Archive of SID.ir

# Identification of Hot Topics and Trends in Knowledge and Information Science, based on Text Mining Techniques

**Elahe Akhavanhariri**
Master of Knowledge and Information Science; M.A. in Knowledge and Information Science; Department of Knowledge and Information Science; University of Isfahan; Isfahan, Iran;
Email: Elaheakhavanhariri@yahoo.com

**Ali Mansouri***
PhD in Knowledge and Information Science; Associate Professor; Department of Knowledge and Information Science; University of Isfahan; Isfahan, Iran Email: a.mansouri@edu.ui.ac.ir

**Hossein Karshenas Najafabadi**
PhD in Artificial Intelligence; Assistant Professor; Faculty of Computer Engineering; University of Isfahan; Isfahan, Iran Email: h.karshenas@eng.ui.ac.ir

Iranian Journal of Information Processing and Management

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 38 | No. 2 | pp. 369-396
Winter 2023
https://doi.org/10.35050/JIPM010.2022.028

**Received: 06, Nov. 2021** **Accepted: 26, Apr. 2022**

**Abstract:** Identification of hot topics in research areas has always been of interest. Making smart decisions about what is needed to be studied is always a fundamental factor for researchers and can be challenging for them. The goal of this study is to identify hot topics and thematic trend analysis of articles indexed in Scopus database in the field of Knowledge and Information Science (KIS), between 2010 and 2019, using Text Mining techniques.

The population consists of 50995 articles published in 249 journals indexed in Scopus database in the field of KIS from 2010 to 2019. To identify thematic clusters, algorithms of Latent Dirichlet Allocation (LDA) technique were used and the data were analyzed using libraries in Python software. To do this, by implementing the word weighting algorithm, using the TF-IDF method, and weighting all of the words and forming a text matrix, the topics in the documents and the coefficients for assigning each document to each topic (theta) were determined. The output of the LDA algorithm led to the identification of the optimal number of 260 topics. Each topic was labeled based on the words with the highest weight assigned to each topic and with considering of experts' opinions. Then, Topic clustering, keywords and topics identification were done. By performing

* Corresponding Author

Archive of SID.ir

Archive of SID.ir

calculations with 95% confidence, 63 topics were selected from 260 main topics. By calculating the average theta in years, 24 topics with a positive trend or slope (hot topic) and 39 topics with a negative trend or negative slope (cold topic) were determined. According to the results, measurement studies, e-management/ e-marketing, content retrieval, data analysis and e-skills, are considered as hot topics and training, archive, knowledge management, organization and librarians' health, were identified as cold topics in the field of KIS, in the period 2010 to 2019. The analysis of the findings shows that due to the interest of the most researchers in the last 10 years in using of emerging technologies, technology-based topics have attracted them more. In contrast, basic issues are less considered to be developed.

**Keywords:** Hot Topics, Latent Dirichlet Allocation (LDA), Text Mining, Knowledge and Information Science

Archive of SID.ir

Archive of SID.ir

# شناسایی موضوعات داغ و روندها در علم اطلاعات و دانش‌شناسی با استفاده از تکنیک‌های متن‌کاوی

**الهه اخوان حریری**
دانش‌آموختهٔ کارشناسی ارشد علم اطلاعات و دانش‌شناسی؛ دانشگاه اصفهان؛ اصفهان، ایران؛
Elaheakhavanhariri@yahoo.com

**علی منصوری**
دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛ گروه علم اطلاعات و دانش‌شناسی؛ دانشکدهٔ علوم تربیتی و روان‌شناسی؛ دانشگاه اصفهان؛ اصفهان، ایران؛
پدیدآور رابط a.mansouri@edu.ui.ac.ir

**حسین کارشناس نجف‌آبادی**
دکتری هوش مصنوعی؛ استادیار؛ گروه هوش مصنوعی؛ دانشکدهٔ مهندسی کامپیوتر؛ دانشگاه اصفهان؛
اصفهان، ایران h.karshenas@eng.ui.ac.ir

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۴۰۰/۰۸/۱۵ | پذیرش: ۱۴۰۱/۰۲/۰۶ | مقاله برای اصلاح به مدت ۱۲ روز نزد پدیدآوران بوده است.

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۳۸ | شمارهٔ ۲ | ص‌ص ۳۶۹-۳۹۶
زمستان ۱۴۰۱
https://doi.org/10.35050/JIPM010.2022.028

**چکیده:** شناسایی موضوعات داغ در حوزه‌های پژوهشی همواره مورد توجه بوده است. تصمیم‌گیری هوشمندانه در مورد آنچه باید مورد مطالعه قرار گیرد، همواره عاملی اساسی برای محققان بوده و می‌تواند برای پژوهشگران موضوعی چالش‌برانگیز باشد. هدف پژوهش حاضر شناسایی موضوعات داغ در مقالات نمایه‌شده در پایگاه «اسکوپوس» در حوزهٔ علم اطلاعات و دانش‌شناسی در سال‌های ۲۰۱۰ تا ۲۰۱۹ با استفاده از تکنیک‌های متن‌کاوی است. جامعهٔ آماری شامل ۵۰۹۹۵ مقالهٔ منتشرشده در ۲۴۹ عنوان مجلهٔ نمایه‌شده در پایگاه «اسکوپوس» در این حوزه در بازهٔ زمانی ۲۰۱۰ تا ۲۰۱۹ است. برای شناسایی خوشه‌های موضوعی این حوزه از الگوریتم‌های تکنیک تخصیص «دیریکلت پنهان» (LDA) استفاده شد و داده‌ها با استفاده از کتابخانه‌های متن‌کاوی در نرم‌افزار «پایتون» مورد تحلیل قرار گرفت. برای این کار با اجرای الگوریتم وزن‌دهی لغات به روش TF-IDF و تشکیل ماتریس متنی، موضوعات در پیکرهٔ اسناد تعیین و ضرایب اختصاص هر سند به هر موضوع (تتا) مشخص شد. خروجی اجرای الگوریتم LDA منجر به شناسایی تعداد ۲۶۰ موضوع شد. تخصیص



Archive of SID.ir

Archive of SID.ir



برچسب به هر یک از خوشه‌های موضوعی بر اساس کلمات با بالاترین وزن اختصاص‌یافته به هر موضوع و با نظر کارشناسان حوزهٔ موضوعی انجام گرفت. با انجام محاسبات با ضریب اطمینان ۹۵ درصد، تعداد ۶۳ موضوع از بین ۲۶۰ موضوع اصلی انتخاب شد. با محاسبهٔ میانگین تتا بر حسب سال، تعداد ۲۴ موضوع دارای روند مثبت (موضوع داغ) و تعداد ۳۹ موضوع دارای روند منفی (موضوع سرد) تعیین شد. بر اساس نتایج، موضوعات مطالعات سنجشی، مدیریت الکترونیک/ بازاریابی الکترونیک، بازیابی محتوا، تحلیل داده و مهارت الکترونیک، از جملهٔ موضوعات داغ بود و موضوعات آموزش، آرشیو، مدیریت دانش، سازماندهی و سلامت کتابداران از جملهٔ موضوعات سرد در حوزهٔ علم اطلاعات و دانش‌شناسی در بازهٔ زمانی ۲۰۱۰ تا ۲۰۱۹ شناسایی شد. تحلیل یافته‌ها نشانگر آن است که به‌دلیل رویکرد اغلب پژوهشگران در ۱۰ سال گذشته به استفاده از فناوری‌های نوپدید، موضوعات مبتنی بر فناوری بیشتر باعث جذب آنان شده و در مقابل، موضوعات پایه‌ای این حوزه کمتر مورد توجه قرار گرفته‌اند.

**کلیدواژه‌ها:** موضوعات داغ، تخصیص دیریکلت پنهان، متن‌کاوی، علم اطلاعات و دانش‌شناسی

## ۱. مقدمه

بررسی گذشتهٔ علم بشری نشان‌دهندهٔ آن است که با گذشت زمان و پیدایش زمینه‌های جدید، زمینه‌های علمی و تحقیقاتی نوینی به‌وجود آمده و تکامل یافته (Lässig 2016) و با توسعهٔ جوامع بشری و پدید آمدن کشفیات جدید و مبتنی بر نیاز بشر، حوزه‌های علمی نیز گسترش یافته است. این تکامل و بالندگی به‌گونه‌ای بوده است که انقلاب فناوری ارتباطی و اطلاعاتی در آن سبب شده در عصر حاضر، شاهد رشد انبوهی از اطلاعات دیجیتال (Al-Azmi and Rashid 2013) در قالب‌های گوناگون از جمله مقاله، کتاب و غیره باشیم.

بر اساس تعداد انتشارات و مراجع نمایه‌شده در پایگاه‌های مختلف، آهنگ رشد علوم مدرن در سال ۲۰۱۰، نرخی برابر ۸ تا ۹ درصد داشته است. این آهنگِ رشد، هر ۹ سال تقریباً دو برابر می‌شود (Bornmann and Mutz 2015) و پیش‌بینی می‌شود که میزان انتشارات علمی جهانی در قالب مقاله در پایگاه «وب‌آوساینس» از حدود ۳ میلیارد به حدود ۱۰ میلیارد در سال ۲۰۴۰ افزایش یابد (Wang and Barabási 2020). همانند نرخ رشد علمی سریع و قابل توجه در سایر رشته‌های علمی، حوزه‌های موضوعی و پژوهشی در رشتهٔ علم اطلاعات و دانش‌شناسی نیز همواره در حال بالندگی و تکامل بوده و با گذشت زمان موضوعات فرعی بسیاری را در ذیل مجموعهٔ خود ایجاد می‌کنند؛ به‌گونه‌ای که در بعد جهانی، پژوهشگران این حوزه با توجه به روند روبه‌رشد نسبی انتشارات علمی سالیانه در



Archive of SID.ir

Archive of SID.ir



این حوزه (Vinay and Basavaraja 2019)، با حجم عظیمی از اطلاعات علمی و بروندادهای پژوهشی و موضوعات با حداکثر توجه با عنوان موضوعات داغ روبه‌رو هستند. موضوعات داغ در هر دورهٔ زمانی، آن دسته از زمینه‌های پژوهشی هستند که از اهمیت ویژه‌ای در میان زمینه‌های فعّال در یک عنوان پژوهشی مشخص برخوردار هستند (Erdfelder and Bosnjak 2016) و باعث جذب علاقه و توجه محققان به خود می‌شوند و ممکن است در میان موضوعات علمی موجود، بیشتر زمینه‌های پژوهشی را به خود اختصاص دهند (Burrill & Ben-Zvi 2019). از طرف دیگر، موضوعات سرد، زمینه‌های پژوهشی هستند که از مرکز توجهات به‌طور کلی خارج‌شده یا در مقایسه با موضوعات داغ، کمتر مورد توجه قرار می‌گیرند. این موضوعات، در مقابل نقاط داغ قرار دارند و ممکن است بعد از اتمام نیمه‌عمرشان، به دست فراموشی سپرده شوند (Gal et al. 2019). وجود این دو گروه از موضوعات در هر حوزهٔ علمی از جمله علم اطلاعات و دانش‌شناسی باعث رشد ناهمگون تحقیقات در یک موضوع خاص شده و اصل توازن میان موضوعات یک زمینهٔ پژوهشی خاص را برهم می‌زند. در همین راستا، شناسایی موضوعات داغ و سرد در این حوزه همواره مورد توجه بوده است، زیرا تصمیم‌گیری هوشمندانه در مورد آنچه باید مورد مطالعه قرار گیرد، توانایی شناسایی موضوعات داغ در هر حوزه، افزون بر هدایت صحیح برای انجام یک پژوهش موفق، باعث شناخت موفق‌ترین پژوهشگران و انتشارات آن حوزه شده و دلایل قانع‌کننده‌ای را برای خوانندگان در راستای ضرورت و اهمیت انجام یک پژوهش فراهم می‌آورد.

در همین راستا به‌منظور شناسایی توسعهٔ موضوعی رشتهٔ علم اطلاعات و دانش‌شناسی[1] در سطح بین‌المللی و تعیین موضوعات داغ و سرد به‌منظور شناسایی اولویت‌های پژوهشی و کمک به سیاست‌گذاری در حوزهٔ مورد نظر و همچنین با توجه به اهمیت و کاربرد گستردهٔ روش متن‌کاوی، پژوهش حاضر با استفاده از تکنیک‌های متن‌کاوی در صدد شناسایی حوزه‌های موضوعی اصلی علم اطلاعات و دانش‌شناسی و موضوع‌های داغ و سرد است.

۱. به‌منظور هماهنگی با پژوهش‌های بین‌المللی و مقایسهٔ نتایج، از این به بعد به‌جای علم اطلاعات و دانش‌شناسی از عنوان علوم کتابداری و اطلاع‌رسانی استفاده خواهد شد.



Archive of SID.ir

Archive of SID.ir



## ۲. پیشینهٔ پژوهش

در پژوهش‌های مختلف از روش‌های متفاوتی برای شناسایی موضوعات یک اثر یا مجموعه‌ای از آثار مرتبط با علوم کتابداری و اطلاع‌رسانی و علم اطلاعات و دانش‌شناسی استفاده شده است. از جملهٔ این روش‌ها، روش تحلیل هم‌واژگانی و هم‌رخدادی واژگان، تحلیل کتاب‌سنجی، تحلیل محتوا یا تحلیل موضوعی، تحلیل دامنه، تحلیل ترکیبی کتاب‌سنجی و هم‌استنادی، تحلیل ترکیبی کتاب‌سنجی و تحلیل دامنه، تحلیل ترکیبی هم‌واژگانی و تحلیل شبکه، روش تحلیل ترکیبی کتاب‌سنجی و تحلیل محتوا و روش مدل‌سازی موضوعی LDA است (باغ‌محمد، منصوری و چشمه‌سهرابی ۱۳۹۹). اما در دههٔ گذشته، به‌منظور کشف موضوعات، شناخت الگوی دانشی و تحلیل روند از روش متن‌کاوی از زیرمجموعهٔ داده‌کاوی به‌صورت گسترده در پژوهش‌های مختلف استفاده شده است. این روند نیز نرخ افزایشی دارد (Katsurai & Joo 2021; Jung & Lee 2020).

روش متن‌کاوی از جمله روش‌های بین رشته‌ای است که در سال‌های گذشته در پژوهش‌های مختلف به‌منظور کشف الگوی دانشی و موضوعات نهان و آشکار موضوعات و حوزه‌های مختلف موضوعات گوناگون از جمله سامانه‌های اطلاعاتی، مدیریت فناوری، آموزش و پرورش، علوم کتابداری و اطلاع‌رسانی، روان‌شناسی، جامعه‌شناسی، پزشکی، کامپیوتر و زبان‌شناسی و سایر حوزه‌ها مورد استفاده قرار می‌گیرد. در این نوع پژوهش‌ها پویایی و رشد علوم مختلف با استفاده از متن‌کاوی مورد کنکاش قرار می‌گیرد. از جمله موضوعاتی که با استفاده از متن‌کاوی مورد تحلیل قرار گرفته، علم رایانه و اطلاعات است (Kim et al. 2015) که چهار الگوی متفاوت در روندهای موضوعی این حوزه شناسایی شده و الگوهای رشدیافته، در حال افزایش یا بالغ (موضوعات مربوط به شبکه‌های رایانه‌ای)، الگوهای در حال کاهش (موضوعات مربوط به داده‌کاوی)، الگوهای در حال پیشرفت (وب، بازیابی اطلاعات با متن‌کاوی و موضوعات مرتبط با پایگاه‌های داده) و الگوهای در حال نوسان و تغییر (سامانه‌های اطلاعاتی و موضوعات مربوط به سامانه‌های چندرسانه‌ای) است. دیگر موضوعی که مورد توجه پژوهشگران حوزهٔ داده‌کاوی و متن‌کاوی قرار گرفته روان‌شناسی است (Bitterman & Fischer 2018) که با استفاده از لغات کلیدی آن حوزه به شناسایی موضوعات اصلی و داغ و روند آن‌ها پرداخته است. حوزهٔ پزشکی یکی از حوزه‌هایی است که به‌صورت گسترده از این تکنیک برای شناسایی موضوعات و تحلیل روند استفاده کرده است. حوزهٔ بیماری‌های قلب (Gal et al. 2019) بیماری کوید-۱۹



Archive of SID.ir

Archive of SID.ir



Wang & Lo 2021; Carracedo, Puertas & Marti 2021; Tworowski et al. 2021; Zengul et al.) (2021; حوزهٔ سلامت و مراقبت‌های بهداشتی (Popowich 2005)؛ و حوزهٔ بیماری (Chen et al. 2020) نمونه‌های کوچکی از این پژوهش‌ها هستند. رسانه‌های اجتماعی (Manoharan 2020)؛ نانوتکنولوژی (Li et al. 2020) از دستهٔ دیگر پژوهش‌هایی است که از این تکنیک استفاده کرده‌اند. در ایران نیز در سال‌های گذشته، این تکنیک در پژوهش‌های مختلفی مورد استفاده بوده است. همچنین، یادگیری الکترونیکی (کیهانی و آقا کاردان ۱۳۹۱)، قرآن کریم (صالحی شهرودی، مینایی و اشرفی ۱۳۹۲؛ اصلانی و اسماعیلی ۱۳۹۷؛ آذری، شکراللهی‌فر و لسانی فشارکی ۱۳۹۹)، مدیریت دانش (رضایی نور و شیخ‌بهایی ۱۳۹۶)، خوشه‌بندی مقالات (سلیمانی‌نژاد، سلاجقه و طیبی ۱۳۹۷)، تشخیص حس در متن‌های فارسی (گرشاسبی، رئیس روحانی و کاباران‌زاده قدیم ۱۳۹۷)، اینترنت اشیا (برنا، فتحی و مؤمنی ۱۳۹۷)، و حوزهٔ سلامت (شکوهیان و همکاران ۱۳۹۸) فهرستی کوتاه از پژوهش‌هایی است که به‌صورت گسترده از تکنیک متن‌کاوی استفاده کرده‌اند.

در حوزهٔ علوم کتابداری و اطلاع‌رسانی انتخاب مجلات مناسب (Jung and Lee 2020)، سازماندهی اطلاعات (پرئی و حمیدی ۱۳۹۶)، بازیابی اطلاعات (جعفری پاورسی و همکاران ۱۳۹۸) در کنار پژوهش‌هایی که به‌طور ویژه به بررسی موضوعات داغ و سرد در این حوزه پرداخته‌اند، قرار دارند؛ اما با توجه به ماهیت پژوهش‌هایی که با استفاده از روش متن‌کاوی به بررسی موضوعات داغ و سرد در این حوزه می‌پردازند، تعداد محدودی از پژوهش‌ها به این موضوع پرداخته‌اند. «لی، کیم و کیم» در پژوهشی به بررسی موضوعات حوزهٔ کتابخانهٔ دیجیتال پرداخته و سه حوزهٔ کلیِ پژوهش‌های پایه، مطالعات کاربران، و فراداده را شناسایی کردند (Lee, Kim & Kim 2010). موضوعات داغ شناسایی‌شده در حوزهٔ علوم کتابداری و اطلاع‌رسانی کرهٔ جنوبی در یک فرایند زمانی طولانی (۱۹۷۰–۲۰۱۹) را ارزیابی کتابخانه در حوزهٔ مدیریت کتابخانه، خدمات اطلاعاتی در حوزهٔ خدمات کتابخانه و مدارک قدیمی در حوزهٔ کتاب‌شناسی تشکیل می‌دهند (Park & Song 2013).

پژوهش «باغ محمد، منصوری و چشمه‌سهرابی» و همکاران با عنوان «توسعه و روند موضوعی در حوزهٔ علوم کتابداری و اطلاع‌رسانی ایران» نشان داد که بیشتر موضوعات خدمات کتابخانه‌ای در شبکه‌های اجتماعی، مدل‌های پژوهش، سرمایهٔ اجتماعی، پایگاه‌های اطلاعاتی پزشکی، داده‌کاوی، روند تولید علمی، موضوع‌های بین رشته‌ای، الگوریتم‌های فضای مجازی، مدیریت دانش، مطالعات شبکه‌های اجتماعی، رویکردهای



Archive of SID.ir

Archive of SID.ir



پژوهشی و آینده‌پژوهی قرار دارند (۱۳۹۹). پژوهش «تیماکوم، کیم و سونگ» با بررسی ۶ مجلهٔ برتر حوزهٔ علوم کتابداری و اطلاع‌رسانی در محدودهٔ سال‌های ۱۹۹۷-۲۰۱۶ نشان داد که موضوعات کتابخانهٔ دانشگاهی، کتابخانهٔ دیجیتال، روش‌شناسی پژوهش، بازاریابی کتابخانه، بازیابی اطلاعات، اطلاعات دیجیتال و استناد به مدارک از جملهٔ موضوعات داغ بوده است (Timakum, Kim and Song 2020). تحلیل محتوایی آثار مرتبط با کتابداری پزشکی با استفاده از متن‌کاوی نشان داد که موضوعات علم‌سنجی، سواد اطلاعاتی و اطلاعات سلامت بیشترین انتشارات را دارند و در زمرهٔ موضوعات داغ هستند (Dastani, Ziaei & Delghandi 2020).

بررسی پیشینه‌های مرتبط با علوم کتابداری و اطلاع‌رسانی نشان می‌دهد که جامعهٔ مورد بررسی پژوهش‌هایی که در حوزهٔ کتابداری و اطلاع‌رسانی انجام شده، محدود بوده و اغلب مقالات مجلات خاص را پوشش داده یا موضوع خاص یا حوزهٔ جغرافیایی مشخصی را بررسی کرده‌اند. بررسی پیشینه‌ها به‌طور عام نیز نشان می‌دهد که بیشتر پژوهش‌هایی که با استفاده از روش متن‌کاوی به شناسایی موضوعات داغ و سرد و یا موضوعات اصلی این حوزه پرداخته‌اند، از روش مدل‌سازی موضوعی و از روش تخصیص «دیریکلت پنهان»[1] استفاده کرده‌اند. برای وزن‌دهی کلمات از روش TF-IDF استفاده شده و نتیجهٔ مشترک بیشتر آن‌ها این است که روش متن‌کاوی از زیرمجموعهٔ داده‌کاوی، تکنیک مناسبی برای کشف موضوعات و الگوی دانشی موجود در متون بدون ساختار علمی به‌ویژه در مقاله است. بنابراین، پژوهش حاضر با استفاده از تکنیک متن‌کاوی با هدف شناسایی موضوعات داغ و سرد حوزهٔ علوم کتابداری و اطلاع‌رسانی در سطح بین‌المللی انجام شده است.

## ۳. روش پژوهش

پژوهش حاضر از نظر هدف، کاربردی و از نظر ماهیت، توصیفی اکتشافی است. از روش متن‌کاوی و زیرمجموعهٔ آن، روش مدل‌سازی موضوعی، و الگوریتم LDA برای شناسایی موضوع‌های داغ و سرد استفاده شد. جامعهٔ آماری شامل تمامی مقالات منتشرشده در حوزهٔ علم اطلاعات و دانش‌شناسی (علوم کتابداری و اطلاع‌رسانی) است که در بازهٔ

1. latent Dirichlet allocation (LDA)



Archive of SID.ir

Archive of SID.ir



زمانی ۲۰۱۰ تا ۲۰۱۹ در ۲۴۸ مجلهٔ مرتبط با این حوزه در پایگاه «اسکوپوس» نمایه شده است. برای تحلیل داده‌ها، عنوان و چکیدهٔ مقالات استخراج و مورد تحلیل قرار گرفت.

بعد از گردآوری و آماده‌سازی داده‌ها و دیگر مراحل لازم برای تحلیل نتایج، به شرح فرایند زیر (نمودار ۱) انجام گرفت:

- **آماده‌سازی:** در این مرحله بر روی داده‌های ابتدایی و خام، عملیات پیش‌پردازش انجام شد؛ چراکه این مرحله در پژوهش‌هایی که مبانی آنها متن‌کاوی است، به‌عنوان اساس و پایه نام برده می‌شود و از اهمیت و ضرورت خاصی در فرایند پژوهش برخوردار است (Tseng & Lin 2003). این مرحله، شامل مراحل سه‌گانهٔ الف) توکن‌سازی به‌عنوان اولین مرحله از نمایه‌سازی متن، وظیفهٔ تکه‌تکه کردن متن به کلمات تشکیل‌دهندهٔ آن و تولید متن‌هایی که به زبان طبیعی نوشته شده‌اند، اعمال می‌شود (Taeho Jo 2019)، ب) استیمینگ[1]، یعنی تبدیل هر واژه یا عبارتی که از مرحلهٔ قبل تولید شده به شکل ریشه‌ای خود (Kowalski, Maybury 2000)، و ج) حذف کلمات نامعتبر[2] برای محاسبه و تعیین وزن کلمات در یک متن است. این مراحل باید به‌صورت سلسله‌مراتبی انجام شود:
- **وزن‌دهی کلمات:** بعد از پیش‌پردازش، برای وزن‌دهی به کلمات از روش TF-IDF که احتمال یا درصد وقوع هر کلمه در متن است، به‌عنوان معیار وزن‌دهی کلمات استفاده شد. در این مرحله کلماتی که دارای بیشترین وزن بودند، شناسایی و به‌عنوان کلمات کلیدی انتخاب شدند.
- **ایجاد ماتریس سند-کلمه:** برای تحلیل متن با استفاده از تکنیک‌های متن‌کاوی، نیاز است که متنِ بدون ساختار را به مجموعه‌ای از داده‌های ساختارمند تبدیل کرد. این فرمت استاندارد ماتریس سند-کلمه است که تعداد تکرار هر کلمه در هر سند مشخص می‌گردد. در این مرحله پس از تعیین وزن کلمات با استفاده از روش TF-IDF، ماتریس مورد نیاز ایجاد شد.
- **تعیین خوشهٔ موضوعی:** وزن کلمات شناسایی‌شده از متن نسبت به همدیگر متفاوت بوده و این تفاوت ممکن است نسبت به هر خوشه‌ای که کلمه به آن تعلق می‌گیرد، متفاوت باشد. خوشهٔ موضوعی شامل تعدادی کلیدواژه است که به‌نوعی با همدیگر

1. steaming
2. stop list



Archive of SID.ir

Archive of SID.ir



ارتباط دارند. به‌منظور تعیین رابطه بین کلیدواژه‌ها و تعلق هر کلیدواژه به خوشهٔ موضوعی از روش مدل‌سازی موضوعی و به‌طور خاص از الگوریتم LDA استفاده شد. LDA یکی از معروف‌ترین و پرکاربردترین روش‌های مدل‌سازی موضوعی است. در این روش فرض بر این است که هر مدرک شامل موضوعاتی است و هر موضوع از مجموعه‌ای از کلیدواژه‌ها شکل گرفته است (Silge & Robinson 2017). پس از استخراج کلیدواژه‌ها و تعیین خوشه‌ها، ۱۵ کلمهٔ برتر هر موضوع (خوشه) شناسایی و با کمک متخصصان حوزهٔ موضوعی به این خوشه‌ها برچسب موضوعی مناسب داده شد. به‌منظور تعیین تراکم کلمات در یک موضوع، از کمیت تتا استفاده شد. در این مرحله، میانگین تتا (میزان توزیع موضوعات در اسناد بر اساس سال انتشار آن‌ها) با اعمال آلفای برابر ۰/۰۵ (وزن موضوعات در اسناد) محاسبه شد و موضوعات دارای بالاترین مقدار میانگین تتا به‌عنوان موضوعات داغ و موضوعات با کمترین مقدار میانگین تتا در طول ۱۰ سال به‌عنوان موضوعات سرد شناسایی شد. برای اجرای محاسبات و استخراج یافته‌ها از زبان برنامه‌نویسی «پایتون» و مجموعهٔ کتابخانه‌های مرتبط با داده‌کاوی و متن‌کاوی در این زبان برنامه‌نویسی و به‌طور خاص از کتابخانهٔ «جنسیم»[1] (Rehurek & Sojka 2010) استفاده شد.

برای خوشه‌بندی از الگوریتم K-میانگین[2] استفاده شد. خوشه‌بندی با استفاده از الگوریتم k-میانگین مهم‌ترین و کاربردی‌ترین نوع خوشه‌بندی از نوع بخش‌بندی است که استفاده از آن برای مجموعهٔ داده‌های بزرگ مناسب است و الگوریتم آن با تعدادِ از قبل مشخص‌شده‌ای از خوشه‌های اسناد (به‌عنوان مثال، تعداد خوشه‌های k) شروع می‌شود (Salton & Buckley 1988). در این روش با طی مراحل خوشه‌بندی تفکیکی، هر شیء یا عضو یا مشاهده‌گر، فقط و فقط به یک خوشه تعلق خواهد داشت و هیچ خوشه‌ای بدون عضو نخواهد بود. در این روش خوشه‌بندی عملیات خوشه‌بندی بر اساس n مشاهده‌گر و k گروه صورت می‌گیرد. از این رو، الگوریتم K-میانگین فاصلهٔ داخل خوشه را به حداقل می‌رساند (Hamerly & Elkan 2002). الگوریتم این روش با تعداد K مرکزیت برای خوشه‌ها شروع می‌شود. سپس، هر الگو در مجموعهٔ داده‌ها به نزدیک‌ترین خوشه (نزدیک‌ترین مرکزیت) اختصاص داده می‌شود. سرانجام، مرکزیت‌ها طبق الگوهای مرتبط با آن

1. Gensim
2. K-means



Archive of SID.ir

Archive of SID.ir



مجدد محاسبه می‌شوند. این روند تا رسیدن به یک همگرایی مشخص (رسیدن به یک معیار عددی که به‌طور معمول، توسط کاربر به الگوریتم داده می‌شود؛ مانند معیارهای متوقف‌کننده) تکرار می‌شود. به‌عبارت‌دیگر، الگوریتم جابه‌جایی مدارک و محاسبهٔ فاصلهٔ آن تا خوشهٔ مرکزی آن‌قدر تکرار می‌شود که به‌دلیل جابه‌جایی مدارک، تغییری در خوشهٔ مرکزی ایجاد نشود.

برای کشف مراکز اولیهٔ خوشه‌ها و انجام خوشه‌بندی موضوعات و سرانجام، بررسی کلمات کلیدی و موضوعات، لازم است محدوده و چارچوبی را در این خصوص مشخص نمود. این کار از طریق انتخاب مناسب پارامترهای آلفا و بتا انجام می‌شود. در واقع، از این دو پارامتر (که به انتخاب ما بستگی دارند)، برای تنظیم موضوعات مورد انتظار استفاده می‌شود و هر دوی این پارامترها برای تعریف پیشینه‌های LDA که برای محاسبات توزیع‌های پسین مربوط به آن استفاده می‌شوند، به‌کار می‌روند. به‌طور دقیق‌تر، عدد آلفا نشان‌دهندهٔ میزان پراکندگی موضوعات موجود در مدارک است. واضح است که این پراکندگی هرچه بیشتر باشد، مطلوب‌تر است؛ چرا که تعداد زیادی مدرک شامل یک موضوع است. مقدار آلفای بزرگ به معنای تأثیر کمتر همان مدرک در پراکندگی موضوعات است؛ یعنی یک مدرک شامل تعداد زیادی از موضوعات است و به‌عبارتی آن مدرک تخصصی نیست؛ در حالی که آلفای کوچک بدین معناست که یک مدرک شامل تعداد خاص و محدودی از موضوعات است. به همین دلیل است که اغلب عدد آلفا به‌صورت کسری از تعداد موضوعات K تنظیم می‌شود. در این پژوهش نیز مقدار آلفا به‌صورت پیش‌فرض با در نظر گرفتن فرض توزیع متقارن (سیمتریک) برای مدل LDA به‌صورت 1/k انتخاب شده است.

معیار عدد بتا چگالی کلمات در موضوع است و بر پایهٔ پراکنده بودن کلمات در یک موضوع تعریف می‌شود. واضح است که این مقدار هرچه کمتر باشد، بهتر است؛ زیرا مقدار بتای بالا به معنای تعلق کلمات بیشتر در یک حوزهٔ (خوشهٔ) موضوعی و به‌تبع آن، پراکندگی موضوعی بالای کلمات است؛ در حالی که انتظار داریم که همهٔ کلمات داخل یک موضوع تا حد امکان مشابه و نزدیک به هم باشند. بنابراین، به‌علت تعداد زیاد موضوعات و لحاظ شدن کلمات کمتر دیده شده در موضوعات، از کمیت بتای کوچک برای یافتن کلمات بیشتر متعلق به موضوعات خاص‌تر استفاده می‌شود. در این پژوهش، این پارامتر برابر ۰/۰۰۱ در نظر گرفته شده است که عدد مناسبی است.



Archive of SID.ir

Archive of SID.ir



### تعیین تعداد خوشه‌های موضوعی

مدل‌سازی موضوعی، موضوعات را به‌صورت خودکار از اسناد بدون برچسب و به روش غیرنظارتی استخراج می‌کند. به همین منظور ابتدا لازم است تعداد خوشه‌های موضوعی تعیین شود. برای انجام این مرحله ضروری است تعدادی مقدماتی از موضوعات منظور شود. مشورت با متخصصان موضوعی و استفاده‌کردن از برخی استانداردها از قبیل سرعنوان‌های موضوعی و روش‌های آماری از جمله روش‌های تعیین تعداد اولیهٔ موضوعات است. در این پژوهش برای به‌دست آوردن تعدادی خوشهٔ مطلوب، ابتدا بررسی پایگاه معتبر «لیزا»[1] که جامع‌ترین پایگاه مقالات در حوزهٔ علوم کتابداری و اطلاع‌رسانی است، مبنا قرار گرفت. با استخراج و بررسی کلیدواژه‌های موضوعی مقالات این حوزه در پایگاه مذکور و محدودسازی آن‌ها با نظر متخصصان موضوعی، تعداد ۲۷۰ موضوع در مرحلهٔ اول شناسایی شد.

در مرحلهٔ بعد و برای به‌دست آوردن بهینه‌ترین تعداد موضوعات، طبق مدل نمونه‌گیری گیبس[2] تعداد موضوعات در دامنهٔ ۲۵۰-۲۹۰ مورد آزمون قرار گرفت و با استفاده از الگوریتم‌های ارزیابی مدل‌های موضوعی سه‌گانه شامل انسجام[3] یا همبستگی، احتمال[4] یا تناسب و عدم یکدستی[5] (به‌عنوان پارامترهای LDA) اجرا شد. متوسط ضریب انسجام که میزان مشابهت معنایی میان لغات با بیشترین دامنهٔ تکرار در موضوعات را تعیین می‌کند، می‌تواند به تمایز میان موضوعاتی که از لحاظ معنایی قابل تفسیر بوده و موضوعاتی که ساختهٔ استنباط‌های آماری هستند، کمک نماید. پارامتر عدم یکدستی تنها یک ابزار است که برای نزدیک شدن به تعداد مناسب موضوعات در یک مجموعه لغات مفید است. این معیار توضیح می‌دهد که مدل نمونه را به چه خوبی پیش‌بینی می‌کند؛ یعنی مدل چقدر توسط نمونه‌ای از داده‌های مشاهده‌شده دچار عدم یکدستی می‌شود. احتمال یا میزان تناسب و مشابهت موضوعات در یک خوشه هرچه بالاتر باشد، آن موضوعات تناسب بیشتری با یکدیگر خواهند داشت (Stevens et al. 2012). نتایج ارزیابی تعداد حوزه‌های موضوعی در شکل‌های ۱ الی ۳ آمده است.

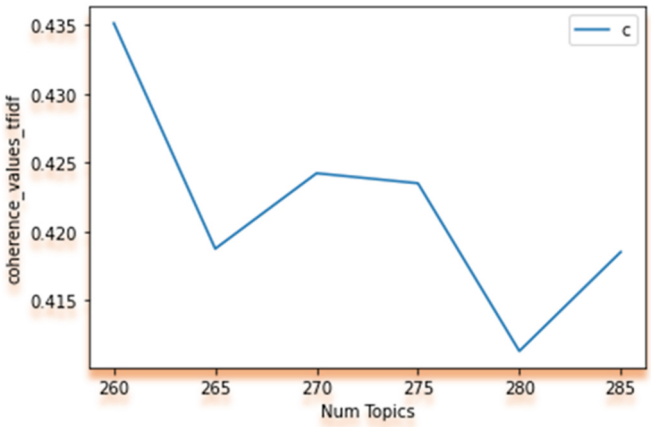
1. LISA
2. Gibbs Sampling
3. coherence
4. likelihood
5. perplexity



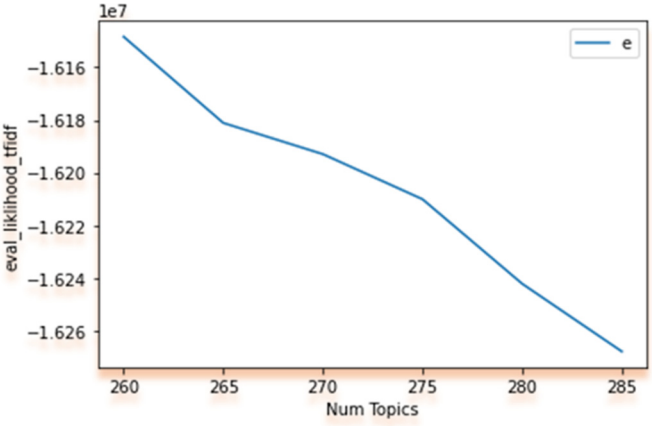
Archive of SID.ir

Archive of SID.ir






**شکل ۱. اجرای الگوریتم انسجام**



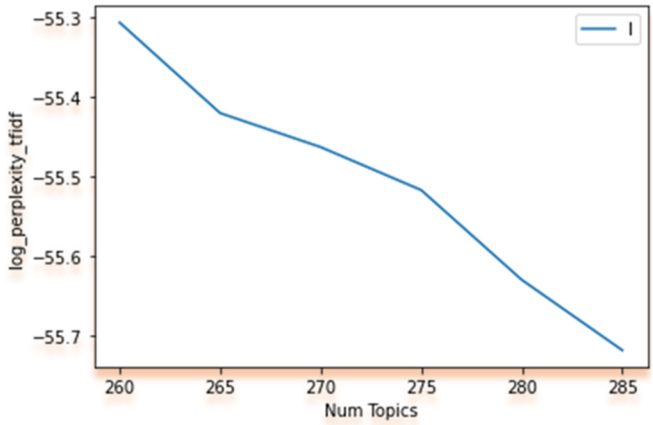

**شکل ۲. اجرای الگوریتم احتمال/ تناسب**



Archive of SID.ir

Archive of SID.ir






**شکل ۳. اجرای الگوریتم عدم یکدستی**

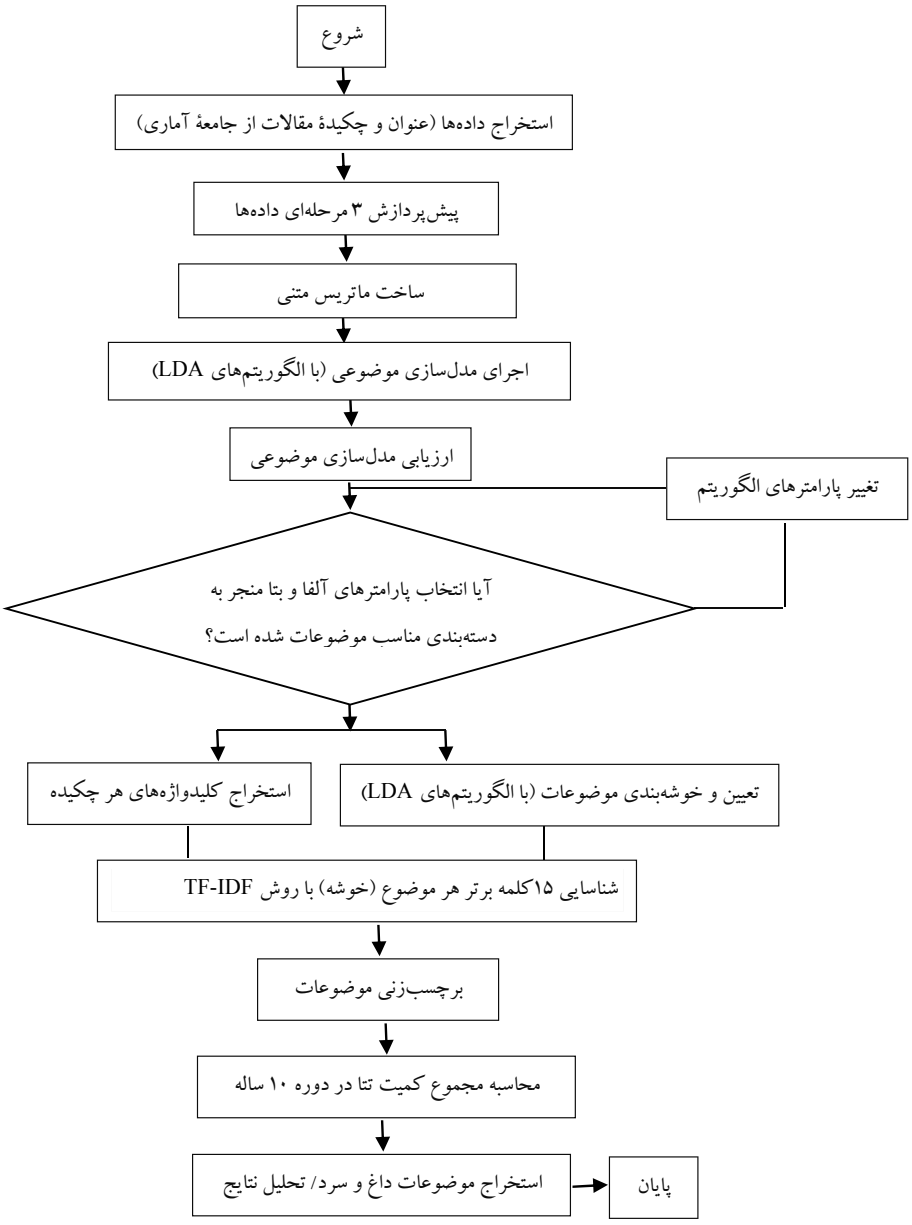
نتایج بالا بیان می‌دارد که بالاترین نمرهٔ انسجام (۰/۴۳۵)، بالاترین احتمال تناسب میان موضوعات خوشه (۱/۶۱۴-) و کمترین عدم یکدستی (۵۵/۳۳-) در حوالی ۲۶۰ موضوع است. بنابراین، ۲۶۰ خوشهٔ موضوعی برای پژوهش انتخاب شد.



Archive of SID.ir

Archive of SID.ir






**نمودار ۱. نمایش فرایند انجام پژوهش**



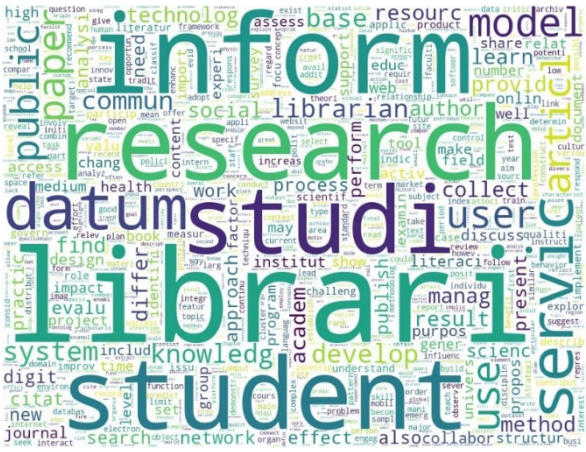
Archive of SID.ir

Archive of SID.ir



### ٤. تحلیل نتایج و یافته‌ها

با استفاده از کتابخانهٔ «وردکلود»[1] نمایش تودهٔ ابر کلمات به‌دست آمد. درجهٔ وزنی لغات با اندازهٔ نمایش آن‌ها به‌صورت مستقیم ارتباط دارد؛ یعنی لغاتی که وزن بیشتری در پژوهش‌های حوزهٔ علم اطلاعات و دانش‌شناسی (علوم کتابداری و اطلاع‌رسانی) داشته‌اند، با اندازهٔ بزرگ‌تر و رنگ مشخص‌تر نمایش داده شده است. در شکل ۴، نمونه‌ای از ابر کلمات نمایش داده شده است.




**شکل ٤. نمایش تودهٔ ابر لغات برای ۱۰۰۰ کلمه با بالاترین وزن به روش TF-IDF**

در نمودار ۲، تعداد چکیدهٔ مقالات بر حسب سال‌های مختلف مورد پژوهش آورده شده است. نتایج بیانگر آن است که تعداد مدارک در سال ۲۰۱۹، نسبت به سال قبل دارای بیشترین مقاله (۳۰ درصد رشد نسبت به سال قبل) و در سال ۲۰۱۱، دارای کمترین تعداد مدرک (۱۰ درصد رشد نسبت به سال قبل) است.

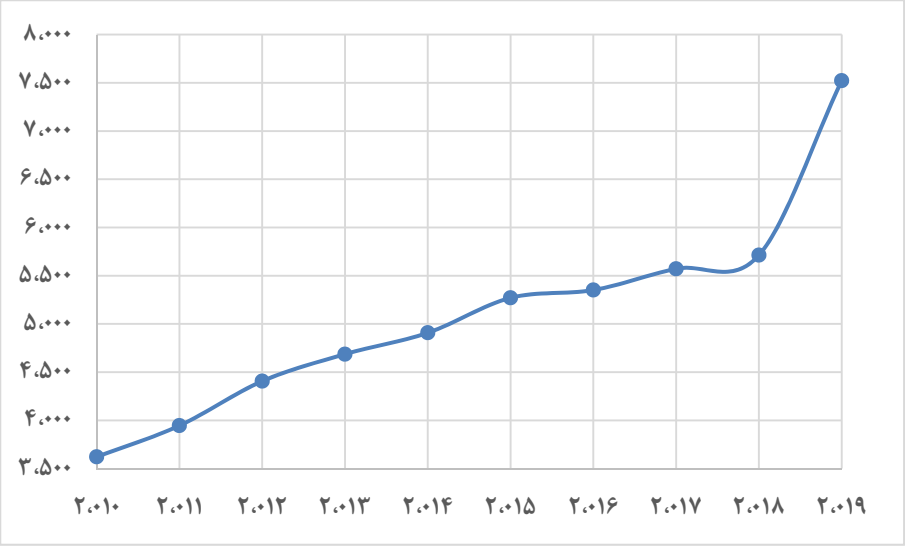
1. Wordcloud Package



Archive of SID.ir

Archive of SID.ir






**نمودار ۲. روند انتشار مدارک در هر سال**

در ادامه، به‌منظور تعیین موضوعات (خوشه‌ها) میانگین تتا (به معنای میزان توزیع موضوعات در اسناد بر اساس سال انتشار آن‌ها)، با انتخاب کوچک کمیت آلفا (به مفهوم وزن موضوعات در اسناد) و با هدفِ محدود نمودن یک سند به تعداد خاصی از موضوعات، برابر با عدد $1/k = 1/260 = 0.003$ (k تعداد بهینه‌شدهٔ موضوعات) محاسبه شد و یک شکل خطی از میانگین ضرایب اشتراک موضوعات (تتا) بر حسب سال رسم و محاسبات با ضریب اطمینان ۹۵ درصد منجر به انتخاب تعداد ۶۳ موضوع (به‌عنوان موضوعات داغ و سرد) از بین ۲۶۰ موضوع شد. از بین این موضوعات، تعداد ۲۴ موضوع دارای بالاترین مقدار میانگین تتا و به‌عنوان موضوعات داغ شناسایی گردید. از بین ۲۴ موضوع داغ به‌دست‌آمده، موضوعات دگرسنجی/ مطالعات‌سنجشی/ سنجش و اندازه‌گیری (مجموع تتا برابر با ۰/۱۰۴۷)، مدیریت الکترونیک/ بازاریابی دیجیتال (مجموع تتا برابر با ۰/۰۹۶۰)، بازیابی محتوا/ متن‌کاوی (مجموع تتا برابر با ۰/۰۸۸۹) و شبکه و تحلیل داده (مجموع تتا برابر با ۰/۰۸۶۶) و مهارت الکترونیک (مجموع تتا برابر با ۰/۰۷۸۸) با دارا بودن بالاترین مقادیر مجموع میانگین تتا، به‌عنوان ۵ موضوع برتر داغ در علوم کتابداری و اطلاع‌رسانی در بازهٔ زمانی ۲۰۱۰ تا ۲۰۱۹ شناسایی شد. در جدول ۱، عدد محاسبه شده (یعنی مجموع میانگین ضرایب تتا در ۱۰ سال) برای ۲۱ نمونه موضوع داغ اول آمده است.



Archive of SID.ir

Archive of SID.ir



لازم به ذکر است که از این به بعد، هر جا که لازم باشد موضوع به همراه شماره آمده و در بعضی از قسمت‌ها فقط شماره موضوع عنوان شده است.

**جدول ۱. مجموع میانگین پارامتر تتا در سال‌های ۲۰۱۰ تا ۲۰۱۹ برای موضوعات داغ**

| مجموع میانگین ضرایب تتا | شماره و عنوان موضوع داغ | مجموع میانگین ضرایب تتا | شماره و عنوان موضوع داغ | مجموع میانگین ضرایب تتا | شماره و عنوان موضوع داغ |
|---|---|---|---|---|---|
| ۰/۰۵۲۵ | ۲۳۱ (حریم خصوصی کاربران) | ۰/۰۶۷۹ | ۱۶۸ (رضایتمندی خدمات) | ۰/۱۰۴۷ | ۱۹۰ (سنجش و اندازه‌گیری) |
| ۰/۰۵۱۱ | ۲۲۴ (رضایتمندی شغلی) | ۰/۰۶۷۳ | ۱۲۰ (مصورسازی اطلاعات) | ۰/۰۹۶۰ | ۲۲۷ (مدیریت و بازاریابی الکترونیکی) |
| ۰/۰۴۴۳ | ۲۴۰ (مطالعهٔ جامعهٔ آماری) | ۰/۰۶۳۴ | ۲۰۹ (آموزش عمومی) | ۰/۰۸۸۹ | ۲۵۵ (بازیابی محتوا/ متن‌کاوی |
| ۰/۰۴۳۶ | ۶۸ (مطبوعات الکترونیک) | ۰/۰۶۱۷ | ۱۹۱ (فناوری اطلاعات در آموزش روستاییان و عشایر) | ۰/۰۸۶۸ | ۱۶۹ (یادگیری الکترونیکی) |
| ۰/۰۴۳۳ | ۱۱۵ (پردازش داده‌ها) | ۰/۰۵۹۱ | ۱۵۲ (وب‌سنجی) | ۰/۰۸۶۶ | ۱۰۰ (تحلیل داده) |
| ۰/۰۴۰۶ | ۱۸۰ (استناددهی) | ۰/۰۵۷۵ | ۱۳۵ (نظرسنجی) | ۰/۰۷۸۸ | ۲۱۲ (مهارت الکترونیکی) |
| ۰/۰۳۹۴ | ۶۲ (عضویت کتابخانه‌ای) | ۰/۰۵۷۱ | ۱۶۳ (فناوری اطلاعات) | ۰/۰۷۱۷ | ۲۰۶ (برنامه‌نویسی و کامپیوتر) |
| ۰/۰۲۳۳ | ۱۰۹ (خدمات الکترونیک) | ۰/۰۵۵۳ | ۲۶ (خدمات کتابخانه‌ای) | ۰/۰۶۸۹ | ۱۴۹ (کتابخانهٔ قفسه باز) |

مشابه روشی که برای شناسایی موضوعات داغ انجام شد، برای شناسایی موضوعات سرد نیز استفاده شد. به‌عبارت دیگر، از بین ۶۳ موضوع مورد اشاره، ۳۹ موضوع کمترین مقدار میانگین تتا را داشته‌اند. جدول ۲، مجموع میانگین پارامتر تتا را در سال‌های ۲۰۱۰ تا ۲۰۱۹ برای موضوعات سرد نشان می‌دهد.

بر پایهٔ نتایج ارائه‌شده در جدول ۲، تعداد ۵ حوزه از به‌اصطلاح سردترین موضوعات که به‌ترتیب دارای کمترین مقدار مجموع میانگین پارامتر تتا برای سال‌های ۲۰۱۰ تا ۲۰۱۹ هستند، به‌ترتیب عبارت‌اند از: موضوع سلامت کتابداران (مجموع تتا برابر با ۰/۰۳۳۰)،



Archive of SID.ir

Archive of SID.ir



موضوع سازماندهی (مجموع تتا برابر با ۰/۰۳۷۷)، موضوع مدیریت دانش/ درخت تصمیم (مجموع تتا برابر با ۰/۰۳۸۷)، موضوع گردآوری اطلاعات/ آرشیو (مجموع تتا برابر با ۰/۰۴۲۷) و آموزش (مجموع تتا برابر با ۰/۰۴۹۳).

**جدول ۲. مجموع میانگین پارامتر تتا در سال‌های ۲۰۱۰ تا ۲۰۱۹ برای موضوعات سرد**

| مجموع میانگین ضرایب تتا | شماره و عنوان موضوع سرد | مجموع میانگین ضرایب تتا | شماره و عنوان موضوع سرد | مجموع میانگین ضرایب تتا | شماره و عنوان موضوع سرد |
|---|---|---|---|---|---|
| ۰/۰۲۷۲ | ۱۸۸ (فهرست‌نویسی) | ۰/۰۲۳۵ | ۱۸ (مدیریت کتابخانه) | ۰/۰۱۶۰ | ۲۲ (اطلاع‌رسانی) |
| ۰/۰۲۷۴ | ۱۷۶ (ارزیابی اطلاعات) | ۰/۰۲۳۵ | ۱۷۱ (نمونه‌گیری) | ۰/۰۱۶۱ | ۶۷ (رفتار اطلاعاتی) |
| ۰/۰۲۷۸ | ۱۱۶ (توسعه مطالعات کتابداران) | ۰/۰۲۴۰ | ۱۶۰ (پردازش داده) | ۰/۰۱۷۰ | ۱۵ (مباحث نظری و عمومی اطلاعات) |
| ۰/۰۲۸۰ | ۱۲۷ (بازیابی اطلاعات) | ۰/۰۲۴۸ | ۲۱۰ (بازیابی اجتماعی) | ۰/۰۱۸۲ | ۱۲ (روش تحقیق) |
| ۰/۰۲۹۶ | ۲۴۶ (اطلاع‌سنجی) | ۰/۰۲۵۶ | ۲۱۶ (رفتار کتابداران) | ۰/۰۱۸۷ | ۱۳۳ (موتورهای جست‌وجو) |
| ۰/۰۳۱۴ | ۱۳۰ (هستی‌شناسی) | ۰/۰۲۵۶ | ۱۳۹ (اخلاق حرفه‌ای) | ۰/۰۱۸۸ | ۱۹۴ (نظارت بر اطلاع‌رسانی) |
| ۰/۰۳۲۶ | ۱۳۷ (اطلاع‌رسانی الکترونیک) | ۰/۰۲۵۶ | ۲۴۲ (طبقه‌بندی داده‌ها) | ۰/۰۱۹۶ | ۱۳۱ (منابع آموزشی) |
| ۰/۰۳۳۰ | ۱۹ (سلامت کتابداران) | ۰/۰۲۵۸ | ۲۲۶ (نمایه‌سازی) | ۰/۰۱۹۸ | ۱۵۰ (پایگاه داده، اطلاع‌رسانی) |
| ۰/۰۳۵۳ | ۲۱۷ (استناددهی) | ۰/۰۲۶۰ | ۴۱ (جست‌وجوی اطلاعات) | ۰/۰۲۰۵ | ۶۰ (ارتباط الکترونیک) |
| ۰/۰۳۷۷ | ۲۵۶ (سازماندهی) | ۰/۰۲۶۳ | ۲۵ (مدل‌های سیستمی پویا) | ۰/۰۲۲۷ | ۲۳۳ (وبلاگ‌نویسی) |
| ۰/۰۳۸۷ | ۱۰۲ (مطالعات یادگیری) | ۰/۰۲۶۳ | ۱۰۶ (جمع‌آوری اطلاعات) | ۰/۰۲۲۷ | ۲۵۹ (رابط کاربر) |
| ۰/۰۴۲۷ | ۳۸ (آرشیو) | ۰/۰۲۶۷ | ۲۸ (بازیابی معنایی) | ۰/۰۲۳۲ | ۱۶ (آموزش اطلاعاتی) |
| ۰/۰۴۹۳ | ۱۶۷ (کتابخانهٔ آنلاین) | ۰/۰۲۶۷ | ۲۰۴ (ارائهٔ اطلاعات) | ۰/۰۲۳۴ | ۷۵ (وب‌سنجی) |



Archive of SID.ir

Archive of SID.ir



### ۵. بحث و نتیجه‌گیری

تحلیل داده‌های پژوهش حاضر چشم‌انداز کلی از پژوهش‌های حوزهٔ علوم کتابداری و اطلاع‌رسانی را به‌دست می‌دهد. بعضی از حوزه‌هایی که بیش از سایر موضوعات در پژوهش‌های مختلف مورد توجه قرار گرفت، به‌عنوان موضوعات داغ شناسایی شد (جدول ۱)، و بعضی از موضوعات که کمترین میزان توجه به آن‌ها معطوف شده بود، با نام موضوعات سرد معرفی شد (جدول ۲).

در پژوهش حاضر با استفاده از مدل‌سازی موضوعی با الگوریتم LDA و محاسبهٔ پارامترهای این الگوریتم، ۲۶۰ موضوع مورد پژوهش حوزهٔ علوم کتابداری و اطلاع‌رسانی شناسایی شد. نتایج پژوهش حاصل از به‌کارگیری فن مدل‌سازی موضوعی و به‌طور خاص الگوریتم LDA نشان داد که موضوع‌های شناسایی‌شده در مدل‌سازی موضوعی نسبت به موضوع‌های به‌دست‌آمده از سایر پژوهش‌ها از نظر کمیت و تعداد بیشتر بوده است (به‌عنوان مثال، در پژوهش Lee, Kim & Kim (2010) تعداد موضوعات شناسایی‌شده ۱۰ موضوع بود). تفاوت تعداد موضوعات پژوهش حاضر با سایر پژوهش‌ها نشان می‌دهد که این تکنیک می‌تواند در شناسایی موضوعات پنهان و حوزه‌های موضوعی مختلف مفید باشد. همچنین، استفاده از روش متن‌کاوی و الگوریتم‌های مرتبط با آن در پژوهش حاضر حاکی از آن است که این روش توانایی تشخیص بیشینهٔ موضوعات از متون در علوم کتابداری و اطلاع‌رسانی را دارد.

در ادامه، روند تغییرات تعداد چکیدهٔ مقالات در حوزهٔ علوم کتابداری و اطلاع‌رسانی در بازهٔ ۲۰۱۰ تا ۲۰۱۹ در نمودار ۳، آمده است. نتایج بیانگر آن است که تعداد مدارک در سال ۲۰۱۹ نسبت به سال قبل دارای بیشترین رشد در تعداد مقاله (۳۰ درصد رشد نسبت به سال قبل) و سال ۲۰۱۱ دارای کمترین تعداد مدرک (۱۰ درصد رشد نسبت به سال قبل) است.

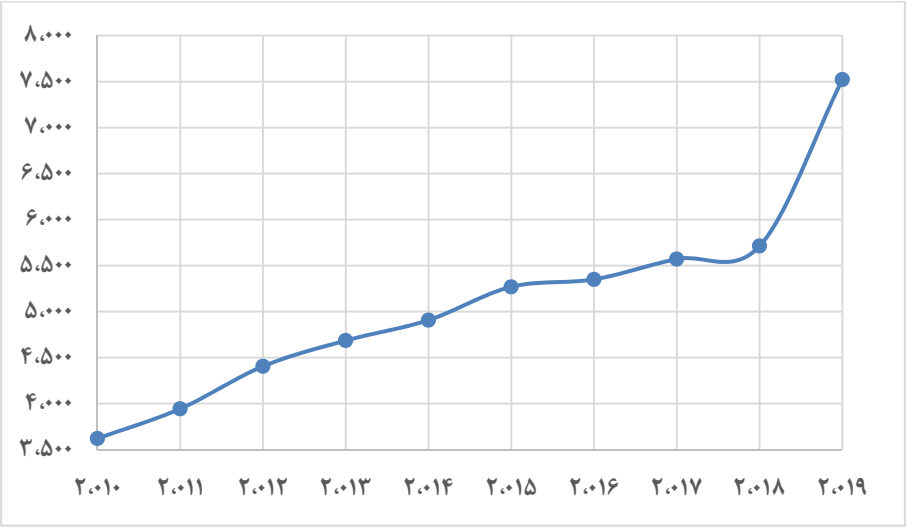
تغییرات شیب در این نمودار در فاصلهٔ سال‌های مختلف، نشانگر آهنگ متغیر رشد انتشارات است. سیر روبه‌رشد انتشارات این حوزه در سال ۲۰۱۰ تا ۲۰۱۹ روند یکنواختی را نشان نمی‌دهد. این روند از سال ۲۰۱۰ تا ۲۰۱۸ با آهنگ رشد تقریباً ثابتی در حال حرکت بود؛ به‌صورتی که تعداد انتشارات رشد تقریبی ۶۵ درصد را نشان می‌دهد. تعداد مدارک در سال ۲۰۱۹ (با حدود ۷۵۰۰ مدرک) نسبت به سال قبل از آن رشدی برابر ۳۰ درصد و نسبت به سال ۲۰۱۰ با داشتن تعداد تقریبی ۳۵۰۰ انتشارات، جهشی بیش از دو برابر را تجربه نموده است.



Archive of SID.ir

Archive of SID.ir






**نمودار ۳. تغییرات در تعداد انتشارات حوزهٔ علوم کتابداری و اطلاع‌رسانی در بازهٔ ۲۰۱۰ تا ۲۰۱۹**

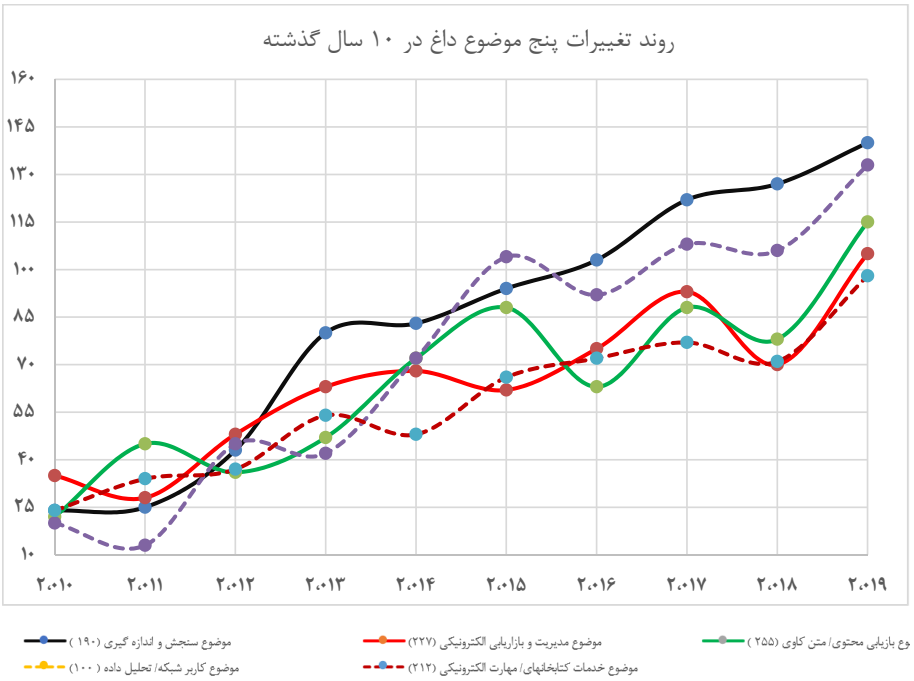
بررسی روند شکل‌گیری و توسعهٔ موضوعات حوزهٔ کتابداری و اطلاع‌رسانی در گذر تاریخ نشان می‌دهد که این موضوعات تحت تأثیر پیشرفت‌ها، توسعهٔ فناوری، جهت‌گیری جامعه و نیازهای کاربران و شرایط کتابخانه و این حوزه در کشورها بوده است. تعیین موضوعات در سال‌های ۱۹۵۷-۸۶ نشان داد که مجموعه‌سازی، ساختمان و تجهیزات، مدیریت کتابخانه، شبکه و به‌اشتراک‌گذاری منابع، آموزش کتابداران، چندرسانه‌ای‌ها، خودکارسازی فرایند کتابخانه و سیاست کتابخانه بوده است (Kajberg 1996). ولی در گذر زمان و بعد از حدود نیم قرن، موضوعات به گونهٔ دیگر تغییر ماهیت داده و این سرعت تغییر در سال‌های اخیر به‌دلیل سرعت تغییر در جامعه، نیازها و ابزارها و امکانات بیشتر شده است. در مقایسه با پژوهش Kajberg (1996)، نتایج پژوهش حاضر نشان می‌دهد که اغلب موضوعات پژوهشی سال‌های ۱۹۵۷-۸۶ دغدغهٔ پژوهشی فعلی کتابداران و پژوهشگران این حوزه نیست، بلکه موضوعات بر اساس پیشرفت فناوری و الزامات و نیازهای کاربران پیش رفته است (نمودار ۴). این تغییر در پژوهش « باغ محمد، منصوری، چشمه‌سهرابی» (۱۳۹۹) و «دستانی» و همکاران (۲۰۲۰) مورد تأیید قرار گرفته است.



Archive of SID.ir

Archive of SID.ir






**نمودار ٤. تغییرات در تعداد انتشارات مربوط به موضوعات داغ در حوزهٔ علوم کتابداری و اطلاع‌رسانی در بازهٔ ۲۰۱۰ تا ۲۰۱۹**

همچنین، بررسی نتایج پژوهش حاضر در مقایسه با سایر پژوهش‌ها نشان داد که هرچه به سمت عصر فناوری پیش رفته‌ایم، اغلب موضوعات پایهٔ علوم کتابداری و اطلاع‌رسانی کم‌رنگ شده و حتی جزء موضوعات سرد قرار گرفته است (جدول ۲). نگاهی مجدد به نتایج پژوهش Kajberg (1996) نشان می‌دهد که موضوعات در دههٔ ۸۰ قرن بیستم به‌طور کامل بر حوزه‌های تخصصی و پایهٔ کتابداری از جمله مجموعه‌سازی، سازماندهی، و خدمات کتابخانه متمرکز بود و در اوایل قرن ۲۱ موضوعات خودکارسازی کتابخانه، کتابخانهٔ دیجیتال، مطالعات کاربران و تاریخ کتابخانه‌ها (Hjørland 2000)، بازیابی اطلاعات، عملکرد کتابخانه، تحقیقات در علوم کتابداری و اطلاع‌رسانی، جست‌وجوی اطلاعات، انواع کتابخانه‌ها و وب‌سایت (Han 2020) جزء موضوعات مهم پژوهش قرار گرفته است. ولی نتایج پژوهش حاضر نشان داد که روند پژوهش‌ها به موضوعات بین رشته‌ای تمایل پیدا کرده و این به شرایط استفاده و توسعه از فناوری‌ها و علوم مرتبط بستگی دارد. از جملهٔ این موضوعات می‌توان از علم‌سنجی، داده‌کاوی، بازاریابی، بازاریابی



Archive of SID.ir

Archive of SID.ir



الکترونیک، تولید محتوا و موارد مرتبط با فناوری الکترونیک و دیجیتال نام برد و این موضوعات نیز در پژوهش‌های سال‌های اخیر مورد تأیید بوده است. در پژوهش Jabeen et al. (2015) موضوعاتی از قبیل اینترنت و بازاریابی الکترونیک و دیجیتال‌سازی از موضوعات اصلی پژوهش‌ها بوده است.

همهٔ این‌ها نشان می‌دهد که فناوری و علوم نوپدید بین رشته‌ای همچون سایر رشته‌ها به حوزهٔ علوم کتابداری و اطلاع‌رسانی نیز رسوخ پیدا کرده و شاید به‌عبارت بهتر، کتابداری و اطلاع‌رسانی برای حفظ خود و حضور در جامعه ناگزیر از استفاده از این علوم شده است. علومی همچون بازاریابی الکترونیک، داده‌کاوی، ارزیابی، سنجش، علم‌سنجی، یادگیری الکترونیک و امثال آن که همگی ماهیت فناورانه دارند، جزء علوم بین رشته‌ای محسوب شده و هر رشته به‌تناسب نیاز خود از این علوم استفاده می‌کند و علوم کتابداری و اطلاع‌رسانی نیز از این قاعده مستثنا نیست. موضوع دیگری که در نتایج پژوهش حاضر قابل تأکید است این‌که، موضوعات پژوهش‌ها از ماهیت نظری به موضوعات عملی و کاربردی گرایش پیدا کرده و به نظر می‌رسد که این امر برای همراهی با نیازها و خواسته‌های کاربران است.

به‌طور کلی، نتایج نشان داد که یک رابطهٔ قوی و محکم بین حوزه‌های موضوعی علوم کتابداری و اطلاع‌رسانی با حوزهٔ کامپیوتر و فناوری‌ها و جنبه‌های کاربردی آن وجود دارد. سنجش و اندازه‌گیری برترین حوزهٔ مورد توجه در پژوهش‌هاست (موضوع ۱۹۰ با بالاترین مقدار میانگین تتا ۰/۱۰۴۷). دلیل این امر را شاید بتوان در اهمیت این موضوع در سنجش و ارزیابی بروندادهای پژوهشی، سازمان‌ها، دانشگاه‌ها، نویسندگان و کشورها و فناوری‌ها از یک سو و اهمیت ارائهٔ شاخص‌های متناسب با نیاز روز دانست. پژوهش‌های مختلف نشان‌دهندهٔ این است که انتشارات حوزهٔ موضوعی مورد پژوهش روندی روبه‌رشد دارد و این رشد در سال‌های اخیر روندی تصاعدی داشته است. اهمیت حوزه موضوعی علوم کتابداری و اطلاع رسانی در پژوهش‌های مختلف نیز مورد تأکید بوده است (Åström 2002).

حوزه‌های دیگری که با توسعهٔ داده‌ها و ایجاد دیتاست‌ها و لزوم دسترسی و تحلیل آن‌ها ضرورت پیدا کرده، حوزه‌های داده‌کاوی، تحلیل داده و پردازش داده است که می‌توان آن‌ها را جزء حوزه‌های نوپدید قلمداد کرد؛ حوزه‌ای که علم اطلاعات و دانش‌شناسی (علوم کتابداری و اطلاع‌رسانی) نیز با توجه به رسالت و ماهیت خود با این



Archive of SID.ir

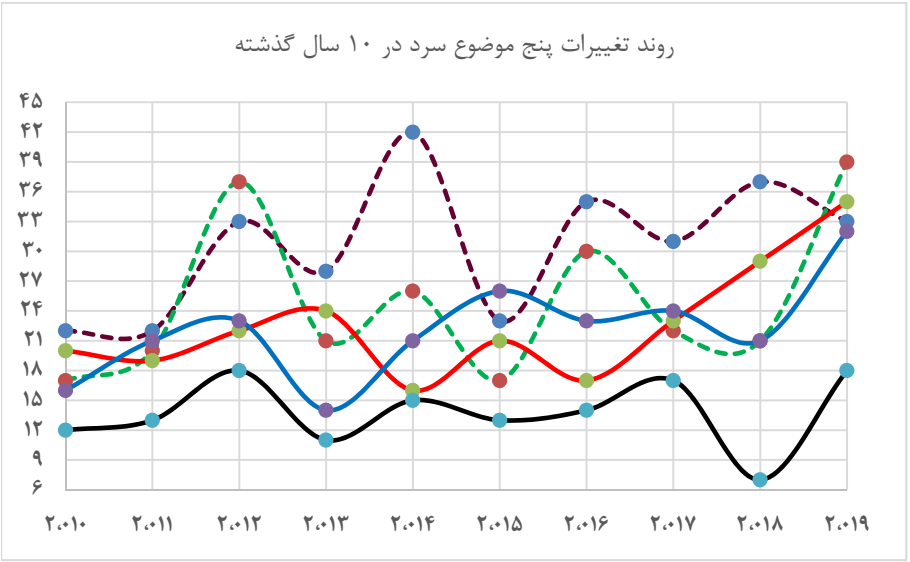
Archive of SID.ir



حوزه‌ها ارتباطاتی دارد. اختصاص بالاترین نرخ رشد به موضوع تحلیل داده‌ها (موضوع شماره ۱۰۰) در سال ۲۰۱۹ نسبت به سال ۲۰۱۰ تأییدکنندهٔ این موضوع است.

در میان موضوعات سرد، آموزش (۱۶۷)، کمترین میزان توجه پژوهشگران را به خود جلب کرده است؛ به‌صورتی که در برخی سال‌ها (۲۰۱۳ تا ۲۰۱۵) این موضوع به‌صورت کلی به فراموشی شمرده شده است. مباحث آرشیو (۳۸)، مدیریت دانش (۱۰۲)، سازماندهی (۲۵۶) و سلامت کتابداران (۱۹) در درجهٔ بعدی از موضوعاتی است که کمتر مورد توجه قرار گرفته است. نمودار ۵، در مورد تغییرات انتشارات موضوعات سرد نشان‌دهندهٔ روندی ثابت با تغییرات ملایم در میان سال‌های ۲۰۱۰ تا ۲۰۱۹ است. این روند در مورد دیگر موضوعات سرد نیز قابل مشاهده است. بعضی دیگر از این موضوعات، مانند مباحث مرتبط با آموزش، و میزان رشد تحقیقات آن نیز با روندی کاهشی یا تقریباً ثابت روبه‌رو بوده است.

بررسی نتایج به‌دست‌آمده از روند صعود یا نزول اشتیاق پژوهشگران به موضوعات مختلف علم اطلاعات و دانش‌شناسی در ۱۰ سال گذشته بیانگر آن است که در مجموع، در اکثر حوزه‌های موضوعی در علم اطلاعات و دانش‌شناسی شاهد روندی افزایشی (حدود ۷۹ درصد) بوده و در بقیهٔ حوزه‌ها (حدود ۲۱ درصد) شاهد روندی نزولی یا بدون تغییر بوده‌ایم.




**نمودار ۵. تغییرات در تعداد انتشارات مربوط به موضوعات سرد در حوزهٔ علوم کتابداری و اطلاع‌رسانی در بازهٔ ۲۰۱۰ تا ۲۰۱۹**



Archive of SID.ir

Archive of SID.ir



همچنین، نگاهی به موضوعات شناسایی‌شده به‌طور کلی نشان داد که این رشته تحت تأثیر چند عامل یا اتفاق قرار دارد. خوشه‌های موضوعی فناوری‌محور حاکی از این است که این رشته تحت تأثیر فناوری و موضوعات مرتبط قرار دارد و این تأثیرپذیری بسیار زیاد است. برخی خوشه‌های موضوعی از نیاز روز کاربران تأثیرپذیر بوده و با نیاز آموزشی و عادت‌های نیازهای اطلاعاتی آن‌ها مرتبط هستند. همچنین، بعضی از خوشه‌ها همچون علم‌سنجی، ارزیابی و ارزشیابی برخواسته از انتظارات خدماتی این رشته است. بعضی دیگر از خوشه‌ها برگرفته از خدمات ذاتی کتابخانه بوده و تحت تأثیر شرایط حاکم بر جامعه، توسعهٔ فناوری و نیازهای کاربران است.

مطالعه و به‌کارگیری نتایج این پژوهش می‌تواند برای پژوهشگران علاقه‌مند به حوزهٔ علوم کتابداری و اطلاع‌رسانی، سیاست‌گذاران، تدوین‌کنندگان متون درسی پیشنهادی ارزشمند بوده و دیدی جزئی از یک کل ارائه دهد. در این پژوهش بسیاری از موضوعات جدید شناسایی شد که در پژوهش‌های مشابه دیگر شناسایی نشده بود. بهتر است سیاست‌گذاران در سیاست‌گذاری آموزشی، پژوهشگران در فعالیت‌های پژوهشی، و دانشکده‌ها و گروه‌های کتابداری در آموزش خود به نتایج این پژوهش توجه نمایند. از سوی دیگر، نتایج این پژوهش نشان داد که حوزه‌های پایه و اساسی رشته کمتر مورد توجه قرار گرفته و جزء حوزه‌های داغ قرار نگرفته‌اند و این موضوع می‌تواند برای رشته و توسعهٔ اساسی مبانی و مفاهیم این رشته مشکل‌ساز و نگران‌کننده باشد. پیشنهاد می‌شود در پژوهش‌های بعدی به این حوزه جدی‌تر توجه شود. همچنین، پیشنهاد می‌شود روند تغییرات در موضوعات علوم کتابداری و اطلاع‌رسانی با استفاده از روش تحلیل محتوا و تحلیل روند و به‌کارگیری نتایج پژوهش‌های قبلی تاکنون بررسی شود تا از روند تغییر و توسعهٔ موضوعات پژوهشی این حوزه آشنایی دقیق حاصل آید.

### ٦. پیشنهادهای پژوهشی

در پژوهش حاضر به بررسی مقالات حوزهٔ علم اطلاعات و دانش‌شناسی در سطح بین‌الملل که در سال‌های ۲۰۱۰ تا ۲۰۱۹ در پایگاه استنادی «اسکوپوس» نمایه شده، پرداخته شده است. پیشنهادهای زیر برای گسترش این پژوهش ارائه می‌شود:

- با توجه به قابلیت‌های متن‌کاوی و کاربردهای آن در زمینه‌ها و رشته‌های مختلف و به‌منظور گسترش دامنهٔ ابزار پژوهش در حوزهٔ علم اطلاعات و دانش‌شناسی پیشنهاد



Archive of SID.ir

Archive of SID.ir



می‌شود که از این روش در پژوهش‌های مختلف استفاده شود؛

- دلایل رشد و افول (توجه و عدم توجه به) حوزه‌های مختلف علم اطلاعات و دانش‌شناسی در پژوهش‌ها مورد بررسی قرار گیرد؛
- نقش فناوری‌ها در توسعهٔ حوزهٔ علم اطلاعات و دانش‌شناسی تبیین شود؛
- عدم توجه به مبانی نظری و پایه‌ای حوزهٔ علم اطلاعات و دانش‌شناسی بررسی شود؛
- دلایل تغییر در روند توسعهٔ موضوعات پژوهشی حوزهٔ علم اطلاعات و دانش‌شناسی تحلیل شود.

## فهرست منابع

آذری، سلمان، محمود شکراللهی‌فر، و محمدعلی لسانی فشارکی. ۱۳۹۹. الگوپردازی مفهوم‌یابی از آیات قرآن کریم با استفاده از دانش متن‌کاوی رایانشی. *مجله ذهن* ۸۲ (۲): ۸۱-۱۰۸.

اصلانی، اکرم و مهدی اسماعیلی. ۱۳۹۷. یافتن الگوهای مکرّر در قرآن کریم به کمک روش‌های متن‌کاوی. *پردازش علائم و داده‌ها* (۳) ۱۵: ۸۹-۱۰۰.

باغ محمد، مریم، علی منصوری، و مهرداد چشمه سهرابی. ۱۳۹۹. بررسی توسعه و روند موضوعی حوزه علم اطلاعات و دانش‌شناسی بر اساس مدل موضوع LDA. *فصلنامه پردازش و مدیریت اطلاعات* ۳۶ (۲): ۲۹۷-۳۲۷.

برنا، کیوان، فرهاد فتحی، و عصمت مؤمنی. ۱۳۹۷. کشف دانش و کاربرد آن در اینترنت اشیا. *فصلنامه مطالعات دانش‌شناسی* ۵ (۱۷): ۱۲۵-۱۶۴.

پرئی، اعظم‌السادات، و حجت‌الله حمیدی. ۱۳۹۶. ارائه رویکردی برای مدیریت و سازمان‌دهی اسناد متنی با استفاده از تجزیه و تحلیل هوشمند متن. *پژوهشنامه پردازش و مدیریت اطلاعات* (۴) ۳۲: ۱۱۷۱-۱۲۰۲.

جعفری پاورسی، حمیده، نجلا حریری، مهدی علیپور حافظی، فهیمه باب‌الحوائجی، و مریم خادمی، مریم. ۱۳۹۸. نمایه‌سازی ماشینی مدارک حوزه بازیابی اطلاعات با استفاده از متن‌کاوی در نرم‌افزار رپیدماینر. *فصلنامه پردازش و مدیریت اطلاعات* ۳۵ (۲): ۳۴۹-۳۷۴.

دستانی، میثم، افشین موسوی چلک، افشین، ثریا ضیایی، و فایزه دلقندی. ۱۳۹۹. تجزیه و تحلیل موضوعی مقالات منتشرشده کتابداری و اطلاع‌رسانی پزشکی در ایران با استفاده از فنون متن‌کاوی. *فصلنامه مرکز تحقیقات مدیریت خدمات بهداشتی درمانی دانشگاه علوم پزشکی و خدمات درمانی تبریز* ۱۱ (۴): ۳۵۵-۳۷۵.

رضایی‌نور جلال، و محمدرضا شیخ‌بهائی. ۱۳۹۶. کاربردهای داده‌کاوی متنی در حوزه مدیریت دانش زنجیرهٔ خدمات دولت الکترونیکی. *فصلنامه مدیریت فناوری اطلاعات* ۹ (۱): ۳۹-۶۰.



Archive of SID.ir

Archive of SID.ir



سلیمانی‌نژاد، عادل، مژده سلاجقه، و الهام طیبی. ۱۳۹۷. خوشه‌بندی مقالات علمی بر پایهٔ الگوریتم k_means مطالعهٔ موردی: پایگاه پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)». *پردازش و مدیریت اطلاعات* ۳۴ (۲): ۸۷۱-۹۸۶.

شکوهیان،؟، عاصفه عاصمی، احمد شعبانی، مظفر چشمه‌سهرابی. ۲۰۱۹. تحلیل ترکیبی کتاب‌سنجی و متن‌کاوی تولیدات علمی حوزه پرونده الکترونیک سلامت در پایگاه PubMed. *مدیریت اطلاعات سلامت* ۱۶ (۴): ۱۹۰-۱۹۶.

صالحی شهرودی، محمدحسین، بهروز مینایی، و امیررضا اشرفی. ۱۳۹۲. متن‌کاوی موضوعی رایانه‌ای قرآن کریم برای کشف ارتباطات معنایی میان آیات بر مبنای تفسیر المیزان. *قرآن شناخت* ۱۲ (۲): ۱۱۷-۱۵۲.

کاردان، احمدآقا، و مینا کیهانی‌نژاد. ۱۳۹۱. ارائه مدلی برای استخراج اطلاعات از مستندات متنی مبتنی بر متن‌کاوی در حوزه یادگیری الکترونیکی. *فناوری اطلاعات و ارتباطات ایران* ۴ (۱۱-۱۲): ۴۷-۵۴.

گرشاسبی، مسعود، آناهید رئیس‌روحانی، و محمدرضا کاباران‌زاده قدیم. ۱۳۹۷. ارائه الگوریتم متن‌کاوی به‌منظور تشخیص حس در متن‌های فارسی. مجله *مدیریت فناوری اطلاعات* ۳۵: ۳۷۵-۳۸۹.

## References

Al-Azmi, Abdul-Aziz Rashid. 2013. Data, text and web mining for business intelligence: a survey. *International Journal of Data Mining & Knowledge Management Process* (IJDKP) 3 (2): 1-21. https://doi.org/10.5121%2Fijdkp.2013.3201 .

Åström, F. 2002. *Visualizing library and information science concept spaces through keyword and citation based maps and clusters*. In Emerging frameworks and methods: Proceedings of the fourth international conference on conceptions of Library and Information Science (CoLIS4) (pp. 185-197). Greenwood Village: Libraries Unlimited.

Ayyoubzadeh, S. M., S. M. Ayyoubzadeh, H. Zahedi, M. Ahmadi, & S. R. N. Kalhori. 2020. Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. *JMIR public health and surveillance* 6 (2): e18828.

Bittermann, André, and Andreas Fischer. 2018. How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie* 226: 3-13.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022. https://doi.org/10.1016/b978-0-12-411519-4.00006-9.

Bornmann. L., and R. Mutz. 2015. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *The Journal of the Association for Information Science and Technology (JASIST)* 66 (11): 2215-2222.

Burrill, G., & D. Ben-Zvi. 2019. Topics and Trends in Current Statistics Education Research. *Springer International Publishing.* https://doi.org/10.1007/978-3-030-03472-6.

Buscema, P. M., F. Della Torre, M. Breda, G. Massini, and E. Grossi. 2020. COVID-19 in Italy and extreme data mining. Physica A: *Statistical Mechanics and its Applications* 557: 124991.

Carracedo, P., R. Puertas, and L. Marti. 2021. Research lines on the impact of the COVID-19 pandemic on business. A text mining analysis. *Journal of Business Research* 132: 586-593.

Chen, S., X. Guo, T. Wu, and X. Ju. 2020. Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis. *Information Processing & Management* 57 (5): 102253.



Archive of SID.ir

Archive of SID.ir



Dastani, M., S. Ziaei, and F. Delghandi. 2020. Identifying Emerging Trends in Scientific Texts Using TF-IDF Algorithm: A Case Study of Medical Librarianship and Information Articles. *Health Technology Assessment in Action* 4 (2): ), .1-8 https://doi.org/10.18502/htaa.v4i2.6231

Erdfelder, Edgar, and Michael Bosnjak. 2016. Hotspots in Psychology: A New Format for Special Issues of the Zeitschrift für Psychologie *Zeitschrift für Psychologie* 224 (3): 141-144, DOI: 10.1027/2151-2604/a000249.

Gal, D., B. Thijs, W. Glänzel, and K. R. Sipido. 2019. Hot topics and trends in cardiovascular research. *European heart journal* 40 (28): 2363-2374.

Hamerly, Greg, & Charles Elkan. 2002. Alternatives to the k-means algorithm that find better clusterings. *Paper presented at the Proceedings of the eleventh international conference on Information and knowledge management*. pp 25-38 Virginia USA: McLean.

Han, X. 2020. Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model. *Scientometrics* 125 (32561-2595 ) :.

Hjørland, B. 2000. Library and information science: practice, theory, and philosophical basis. *Information processing & management* 36 (3): 501-531.

Jabeen, M., L. Yun, M. Rafiq, M. Jabeen & M. A. Tahir. 2015. Scientometric analysis of library and information science journals 2003–2012 using Web of Science. *International Information & Library Review* 47 (3-4): 71-82.

Jo, Taeho. 2018. *Text mining: Concepts, implementation, and big data challenge* (Vol. 45)?: Springer, pp 230-286.

Jung, H., & B. G. Lee. 2020. Research trends in text mining: Semantic network and main path analysis of selected journals. *Expert Systems with Applications* 162: 113851.

Kajberg, Leif. 1996. A content analysis of library and information science serial literature published in Denmark, 1957-198*Library and Information Science Research* 18 (1): 25-52.

Katsurai, M., and S. Joo. 2021. Adoption of Data Mining Methods in the Discipline of Library and Information Science. *Journal of Library and Information Studies* 19 (1): 1-17.

Kim, S. Y., S. J. Song, and M. Song. 2019. Investigation of Topic Trends in Computer and Information Science by Text Mining Techniques: From the Perspective of Conferences in DBLP. *Journal of the Korean Society for information Management* 32 (1): 135-152.

Kowalski, G. J., & M. T. Maybury. 2000. Information storage and retrieval systems: theory and implementation (Vol. 8). ?: Springer Science & Business Media.

Larsen. P. O., and M. V Ins. 2010. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index, *Scientometrics* 84 (3): 575–603. DOI: 10.1007/s11192-010-0202-z.

Lässig, Simone. 2016. The History of Knowledge and the Expansion of the Historical Research Agenda, *Bulletin of the German Historical Institute* 59 (4): 29–59.

Lee, J. Y., H. Kim, and P. J. Kim. 2010. Domain analysis with text mining: Analysis of digital library research trends using profiling methods. *Journal of Information Science* 36 (2): 144-161.

Li, X., M. Fan, Y. Zhou, J. Fu, F. Yuan, and L. Huang. 2020. Monitoring and forecasting the development trends of nanogenerator technology using citation analysis and text mining. *Nano Energy* 71: 104636.

Luhn, Hans Peter. 1957. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of research and development* 1 (4): 309-317.

Manoharan, S. 2020. Geospatial and social media analytics for emotion analysis of theme park visitors using text mining and *Journal of Information Technology* 2 (02): 100-107.



Archive of SID.ir

Archive of SID.ir



Natarajan, M. 2005. Role of text mining in information extraction and information management. DESIDOC *Journal of Library & Information Technology* 25 (4): 31-38.

Park, J. H., and M. Song. 2013. A study on the research trends in library & information science in Korea using topic modeling. *Journal of the Korean Society for information, koreascience.or.kr, Journal of the Korean Society for information Management* 30 (1): 7-32.

Popowich, F. 2005. Using text mining and natural language processing for health care claims processing. *ACM SIGKDD Explorations Newsletter* 7 (1)59-66 :.

Rehurek, R., and P. Sojka. 2010. Software framework for topic modelling with large corpora. *In Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks* (pp. 45–50). Valletta, Malta.

Salton, G., & C. Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management* 24 (5)513-523 :.

Silge, J., & D. Robinson. 2017. Text mining with R: A tidy approach. Newton, Massachusetts, United States, *O'Reilly Media, Inc*.

Stevens, K., P. Kegelmeyer, D. Andrzejewski, and D. Buttler. 2012. Exploring topic coherence over many models and many topics. In EMNLP-CoNLL 2012–*2012 joint conference on empirical methods in natural language processing and computational natural language learning* (pp. 952–961). Valletta, Malta.

Tseng, Yuen-Hsien, Chi-Jen Lin, and Yu-I Lin. 2006. Text mining techniques for patent analysis, *Information Processing & Management* 43 (5): 1216-1247.

Tworowski, D., A. Gorohovski, S. Mukherjee, G. Carmi, E. Levy, R. Detroja, and M. Frenkel-Morgenstern. 2021. COVID19 Drug Repository: text-mining the literature in search of putative COVID19 therapeutics. *Nucleic acids research* 49 (1): D1113-D1121.

Van Dijk, W. B., A. T. Fiolet, E. Schuit, A. Sammani, T. K. J. Groenhof, R. van der Graaf, ... and A. Mosterd. 2021. Text-mining in electronic healthcare records can be used as efficient tool for screening and data collection in cardiovascular trials: a multicenter validation study. *Journal of Clinical Epidemiology* 132: 97-105.

Vinay, R. S., and M. T. Basavaraja. 2019. Trends in Library and Information Science Research: A Scientometric Analysis, *9th* KSCLA National Conference on Library in the Life of the User, 1-2 March 2019 At: Tumakuru. India.

Wang, Dashun and Albert-László Barabási. 2020. *The Science of Science*. United Kingdom *Cambridge University Press*. Doi: 10.1017/9781108610834.

Wang, L., and K. Lo. 2021. Text mining approaches for dealing with the rapidly expanding literature on COVID-19. *Briefings in Bioinformatics* 22 (2): 781-799.

Zengul, F. D., A. G. Zengul, M. Mugavero, N. Oner, B. Ozaydin, B. Delen and J. Cimino. 2021. A critical analysis of COVID-19 research literature: Text mining approach. Intelligence-based medicine, 100036. https://doi.org/10.1016/j.ibmed.2021.100036.

**الهه اخوان حریری**

دارای مدرک تحصیلی کارشناسی ارشد علم اطلاعات و دانش‌شناسی، گرایش مدیریت کتابخانه‌های دانشگاهی از دانشگاه اصفهان است. مدیریت اطلاعات و داده‌کاوی از جمله علایق پژوهشی وی است.

Archive of SID.ir

Archive of SID.ir

**علی منصوری**

دارای مدرک دکتری در رشتهٔ علوم کتابداری و اطلاع سانی از دانشگاه شهید چمران اهواز است. ایشان هم‌اکنون دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است.

علم‌سنجی، فناوری سنجی، مدیریت اطلاعات و داده‌کاوی از جمله علایق پژوهشی وی است.

**حسین کارشناس نجف‌آبادی**

دارای مدرک تحصیلی دکتری در رشتهٔ هوش مصنوعی از دانشگاه پلی‌تکنیک مادرید، اسپانیاست. ایشان هم‌اکنون استادیار گروه هوش مصنوعی دانشگاه اصفهان است.

هوش محاسباتی، تحلیل و مدل‌سازی داده‌ها، یادگیری ماشین و سامانه‌های هوشمند سلامت از جمله علایق پژوهشی وی است.



Archive of SID.ir