

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

الهه اخوان حریری

کارشناس 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،  
دانشکده علوم تربیتی و روان‌شناسی، دانشگاه اصفهان، اصفهان، ایران  
a.mansouri@edu.ui.ac.ir

حسین کارشناس نجف‌آبادی

دکتری هوش مصنوعی استادیار، گروه هوش مصنوعی، دانشکده مهندسی  
کامپیوتر، دانشگاه اصفهان، اصفهان، ایران  
h.karshenas@eng.ui.ac.ir

مقاله برای اصلاح به مدت ۱۲ روز نزد پدید آوران بوده است

دریافت: ۱۴۰۰/۰۸/۱۵ | پذیرش: ۱۴۰۱/۰۲/۰۶

نشریه علمی (رتبه بین‌المللی)  
پژوهشگاه علوم و فناوری اطلاعات ایران  
شاپا (چاپی) ۸۲۲۳-۲۲۵۱  
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱  
نمایه در SCOPUS، LISTA و ISC  
http://jipm.irandoc.ac.ir  
دوره XX | شماره X | صص XX-XX  
۱۳XX X

نوع مقاله: پژوهشی

به این مقاله به شکل زیر استناد کنید:

درون متن:

(الهه اخوان حریری، علی منصوری و حسین کارشناس نجف‌آبادی)

در فهرست منابع:

اخوان حریری، الهه. منصوری، علی و کارشناس نجف‌آبادی، حسین. شناسایی موضوعات داغ و روندها در علم اطلاعات و دانش‌شناسی با استفاده از تکنیک‌های متن‌کاوی. پژوهشنامه پردازش و مدیریت اطلاعات.

http://Jipm.irandoc.ac.ir (دسترسی در

روز/ماه/سال)

چکیده: شناسایی موضوعات داغ در حوزه‌های پژوهشی همواره مورد توجه بوده است؛ تصمیم‌گیری هوشمندانه در مورد آنچه باید مورد مطالعه و قرار بگیرد، همواره عاملی اساسی برای محققان بوده و می‌تواند برای پژوهشگران موضوعی چالش‌برانگیز باشد. هدف پژوهش حاضر، شناسایی موضوعات داغ در مقالات نمایه شده در پایگاه اسکوپوس در حوزه علم اطلاعات و دانش‌شناسی در سال‌های ۲۰۱۰ تا ۲۰۱۹ با استفاده از تکنیک‌های متن‌کاوی است. جامعه آماری، شامل ۵۰۹۹۵ مقاله منتشر شده در ۲۴۹ عنوان مجله نمایه شده در پایگاه اسکوپوس در این حوزه در بازه 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 بیکره اسناد تعیین و ضرایب اختصاص هر سند به هر موضوع (تا) مشخص شد. خروجی اجرای الگوریتم LDA، منجر به شناسایی تعداد ۲۶۰ موضوع شد. تخصیص برچسب به هر یک از خوشه‌های موضوعی، بر اساس کلمات بالاترین وزن اختصاص یافته به هر موضوع و با نظر کارشناسان حوزه موضوعی، انجام گرفت. با انجام محاسبات با ضریب اطمینان ۹۵٪، تعداد ۶۳ موضوع از بین ۲۶۰ موضوع اصلی انتخاب شد. با محاسبه میانگین تترابرجسب 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 الکترونیکی / بازاریابی الکترونیکی، بازاریابی محتوا، تحلیل داده و مهارت الکترونیکی، از جمله موضوعات داغ بود و موضوعات آموزش، آرشیو،

مدیریت دانش، سازمان‌دهی و سلامت کتابداران از جمله موضوعات سرد در حوزه علم اطلاعات و دانش‌شناسی در بازه زمانی ۲۰۱۰ تا ۲۰۱۹ شناسایی شد. تحلیل یافته‌ها نشانگر آن است که به دلیل رویکرد اغلب پژوهشگران در ۱۰ سال گذشته به استفاده از فناوری‌های نوظهور، 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، متن کاوی، علم اطلاعات و دانش‌شناسی

\* علی منصور، a.mansouri@edu.ui.ac.ir

## ۱. مقدمه

بررسی گذشته علم بشری نشان‌دهنده آن است که با گذشت زمان و پیدایش زمینه‌های جدید، زمینه‌های علمی و تحقیقاتی نوینی به وجود آمده و تکامل یافته (لاسیگ<sup>۱</sup>، ۲۰۱۶) و با توسعه جوامع بشری و ظهور کشفیات جدید و مبتنی بر نیاز بشر، حوزه‌های علمی نیز گسترش یافته است. این تکامل و بالندگی به گونه‌ای بوده است که انقلاب فناوری ارتباطی و اطلاعاتی در آن سبب شده تا عصر حاضر، شاهد رشد انبوهی از اطلاعات دیجیتال (العظمی<sup>۲</sup>، ۲۰۱۳) در قالب‌های گوناگون مقاله، کتاب و سایر باشیم.

بر اساس تعداد انتشارات و مراجع نمایه شده در پایگاه‌های مختلف، آهنگ رشد علوم مدرن در سال ۲۰۱۰، نرخ برابر ۸ تا ۹٪ داشته است این آهنگ رشد، هر ۹ سال تقریباً دو برابر می‌شود (بورنمن و موتز<sup>۳</sup>، ۲۰۱۵) و پیش‌بینی می‌شود که میزان انتشارات علمی جهانی در قالب مقاله در پایگاه وب آو ساینس از حدود ۳ میلیارد به حدود ۱۰ میلیارد در سال ۲۰۴۰ افزایش یابد (ونگ<sup>۴</sup>، ۲۰۲۰)؛ همانند نرخ رشد علمی سریع و قابل توجه در سایر رشته‌های علمی، حوزه‌های موضوعی و پژوهشی در رشته علم اطلاعات و دانش‌شناسی نیز، همواره در حال بالندگی و تکامل بوده و با گذشت زمان موضوعات فرعی بسیاری را در ذیل مجموعه خود ایجاد می‌کنند، به گونه‌ای که در بعد جهانی، پژوهشگران این حوزه با توجه به روند رو به رشد نسبی انتشارات علمی سالیانه در این حوزه (وینای و باساواراجا<sup>۵</sup>، ۲۰۱۹)، با حجم عظیمی از اطلاعات علمی و برون‌دادهای پژوهشی و موضوعات با حداکثر توجه با عنوان موضوعات داغ روبرو هستند. در هر دوره زمانی، موضوعات داغ، آن دسته از زمینه‌های پژوهشی هستند که از اهمیت ویژه‌ای در میان زمینه‌های فعال در یک عنوان پژوهشی مشخص برخوردار هستند (اردفلدر و بوسنچاک<sup>۶</sup>،

<sup>۱</sup> Lässig

<sup>۲</sup> Al-Azmi

<sup>۳</sup> Bornmann and Mutz

<sup>۴</sup> Wong

<sup>۵</sup> Vinay and Basavaraja

<sup>۶</sup> Erdfelder & Bošnjak

۲۰۱۶) 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 (بوریل، گیل، بنزوی و دنی<sup>۱</sup>، ۲۰۱۹). از طرف دیگر، موضوعات سرد، زمینه‌های پژوهشی هستند که از مرکز توجهات به‌طور کلی خارج شده 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 (گال، تیجس و گلانزی<sup>۲</sup>، ۲۰۱۹). وجود این دو گروه از موضوعات در هر حوزه علمی از جمله علم اطلاعات و دانش‌شناسی، باعث رشد ناهمگون تحقیقات در یک موضوع خاص شده و اصل توازن میان موضوعات یک زمینه 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 بگیرد، توانایی شناسایی موضوعات داغ در هر حوزه علاوه بر هدایت صحیح برای انجام یک پژوهش موفق، باعث شناخت موفق-ترین 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 توسعه موضوعی رشته علم اطلاعات و دانش‌شناسی<sup>۳</sup> در سطح بین‌المللی و تعیین موضوعات داغ و سرد در راستای شناسایی اولویت‌های پژوهشی و کمک به سیاست‌گذاری در حوزه علم اطلاعات و دانش‌شناسی و نظر به اهمیت و کاربرد گسترده روش متن کاوی، پژوهش حاضر با استفاده از تکنیک‌های متن کاوی در صدد شناسایی حوزه‌های موضوعی اصلی علم اطلاعات و دانش‌شناسی و موضوع‌های داغ و سرد است.

## ۲. پیشینه پژوهش

در پژوهش‌های مختلف، 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 جمله این روش‌ها، روش تحلیل هم‌واژگانی و هم‌رخدادی واژگان، تحلیل کتاب‌سنجی، تحلیل محتوا - 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 (باغ محمد، منصوری و چشمه سهرابی، ۱۳۹۹)؛ اما در دهه اخیر روش متن کاوی از زیرمجموعه داده کاوی به‌صورت گسترده در پژوهش‌های

<sup>۱</sup> Burrill, Gail, Ben-Zvi and Dani

<sup>۲</sup> Gal, Thijs, Glanzel

<sup>۳</sup> به‌منظور هماهنگی با پژوهش‌های بین‌المللی و مقایسه نتایج، از این به بعد به‌جای علم اطلاعات و دانش‌شناسی، از عنوان علوم کتابداری و اطلاع‌رسانی استفاده خواهد شد.

مختلف به منظور کشف موضوعات، شناخت الگوی دانشی و تحلیل روند استفاده شده است و این روند نیز نرخ صعودی دارد (جونگا و گیون لی<sup>۱</sup>، ۲۰۲۰؛ کاتسورای و جو<sup>۲</sup>، ۲۰۲۱).

روش متن کاوی از جمله روش‌های بین‌رشته‌ای است که در سال‌های اخیر در پژوهش‌های مختلف به منظور کشف الگوی دانشی و موضوعات نهان و آشکار موضوعات و حوزه‌های مختلف موضوعی مختلف 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 متن کاوی مورد کنکاش قرار می‌گیرد. از جمله موضوعاتی که با استفاده از متن کاوی مورد تحلیل قرار گرفته، علم رایانه و اطلاعات (کیم<sup>۳</sup> و همکارانش، ۲۰۱۵) است که چهار الگوی متفاوت در روندهای موضوعی این حوزه شناسایی شده که الگوهای رشد یافته، در حال افزایش یا بالغ (موضوعات مربوط به شبکه‌های رایانه‌ای)، الگوهای در حال کاهش (موضوعات مربوط به داده کاوی)، الگوهای در حال پیشرفت (وب، بازیابی اطلاعات با متن کاوی و موضوعات مرتبط با پایگاه‌های داده) و الگوهای در حال نوسان و تغییر (سامانه‌های اطلاعاتی و موضوعات مربوط به سامانه‌های چندرسانه‌ای) است. دیگر موضوع که مورد توجه پژوهشگران حوزه داده کاوی و متن کاوی قرار گرفته روان‌شناسی است (بیترمن و آندره فیشر<sup>۴</sup>، ۲۰۱۸) که با استفاده از لغات کلیدی آن حوزه به شناسایی موضوعات اصلی و داغ و روند آن‌ها پرداخته است. حوزه پزشکی یکی از حوزه‌هایی است که به صورت گسترده از این تکنیک برای شناسایی موضوعات و تحلیل روند استفاده کرده است. حوزه‌ی بیماری‌های قلب (گال، تیجس و گلانزی، ۲۰۱۹) بیماری کوید-۱۹ (وانگ و لو، ۲۰۲۱؛ کاراسدو، پویرتاس و مارتی<sup>۵</sup>، ۲۰۲۱؛ توروووسکی<sup>۶</sup> و همکاران ۲۰۲۱؛ زنگول<sup>۷</sup> و همکاران، ۲۰۲۱) حوزه‌ی سلامت (ووتر<sup>۸</sup> و همکاران ۲۰۲۱)؛ بیمار (چن، گوو، وو<sup>۹</sup> و جو، ۲۰۲۰) نمونه‌های کوچکی از این پژوهش‌ها هستند. رسانه‌های اجتماعی (مانوهاران<sup>۱۰</sup>، ۲۰۲۰)؛ نانو تکنولوژی (لی، فن، ژو<sup>۱۱</sup> و همکاران، ۲۰۲۰) از جمله دیگر پژوهش‌هایی است که از این تکنیک استفاده کرده است. در ایران نیز در سال‌های اخیر این تکنیک در پژوهش‌های مختلفی مورد استفاده گرفته است. یادگیری الکترونیکی (کیهانی

<sup>1</sup> H. Junga, B.G. Lee

<sup>2</sup> M. Katsurai, S. Joo

<sup>3</sup> Kim

<sup>4</sup> Andre Bitterman and Andre Fischer

<sup>5</sup> Carracedo, Puertas and Marti

<sup>6</sup> Tworowski

<sup>7</sup> Zengul

<sup>8</sup> Wouter

<sup>9</sup> Chen, Guo, Wu and Ju

<sup>10</sup> Manoharan

<sup>11</sup> Li, Fan, Zhou

و آقا کاردان، ۱۳۹۱؛ قرآن کریم (صالحی شهرودی، مینایی و اشرفی، ۱۳۹۲؛ اصلانی و اسماعیلی، ۱۳۹۷؛ آذری، شکراللهی فر و لسانی فشارکی، ۱۳۹۹) مدیریت دانش (رضایی نور و شیخ بهایی، ۱۳۹۶)؛ خوشه‌بندی مقالات (سلیمانی نژاد، سلاجقه و طیبی، ۱۳۹۷)؛ تشخیص حس در متن‌های فارسی (گرشاسبی، رئیس روحانی و کاباران زاده قدیم، ۱۳۹۷)؛ اینترنت اشیاء (برنا، فتحی و مؤمنی ۱۳۹۷) حوزه سلامت (شکوهیان و همکاران، ۱۳۹۸) فهرستی کوتاه از پژوهش‌هایی است که به صورت گسترده از تکنیک متن کاوی استفاده کرده‌اند.

در حوزه علوم کتابداری و اطلاع‌رسانی انتخاب مجلات مناسب (جونگ و لی<sup>۱</sup>، ۲۰۲۰)؛ سازمان‌دهی اطلاعات (پرئی و حمیدی، ۱۳۹۶)؛ بازیابی اطلاعات (جعفری پاورسی و همکاران، ۱۳۹۷) در کنار پژوهش‌هایی که به طور ویژه به بررسی موضوعات داغ و سرد در این حوزه پرداخته‌اند، قرار دارد؛ اما با توجه به ماهیت پژوهش‌هایی که به بررسی موضوعات داغ و سرد با استفاده از روش متن کاوی در این حوزه می‌پردازند، تعداد محدودی پژوهش به این موضوع پرداخته است. لی، کیم و کیم<sup>۲</sup> (۲۰۱۰) در پژوهشی به بررسی موضوعات حوزه کتابخانه دیجیتال پرداخته که سه حوزه کلی پژوهش‌های پایه، مطالعات کاربران و فراداده شناسایی شد. موضوعات داغ شناسایی شده در حوزه علوم کتابداری و اطلاع‌رسانی کره جنوبی در یک فرایند زمانی طولانی (۱۹۷۰-۲۰۱۹) را ارزیابی کتابخانه در حوزه مدیریت کتابخانه، خدمات اطلاعاتی در حوزه خدمات کتابخانه و مدارک قدیمی در حوزه کتاب‌شناسی تشکیل می‌دهند (پارک<sup>۳</sup> و همکاران ۲۰۱۳).

پژوهش توسعه و روند موضوعی در حوزه علوم کتابداری و اطلاع‌رسانی ایران (باغ محمد، منصوری و چشمه سهرابی، ۱۳۹۹) نشان داد که غالب موضوعات خدمات کتابخانه‌ای در شبکه‌های اجتماعی، مدل‌های پژوهش، سرمایه اجتماعی، پایگاه‌های اطلاعاتی پزشکی، داده کاوی، روند تولید علمی، موضوع‌های بین‌رشته‌ای، الگوریتم‌های فضای مجازی، مدیریت دانش، مطالعات شبکه‌های اجتماعی، رویکردهای پژوهشی و آینده پژوهی قرار دارند. پژوهش تیماکوم، کیم و سونگ<sup>۴</sup> (۲۰۲۰) با بررسی ۶ مجله برتر حوزه علوم کتابداری و اطلاع‌رسانی در محدوده سال‌های ۱۹۹۷-۲۰۱۶ نشان داد که موضوعات کتابخانه دانشگاهی، کتابخانه دیجیتال، روش‌شناسی پژوهش، بازاریابی کتابخانه، بازیابی اطلاعات، اطلاعات دیجیتال و استناد به مدارک از جمله موضوعات داغ بوده است. تحلیل محتوایی آثار مرتبط با کتابداری پزشکی با استفاده از متن کاوی نشان داد که موضوعات علم‌سنجی، سواد اطلاعاتی و اطلاعات سلامت بیشترین انتشارات را دارند و در زمره موضوعات داغ قرار دارند (دستانی، موسوی چیلک، ضیایی و دل‌قندی، ۲۰۲۰).

<sup>1</sup> Jung and Lee

<sup>2</sup> Lee, Kim and Kim

<sup>3</sup> Park

<sup>4</sup> Timakum, Kim and Song

بررسی پیشینه‌های مرتبط با علوم کتابداری و اطلاع‌رسانی نشان می‌دهد جامعه موردبررسی پژوهش‌هایی که در حوزه کتابداری و اطلاع‌رسانی انجام شده محدود بوده و اغلب مقالات مجلات خاص را پوشش داده یا موضوع خاص یا حوزه جغرافیایی مشخصی را بررسی کرده‌اند. بررسی پیشینه‌ها به‌طور عام نیز نشان می‌دهد که عمده پژوهش‌هایی که از روش متن‌کاوی به‌منظور شناسایی موضوعات داغ و سرد و یا موضوعات اصلی این حوزه پرداخته‌اند از روش مدل‌سازی موضوعی و از روش تخصیص دیریکلت پنهان (LDA) استفاده کرده‌اند. برای وزن‌دهی کلمات از روش TF-IDF استفاده شده و نتیجه مشترک اغلب آن‌ها این است که روش متن‌کاوی از زیرمجموعه داده‌کاوی تکنیک مناسبی برای کشف موضوعات و الگوی دانشی موجود در متون بدون ساختار علمی بخصوص مقاله است؛ بنابراین پژوهش حاضر با استفاده از تکنیک متن‌کاوی با هدف شناسایی موضوعات داغ و سرد حوزه 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 زیرمجموعه آن روش مدل‌سازی موضوعی و الگوریتم LDA برای شناسایی موضوع‌های داغ و سرد استفاده شد. جامعه آماری شامل تمامی مقالات منتشرشده حوزه علم اطلاعات و دانش‌شناسی (علوم کتابداری و اطلاع‌رسانی) است که در بازه زمانی ۲۰۱۰ تا ۲۰۱۹ در ۲۴۸ مجله مرتبط با این حوزه در پایگاه اسکوپوس نمایه شده است. برای تحلیل داده‌ها، عنوان و چکیده مقالات استخراج و مورد تحلیل قرار گرفت.

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

- **آماده‌سازی:** در این مرحله بر روی داده‌های ابتدایی و خام، عملیات پیش‌پردازش انجام شد، چراکه این مرحله به‌عنوان اساس و پایه در پژوهش‌هایی که مبانی آن‌ها متن‌کاوی است، نام برده می‌شود و از اهمیت و ضرورت خاصی در فرایند پژوهش برخوردار است (تسنگ و لین<sup>۱</sup>، ۲۰۰۳). این مرحله، شامل مراحل سه‌گانه‌ی توکن‌سازی به‌عنوان اولین مرحله از نمایه‌سازی متن، وظیفه تکه‌تکه کردن متن به کلمات تشکیل‌دهنده آن و تولید متن‌هایی که به زبان طبیعی نوشته شده‌اند اعمال می‌شود (تهو جو<sup>۲</sup>، ۲۰۱۹)، استیمینگ<sup>۳</sup> یعنی تبدیل هر واژه یا عبارتی که از مرحله

<sup>1</sup> Tseng and Lin

<sup>2</sup> Taeho Jo

<sup>3</sup> Steaming

قبل تولید شده است به شکل ریشه‌ای خود (کوالسکی، میبوری<sup>۱</sup>، ۲۰۰۰) و حذف کلمات نامعتبر<sup>۲</sup> برای محاسبه و تعیین وزن کلمات در یک متن می‌شود. این مراحل باید به صورت سلسله‌مراتبی انجام شود (تهئو جو، ۲۰۱۹):

- **وزن دهی کلمات:** بعد از پیش‌پردازش، برای وزن‌دهی به کلمات از روش TF-IDF که احتمال یا درصد وقوع هر کلمه در متن است، به عنوان معیار وزن‌دهی کلمات استفاده شد. در این مرحله کلماتی که دارای بیشترین وزن بودند شناسایی و به عنوان کلمات کلیدی انتخاب شدند.

- **ایجاد ماتریس سند - کلمه:** برای تحلیل متن با استفاده از تکنیک‌های متن کاوی نیاز است که متن 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 خوشه موضوعی:** وزن کلمات شناسایی شده از متن نسبت به همدیگر متفاوت بوده و این تفاوت ممکن است نسبت به هر خوشه‌ای که کلمه به آن تعلق می‌گیرد متفاوت باشد. خوشه موضوعی شامل تعدادی کلیدواژه است که به نوعی با همدیگر ارتباط دارند. به منظور تعیین رابطه بین کلیدواژه‌ها و تعلق هر کلیدواژه به خوشه موضوعی از روش مدل‌سازی موضوعی و به طور خاص از الگوریتم LDA استفاده شد. LDA یکی از معروف‌ترین و پرکاربردترین روش 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 (سیلیج و رویسنسون<sup>۳</sup>، ۲۰۱۷). پس از استخراج کلیدواژه‌ها و تعیین خوشه‌ها، ۱۵ کلمه برتر هر موضوع (خوشه) شناسایی و با کمک متخصصان حوزه موضوعی به این خوشه‌ها برچسب موضوعی مناسب داده شد. به منظور تعیین تراکم کلمات در یک موضوع، از کمیت تنا استفاده شد. در این مرحله، میانگین تنا (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 تنا به عنوان موضوعات داغ و موضوعات با کمترین مقدار میانگین تنا در طول ۱۰ سال به عنوان موضوعات سرد شناسایی شد. برای اجرای محاسبات و استخراج یافته‌ها از زبان برنامه‌نویسی پایتون و مجموعه کتابخانه‌های مرتبط با داده کاوی و متن کاوی در این زبان برنامه‌نویسی و به طور خاص از کتابخانه جنسیم<sup>۴</sup> (رهورک و سوچنش<sup>۵</sup>، ۲۰۱۰) استفاده شد.

<sup>1</sup> Kowalski, G.J. Maybury, M

<sup>2</sup> Stop List

<sup>3</sup> Silge and Robinson

<sup>4</sup> Gensim library

<sup>5</sup> Rehurek & Sojka

برای خوشه‌بندی از الگوریتم K- میانگین<sup>۱</sup> استفاده شد. خوشه‌بندی با استفاده از الگوریتم K- میانگین مهم‌ترین و کاربردی‌ترین نوع خوشه‌بندی از نوع بخش‌بندی است که استفاده از آن برای مجموعه داده‌های بزرگ مناسب است و الگوریتم آن با تعداد از قبل مشخص شده‌ای از خوشه‌های اسناد (به‌عنوان مثال تعداد خوشه‌های k) شروع می‌شود (سالتون و بوکلی<sup>۲</sup>، ۱۹۸۸). 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. این روش خوشه‌بندی، بر اساس n مشاهده‌گر و k گروه، عملیات خوشه‌بندی صورت می‌گیرد. از این رو، الگوریتم K- میانگین فاصله‌ی داخل خوشه را به حداقل می‌رساند (همرلی و الکان<sup>۳</sup>، ۲۰۰۲). الگوریتم این روش، با تعداد K مرکزیت برای خوشه‌ها شروع می‌شود، سپس، هر الگو در مجموعه داده‌ها به نزدیک‌ترین خوشه (نزدیک‌ترین مرکزیت) اختصاص داده می‌شود. سرانجام، مرکزیت‌ها طبق الگوهای مرتبط با آن، مجدداً محاسبه می‌شوند. این روند تا رسیدن به یک همگرایی مشخص (رسیدن به یک معیار عددی که معمولاً توسط کاربر به الگوریتم داده می‌شود، مانند معیارهای متوقف کننده) تکرار می‌شود. به عبارت دیگر، الگوریتم جابجایی مدارک و محاسبه فاصله آن تا خوشه مرکزی آن قدر تکرار می‌شود تا به دلیل جابجایی مدارک، تغییری در خوشه مرکزی ایجاد نشود.

برای کشف مراکز اولیه خوشه‌ها و انجام خوشه‌بندی موضوعات و در نهایت 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، بکار می‌روند. به‌طور دقیق‌تر، عدد آلفا نشان‌دهنده میزان پراکندگی موضوعات موجود در مدارک است. واضح است که این پراکندگی هرچه بیشتر باشد، مطلوب‌تر است، چراکه 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 به عبارتی آن مدرک تخصصی نیست، درحالی‌که آلفای کوچک بدین معنی است که یک مدرک شامل تعداد خاص و محدودی از موضوعات است، به همین دلیل است که اغلب عدد آلفا به صورت کسری از تعداد موضوعات k تنظیم می‌شود. در این پژوهش نیز، مقدار آلفا به صورت پیش فرض، با در نظر گرفتن فرض توزیع متقارن (سیمتریک) برای مدل LDA به صورت  $1/k$  انتخاب شده است.

معیار عدد بتا، چگالی کلمات در موضوع است و بر اساس 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 حوزه

<sup>1</sup> K-means

<sup>2</sup> Salton.G, Buckley. C

<sup>3</sup> Hamerly, Elkan



(خوشه) موضوعی و به تبع پراکندگی موضوعی بالای کلمات است، درحالی که ما انتظار داریم که همه کلمات داخل یک موضوع تا حد امکان مشابه و نزدیک به هم باشند؛ بنابراین به علت تعداد زیاد موضوعات و لحاظ شدن کلمات کمتر دیده شده در موضوعات، از کمیت بتای کوچک برای یافتن بیشتری از کلمات متعلق به موضوعات خاص تری استفاده می‌شود. در این پژوهش، این پارامتر برابر  $0/001$  در نظر گرفته شده است که عدد مناسبی است.

### تعیین تعداد خوشه‌های موضوعی

مدل‌سازی موضوعی، موضوعات را به صورت خودکار از اسناد بدون برچسب و به روش غیر نظارتی استخراج می‌کند. به همین منظور، در ابتدا لازم است تعداد خوشه‌های موضوعی تعیین شود. برای انجام این مرحله، ضروری است که مقدار اولیه‌ای در خصوص تعداد موضوعات منظور شود. مشورت با متخصصان موضوعی، استفاده کردن از برخی استانداردها از قبیل سر عنوان موضوعی و روش‌های آماری از جمله روش‌های تعیین مقدار اولیه تعداد موضوعات است. در این پژوهش جهت به دست آوردن تعداد خوشه مطلوب، ابتدا بررسی پایگاه معتبر LISA که جامع‌ترین پایگاه مقالات در حوزه علوم کتابداری و اطلاع‌رسانی، 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 مرحله ابتدایی شناسایی شد.

در مرحله بعد برای به دست آوردن بهینه‌ترین تعداد موضوعات، طبق مدل نمونه‌گیری گیبس<sup>۱</sup>، تعداد موضوعات در دامنه‌ی ۲۵۰-۲۹۰ مورد آزمون قرار گرفت و با استفاده از الگوریتم‌های ارزیابی مدل‌های موضوعی سه‌گانه شامل انسجام<sup>۲</sup> یا همبستگی، احتمال<sup>۳</sup> یا تناسب و عدم یکدستی<sup>۴</sup> (به عنوان پارامترهای LDA) اجرا شد. متوسط ضریب انسجام که میزان مشابهت معنایی میان لغات با بیشترین دامنه تکرار در موضوعات را تعیین می‌نماید، می‌تواند به تمایز میان موضوعاتی که از لحاظ معنایی قابل تفسیر هستند و موضوعاتی که ساخته‌ی 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 به چه خوبی یک نمونه را پیش‌بینی می‌کند، یعنی مدل چقدر توسط نمونه‌ای از داده‌های مشاهده شده 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 (استیونس<sup>۵</sup> و همکاران، ۲۰۱۲). نتایج ارزیابی تعداد حوزه موضوعی، در شکل‌های ۱ الی ۳ آمده است.

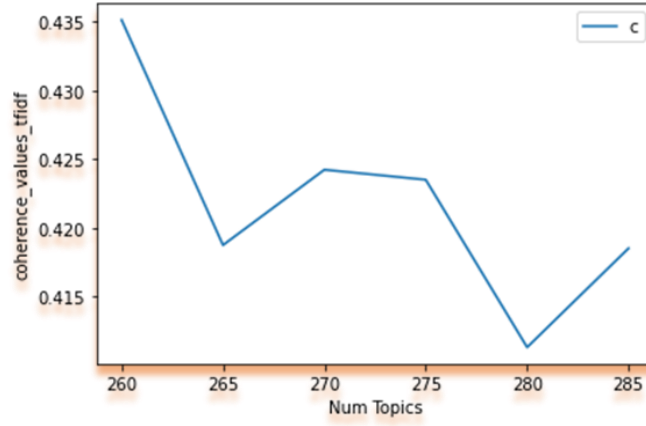
<sup>1</sup> Gibbs Sampling

<sup>2</sup> Coherence

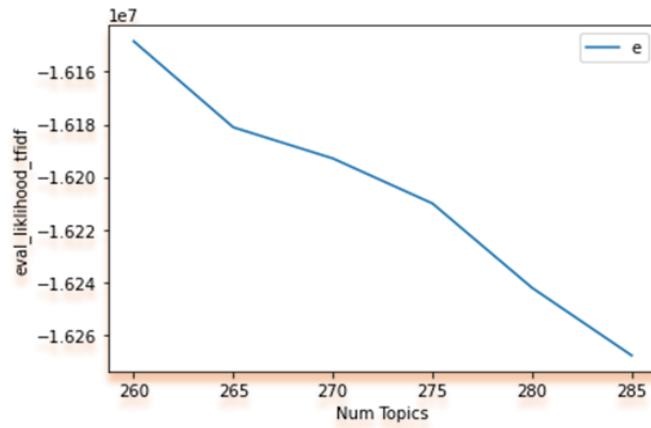
<sup>3</sup> Likelihood

<sup>4</sup> Perplexity

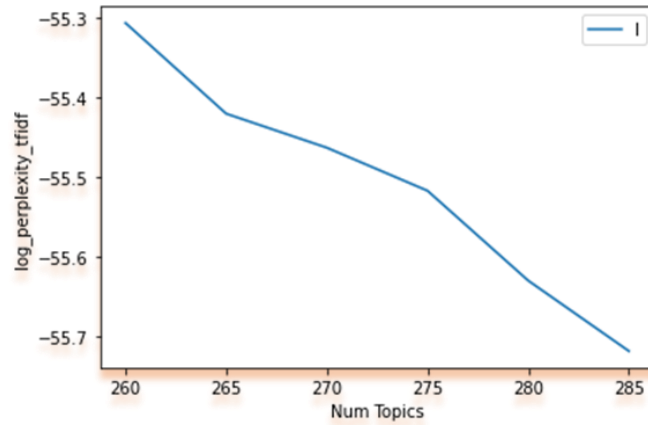
<sup>5</sup> Stevens



شکل ۱- اجرای الگوریتم انسجام

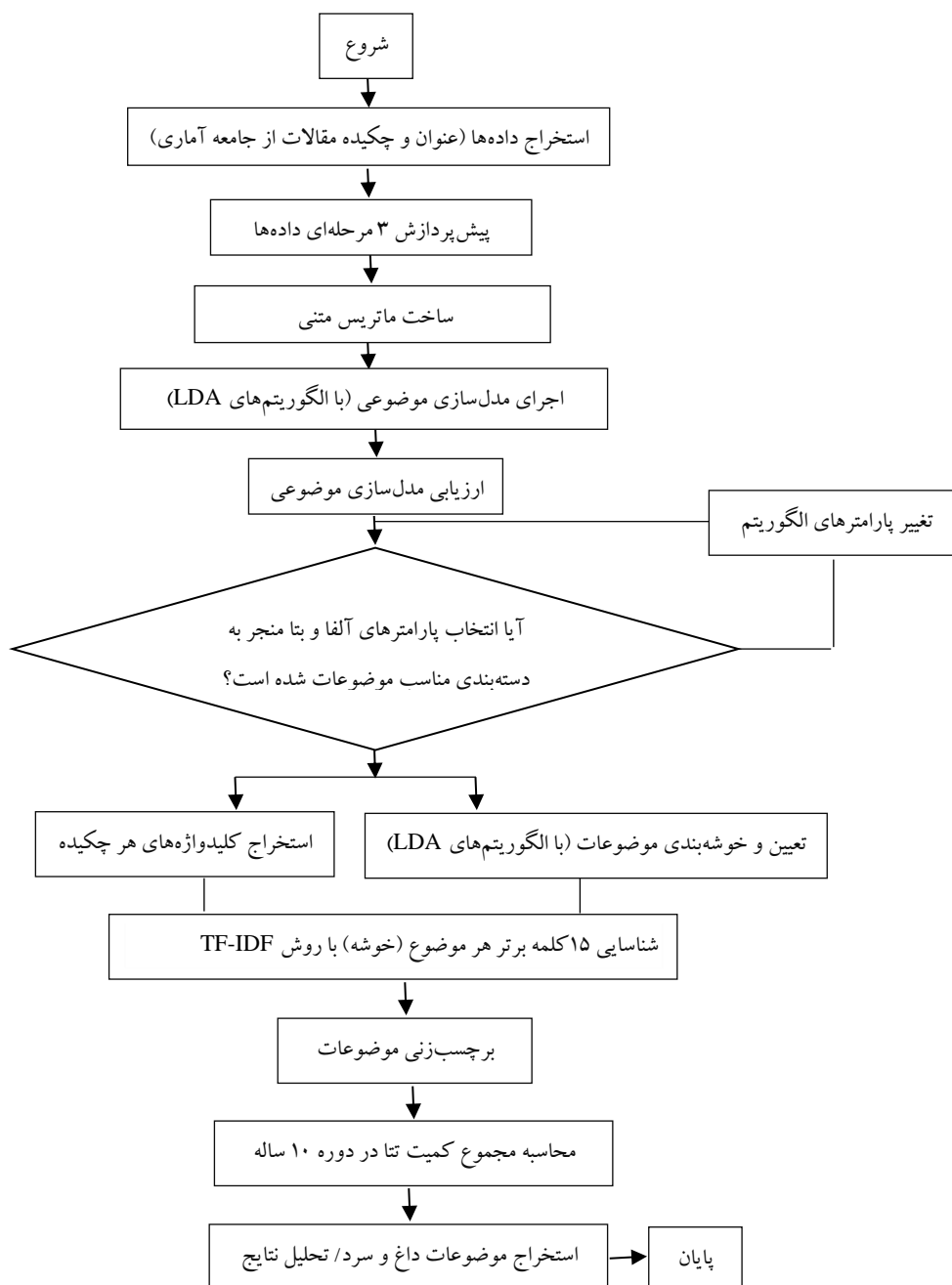


شکل ۲- اجرای الگوریتم احتمال / تناسب



شکل ۳- اجرای الگوریتم عدم یکدستی

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



نمودار ۱- نمایش فرایند انجام پژوهش



در ادامه به منظور تعیین موضوعات (خوشه‌ها) میانگین تتا (به معنی میزان توزیع موضوعات در اسناد بر اساس سال انتشار آن‌ها)، با انتخاب کوچک کمیت آلفا (به مفهوم وزن موضوعات در اسناد) و با هدف محدود نمودن یک سند به تعداد خاصی از موضوعات، برابر با عدد  $1/k = 1/260 = 0.003$  (تعداد بهینه شده موضوعات) محاسبه شد و یک شکل خطی از میانگین ضرایب اشتراک موضوعات (تتا) بر حسب سال رسم و محاسبات با ضریب اطمینان ۹۵٪ منجر به انتخاب تعداد ۶۳ موضوع (به عنوان موضوعات داغ و سرد) از بین ۲۶۰ موضوع شد. از بین این موضوعات، تعداد ۲۴ موضوع دارای بالاترین مقدار میانگین تتا و به عنوان موضوعات داغ شناسایی گردید. از بین ۲۴ موضوعات داغ به دست آمده، موضوعات دگر سنجی / مطالعات سنجشی / سنجش و اندازه گیری (مجموع تتا برابر با ۰/۱۰۴۷)، مدیریت الکترونیکی / بازاریابی دیجیتالی (مجموع تتا برابر با ۰/۰۹۶۰)، بازیابی محتوا / متن کاوی (مجموع تتا برابر با ۰/۰۸۸۹) و شبکه و تحلیل داده (مجموع تتا برابر با ۰/۰۸۶۶) و مهارت الکترونیکی (مجموع تتا برابر با ۰/۰۷۸۸) با دارا بودن بالاترین مقادیر مجموع میانگین تتا، به عنوان ۵ موضوع برتر داغ در علوم کتابداری و اطلاع‌رسانی در بازه 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.

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

شماره و عنوان موضوع داغ	مجموع میانگین ضرایب تتا	شماره و عنوان موضوع داغ	مجموع میانگین ضرایب تتا	شماره و عنوان موضوع داغ	مجموع میانگین ضرایب تتا
۱۹۰ (سنجش و اندازه گیری)	۰/۱۰۴۷	۱۶۸ (رضایتمندی خدمات)	۰/۰۶۷۹	۲۳۱ (حریم خصوصی کاربران)	۰/۰۵۲۵
۲۲۷ (مدیریت و بازاریابی الکترونیکی)	۰/۰۹۶۰	۱۲۰ (مصورسازی اطلاعات)	۰/۰۶۷۳	۲۲۴ (رضایتمندی شغلی)	۰/۰۵۱۱
۲۵۵ (بازیابی محتوا / متن کاوی)	۰/۰۸۸۹	۲۰۹ (آموزش عمومی)	۰/۰۶۳۴	۲۴۰ (مطالعه جامعه آماری)	۰/۰۴۴۳
۱۶۹ (یادگیری الکترونیکی)	۰/۰۸۶۸	۱۹۱ (فناوری اطلاعات در آموزش روستاییان و عشایر)	۰/۰۶۱۷	۶۸ (مطبوعات الکترونیکی)	۰/۰۴۳۶
۱۰۰	۰/۰۸۶۶	۱۵۲	۰/۰۵۹۱	۱۱۵	۰/۰۴۳۳

(تحلیل داده)		(وب سنجی)		(پردازش داده‌ها)	
۲۱۲ (مهارت الکترونیکی)	۰/۰۷۸۸	۱۳۵ (نظرسنجی)	۰/۰۵۷۵	۱۸۰ (استاددهی)	۰/۰۴۰۶
۲۰۶ (برنامه‌نویسی و کامپیوتر)	۰/۰۷۱۷	۱۶۳ (فناوری اطلاعات)	۰/۰۵۷۱	۶۲ (عضویت کتابخانه‌ای)	۰/۰۳۹۴
۱۴۹ (کتابخانه قفسه باز)	۰/۰۶۸۹	۲۶ (خدمات کتابخانه‌ای)	۰/۰۵۵۳	۱۰۹ (خدمات الکترونیک)	۰/۰۲۳۳

مشابه روشی که برای شناسایی موضوعات داغ انجام شد، برای شناسایی موضوعات سرد نیز از همان روش استفاده شد. به عبارت دیگر، از بین ۶۳ موضوع مورداشاره، ۳۹ موضوع کمترین مقدار میانگین تا را داشته‌اند. جدول ۲ مجموع میانگین پارامتر تا را در سال‌های مختلف (۲۰۱۰ تا ۲۰۱۹) برای موضوعات سرد نشان می‌دهد.

بر اساس نتایج ارائه‌شده در جدول ۲، تعداد ۵ حوزه از اصطلاحاً سردترین موضوعات که به ترتیب دارای کم‌ترین مقدار مجموع میانگین پارامتر تا برای سال‌های ۲۰۱۰ تا ۲۰۱۹ می‌باشند، به ترتیب عبارت‌اند از: موضوع سلامت کتابداران (مجموع تا برابر با ۰/۰۳۳۰)، موضوع سازمان‌دهی (مجموع تا برابر با ۰/۰۳۷۷)، موضوع مدیریت دانش / درخت تصمیم (مجموع تا برابر با ۰/۰۳۸۷)، موضوع گردآوری اطلاعات / آرشیو (مجموع تا برابر با ۰/۰۴۲۷) و آموزش (مجموع تا برابر با ۰/۰۴۹۳).

جدول ۲- مجموع میانگین پارامتر تا در سال‌های ۲۰۱۰ تا ۲۰۱۹ برای موضوعات سرد

شماره و عنوان موضوع سرد	مجموع میانگین ضرایب تا	شماره و عنوان موضوع سرد	مجموع میانگین ضرایب تا	شماره و عنوان موضوع سرد	مجموع میانگین ضرایب تا
۲۲ (اطلاع‌رسانی)	۰/۰۱۶۰	۱۸ (مدیریت کتابخانه)	۰/۰۲۳۵	۱۸۸ (فهرست‌نویسی)	۰/۰۲۷۲
۶۷ (رفتار اطلاعاتی)	۰/۰۱۶۱	۱۷۱ (نمونه‌گیری)	۰/۰۲۳۵	۱۷۶ (ارزیابی اطلاعات)	۰/۰۲۷۴
۱۵ (مباحث نظری و عمومی اطلاعات)	۰/۰۱۷۰	۱۶۰ (پردازش داده)	۰/۰۲۴۰	۱۱۶ (توسعه مطالعات کتابداران)	۰/۰۲۷۸
۱۲ (روش تحقیق)	۰/۰۱۸۲	۲۱۰ (بازیابی اجتماعی)	۰/۰۲۴۸	۱۲۷ (بازیابی اطلاعات)	۰/۰۲۸۰
۱۳۳ (موتورهای جستجو)	۰/۰۱۸۷	۲۱۶ (رفتار کتابداران)	۰/۰۲۵۶	۲۴۶ (اطلاع‌سنجی)	۰/۰۲۹۶

۱۹۴	۰/۰۱۸۸	۱۳۹	۰/۰۲۵۶	۱۳۰	۰/۰۳۱۴
(نظارت بر اطلاع‌رسانی)		(اخلاق حرفه‌ای)		(هستی‌شناسی)	
۱۳۱	۰/۰۱۹۶	۲۴۲	۰/۰۲۵۶	۱۳۷	۰/۰۳۲۶
(منابع آموزشی)		(طبقه‌بندی داده‌ها)		(اطلاع‌رسانی الکترونیکی)	
۱۵۰	۰/۰۱۹۸	۲۲۶	۰/۰۲۵۸	۱۹	۰/۰۳۳۰
(پایگاه داده، اطلاع‌رسانی)		(نمایه‌سازی)		(سلامت کتابداران)	
۶۰	۰/۰۲۰۵	۴۱	۰/۰۲۶۰	۲۱۷	۰/۰۳۵۳
(ارتباط الکترونیکی)		(جستجوی اطلاعات)		(استناددهی)	
۲۳۳	۰/۰۲۲۷	۲۵	۰/۰۲۶۳	۲۵۶	۰/۰۳۷۷
(وبلاگ‌نویسی)		(مدل‌های سیستمی پویا)		(سازمان‌دهی)	
۲۵۹	۰/۰۲۲۷	۱۰۶	۰/۰۲۶۳	۱۰۲	۰/۰۳۸۷
(رابط کاربر)		(جمع‌آوری اطلاعات)		(مطالعات یادگیری)	
۱۶	۰/۰۲۳۲	۲۸	۰/۰۲۶۷	۳۸	۰/۰۴۲۷
(آموزش اطلاعاتی)		(بازیابی معنایی)		(آرشیو)	
۷۵	۰/۰۲۳۴	۲۰۴	۰/۰۲۶۷	۱۶۷	۰/۰۴۹۳
(وب‌سنجی)		(ارائه اطلاعات)		(کتابخانه آنلاین)	

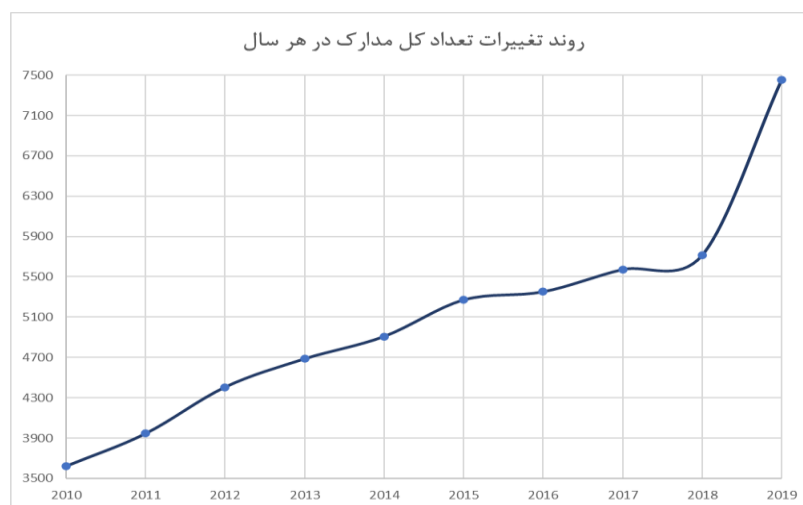
## ۵. بحث و نتیجه‌گیری

تحلیل داده‌های پژوهش حاضر چشم‌انداز کلی از پژوهش‌های حوزه علوم کتابداری و اطلاع‌رسانی ارائه کرد. بعضی از حوزه‌هایی که بیش از سایر موضوعات در پژوهش‌های مختلف مورد توجه قرار گرفتند به‌عنوان موضوعات داغ شناسایی شد (جدول ۱) و بعضی از موضوعات که کمترین میزان توجه به آن‌ها شده بود با نام موضوعات سرد معرفی شد (جدول ۲).

در پژوهش حاضر با استفاده از مدل‌سازی موضوعی با الگوریتم LDA و محاسبه پارامترهای این الگوریتم، ۲۶۰ موضوع مورد پژوهش حوزه علوم کتابداری و اطلاع‌رسانی شناسایی شد. نتایج پژوهش حاصل از به‌کارگیری فن مدل‌سازی موضوعی و به‌طور خاص الگوریتم LDA نشان داد که موضوع‌های شناسایی شده در مدل‌سازی موضوعی نسبت به موضوع‌های به‌دست آمده از سایر پژوهش‌ها از نظر کمیت و تعداد بیشتر بوده است (به عنوان مثال در پژوهش لی، کیم و کیم (۲۰۱۰) تعداد موضوعات شناسایی شده 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 حداکثر موضوعات از متون در علوم کتابداری و اطلاع‌رسانی را دارد.

در ادامه، روند تغییرات تعداد چکیده مقالات در حوزه علوم کتابداری و اطلاع‌رسانی در بازه 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٪ رشد نسبت به سال قبل) و سال ۲۰۱۱ دارای کمترین تعداد مدرک (۱۰٪ رشد نسبت به سال قبل) است. تغییرات شیب در این نمودار در فاصله سال‌های مختلف، نشانگر آهنگ متغیر رشد انتشارات است. سیر رو به رشد 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 بوده به صورتی که تعداد انتشارات رشد تقریبی ۶۵٪ را نشان می‌دهد. تعداد مدارک در سال ۲۰۱۹ (با حدود ۷۵۰۰ مدرک) نسبت به سال قبل از آن (۲۰۱۸) رشدی برابر ۳۰٪ و نسبت به سال ۲۰۱۰ با داشتن تعداد تقریبی ۳۵۰۰ تعداد انتشارات، جهشی بیش از دو برابری را تجربه نموده است.

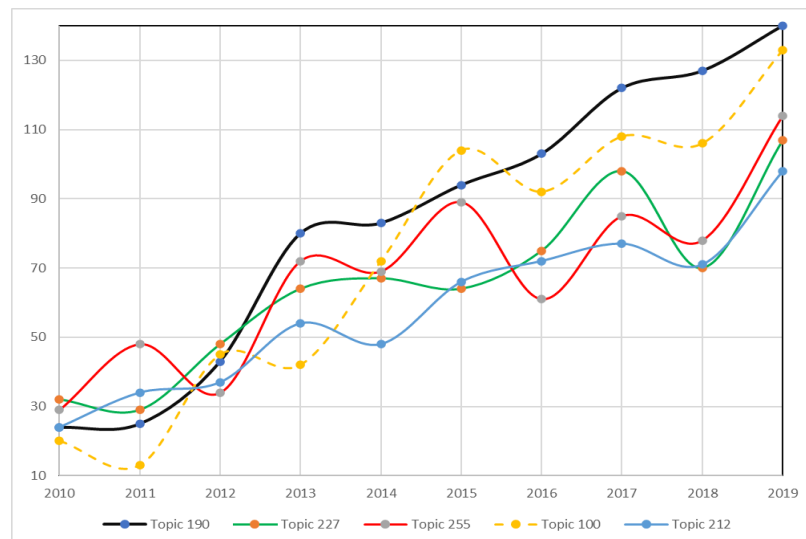


نمودار ۳- تغییرات در تعداد انتشارات حوزه علوم کتابداری و اطلاع‌رسانی در بازه ۲۰۱۰ تا ۲۰۱۹

بررسی روند شکل‌گیری و توسعه موضوعات حوزه کتابداری و اطلاع‌رسانی در گذر تاریخ نشان می‌دهد که این موضوعات تحت تأثیر پیشرفت‌ها، توسعه 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 گونه دیگر تغییر ماهیت داده و این سرعت تغییر در سال‌های اخیر به دلیل سرعت تغییر در جامعه، نیازها و ابزارها و امکانات بیشتر شده است. در مقایسه با پژوهش کاجبرگ (۱۹۹۶)، نتایج پژوهش حاضر نشان می‌دهد که اغلب موضوعات پژوهشی سال‌های ۱۹۵۷-۸۶ دغدغه 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 مریم باغ محمد، علی منصوری، مهرداد چشمه سهرابی؛ ۱۳۹۹ و دستانی، موسوی چیلک، ضیایی و دل قندی، ۲۰۲۰ مورد تأیید قرار گرفته است.



نمودار ۴- تغییرات در تعداد انتشارات مربوط به موضوعات داغ در حوزه علوم کتابداری و اطلاع‌رسانی در بازه ۲۰۱۰ تا ۲۰۱۹

همچنین بررسی نتایج پژوهش حاضر در مقایسه با سایر پژوهش نیز نشان داد که هرچه به سمت عصر فناوری پیش رفته‌ایم، اغلب موضوعات پایه موضوعات علوم کتابداری و اطلاع‌رسانی کم‌رنگ شده و حتی جزء موضوعات سرد قرار گرفته است (جدول ۲). با نگاهی مجدد به نتایج پژوهش کاجبرگ (۱۹۹۶) نشان می‌دهد که در دهه ۸۰ قرن ۲۰ موضوعات کاملاً متمرکز بر حوزه‌های تخصصی و پایه کتابداری از جمله مجموعه‌سازی، سازمان‌دهی، خدمات و کتابخانه بوده و در اوایل قرن ۲۱ موضوعات خودکارسازی کتابخانه، کتابخانه دیجیتال، مطالعات کاربران و تاریخ کتابخانه‌ها (Hjørland, 2000)، بازیابی اطلاعات، عملکرد کتابخانه، تحقیقات در علوم کتابداری و اطلاع‌رسانی، جستجوی اطلاعات، انواع کتابخانه‌ها و وبسایت (konfogiannakis, 2004) جزء موضوعات مهم مورد پژوهش

قرار گرفته است، ولی نتایج پژوهش حاضر نشان داد که روند پژوهش‌ها به موضوعات بین‌رشته‌ای تمایل پیدا کرده و این وابسته به شرایط استفاده و توسعه از فناوری‌ها و علوم مرتبط است. از جمله این موضوعات می‌توان از علم‌سنجی، داده‌کاوی، بازاریابی، بازاریابی الکترونیکی، تولید محتوا و موارد مرتبط با فناوری الکترونیکی و دیجیتالی نام برد و این موضوعات نیز در پژوهش‌های سال‌های اخیر مورد تأیید بوده است. در پژوهش جابین و همکاران (۲۰۱۵) موضوعاتی از قبیل اینترنت و بازاریابی الکترونیکی و دیجیتال‌سازی از موضوعات اصلی پژوهش‌ها بوده است (Jabeen et al, 2015).

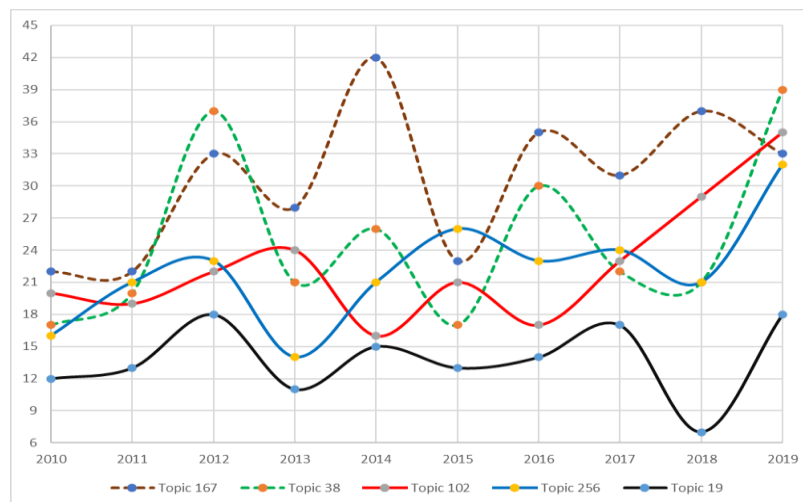
همه این‌ها نشان می‌دهد که فناوری و علوم نوظهور بین‌رشته‌ای همچون سایر رشته‌ها به حوزه علوم کتابداری و اطلاع‌رسانی نیز رسوخ پیدا کرده و شاید به عبارت بهتر کتابداری و اطلاع‌رسانی برای حفظ خود و حضور در جامعه ناگزیر از استفاده از این علوم شده است. علمی همچون بازاریابی الکترونیکی، داده‌کاوی، ارزیابی، سنجش، علم‌سنجی، یادگیری الکترونیکی و امثال آن که همگی ماهیت فناورانه دارند، جزء علوم بین‌رشته‌ای محسوب شده و هر رشته‌ای به تناسب نیاز از این علوم استفاده می‌کند و علوم کتابداری و اطلاع‌رسانی نیز از این قاعده مستثنی نیست. موضوع دیگری که در نتایج پژوهش حاضر قابل تأکید است این است که موضوعات پژوهش‌ها از ماهیت نظری به موضوعات عملی و کاربردی گرایش پیدا کرده و به نظر می‌رسد این همراهی با نیازها و خواسته‌های کاربران است. به‌طور کلی نتایج نشان داد که یک رابطه قوی و محکم بین حوزه‌های موضوعی با حوزه کامپیوتر و فناوری‌ها و جنبه‌های کاربردی آن وجود دارد. حوزه سنجش و اندازه‌گیری برترین حوزه مورد توجه در پژوهش‌های این حوزه قرار دارد (موضوع ۱۹۰ با بالاترین مقدار میانگین تا ۱۰۴۷/۰). دلیل این امر را شاید بتوان در اهمیت این موضوع در سنجش و ارزیابی برون‌داده‌های پژوهشی، سازمان‌ها، دانشگاه‌ها، نویسندگان و کشورها و فناوری‌ها از یک سو و اهمیت ارائه شاخص‌هایی متناسب با نیاز به‌روز دانست. پژوهش‌های مختلف نشان‌دهنده این است که انتشارات این حوزه روندی رو رشد دارد و این رشد در سال‌های اخیر روندی تصاعدی داشته است. اهمیت این حوزه در پژوهش‌های مختلف نیز مورد تأکید بوده است (Åström, 2002).

حوزه‌های دیگری با توسعه داده‌ها و ایجاد دیتاست‌ها و لزوم دسترسی و تحلیل آن ضروری شده، حوزه داده‌کاوی، تحلیل داده و پردازش داده است که جزء حوزه‌های نوظهور می‌توان قلمداد کرد که علم اطلاعات و دانش‌شناسی (علوم کتابداری و اطلاع‌رسانی) نیز با توجه به رسالت و ماهیت خود با این حوزه‌ها ارتباطاتی دارد. اختصاص بالاترین نرخ رشد به موضوع تحلیل داده‌ها (موضوع شماره ۱۰۰) در سال ۲۰۱۹ نسبت به سال ۲۰۱۰ تأیید کننده این موضوع است.

در میان موضوعات سرد، آموزش (۱۶۷)، کمترین میزان جلب توجه پژوهشگران را به خود اختصاص داده است. به صورتی 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 درجه بعدی از موضوعات کمتر مورد توجه قرار گرفته، واقع شده است. نمودار ۵ در مورد تغییرات انتشارات موضوعات سرد نشان‌دهنده روندی ثابت با تغییرات ملایم در میان سال‌های ۲۰۱۰ تا ۲۰۱۹ است. این روند در مورد دیگر موضوعات سرد نیز قابل مشاهده است. در خصوص برخی دیگر از این موضوعات، مانند مباحث مرتبط با آموزش، میزان رشد تحقیقات آن نیز با روندی کاهشی یا تقریباً ثابت روبرو بوده است.

بررسی نتایج به‌دست آمده 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٪) بوده و در بقیه حوزه‌ها (حدود ۲۱٪) روندی نزولی یا بدون تغییر بوده‌ایم.



نمودار ۵- تغییرات در تعداد انتشارات مربوط به موضوعات سرد در حوزه علوم کتابداری و اطلاع‌رسانی در بازه ۲۰۱۰ تا ۲۰۱۹

همچنین نگاهی به موضوعات شناسایی شده به طور کلی نشان داد که این رشته تحت تأثیر چند عامل یا اتفاق قرار دارد. خوشه‌های موضوعی فناوری محور حاکی از تحت تأثیر فناوری و موضوعات مرتبط قرار داشتن این رشته است و این تأثیرپذیر و تأثیرگذاری بسیار زیاد است. خوشه‌های موضوعی وجود دارد که متأثر از نیاز روز کاربران بوده و مرتبط با نیاز آموزشی و عادت‌های نیازهای اطلاعاتی آنها است. همچنین خوشه‌هایی همچون علم‌سنجی، ارزیابی و ارزشیابی وجود دارد که بر خواسته از انتظارات خدماتی این رشته است. بعضی دیگر از خوشه‌ها برگرفته از خدمات ذاتی کتابخانه بوده که تحت تأثیر شرایط حاکم بر جامعه، توسعه فناوری و نیازهای کاربران است.

مطالعه و به کارگیری نتایج این پژوهش می‌تواند برای پژوهشگران علاقه‌مند به حوزه علوم کتابداری و اطلاع‌رسانی، سیاست‌گذاران، تدوین‌کنندگان متون درسی پیشنهادی ارزشمند باشد. همچنین می‌تواند در ارائه 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 توجه نمایند. از سوی دیگر حوزه‌های پایه و اساسی رشته کمتر مورد توجه قرار گرفته و جزء حوزه‌های داغ قرار نگرفته است که این موضوع می‌تواند برای رشته و توسعه اساسی مبانی و مفاهیم مشکل‌ساز و نگران‌کننده باشد که بهتر است در پژوهش‌ها به این حوزه توجه جدی‌تری انجام گیرد. همچنین پیشنهاد می‌شود با استفاده از روش تحلیل محتوا و تحلیل روند و به کارگیری نتایج پژوهش‌های قبلی روند تغییرات در موضوعات علوم کتابداری و اطلاع‌رسانی از قبل تاکنون بررسی شود تا از روند تغییر و توسعه موضوعات پژوهشی این حوزه آشنایی دقیق پیدا کرد.

#### ۶. پیشنهادهای پژوهشی

در پژوهش حاضر به بررسی مقالات حوزه علم اطلاعات و دانش‌شناسی در حوزه بین‌المللی که در پایگاه استنادی اسکوپوس در سال‌های ۲۰۱۰ تا ۲۰۱۹ نمایه شده است، پرداخته شده است. پیشنهادهای موجود برای گسترش این پژوهش شامل موارد زیر است:

- با توجه به قابلیت‌های متن‌کاوی و کاربردهای آن در زمینه‌ها و رشته‌های مختلف و به‌منظور گسترش دامنه ابزار پژوهش در حوزه علم اطلاعات و دانش‌شناسی پیشنهاد می‌شود 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 توسعه حوزه علم اطلاعات و دانش‌شناسی تبیین شود.
- عدم توجه به مبانی نظری و پایه‌ای حوزه علم اطلاعات و دانش‌شناسی بررسی شود.
- دلایل تغییر در روند توسعه موضوعات پژوهشی حوزه علم اطلاعات و دانش‌شناسی تحلیل شود.

#### مراجع و منابع

- اصلائی، اکرم و اسماعیلی، مهدی. ۱۳۹۷. یافتن الگوهای مکرر در قرآن کریم به کمک روش‌های متن‌کاوی. *پردازش علائم و داده‌ها*. ۱۵(۳)، ۸۹-۱۰۰.
- باغ محمد، مریم؛ منصوری، علی و چشمه سهرابی، مهرداد. 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.

- پرنی، اعظم السادات و حمیدی، حجت‌الله. ۱۳۹۶. ارائه رویکردی برای مدیریت و سازمان‌دهی اسناد متنی با استفاده از تجزیه و تحلیل هوشمند متن. *پژوهشنامه پردازش و مدیریت اطلاعات*. ۱۲۰۲-۱۱۷۱: ۳۲(۴).
- جعفری پاورسی، حمیده؛ حریری، نجلا؛ علیپور حافظی، مهدی؛ باب‌الحوائجی، فهیمه و خادمی، مریم. ۱۳۹۸. نمایه‌سازی ماشینی مدارک حوزه بازیابی اطلاعات با استفاده از متن کاوی در نرم‌افزار رپیدمایتر. *فصلنامه پردازش و مدیریت اطلاعات*. ۳۷۴-۳۴۹: ۳۵(۲).
- دستانی، میثم؛ موسوی چیلک، افشین؛ ضیایی، ثریا و دل‌قندی، فایزه. ۱۳۹۹. تجزیه و تحلیل موضوعی مقالات منتشر شده کتابداری و اطلاع‌رسانی پزشکی در ایران با استفاده از فنون متن کاوی. *فصلنامه مرکز تحقیقات مدیریت خدمات بهداشتی درمانی دانشگاه علوم پزشکی و خدمات درمانی تبریز*. ۳۷۵-۳۵۵، ۴(۱۱).
- رضایی نور جلال و شیخ بهائی، محمدرضا. ۱۳۹۶. کاربردهای داده کاوی متنی در حوزه مدیریت دانش زنجیره خدمات دولت الکترونیکی. *فصلنامه مدیریت فناوری اطلاعات*. ۶۰-۳۹، ۹(۱).
- آذری، سلمان؛ شکراللهی، فرهاد و محمود و لسانی فشارکی، محمدعلی. ۱۳۹۹. الگوبرداری مفهوم یابی از آیات قرآن کریم با استفاده از دانش متن کاوی رایانشی. *مجله ذهن*. ۱۰۸-۸۱، ۸۲(۲).
- سلیمانی نژاد، عادل؛ سلاجقه، مزده و طیبی، الهام. ۱۳۹۷. خوشه‌بندی مقالات علمی بر پایه الگوریتم *k-means* مطالعه موردی: پایگاه پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک). *پردازش و مدیریت اطلاعات*. ۹۸۶-۸۷۱، ۳۴(۲).
- صالحی شهرودی، محمدحسین؛ مینایی، بهروز و اشرفی، امیررضا. ۱۳۹۲. متن کاوی موضوعی رایانه‌ای قرآن کریم، برای کشف ارتباطات معنایی میان آیات، بر مبنای تفسیر المیزان. *قرآن شناخت*. ۱۵۲-۱۱۷، ۱۲(۲).
- کاردان، احمد آقا و کیهانی نژاد، مینا. ۱۳۹۱. ارائه مدلی برای استخراج اطلاعات از مستندات متنی، مبتنی بر متن کاوی در حوزه یادگیری الکترونیکی. *فناوری اطلاعات و ارتباطات ایران*. ۴(۱۱-۱۲)، ۴۷-۵۴.
- برنا، کیوان؛ فتحی، فرهاد و مؤمنی، عصمت. ۱۳۹۷. کشف دانش و کاربرد آن در اینترنت اشیا. *فصلنامه مطالعات دانش‌شناسی*، سال پنجم زمستان ۱۳۹۷ شماره ۱۷، ۱۶۴-۱۲۵.
- گرشاسی، مسعود؛ رئیس روحانی، آناهید و کابارن‌زاده قدیم، محمدرضا. ۱۳۹۷. ارائه الگوریتم متن کاوی به منظور تشخیص حس در متن‌های فارسی. *مجله مدیریت فناوری اطلاعات*، شماره ۳۵، ۳۷۵-۳۸۹.

- Al-Azmi and Abdul-Aziz Rashid. 2013. Data, text and web mining for business intelligence: a survey, *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, Vol.3, No.2.
- 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., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & Kalhori, S. R. N. 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 Fischer, Andreas. 2018. How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, 226, pp 3-13.
- Blei, D. M., Ng, A. Y. and Jordan, M. I. 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 Mutz, R. 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)*, Volume66, Issue11, November 2015, Pages 2215-2222.
- Burrill, Gail, Ben-Zvi and Dani (Eds.). 2019. Topics and Trends in Current Statistics Education Research, *European Heart Journal* 40(28). DOI: 10.1093/eurheartj/ehz282. ISBN 978-3-030-03472-6.
- Carracedo, P., Puertas, R. and Marti, L. 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., Guo, X., Wu, T. and Ju, X. 2020. Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis. *Information Processing & Management*, 57(5), 102253.

- Dastani, M., Ziaei, S. and Delghandi, F. 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).
- Erdfelder, Edgar and Bosnjak, Michael .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., Thijs, B., Glänzel, W. and Sipido, K. R. 2019. Hot topics and trends in cardiovascular research. *European heart journal*, 40(28), 2363-2374.
- Hamerly, Greg, & Elkan, Charles. (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
- Hjørland, B. 2000. Library and information science: practice, theory, and philosophical basis. *Information processing & management*, 36(3), 501-531.
- Jabeen, M., Yun, L., Rafiq, M., Jabeen, M., & Tahir, M. A. (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., & Lee, B. G. 2020. Research trends in text mining: Semantic network and main path analysis of selected journals. *Expert Systems with Applications*, 162, 113851.
- Katsurai, M. and Joo, S. 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, SY, Song, SJ and song, M .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*, Volume 32 Issue 1, Pages.135-152
- Kowalski, Gerald J, and Maybury, Mark T. 2000. Information storage and retrieval systems: theory and implementation, *Springer Science & Business Media*. (Vol. 8).
- 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*. 2010 Sep; 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 (Fall 2016): 29–59.
- Kajberg, Leif. 1996. A content analysis of library and information science serial literature published in Denmark, 1957-198. *Library and Information Science Research*, Vol. 18, No.1, pp. 25-52.
- Luhn, Hans Peter. 1957. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of research and development*, 1(4), pp 309-317.
- Natarajan, M. 2005. Role of text mining in information extraction and information management. *DESIDOC Journal of Library & Information Technology*, 25(4).
- Park, JH and Song, M .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*, Volume 30 Issue 1, Pages.7-32.
- Li, X., Fan, M., Zhou, Y., Fu, J., Yuan, F. and Huang, L. 2020. Monitoring and forecasting the development trends of nanogenerator technology using citation analysis and text mining. *Nano Energy*, 71, 104636.
- Lee, J. Y., Kim, H. and Kim, P. J. 2010. Domain analysis with text mining: Analysis of digital library research trends using profiling methods. *Journal of Information Science*, 36(2), 144-161.
- 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.
- Rehurek, R. and Sojka, P. 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).
- Silge, J., & Robinson, D. 2017. Text mining with R: A tidy approach. *O'Reilly Media, Inc*.
- Stevens, K., Kegelmeyer, P., Andrzejewski, D. and Buttler, D. 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).
- Tseng, Yuen-Hsien, Lin, Chi-Jen and Lin, Yu-I .2006. Text mining techniques for patent analysis, *Information Processing & Management*, 43(5):1216-1247.
- Tworowski, D., Gorohovski, A., Mukherjee, S., Carmi, G., Levy, E., Detroja, R and Frenkel-Morgenstern, M. 2021. COVID19 Drug Repository: text-mining the literature in search of putative COVID19 therapeutics. *Nucleic acids research*, 49(D1), D1113-D1121.
- 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.
- Van Dijk, W. B., Fiolet, A. T., Schuit, E., Sammani, A., Groenhof, T. K. J., van der Graaf, R., ... and Mosterd, A. 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.
- Wang, L. and Lo, K. 2021. Text mining approaches for dealing with the rapidly expanding literature on COVID-19. *Briefings in Bioinformatics*, 22(2), 781-799.



شناسایی موضوعات داغ و روندها در علم اطلاعات و دانش‌شناسی با استفاده از تکنیک‌های متن‌کاوی | الهه اخوان حریری، علی منصوری و دیگران

- Buscema, P. M., Della Torre, F., Breda, M., Massini, G and Grossi, E. 2020. COVID-19 in Italy and extreme data mining. *Physica A: Statistical Mechanics and its Applications*, 557, 124991.
- Zengul, F. D., Zengul, A. G., Mugavero, M., Oner, N., Ozaydin, B., Delen and Cimino, J. 2021. A critical analysis of COVID-19 research literature: Text mining approach. *Intelligence-based medicine*, 100036.

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

**Elahe Akhavanhariri**

Master of 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

### 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, by 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 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

### 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.



### علی منصوری

دارای مدرک دکتری در رشته علوم کتابداری و اطلاع‌سانی از دانشگاه شهید چمران اهواز است. ایشان هم‌اکنون دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است. علم‌سنجی، فناوری‌سنجی، مدیریت اطلاعات و داده‌کاوی از جمله علایق پژوهشی وی است.





### حسین کارشناس نجف‌آبادی

دارای مدرک تحصیلی دکتری در رشته هوش مصنوعی از دانشگاه پلی‌تکنیک مادرید - اسپانیا است. 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.

