررسی چگونگی اتصال این موضوعات به یکدیگر و چگونگی تغییر آن‌ها را با گذر زمان فراهم شود. یک موضوع به عنوان توزیعی از مجموعه‌ای از واژگان تعریف می‌شود. به عنوان مثال موضوع اینترنت اشیا دارای کلماتی مانند هوشمندسازی، رباتیک، تعامل انسان و ماشین با توزیع احتمالی بالا است (Choudhary et al., 2009).

این الگوریتم‌ها شامل تکنیک‌های آماری است که هدف آن‌ها توصیف موضوعاتی است که در اسناد منتشر شده‌اند. این الگوریتم‌ها بر اساس تجزیه و تحلیل آماری کلمات در آن اسناد، اقدام به شناسایی خوشه‌هایی از کلمات مشترک، تشخیص موضوعات مورد بحث، تحلیل روابط بین این موضوعات و پیش‌بینی روند رشد یا افول آن‌ها می‌کنند. چندین مدل مانند تجزیه و

تحلیل معنایی معکوس[1]، تجزیه و تحلیل معنایی بالقوه احتمالی[2] و تخصیص نهفته دیریکله[3] (Blei et al., 2003)؛ و همچنین برخی از مشتقات آن‌ها، مانند پاچینو[4] و یا مدل‌سازی موضوعی رابطه‌ای (Chang & Blei , 2009) در این زمینه کاربرد دارند. مدل‌سازی موضوعی نوعی از مدل آماری برای کشف موضوعات نهفته[5] است که در مجموعه‌ای از اسناد از طریق یادگیری ماشین رخ می‌دهد. در حال حاضر، تخصیص نهفته دیریکله یک رویکرد مدل‌سازی کارا و رایج در این زمینه است.

روش‌های استخراج موضوع بر مبنای این فرض عمل می‌کنند که کلمات موجود در یک متن از نظر معنایی به یکدیگر وابسته هستند و معنایی که در مستندات یک موضوع مسنجم وجود دارد از مجموعه کلمات این مستندات به دست آمده است. به عبارتی صرف‌نظر از معانی یا مکان قرارگیری متن، هم‌رخدادی کلمات مورد تحلیل قرار گرفته و اسناد به‌صورت مجموعه‌ای از کلمات در نظر گرفته می‌شود. در مدل‌سازی موضوعی مفروضات زیر در نظر گرفته می‌شود (Wallach, 2006).

- ترتیب حضور مستندات در پیکره اهمیتی ندارد؛
- ترتیب قرار گرفتن کلمات در هر مستند اهمیتی ندارد؛
- موضوعات توزیع چندجمله‌ای از کلمات موجود در یک واژه‌نامه هستند؛
- کلمات یک مستند از تعدادی موضوع پنهان نشأت گرفته‌اند.

دو فرض اول که ترتیب مستندات در یک پیکره و ترتیب کلمات در یک مستند را نادیده می‌گیرند منجر به خلق مفهومی به نام سبد واژگان[6] می‌شود که نماینده‌ای از یک مستند خواهد بود و اصلی‌ترین ورودی الگوریتم‌های مدل‌سازی موضوعی است. فرض سوم و چهارم تبیین‌کننده مفهوم موضوع است که توزیعی از کلمات موجود در یک واژه‌نامه است (Rabiei et al., 2017). در صورتی‌که به تعیین نام یا برچسب برای یک موضوع نیاز باشد، معمولاً از چند کلمه پرتکراری که بیشترین ارتباط را به آن موضوع دارند و یا از نظر خبرگان استفاده می‌شود (Wallach, 2006).

این مدل فرض می‌کند که موضوعات قبل از اسناد (مقالات) تولید می‌شوند. برای هر سند، کلمات در یک فرایند دومرحله‌ای تولید می‌شوند:

- به طور تصادفی یک توزیع برای موضوع انتخاب می‌شود (توزیع دریکله)؛
- برای هر کلمه آن موضوع، ابتدا به صورت تصادفی یک موضوع از توزیع موضوعات فاز قبل انتخاب شده و سپس به طور تصادفی یک کلمه از توزیع متناظر در فهرست لغات انتخاب می‌شود (Yau et al., 2014)؛

با توجه به توضیحات بالا هدف از انجام این مطالعه:

- استخراج دسته‌بندی از حوزه‌های مفهومی فناوری اطلاعات؛
- مقایسه روندهای حوزه‌های فناوری اطلاعات در سطح بین‌المللی؛

بوده و برای دستیابی به اهداف بالا، این پژوهش در تلاش است به پرسش‌های زیر پاسخ دهد:

- سبد واژگان حوزه‌های فناوری اطلاعات در سطح بین‌المللی شامل چه کلیدواژگانی است؟
- روند کاربرد این واژگان در سطح بین‌المللی چگونه بوده است؟

در بخش بعدی این مقاله، مروری بر مطالعات پیشین این حوزه خواهیم داشت و پس از آن روش انجام پژوهش و گام‌های آن تشریح شده است. سپس روش‌های مطرح شده پیاده‌سازی شده و نتایج حاصل تحلیل شده است، در نهایت در بخش نتیجه‌گیری یافته‌های پژوهش تحلیل شده و بحث‌های مرتبط با مطالعه جاری ارائه شده است.

1. LSA
2. PLSA
3 .Dirichlet (LDA)
4. Pachinco
5. Latent
6. Bag of Words

مبانی نظری پژوهش

به‌منظور استخراج اطلاعات از داده‌های متنی، بسیاری از پژوهشگران در سال‌های اخیر به توسعه روش‌های مبتنی بر آمار و احتمالات (Steyvers et al., 2004) و شناسایی ترکیب‌های مشترک در متن یا بین کلیدواژه‌ها و با تمرکز بر تجزیه و تحلیل داده‌ها (Feldman & Sanger, 2007) پرداختند. نهایتاً مدل‌سازی موضوعی به‌عنوان ابزاری مفید و یکی از روش‌های سیستماتیک (Robinson et al., 2011) در خوشه‌بندی متون و پیش‌بینی روندهای توسعه علم مورد بهره‌برداری قرار گرفت.

در پژوهش انجام شده توسط Figuerola et al. (2017) یک مرور کلی از مطالعه کتابشناختی حوزه کتابداری و اطلاع رسانی با هدف شناسایی مرزهای موضوعی و زمینه‌های اصلی و فرعی موضوعات شناخته شده ارائه شد. در این پژوهش که منابع حوزه کتابداری و اطلاع رسانی را در بین سال‌های ۱۹۷۸ تا ۲۰۱۴ بررسی کرده است، تعداد ۹۲۷۰۵ سند استخراج شد. همچنین مدل‌سازی موضوعی، از طریق تکنیک تخصیص دیریکله به منظور شناسایی موضوعات نهفته و دسته‌های اصلی در مجموعه اسناد، مورد استفاده قرار گرفت. نتایج کمّی، وجود ۱۹ موضوع مهم که می‌تواند به چهار بخش اصلی فرایندها، فناوری اطلاعات، کتابخانه و برنامه‌های اطلاعاتی تقسیم شود را نشان داد.

Debortoli et al. (2016) با توجه به رشد داده‌های غیرساختارمند (مثلاً متن، صدا، تصویر، ویدیو) که بیشتر آن‌ها با زبان طبیعی غنی و مبهم بیان می‌شوند این چالش مورد بحث قرار دادند که با وجود این‌که به طور سنتی، برای تجزیه و تحلیل زبان طبیعی، از رویکردهای تحلیل داده‌های کیفی مانند کدگذاری دستی استفاده می‌شود، ولی اندازه مجموعه داده‌های متنی به‌دست آمده از اینترنت، تجزیه و تحلیل دستی را عملا غیرممکن می‌کند. در این پژوهش، چالش‌هایی را که هنگام استفاده از تکنیک‌های متن‌کاوی خودکار در تحقیقات سیستم‌های اطلاعاتی با آن مواجه می‌شوند، مورد بحث قرار داده و نحوه استفاده از مدل‌سازی موضوع احتمالی را از طریق تخصیص دیریکله پنهان، برای توضیح رضایت کاربر از یک محصول فناوری اطلاعات (با تجزیه و تحلیل خودکار بیش از ۱۲۰۰۰ نظر مشتریان آنلاین) نشان دادند.

در سال ۲۰۱۸ ابوهای و همکاران روشی جایگزین برای مطالعه و پیش‌بینی روند موضوعات تحقیق ارائه دادند که در آن پیش‌بینی روند موضوعات پژوهشی علمی بدون استفاده از قضاوت ذهنی کارشناسان یا تحلیل کمی انجام می‌شد. آن‌ها در این مقاله با استفاده از مدل‌سازی موضوعی، روند موضوعات تحقیق را پیش‌بینی کردند و روش پیشنهادی با استفاده از مجموعه مقالات کنفرانس بین‌المللی علوم محاسباتی[1] که در مجموع شامل ۵۹۸۲ مقاله در طول هفده سال (۲۰۰۱–۲۰۱۷) می‌شد، آزمایش شد. برای کشف موضوعات از روش مدل‌سازی موضوعی استفاده شد و نتیجه به عنوان داده‌های سری زمانی ساختار یافته و برای پیش‌بینی روند موضوعات تحقیق مورد استفاده قرار گرفت (Abuhay et al., 2018).

Kherwa & Bansal (2018) بررسی جامعی بر کاربردهای مدل‌سازی موضوعی انجام داده و این روش را انقلاب جدید در متن‌کاوی نامیدند. ایشان پس از تجزیه و تحلیل حدود ۳۰۰ مقاله تحقیقاتی در مورد مدل‌سازی موضوعی، یک نظرسنجی جامع در مورد آن ارائه دادند که شامل سلسله‌مراتب طبقه‌بندی، روش‌های مدل‌سازی موضوعی، تکنیک‌های استنتاج پسین، مدل‌های تکاملی مختلف تخصیص دیریکله نهفته[2] و کاربردهای آن در حوزه‌های مختلف فناوری از جمله بیوانفورماتیک، مهندسی نرم‌افزار و تحلیل شبکه‌های اجتماعی بود.

در همین سال جیاراج و حسن‌زاده ادعا کردند که تحقیقات سیستم‌های اطلاعاتی به طور پیوسته در طول زمان تکامل یافته است و در حال تغییر است. آن‌ها در پژوهش خود، ۲۹۶۲ مقاله منتشر شده در شش مجله پیشرو بین سال‌های ۲۰۰۳ تا ۲۰۱۷، را با روش مدل‌سازی موضوعی و با استفاده از نمایه‌سازی معنایی پنهان برای کلیدواژه‌های ارائه‌شده توسط نویسنده بررسی کردند. بر اساس این مطالعه، موضوعاتی مانند توسعه سیستم‌های اطلاعاتی، پذیرش فناوری اطلاعات و استفاده از سیستم‌های اطلاعاتی در طول زمان

1. ICCS
2. LDA

ماندگار شده بودند درحالی که موضوعاتی مانند تجارت الکترونیک و برون سپاری فناوری اطلاعات شتاب بیشتری یافته و با گذشت زمان از بین رفته اند. همچنین از دیدگاه ایشان موضوعاتی مانند رسانه‌های اجتماعی، علم طراحی و جوامع آنلاین در زمان‌های اخیر شتاب بیشتری به دست آورده و نتایج نشان داد که تحقیقات سیستم‌های اطلاعاتی به طور قابل‌توجهی در طول زمان تکامل یافته است (Jeyaraj & Hassan Zadeh, 2019).

Lee et al. (2018) ایده‌های جدید در میان دو موضوع تحقیقاتی بسیار پرطرفدار یعنی پایداری و رسانه‌های اجتماعی را مورد مطالعه قرار داده و یک تحلیل تجربی از ۷۶۲ اثر منتشر شده ارائه دادند که عبارت‌های «پایداری» و «رسانه‌های اجتماعی» را در چکیده‌های خود گنجانده‌اند. داده‌های کتابشناختی شامل چکیده‌ها از پایگاه داده اسکوپوس جمع‌آوری شد. به منظور انجام تجزیه و تحلیل، از تخصیص نهفته دیریکله استفاده و۱۰ موضوع اصلی شناسایی شد. با اندازه‌گیری تنوع توزیع‌های موضوعی در طول زمان، موضوعات داغ (روندهای تحقیقاتی که در طول زمان به طور فزاینده‌ای محبوب می‌شوند) و موضوعات سرد نیز شناسایی شدند. رفتار مصرف کننده پایدار، جامعه پایدار و گردشگری پایدار به عنوان موضوعات داغ و آموزش برای پایداری به عنوان تنها موضوع سرد شناسایی شدند (Lee et al., 2018).

Sharma et al. (2021) در حوزه مدیریت اطلاعات، مدیریت فرآیندهای سازمانی را مورد بررسی قرار دادند. مدیریت اطلاعات یک دامنه وسیع و چند رشته‌ای است که زیر دامنه‌های مختلف را به اشتراک می‌گذارد و کاملاً با سایر حوزه‌ها در می‌آمیزد. هدف این مطالعه ارائه مروری جامع در حوزه مدیریت اطلاعات از سال ۱۹۷۰ تا ۲۰۱۹ بود. این مطالعه با استفاده از روش تحلیل موضوعی متن ، سیر تکامل دانش در این حوزه را با بررسی روند انتشار، بر اساس نویسندگان، مؤسسات، کشورها نشان داد. علاوه بر این، این مطالعه یک مدل مولد احتمالی مبتنی بر مدل‌سازی موضوعی ساختاری را برای درک و استخراج مضامین نهفته از مقالات پژوهشی مرتبط با مدیریت اطلاعات پیشنهاد داد و به صورت گرافیکی تغییرات در توسعه موضوع را در دوره ۱۹۷۰ تا ۲۰۱۹ نشان داد. نتایج این پژوهش نشان داد که رایج‌ترین موضوعات مدیریت داده، مدیریت دانش، مدیریت محیطی، مدیریت پروژه، مدیریت خدمات، و مدیریت تلفن همراه و وب بودند. این یافته ها همچنین موضوعاتی مانند مدیریت دانش، مدیریت زیست محیطی، مدیریت پروژه و ارتباطات اجتماعی را به عنوان نقاط داغ دانشگاهی برای تحقیقات آینده مشخص کرد.

هر چند که استفاده از روش‌های ماشینی و تکیه بر تحلیل حجم انبوه اطلاعات، اخیراً در فرایندهای تصمیم‌گیری و برنامه‌ریزی موردنظر قرار گرفته‌اند اما کاربست این روش‌ها در منابع زبان فارسی بسیار اندک بوده و الگوریتم‌ها و ابزارهای توسعه داده‌شده در این روش‌ها نیز بیشتر مبتنی بر متون عمومی مانند روزنامه و وب‌سایت‌های خبری و ... است و کمتر به متون علمی و پژوهشی پرداخته شده است (Rabiei et al., 2021). بنابراین، در این قسمت مروری بر فعالیت‌های پژوهش‌های داخلی کشور داشته‌ایم. در پژوهشی که توسط جلالی‌منش و همایون‌والا انجام شد، ۶۵۰ عنوان پایان‌نامه در زمینه مهندسی صنایع با الگوریتم‌های متن‌کاوی و تحلیل شبکه اجتماعی با هدف استخراج حوزه مهندسی صنایع، مورد پایش قرار گرفت. در این مطالعه ۱۵۰۰ واژه به‌عنوان واژه‌های مرتبط با مهندسی صنایع انتخاب شدند (Jalalimanesh & Homayoon Vala, 2012). در همین سال صدیقی و جلالی‌منش در پژوهشی با استفاده از روش‌های علم‌سنجی و کتاب‌سنجی، در صدد پرداختن به این مساله بودند که مجموعه تولیدات علمی در حوزه مدیریت دانش در سطح بین‌المللی در بازه زمانی ۱۰ ساله (۲۰۰۱-۲۰۱۰)، چه رویکردی داشته و به چه حوزه‌هایی پرداخته است. از نظر ایشان به لحاظ موضوعی، بیشترین گرایش به برخی فنّاوری‌های اطلاعاتی پشتیبان، به‌ویژه «داده‌کاوی» و نیز مفاهیم و جنبه‌های سازمانی و اجتماعی حوزه مدیریت دانش بوده است (Seddiqqi & Jalalimnesh, 2013).

در ادامه این کار Razeqi & Aghajani (2020) به بررسی تولیدات علمی، و تعیین موقعیت ایران در عرصه تولید جهانی علم، طی سال‌های ۲۰۱۵-۱۹۹۶ و پیش‌بینی این روند در سال‌های آتی پرداختند. همچنین از روش تحلیل متون برای اولویت‌گذاری نیازهای اطلاعاتی استفاده شد (Zare-Farashbandi et al., 2020).

Rabiei et al. (2021) مقالات منتشر شده در بازه ۲۰ ساله پایگاه وب آو ساینس را در سه حوزه موضوعی فناوری اطلاعات،

مدیریت و مهندسی صنایع گردآوری کردند و با بکارگیری تکنیک‌های متن‌کاوی و پردازش زبان طبیعی مهم‌ترین موضوعات و حوزه‌های دانشی این سه رشته را شناسایی و بررسی کردند. همچنین روند شکل‌گیری و تحولات موضوعات این سه رشته و چگونگی تاثیرگذاری موضوعات این سه رشته روی یکدیگر را بررسی و با یکدیگر مقایسه کردند. در این مطالعه، موضوع سلامت به عنوان یکی از موضوعات پرکاربرد و روبه رشد در حوزه فناوری اطلاعات شناخته شد.

با مطالعه پژوهش‌های گذشته می‌توان دریافت که این پژوهش‌ها صرف‌نظر از تکنیک و روش انجام کار، عمدتاً نشان می‌دهند که در چه حوزه‌هایی و به چه میزان کار پژوهشی انجام شده است، اما در تعداد محدودی از آن‌ها روندهای آتی پیش‌بینی و تفسیر شده است. همچنین در بسیاری از پژوهش‌های پیشین، تعداد محدودی از منابع اطلاعاتی یا کاربران و یا خبرگان را به‌عنوان جامعه پژوهشی در بازه زمانی محدود در نظر گرفته‌اند و با مصاحبه و یا پرسش از آن‌ها به تعیین کاستی‌های پژوهشی پرداخته‌اند و روش‌های مورد استفاده نیز تنها قابلیت استفاده در فرایندهای گردآوری و تحلیل نظرات گروه‌های محدود و یا تعداد افراد محدود را دارد. حال‌آنکه در این مقاله با استفاده از روش‌های ماشینی و مستقل از خبرگان، حجم انبوه اطلاعات را مورد تحلیل قرار داده و دسته‌بندی موضوعات را با استفاده از این داده‌ها استخراج می‌نماییم.

روش پژوهش

در این مقاله از روش‌مدل‌سازی موضوعی برای کشف محدوده و دسته‌بندی موضوعات حوزه‌ی فناوری اطلاعات استفاده شده است. رویکرد مدل‌سازی موضوعی، رویکردی مبتنی بر متن‌کاوی و یادگیری ماشین است و در این مقاله بر مبنای آن، امکان شناسایی حوزه‌های موضوعی یک سند یا مجموعه‌ای از اسناد فراهم می‌شود. از دیگر کاربردهای استفاده از مدل‌سازی موضوعی در این مقاله می‌توان به ارائه سطح بالا از مفاهیم، کلمات کلیدی حوزه فناوری اطلاعات و زمینه‌هایی که مربوط به موضوع مورد نظر در یک دوره زمانی برای یک حوزه خاص پژوهش (که در این پژوهش فناوری اطلاعات) است، اشاره نمود. خروجی مدل‌سازی موضوعی، مدلی احتمالی است که در آن خروجی، نتیجه یک فرایند آماری است. این فرایند توزیع احتمالی مشترک را در هر دو متغیر تصادفی مشاهده شده و نهفته تعریف می‌کند. مشکل اصلی این پژوهش در بهره از مدل‌سازی موضوعی، نیاز به بالا بودن تعداد رکوردها و کلیدواژه‌ها برای بالاتر بودن صحت مدل است.

الگوریتم‌های مدل‌سازی موضوعی به دو گروه تقسیم می‌شوند که توزیع‌های جایگزین دیگری را پیشنهاد می‌دهند تا تقریباً الگوریتم‌های مبتنی بر نمونه‌برداری را بهبود بخشند:

- گروه اول ارایه یک زنجیره مارکوف (یک توالی از متغیرهای تصادفی)،
- گروه دوم، الگوریتم‌های مبتنی بر نمونه‌گیری[1].

بدین‌منظور و با توجه به مشکل ذکر شده و فراوانی بیش از حد جامعه پژوهش، الگوریتم مبتنی بر نمونه‌گیری در این مقاله مورد استفاده قرار می‌گیرد تا بتوان دسته‌بندی موضوعی از حوزه‌ی فناوری اطلاعات ارائه داده و همپوشانی بین موضوعات استخراج شود.

در نتیجه گام‌های زیر برای شناسایی و تحلیل موضوعات حوزه فناوری اطلاعات طی شدند:

۱) به منظور مطالعه موضوعات فناوری اطلاعات در سطح بین‌المللی، لازم بود تا در ابتدا سبد اولیه از کلیدواژگان مرتبط با حوزه فناوری اطلاعات مشخص گردد تا طبق آن سبد، جستجو در پایگاه‌های علمی کامل شود. این گام پژوهش به چند مرحله تقسیم شد تا تاثیر جهت‌گیری احتمالی در نظر خبرگان حذف شود. خبرگان همکار در این مقاله، شامل ۴ عضو هیات علمی با رشته‌های مرتبط با فناوری اطلاعات بود که آشنایی کافی با مجلات معتبر در حوزه فناوری اطلاعات داشتند و تاکنون چندین مقاله علمی در حوزه فناوری اطلاعات منتشر کرده‌اند. در مرحله ابتدایی به‌جای ارائه کلیدواژگان مد نظر ایشان (و برخلاف تمام مقالات مرور شده پیشین)، از ایشان تنها عنوان مجلات معتبر که می‌تواند موضوعات حوزه‌ی فناوری اطلاعات را به‌خوبی پوشش دهد،

1 . Variational expectation-maximization (VEM) algorithm

مورد پرسش قرار گرفت. همچنین برای به دست آوردن نتیجه مطلوب، توصیفگرهایی مانند اطلاعات مربوط به کلمات کلیدی مرتبط و اطلاعات آماری مفید درباره انتشارمجله شامل توزیع مجلات مقالات در هر موضوع، توزیع مقالات موضوعی در هر سال، کشور و تعداد استنادات و قدمت مجلات در اختیار خبرگان قرار گرفت.

۲) در مرحله دوم، تمامی مقالات مجلات پیشرو (به انتخاب خبرگان) در سال‌های ۲۰۱۸ به بعد استخراج و تمامی کلیدواژه‌های آن‌ها استخراج شدند که این مجموعه به «سبد اولیه کلیدواژگان» نامگذاری شد که شامل بیش از ۳۰۰ کلیدواژه متفاوت بودند. پس در این مرحله سبد واژگان که شامل کلیدواژه‌های تمام مقالات موجود در مجلات پیشرو در زمینه فناوری اطلاعات است، شکل گرفت. کلیدواژه‌های استخراج شده ممکن است دارای فرکانس‌های متفاوت بوده و ممکن است به علوم دیگر نیز مرتبط باشند (مانند مقالات کاربردی فناوری اطلاعات در دیگر علوم نظیر پزشکی). در این مرحله تمامی اولیه کلیدواژگان، توسط خبرگان پالایش شده و از بین آن‌ها «سبد واژگان برگزیده» استخراج شد (شامل ۸۸ کلیدواژه منتخب خبرگان).

همچنین در این گام، سبد اولیه کلیدواژگان شامل کلمات عمومی (و غیر مرتبط با فناوری اطلاعات مانند ... , an , the , a) نیز می‌باشد که می‌بایست پاکسازی گردند. پس از تشکیل سبد واژگان برگزیده، تمام کلمات این سبد در پایگاه علمی WOS، مورد جستجو قرار گرفت (بصورت تصادفی در بازه زمانی بین ۲۰۱۱ تا ۲۰۱۹ میلادی) تا مقالاتی که یکی از کلیدواژگان آن‌ها در سبد جستجو قرار دارند استخراج شوند. در نتیجه نمونه شامل ۱۰هزار فهرست کلیدواژه/مقاله برای بدست آوردن میزان کاربرد و روند استفاده هر واژه از سبد یافت شد. همانگونه که ذکر شد، سبد واژگان برگزیده (که نشان‌دهنده بخش بزرگی از حوزه مربوطه هستند و می‌توانند پژوهش‌های حوزه فناوری اطلاعات را بازیابی نمایند، در کل پایگاه WOS (و نه محدود به مجلات معتبر مرحله قبل)، مورد جستجو قرار گرفت و مطالعات مرتبط(موضوع و چکیده و کلیدواژه‌های مطالعات انجام شده) استخراج شد.

۳) در مرحله سوم آمارهای توصیفی از روند توسعه هر کلیدواژه استخراج و نمایش داده می‌شود. این آمارها بر اساس تعداد تکرار هر کلیدواژه در مقالات منتخب در دوره زمانی مورد بررسی به‌دست آمد و پرکاربردترین کلیدواژه‌ها در حوزه‌ی فناوری اطلاعات شناسایی و همچنین روند تکرار کلیدواژه در طول زمان استخراج شد و با استفاده از این آمار، داغ‌شدن کلیدواژه‌ها در طول زمان رصد شد.

۴) به‌منظور مطالعه موضوعات پژوهش‌ها در حوزه فناوری اطلاعات، مقالات استخراج‌شده با توجه به کلیدواژه‌های آن‌ها تحلیل موضوعی و محتوایی (مدل‌سازی موضوعی) شده و با توجه به احتمال حضور هر کلیدواژه در کنار دیگر کلیدواژگان، موضوعات متفاوت استخراج شد.

در جدول ۱. مراحل مختلف انجام پژوهش تشریح شده است.

جدول ۱. شرح روش پژوهش

مرحله	روش
انتخاب مجلات معتبر	• جستجوی مجلات برتر در حوزه فناوری اطلاعات و استخراج توصیفات هر مجله (مانند قدمت) • انتخاب مجلات پیشرو توسط خبرگان
تهیه سبد واژگان اولیه و برگزیده	• تجمیع کلیدواژگان تمام مقالات مجلات منتخب • پاکسازی عبارات اضافه و تشکیل سبد واژگان اولیه • تشکیل سبد برگزیده شده توسط خبرگان
بررسی روند توسعه کلیدواژه‌ها در بازه زمانی ۱۰ ساله	• جستجو در تمامی در پایگاه اطلاعاتی WOS بدون محدودیت حوزه فناوری اطلاعات در باز زمانی بین ۲۰۱۱ تا ۲۰۱۹ میلادی • بررسی چگونگی گزینش متون خروجی جستجو
مدل‌سازی موضوعی	• بررسی روند کاربرد کلیدواژگان/ ارتباطات کلیدواژگان • دسته‌بندی کلیدواژه‌ها در حوزه‌های مفهومی فناوری اطلاعات با بهره‌گیری از فرکانس استفاده از کلیدواژگان در هر حوزه • دسته‌بندی کلیدواژه‌ها در حوزه‌های مفهومی فناوری اطلاعات با بهره‌گیری از مدل‌سازی موضوعی

مراحل پیاده‌سازی پژوهش با استفاده از زبان برنامه‌نویسی R انجام شده است. مراحل ذکر شده در بخش پیاده‌سازی و تحلیل نتایج شامل آمارهای توصیفی، تحلیل ضریب سرگشتگی و مدل‌سازی موضوعی به این زبان نوشته شده که نتایج آن در ادامه آمده است.

یافته‌های پژوهش

در گام اول این پژوهش، انتخاب مجلات معتبر برای استخراج سبد واژگان اولیه انجام پذیرفت. در طرحواره این پژوهش استفاده از کلیدواژگان مورد استفاده در مجلات برتر در حوزه فناوری اطلاعات به‌عنوان روشی برای تشکیل سبد اولیه در نظر گرفته شد. توصیف‌های متعددی برای انتخاب این مجلات پیش روی محققین قرار دارد. در این مقاله شاخص‌های فعالیت مجلات شامل اچ‌ایندکس[1]، سابقه، میزان ضریب تأثیر[2] و حوزه پوشش مورد ادعای مجله در اختیار خبرگان قرار گرفت. بر اساس جدول زیر، مجلات پر کاربرد در هر حوزه‌ی پوشش مورد ادعای مجلات، انتخاب و بر حسب شاخص اچ مرتب‌سازی شده و توسط خبرگان انتخاب شدند. به خبرگان توصیه شد تا این انتخاب به نحوی صورت گیرد که تمامی حوزه‌های فناوری اطلاعات، در پوشش مورد ادعای مجلات قرار گرفته و همچنین همه دسته‌های حوزه پوشش، حداقل یک مجله به‌عنوان نماینده برای جستجو داشته باشند. در نتیجه در مرحله اول یافتن مجلاتی که در فهرست مجلات تأثیرگذار هر زمینه بودند و همچنین در زمینه‌های مختلف فعالیت داشتند در اولویت انتخاب قرار گرفت. در جدول ۲. ، مجلات ۹ گانه انتخاب شده و حوزه پوشش آن‌ها نیز از اطلاعات مجله مشخص و استخراج شد.

جدول ۲. اطلاعات و حوزه پوشش مجلات منتخب

Table 2. Information and coverage area of selected journals

نام مجله	حوزه‌ی پوشش	ناشر	کشور	اولین انتشار
Bioinformatics	Computational theory and Mathematics Computer Science Application	Elsevier Ltd.	UK	1985
Computer Physics Communications	Hardware and Architecture	Elsevier BV	Netherlands	1969
Computers in Human Behavior	Human Computer Interaction	Elsevier Ltd.	UK	1985
European Journal of Operational Research	Computer Science (miscellaneous)	Elsevier BV	Netherlands	1977
IEEE Communications Magazine	Computer Network and Communication Computer Science Application	EEEI	SAU	1979
IEEE Transactions on Image Processing	Computer Graphics and Computer Aided Design Software	EEEI	USA	1992
IEEE Transactions on Information Theory	Computer Science Application Information Systems	EEEI	USA	1969
IEEE Transactions on Pattern Analysis and Machine Intelligence	AI, Computional Theory and Mathematics Computer Vision and Pattern Recognition Software	EEEI	USA	1978
Pattern Recognition	AI, Signal Processing	Elsevier Ltd.	UK	1968

در مرحله دوم تمامی کلیدواژه‌های مجلات جدول ۲. استخراج شده و پس از کسب نظر تأیید یا رد خبرگان، در سبد واژگان منتخب قرار گرفت (۸۸ کلیدواژه).

1. H-Index
2 . Impact Factor

اقدامات زیر روی واژگان این سبد انجام شد:

- حذف فاصله‌های اضافی موجود در عبارت‌ها،
- تبدیل تمامی کاراکترها به حروف کوچک برای شناسایی عبارت‌های یکسان،
- جداسازی کلیدواژه‌های هر مقاله و تولید ماتریس مستند-واژه[1]،
- شمارش تکرار هر کلیدواژه در کل مستندات.

با توجه به درهم‌تنیدگی مطالعات فناوری اطلاعات با دیگر مطالعات، این ماتریس شامل بیش از ۳۰۰ کلیدواژه است که بسیاری از آن‌ها مختص به فناوری اطلاعات نیستند به همین دلیل، در این مرحله با استفاده از نظرات خبرگان، کلیدواژه‌های مختص به فناوری اطلاعات انتخاب شدند که در مجموع ۸۸ کلیدواژه برای مرحله جستجوی بیشتر تعیین شد. با توجه به فهرست بالای واژگان پرتکرار در دو جستجوی مجلات پیشرو و مجلات منتخب خبرگان، یافتن سبد واژگان منتخب فناوری اطلاعات با قوانینی (به‌عنوان توصیه به خبرگان) در این گام مورد توجه قرار گرفت. قوانین زیر برای اجماع خبرگان درنظر گرفته و به ایشان توصیه شد:

- واژه مختص به فناوری اطلاعات باشد؛
- واژه شامل ابزار حل مساله نباشد؛
- واژه‌های علمی عمومی مانند مدل، مدلسازی، الگو، تحلیل، توصیف در سبد واژگان قرار نگیرد زیر می‌توانند نتایج جستجو را به سمت غیر فناوری اطلاعات منحرف نمایند؛
- جزو واژه‌های پرتکرار باشد (در این پژوهش ۱۰ تکرار).

واژه شامل کاربرد فناوری اطلاعات در زمینه‌های علمی دیگر نباشد. برای در اختیار گذاشتن و قابل مقایسه بودن تمامی واژگان یافت شده در دو جستجوی اخیر، پاکسازی واژگان نیز در دستور کار قرار گرفت. نحوه استفاده از کلیدواژگان در مجلات مختلف ممکن است یکسان نباشد. به طور مثال هر دو واژه IOT و یا Internet of Things ممکن است در کلیدواژگان مقالات استفاده شوند درصورتی که به یک مفهوم اشاره دارند. در ادامه، کلیدواژگان پرتکرار در اختیار خبرگان قرار گرفته و کلیدواژگان منتخب آن‌ها انتخاب شد.

با توجه به انتخاب مجموعه فوق در مرحله سوم، تعداد ۱۰هزار رکورد شامل نام مجله، کلیدواژگان، عنوان و چکیده در بازه زمانی ۲۰۱۱ تا ۲۰۱۹ دریافت شد. بر این اساس ۱۷۴۰۳ کلیدواژه منحصر به فرد از مجموع ۲۶۲۶۸ واژه شناسایی شده و واژگان بالای ۱۰ تکرار در سبد اولیه قرار گرفت. به طور نمونه در جدول ۳، ۱۰ کلیدواژه استخراج شده پرکاربرد آمده است.

جدول ۳. ده کلیدواژه پرکاربرد در مجلات دارای ضریب تاثیر بالا

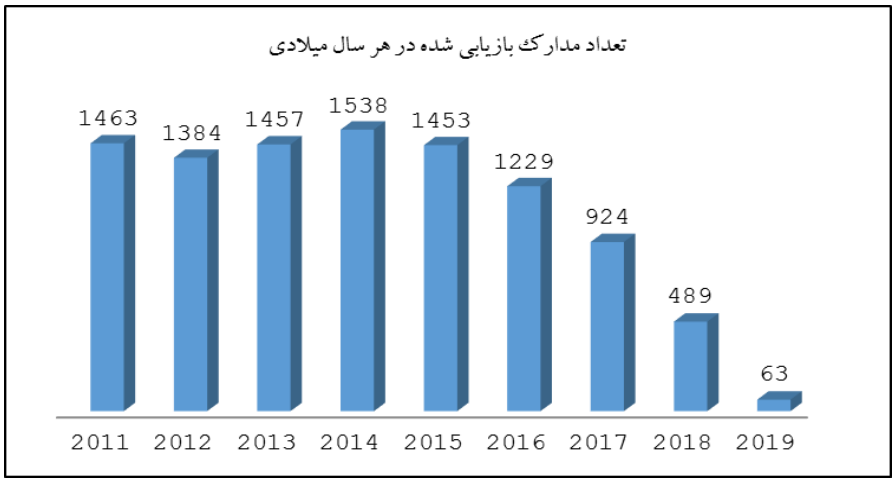
Table 3. Ten most frequently used keywords in high impact journals

ردیف	واژه	تکرار	ردیف	واژه	تکرار
1	Deep Learning	171	6	Game Theory	62
2	Convolutional Neural Network (CNN)	145	7	Combinatorial Optimization	43
3	Supply Chain Management	70	8	Integer Programming	42
4	Social Media	66	9	Optimization	41
5	Data Envelopment Analysis	62	10	Feature Extraction	40

کلیدواژه‌های انتخاب شده در دسته‌بندی فناوری اطلاعات در پایگاه WOS در بازه زمانی بین ۲۰۱۱ تا ۲۰۱۹ میلادی مورد

1. Document-Term-Martix

جستجو قرار گرفت و تعداد ۸۶۹،۷۶۳ رکورد بازیابی شد. با توجه به جستجوی بیش از ۸۰۰ هزار مدرک علمی در این پژوهش، تصمیم بر آن شد که بر اساس ارتباط پژوهش‌ها با حوزه جستجو ۱۰ هزار سند مرتبط انتخاب شوند. این انتخاب بر اساس تعداد کلمات جستجوی موجود در عنوان، میزان ارجاع و شباهت باشد. معیار شباهت بر اساس تعداد کلمات مورد جستجو در عنوان و کلیدواژه است (Web of Science, 2020). در تحقیق جاری از همین معیار برای استخراج مقالات و جستجو در پایگاه استفاده شد. فراوانی سالانه مدارک انتخابی در بازه زمانی ۲۰۱۱ تا ۲۰۱۹ در شکل ۱ آمده است.

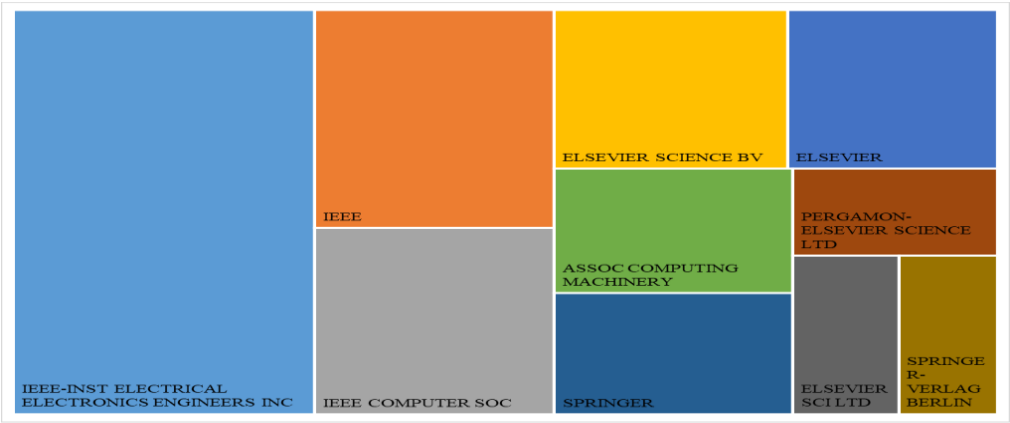


شکل ۱. فراوانی مقالات استخراج شده در سال‌های ۲۰۱۱ تا ۲۰۱۹

Figure 1. Frequency of extracted articles from 2011 to 2019

کمتر بودن منابع در سال‌های اخیر به‌دلیل عدم تکمیل مستندات و عدم نمایه شدن کامل مطالعات در سال‌های اخیر و زمان‌بر بودن آن است. با توجه به تغییرات سریع حوزه‌ی فناوری اطلاعات و احتمال تغییر در دسته‌بندی موضوعات، محدودیتی در جستجوی مقالات بر اساس سال اعمال نشده و نمونه‌گیری کاملا بدون محدودیت اتجام پذیرفت. با توجه به تحلیل توصیفی بر مبنای سال (مشاهده میزان حضور هر کلیدواژه در سال نسبت به دیگر کلیدواژه‌های همان سال) تعداد متفاوت نمونه در سال‌های مختلف، خللی بر نتایج پژوهش جاری ایجاد نمی‌کند.

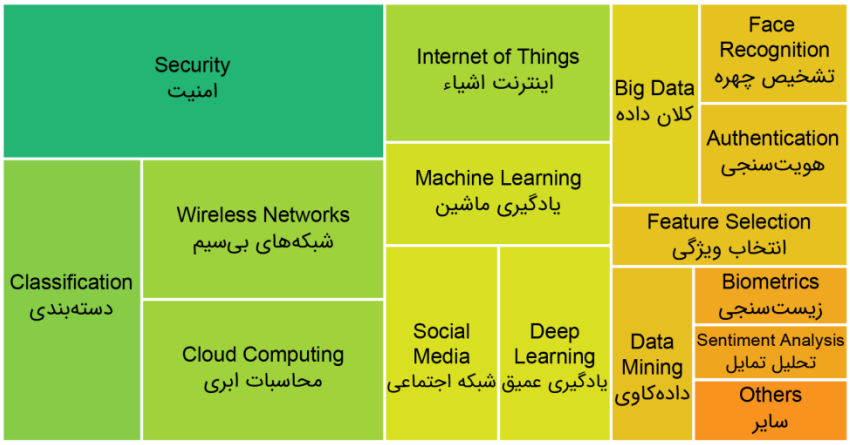
سهم ۱۰ دسته‌بندی انتشارات برتر که بیش از ۸۰ درصد پژوهش‌های منتخب را منتشر نموده‌اند، در شکل ۲. نمایش داده شده‌است:



شکل ۲. انتشارات فعال در زمینه فناوری اطلاعات

Figure 2. Active publications in the field of information technology

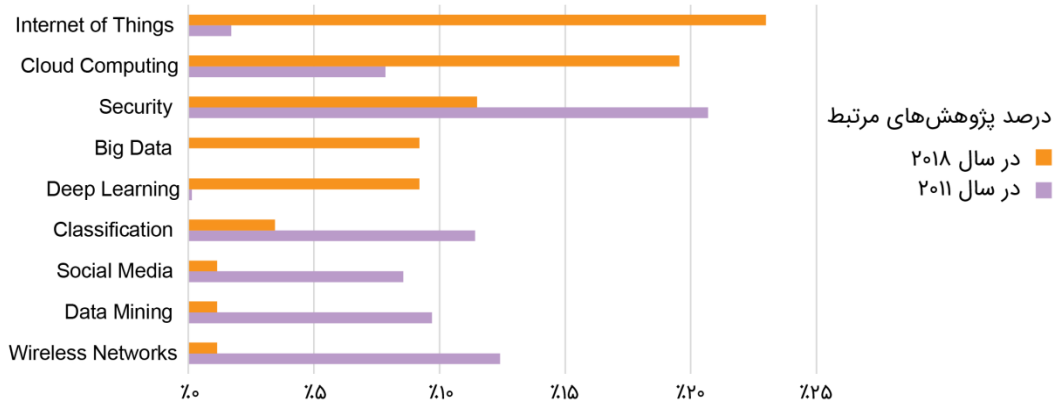
از این میان بیش از ۱۲۰ مجله، (نزدیک به نیمی از مجلات) در مجلات زیر مجموعه‌ی Elsevier و IEEE قرار داشتند که نشان می‌دهد این دو ناشر بیشترین میزان از مقالات حوزه فناوری اطلاعات را پوشش می‌دهند. در شکل زیر می‌توان میزان تکرار کلیدواژه‌های پرتکرار (به نسبت دیگر کلیدواژگان) در زمینه فناوری اطلاعات در دوره ده ساله مورد بررسی را مشاهده نمود.



شکل ۳. سهم کلیدواژه‌های پرتکرار در زمینه فناوری اطلاعات در دوره ۲۰۱۱ تا ۲۰۱۹

Figure 3. Share of frequently used keywords in the field of information technology in the period 2011 to 2019

اما این تکرار با توجه به ماهیت پویای فناوری اطلاعات، در خلال این مدت همواره در حال تغییر بوده است. همانگونه که در شکل زیر مشخص شده است، در طی بازه زمانی مورد بررسی، تمرکز مطالعات بر کلیدواژه‌های مختلف متفاوت شده است و از شبکه‌های بی‌سیم در سال ۲۰۱۰ به سمت اینترنت اشیا در سال ۲۰۱۸ (و مشابه آن در سال ۲۰۱۹)، سوق پیدا کرده است.

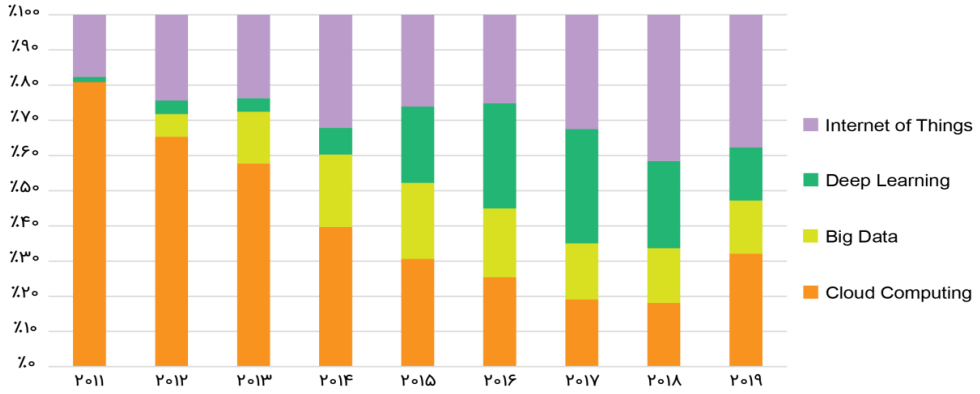


شکل ٤. مقایسه درصد پژوهش‌های مرتبط با کلیدواژه‌های پرکاربرد در دو بازه‌ی زمانی

Figure 4. Comparison of the percentage of research related to frequently used keywords in two time periods

اعداد افقی در نمودار میانی، نشان‌دهنده درصد مطالعات در ۹ کلیدواژه برتر در بازه ۲۰۱۱ تا ۲۰۱۹ نسبت به یکدیگر است. به طور مثال، همانگونه که مشاهده می‌شود در بین پژوهش‌های شامل این ۹ کلیدواژه در سال‌های اخیر، بیش از ۲۰ درصد پژوهش‌ها به اینترنت اشیا معطوف بوده است. این میزان در سال های گذشته متعلق به حوزه شبکه‌های بیسیم با حدود ۱۲ درصد بوده است. اگر این درصد را بخواهیم به چهار حوزه‌ی نوظهور محدود کنیم. در نمودار پایین و در محور عمودی می‌توان نگاهی دقیقتر به درصد

پوشش تحقیقات در این چهار حوزه داشت. به طور مثال میان تحقیقاتی که در سال‌های اخیر در این حوزه تعریف شده است خوزه اینترنت اشیا در حال رشد و حوزه محاسبات ابری در حال افول است (میان این چهار حوزه).



شکل ۵. مقایسه جایگاه نسبی ٤ کلیدواژه پرکاربرد

Figure 5. Comparison of the relative position of 4 frequently used keywords

همان‌گونه که مشاهده می‌شود رشد این فناوری اطلاعات بویژه در موضوعات اینترنت اشیا و یادگیری عمیق بسیار مشهود است و موضوعاتی مانند اینترنت اشیا همچنان در حال رشد است. تعداد کمتر مشاهدات علمی در سال ۲۰۱۹ به علت کامل نشدن تمام انتشارات (یا آنلاین نشدن آن‌ها) است. در ادامه با درنظر گرفتن روابط بین کلیدواژه‌ها (رابطه وجود در یک مقاله) و بدون توجه به دسته‌بندی‌های موضوعی خود پایگاه به بررسی دسته‌بندی بر اساس کمترین احتمال پراکندگی با روش مدل‌سازی موضوعی می‌پردازیم. الگوریتم‌های مدل‌سازی موضوعی با حداقل دخالت انسان این فرایند را انجام می‌دهند و همین نکته باعث جذابیت و استفاده روزافزون آن‌ها در تحلیل حجم عظیمی از متون شده است. در این روش بجای استفاده از عناوین از قبل تعیین شده، تنها با تعیین تعداد موضوعات، زیرمجموعه‌های موضوعی استخراج شده و میزان ارتباط هر سند (شامل مقالات منتشر شده در مجلات پایگاه) با هر زیرمجموعه نیز تعیین می‌شود (John & Petko, 2013).

هر سند از تعدادی واژه تشکیل شده است و میزان ارتباط یک سند با یک موضوع، به میزان تکرار کلمات مرتبط با آن موضوع در سند وابسته است. به عبارتی اگر میزان اهمیت یا وزن یک کلمه در یک سند را با w نشان دهیم و یک سند شامل m واژه باشد که هر یک وزن و اهمیت خاص خود را برای آن سند دارند و با فرض اینکه پیکره متنی موردنظر شامل n مستند است. آنگاه LDA می‌تواند تعداد k موضوع را شناسایی کند به‌نحوی که (Blei, 2012):

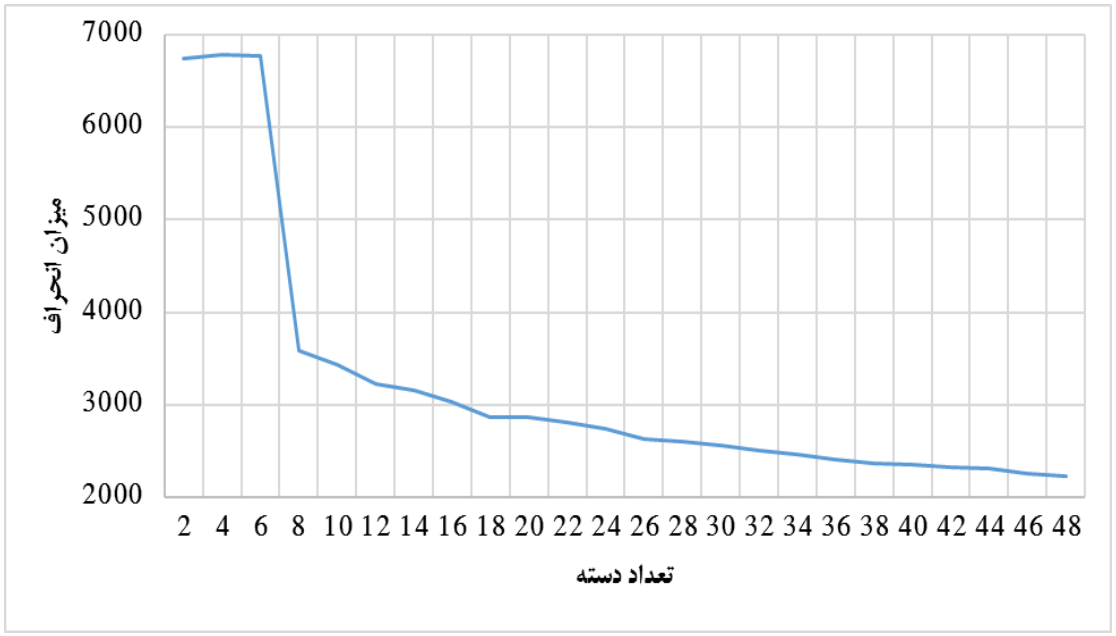
رابطه ۱:

$$\begin{array}{c} \\ Term_1 \\ Term_2 \\ \vdots \\ Term_m \end{array} \begin{array}{c} \begin{array}{cccc} Doc_1 & Doc_2 & \dots & Doc_n \end{array} \\ \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & \dots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,2} & \dots & w_{m,n} \end{bmatrix} \end{array} \xrightarrow{LDA(BoW(Term,Doc),k)}$$

$$\begin{array}{c} \\ Topic_1 \\ Topic_2 \\ \vdots \\ Topic_k \end{array} \begin{array}{c} \begin{array}{cccc} Doc_1 & Doc_2 & \dots & Doc_n \end{array} \\ \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \dots & \theta_{1,n} \\ \theta_{2,1} & \theta_{2,2} & \dots & \theta_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{k,1} & \theta_{k,2} & \dots & \theta_{k,n} \end{bmatrix} \end{array} * \begin{array}{c} \\ Term_1 \\ Term_2 \\ \vdots \\ Term_m \end{array} \begin{array}{c} \begin{array}{cccc} Topic_1 & Topic_2 & \dots & Topic_k \end{array} \\ \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \dots & \beta_{1,k} \\ \beta_{2,1} & \beta_{2,2} & \dots & \beta_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{m,1} & \beta_{m,2} & \dots & \beta_{m,k} \end{bmatrix} \end{array}$$

همان‌طور که در رابطه آمده است، نتیجه مدل‌سازی موضوعی، شناسایی k موضوع است که در هر موضوع کلمات وزن مختص خود را پیدا کرده‌اند. در خروجی حاصل، وزن هر یک از مستندات در یک موضوع مشخص می‌شود. هدف اصلی از مدل‌سازی موضوعی شناسایی موضوعات مستندات (ماتریس k*n) است و ازآنجاکه تعداد موضوعات (k) عددی بسیار کوچک‌تر از تعداد واژه‌ها (m) و تعداد مستندات (n) است، استفاده از ماتریس‌های با اندازه k*n برای شرح وضعیت و تحلیل مستندات، بسیار قابل‌فهم‌تر خواهد بود.

مدل‌سازی با تمامی مقادیر نمونه مورد بررسی در این تحقیق انجام شده است. پارامتر ورودی مهم LDA، مقدار k است که تعداد موضوعات مدنظر را نشان می‌دهد (Hall et al., 2008) . مزیت انتخاب تعداد موضوعات زیاد در این پژوهش، این است که تمامی حوزه‌های موضوعی پژوهش پوشش داده خواهد شد، اما به علت کاربردی بودن این حوزه در علوم دیگر ممکن است حوزه‌های غیر مرتبط با فناوری اطلاعات نیز جایی برای خود در بین موضوعات پیدا کنند. از سوی دیگر، مزیت تعداد حوزه‌های محدود این است که خبرگان و سیاست‌گذاران درک بهتری از آن خواهند داشت و تفسیر و تحلیل موضوعات راحت‌تر خواهد بود. بر اساس ادبیات، میزان درست یا مناسب دسته‌ها می‌تواند از ملاک‌های زیادی پیروی کند. یکی از این معیارها سرگشتگی[1] استکه یک معیار اندازه‌گیری در مدل‌های آماری است که برای بررسی میزان مناسب بودن توزیع احتمالی یک نمونه مورد استفاده قرار می‌گیرد. در این پژوهش برای ارزیابی مدل‌سازی موضوعی از سرگشتگی اسناد مشاهده نشده استفاده شد که تابعی نزولی از لگاریتم احتمال وقوع اسناد مشاهده نشده است و هر چه مقدار آن کمتر باشد نشان می‌دهد که مدل‌سازی قابلیت تعمیم‌پذیری بالاتری دارد (Battisti et al., 2015). به‌صورت معمول کمترین میزان برای انتخاب اندازه دسته مناسب برگزیده می‌شود. شکل زیر میزان این شاخص را برای تعداد دسته‌های ۲ تا ۵۰ نشان می‌دهد.



شکل ٦. میزان انحراف سرگشتگی با توجه به اندازه دسته‌بندی

Figure 6. Confusion deviation rate according to category size

1. Perplexity (Battisti, Ferrara, & Salini , 2015).

همانگونه که از شکل ۶. بر می‌آید، میزان سرگشتگی در تا تعداد دسته ۸ به شدت کاهش یافته و پس از آن میزان کاهش سرگشتگی با افزایش تعداد دسته شیب کمتری دارد. بر این اساس، تعداد گروه ۸ که نقطه شکست نمودار است، به عنوان تعداد موضوعات در مدل‌سازی انتخاب شد. بر این اساس ۸ موضوع استخراج شد که فهرست ۱۰ عبارت پرکاربرد در هر یک از این دسته‌ها در جدول ۴ آمده است. با مشورت با خبرگان و بر اساس عبارت‌های پرکاربرد در هر یک از موضوعات نام پیشنهادی هر موضوع انتخاب شده است.

جدول ٤. عبارت‌های پرکاربرد در موضوعات استخراج شده

Table 4. Frequently used terms in extracted topics

موضوع	نام پیشنهادی خبرگان	عبارت‌های پرکاربرد
موضوع ۱	یادگیری و انتخاب ویژگی	Cloud computing, Face recognition, Online learning, Data mining, Internet of things, Text mining, Classification, Privacy, Feature extraction, Wireless Sensor Network
موضوع ۲	اینترنت اشیا	Wireless sensor networks, Deep learning, Machine learning, Big data, Face recognition, Feature selection, Privacy, Clustering, Neural networks, Recommender Systems,
موضوع ۳	شبکه	Wireless sensor networks, Deep learnig, Internet of things, Cloud computins, Machine learning, Authentication, Multi-task learning, Neural networks, Adaptive control
موضوع ۴	رمزنگاری و امنیت	Classification, Security, Face recognition, Wireless sensor networks, Privacy, Internet of things, Cloud computing, Internet of things (iot), Pattern recognition, Cryptography
موضوع ۵	کاربردهای فناوری اطلاعات و اینترنت اشیا	Machine learning, Cloud computing, Data mining, Feature selection, Feature extraction,Authentication,Sentiment analysis,Social media,Internet of things (IOT) ,Extreme Learning Machine
موضوع ۶	پردازش تصویر	Feature selection, Pattern recognition,biometrics,computer vision, Authentication, Security, Deep learning, Internet of things, Support vector machine, Measurement
موضوع ۷	کلان‌داده	Cloud computing, Big data, Privacy,Internet of things, Deep learning, Security, Social media, Multi-task learning, Sparse representation, Anonymity
موضوع ۸	یادگیری ماشین	Machine learning, Security, Deep learning, Privacy, Feature selection, Wireless sensor networks, Cloud computing, Image classification, Internet of things (IOT), Face Recognition

نکته قابل توجه در عبارت‌های جدول ۴ این است که این عبارت‌ها در طول زمان در حال تغییرند. به طور مثال در دسته سوم (شبکه) در ابتدای بازه زمانی عبارت‌های زیر نقش پررنگ‌تری داشتند:

- Algorithms
- Algebraic coding
- Clustering
- Location privacy
- Wireless sensor networks

ولی در انتهای بازه زمانی، تعدد عبارت‌های زیر در این موضوع در حال افزایش بوده است:

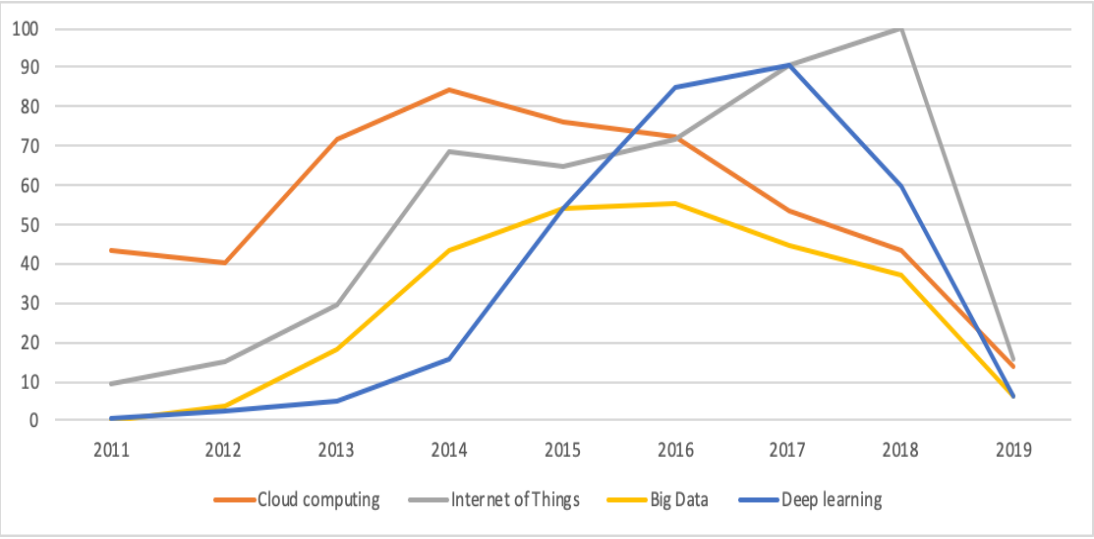
- Hybrid machine learning
- multi-task learning
- Cloud storage
- Spam filtering

با توجه به مباحث فوق می‌توان دریافت که در زمینه فناوری اطلاعات موضوعات در حال ترکیب و در بین دسته‌های مختلف فناوری اطلاعات در حال تغییر است. همچنین عبارت‌های پرکاربرد بصورت پیوسته و با سرعت درحال تغییر هستند.

بحث و نتیجه‌گیری

در این مقاله به این مسئله پرداختیم که سبد واژگان حوزه‌های فناوری اطلاعات در سطح بین‌المللی شامل چه کلیدواژگانی است وآیا این کلیدواژگان همواره در بازه زمانی یکسان بوده یا دارای افت و خیز هستند. رویکرد حل این مسأله تحلیل متن بوده که در آن به دنبال یافتن کلمات یا عباراتی بودیم که بتواند محتوا و ساختار (روابط) موجود در داده‌ها را توضیح دهد.

کاربرد روش‌شناسی فوق در این مقاله در بررسی و مطالعه حوزه‌ی فناوری اطلاعات در مطالعات بین‌المللی بوده است. در این بررسی از روش‌های مختلفی برای دستیابی به نتایج استفاده شد. در بخش اول و در انتخاب مجلات پیشرو در زمینه فناوری اطلاعات، بسیاری از پژوهش‌ها بر توصیفاتی مانند ضریب تاثیر یا قدمت مجله بسنده نمودند؛ اما در این پژوهش توجه به میزان استنادهای مجلات پیشرو (اچ ایندکس بالا)، قدمت مجله و حوزه‌ی مورد ادعای پوشش مجله بود که این اطلاعات به عنوان داده تکمیلی در اختیار خبرگان قرار گرفت تا با کمک آن مجلات معتبر و مدنظر انتخاب شوند. همچنین برای تشکیل سبد واژگان کلیدی برگزیده، از نظرات خبرگان استفاده شد و کلیدواژه‌های مستخرج از مجلات معتبر، دوباره مورد پالایش قرار گرفتند. این گام باعث شد اعتبار سبد واژگان اولیه بیشتر شود ولی از سوی دیگر با افزایش تعداد کلیدواژه‌های استخراج شده، نیاز به پاک‌سازی و انتخاب واژهای اولیه برای جستجو در کل پایگاه بیشتر نمود پیدا کرد. نتایج روندهای رشد و نموی کلیدواژگان، اطلاعات مناسبی را در اختیار محققین این پژوهش قرار داد. با توجه به روش پژوهش جاری و با مقایسه با پژوهش‌های گذشته به نظر می‌رسد پیش‌بینی رشد پژوهش‌ها در زمینه یادگیری ماشین همچنان ادامه‌دار خواهد بود. این پیش‌بینی در پژوهش Abuhay et al. (2018) بر اساس داده‌های بین سال‌های ۲۰۰۱ تا ۲۰۱۷ مورد ادعا قرار گرفت. البته برخلاف ادعای ایشان، پیش‌بینی روند صعودی واژه‌های مرتبط با امنیت، در پژوهش جاری مورد اثبات قرار نگرفت. شاید کاربردهای غیرفناوری این واژه در دیگر علوم مانند امنیت فیزیکی و جغرافیایی، باعث اختلاف در نتیجه جاری باشد. زیرا در روش‌شناسی انتخاب سبد برای مدل‌سازی موضوعی (در این پژوهش) واژه‌های مرتبط با دیگر واژگان فناوری مورد ارزش و اهمیت قرار می‌گیرد. پس همانگونه که در بخش نتایج مشاهده شد، کلیدواژه‌های نوظهوری همانند اینترنت اشیا و محاسبات ابری و داده‌های عظیم در کنار حوزه‌های کاری مانند یادگیری ماشین و یادگیری عمیق تعریف حوزه‌های فناوری اطلاعات را در بازه زمانی جدید تشکیل داده‌اند. دراینجا باتوجه به نیاز به مقایسه پیش‌بینی این ۴ کلیدواژه، درصد حضور این کلیدواژه‌ها در تمامی مقالاتی که دارای ۱۰ کلیدواژه پرتکرار استخراج شده هستند ترسیم شده است. در نتیجه رشد حضور ۴ کلیدواژه جدید اینترنت اشیا، محاسبات ابری، داده‌های بزرگ و یادگیری عمیق را می‌توان در سال‌های منتهی به ۲۰۱۷ میلادی مشاهده نمود:



شکل ۷. روند رشد ٤ حوزه نوظهور کشف شده در پژوهش

Figure 7. Growth trends of 4 emerging areas discovered in the research

همچنین در این مقاله به کمک مدل‌سازی موضوعی ۸ موضوع در حوزه فناوری اطلاعات شناسایی شد که تحلیل آن‌ها نشان داد که وجود حوزه‌هایی مانند امنیت، همواره در فناوری اطلاعات مطرح بوده است. اما کلیدواژه‌های نوظهوری مانند اینترنت اشیا نه تنها در میان کلیدواژگان دیده می‌شوند بلکه توانسته‌اند موضوعات مختص خود را نیز داشته باشند. موضوعاتی مانند اینترنت اشیا و کاربردهای فناوری اطلاعات در علوم دیگر موضوعاتی است که با طیف وسیعی از کلیدواژه‌های استفاده شده مانند داده‌های عظیم و یادگیری ماشین و شبکه‌های اجتماعی عجین شده است.

در پژوهش لی و همکاران ایده‌های جدید در موضوع پایداری و رسانه‌های اجتماعی موضوعات رفتار مشتریان شبکه‌های مجازی، توریست و زنجیره‌تامین این شبکه‌ها به عنوان موضوعات دارای رشد در این حوزه مطرح شد (Lee et al., 2018). این تناسب در موضوع ۵ استخراج شده با عنوان کاربردهای فناوری اطلاعات و اینترنت اشیا در پژوهش جاری نمایان شد. استفاده از فناوری اطلاعات در زمینه‌های منبع‌یابی (به‌طور مثال در جذب توریست) و بهره‌گیری از دانش ذخیره شده از رفتار کاربران شبکه‌های اجتماعی، در این موضوع به چشم می‌خورد.

همچنین در زمینه پژوهش‌های مرتبط با دسته‌بندی حوزه‌های سیستم های اطلاعاتی جیاراج و حسن‌زاده ادعا کردند که توسعه سیستم‌های اطلاعاتی، پذیرش فناوری اطلاعات و استفاده از سیستم‌های اطلاعاتی در طول زمان ماندگار شده بودند (Jeyaraj & Hassan Zadeh, 2019). این درحالی است که دسته‌بندی موضوعی این پژوهش، اینترنت اشیا را به‌عنوان نماینده سیستم‌های اطلاعاتی معرفی کرده و استفاده از داده‌های عظیم را در سیستم‌های پیشنهاد دهنده، به‌عنوان موضوع روز حوزه فناوری درنظر گرفته است. دلیل دیگر این تفاوت را می‌توان از پژوهش شارما و همکارانش در حوزه مدیریت اطلاعات نیز دریافت (Sharma et al., 2021). اگر نگاهی به موضوعات حوزه‌ی اطلاعات در پژوهش ایشان داشته‌باشیم مشخص می‌شود که حوزه اطلاعات در علوم کاربردی بهره‌وری مالی و بانکی، کسب‌وکار بین‌الملل و نوآوری صنعتی توسعه یافته است و پژوهش صرف در زمینه اطلاعات نمود کمرنگتری دارد.

نتاج بدست آمده در مورد روند پژوهش‌های موضوع محاسبات ابری در مقایسه با نتایجی که پیشتر ربیعی و همکاران در این حوزه بدست آورده بودند نشان می‌دهد که این حوزه گرچه تا اواسط دهه پیش (سال ۲۰۱۴) با رشد مواجه بوده اما از آن زمان به بعد مطالعات این حوزه رو به کاهش است (Rabiei et al., 2021) ، همچنین مقایسه این دو پژوهش نشان می‌دهد که رویکرد مدیریت داده‌ها به سمت تحلیل کلان داده[1] و تحلیل‌های مبتنی بر یادگیری ماشین تغییر کرده است.

در انتهای می‌توان ذکر نمود که استخراج موضوعات این پژوهش از بین طیف وسیعی از کلیدواژه‌ها و موضوعات استخراج شده می‌تواند روند آتی فناوری اطلاعات و کاربردهای آن را نمایان سازد. همچنین مقایسه روندهای جهانی در حوزه فناوری اطلاعات با پژوهش‌های داخلی در همین حوزه می‌تواند در پژوهش‌های آتی مورد توجه قرار گیرد و بوسیله این مقایسه می‌توان فاصله پژوهش‌های کشور را با جهان در حوزه فناوری اطلاعات مقایسه کرد و سیاست‌گذاران پژوهشی و مدیران کشور را در تعیین سمت و سو و اولویت‌گذاری پژوهش‌ها یاری رساند.

تعارض منافع

هیچ گونه تعارض منافع از سوی نویسندگان گزارش نشده است.

References

Abuhay, T. M., Nigatie, G. Y., & Kovalchuka, S. V. (2018). Towards Predicting Trend of Scientific Research Topics using Topic Modeling. *Procedia Computer Science, 136*(5), 304–310. DOI:10.1016/j.procs.2018.08.284

1 . Big Data

Battisti, F. D., Ferrara, A., & Salini , S. (2015). A decade of research in statistics: A topic model approach. *Scientometrics, Springer*,*103*(2), 413–433. DOI:10.1007/s11192-015-1554-1

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84. https://doi.org/10.1145/2133806.2133826

Blei, D., Andrew, Y., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3(4-5)*, 993-1022. DOI:10.1162/jmlr.2003.3.4-5.993

Chang, J., & Blei, D. M. (2009). Relational topic models for document networks. *International conference on artificial intelligence and statistics*, *5*, 81-88. https://www.researchgate.net/publication/220321034_Relational_Topic_Models_for_Document_Networks

Choudhary, A., Oluikpe, P. I., Harding, J., & Carrillo, P. M. (2009). The needs and benefits of Text Mining applications on Post-Project Reviews. *Computers in Industry, 60*(9), 728-740. DOI:10.1016/j.compind.2009.05.006

Debortoli, S., Müller, O., Junglas, I. A., & Brocke, J. v. (2016). Text Mining For Information Systems Researchers: An Annotated Topic Modeling Tutorial. *Communications of the Association for Information Systems 39*(1), 110 – 135. DOI:10.17705/1CAIS.03907

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.

Figuerola, C., García Marco, F. J., & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. *Siecntometrics*, 112(12). DOI:10.1007/s11192-017-2432-9

Furner, J. (2015). Information science is neither. *Library Trends, 63(3)*, 362–377.

Ghanadinejad, F., Heidari, G., & Chin Pardaz, R. (2018). Content analysis of texts related to research priorities in information science. *Journal of Library and Information Science, 8 (1)*, 55-74. https://doi.org/10.22067/riis.v0i. (In Persian)

Gru, B., & Hornik, K. (2011). Topicsmodels: An R package for fitting topic models. *Journal of Statistical Software, 40(13)*, 1–30.

Gurcan , F., Cagiltay , N. E., & Cagi, K. (2020). Mapping Human–Computer Interaction Research Themes and Trends from Its Existence to Today: A Topic Modeling-Based Review of past 60 Years. *International Journal of Human–Computer Interaction*, 2-15.

Hall, D. L., Jurafsky, D., & Manning, C. D. (2008). *Studying the history of ideas using topic models.* EMNLP '08: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 363–371. DOI:10.3115/1613715.1613763

Homayounvala, E., & Jalalimanesh, A. (2012). Promoting Research Collaboration Based on Data Mining Techniques in Library Information Systems. *IADIS Collaborative Technologies*, *8*(1), 73-82. Https://www.researchgate.net/publication/258246307_PROMOTING_RESEARCH_COLLABORATION_BASED_ON_DATA_MINING_TECHNIQUES_IN_LIBRARY_INFORMATION_SYSTEMS. (In Persian)

Jalali Manesh, A. (2020). *Design and compilation of research charter, research institutes, and research groups of Iran Institute of Information Science and Technology [In Persian]*. Tehran: Irandoc.

Jeyaraj, A., & Hassan Zadeh, A. (2019). Evolution of Information Systems Research: Insights from Topic Modeling. *Information & Management*, *57*(4), 103207. DOI:10.1016/j.im.2019.103207

John, W, M., & Petko, B. (2013). Introduction-Topic models: What they are and why they matter . *Poetics 41*(6), 545-569. DOI: 10.1016/j.poetic.2013.10.001

Kherwa, P., & Bansal, P. (2018). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, *7*(24), 159623. DOI:10.4108/eai.13-7-2018.159623

Lee, J., Wood, J., & Kim, J. (2021). Tracing the Trends in Sustainability and Social Media Research Using Topic Modeling. *Sustainability 13*, 1-19.

Rabiei, M., Hosseini-Motlagh, S., & Haeri, A. (2017). Using text mining techniques for identifying research gaps and priorities: A case study of the environmental science in Iran. *Scientometrics, 110*(2), 815-842. (In Persian)

Rabiei, M., Hosseini-Motlagh, S., & Haeri, A. (2021). Evolution of IT, Management and Industrial Engineering research: A topic model approach. *Scientia Iranica, 28(3)*, 1830-1852. (In Persian)

Razeqi, N., & Aghajani, H. (2020). Future research of Iranian scientific products until 2030 using ARIMA model. *Library and Information Science Studies, 12(1)*, 13-34. (In Persian)

Robinson, K. A., Saldanha, I. J., & Mckoy, N. A. (2011). Development of a framework to identify research gaps from systematic reviews. *Journal of clinical epidemiology*, *64*(12), 1325-1330. DOI:10.1016/j.jclinepi.2011.06.009

Sajedinejad, A. (2019). *Comparison of research trends in the field of information technology in theses and dissertations of the country and its global trend using the method of text analysis.* Tehran: Irandoc.

Seddiqqi, M., & Jalalimnesh, A. (2013). Study of Research Trend in Knowledge Management Field (2001-2010) and Mapping its Structure. *Journal of Information Processing and Management*, *28*(2), 363-392. https://jipm.irandoc.ac.ir/article_699198.html. (In Persian)

Sharma, A., Rana, N. P., & Nunkoo, R. (2021). Fifty years of information management research: A conceptual structure analysis using structural topic modeling. *International Journal of Information Management 58*(5), 102316. DOI:10.1016/j.ijinfomgt.2021.102316

Steyvers, M., Smyth, P., Rosen-Zvi, M., & Griffiths, T. (2004). *Probabilistic author-topic models for information discovery.* International conference on Knowledge discovery and data mining, Seattle, WA, USA. 306–315. Proceedings of the tenth ACM SIGKDD. https://doi.org/10.1145/1014052.101408

Wallach, H. (2006). *Topic modeling: Beyond bag-of-words.* the 23rd International Conference on Machine Learning, 977-984. Pittsburgh, Pennsylvania, U.S.

Web of Science. (2020). *Keywords Plus® (Web of Science Core Collection and Current Contents Connect only)*. Retrieved 01 29, 2020, from http://images.webofknowledge.com/WOKRS534DR1/help/WOS/hp_advanced_search.html

Yau, C., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics, GTM special issue, 100*(3), 767-786.

Zare-Farashbandi, F., Koohkan, E., Rajabi, G., & Yousofian, S. (2020). Identify and prioritize the information needs of families of children with cancer. *Library and Information Science Studies, 12(1), 208-226.* https://doi.org/10.22055/slis.2019.29524.1603. (In Persian)

Copyrights

© 2024, by the authors. Licensee SCU, Ahvaz, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (https://creativecommons.org/licenses/by/4.0/)