

# Using Conceptual Clustering to Extraction of Key Phrases and Related Terms: A Case Study of Scientific Communication Texts

## **Rajab Kiyani Shahvandi**

PhD Candidate in Knowledge and Information Science;  
University of Isfahan; Isfahan, Iran Email: kiani.rajab@gmail.com

## **Ahmad Shabani\***

PhD in Knowledge and Information Science; Professor;  
University of Isfahan; Isfahan, Iran Email: shabania@edu.ui.ac.ir

## **Asefeh Asemi**

PhD in Information Technology; Associate Professor; Institute of  
Data Analytics and Information Systems; Corvinus University of  
Budapest; Budapest, Hungary Email: asefeh.asefeh@uni-corvinus.hu

## **Morteza Mohammadi Ostani**

PhD in Knowledge and Information Science; Assistant Professor;  
University of Isfahan; Isfahan, Iran Email: m.ostani@edu.ui.ac.ir

Received: 09, Feb. 2024 Accepted: 09, Jun. 2024

**Abstract:** Scientific communication encompasses various types and forms of communication conducted through the use of communication methods and tools, aiming to exchange scientific knowledge and information. To gain a comprehensive understanding of scientific and research communications and enhance them, it is crucial to identify the terms and concepts. Therefore, the main objective of this research is to identify and conceptually cluster key terms in the field of scientific communication using text mining techniques. The present research method is quantitative in terms of approach and practical in terms of purpose and utilized various text mining techniques for identifying and clustering key terms in the field of scientific communication. The research population consists of abstracts of articles related to scientific communication, extracted from databases such as Web of Science and Scopus, totaling 558 articles. The sampling method was census. Initially, all nominal phrases were extracted using available libraries. Each compound phrase was decomposed into its constituent words, and based on GloVe dictionary, the average vectors of those words were calculated, assigning a numerical vector to each compound phrase. The

**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. 39 | No. 4 | pp. 1477-1506

Summer 2024

<https://doi.org/10.22034/ijpm.2024.2022598.1545>



\* Corresponding Author

researchers created an equivalent expression using existing vocabulary to describe unknown terms that did not exist in the GloVe dictionary. The clustering (using the K-means method) was performed on these vectors. The findings revealed that out of 17,930 extracted keywords, 13,651 terms were noun phrases. Also, 16% of terms in the field of scientific communication were single words and 84% of them were compound. After creating vectors of compound terms and performing clustering, 40 conceptual clusters were created from 792 phrases or terms in the field of scientific communication. After adjusting and removing weak clusters, researchers finally identified 22 clusters in the field of scientific communication. Identifying the concepts and components in scientific communication in the form of conceptual clusters and its elements is attributed to the results of this research. One of the most significant findings was the assignment of numerical vectors to composite phrases based on the vectors of their constituent words. These vectors were then used for clustering and categorizing phrases, as well as improving and correcting some clusters. This method pays attention to the semantics aspects and learning in the clustering and categorization of concepts and, will aid to precise analysis of key terms and phrases in various fields.

**Keywords:** Scientific Communication, Text Mining, Conceptual Clustering, Nominal Phrase, GloVe Dictionary, Semantic Approaches



# به کارگیری خوشه‌بندی مفهومی برای استخراج عبارات کلیدی و اصطلاحات مرتبط؛ مطالعه موردی متون حوزه ارتباطات علمی

رجب کیانی شاهوندی

دانشجوی دکتری علم اطلاعات و دانش‌شناسی؛  
دانشگاه اصفهان؛ اصفهان، ایران؛  
kiani.rajab@gmail.com

احمد شعبانی

دکتری علم اطلاعات و دانش‌شناسی؛ استاد؛  
دانشگاه اصفهان؛ اصفهان، ایران؛  
shabania@edu.ui.ac.ir

عاصفه عاصمی

دکتری تکنولوژی اطلاعات؛ دانشیار؛  
دانشگاه کورونیوس بوداپست؛ بوداپست، مجارستان؛  
asemi.asefeh@uni-corvinus.hu

مرتضی محمدی استانی

دکتری علم اطلاعات و دانش‌شناسی؛ استادیار؛  
دانشگاه اصفهان؛ اصفهان، ایران؛  
m.ostani@edu.ui.ac.ir



دریافت: ۱۴۰۲/۱۱/۲۰ | پذیرش: ۱۴۰۳/۰۳/۲۰ | مقاله برای اصلاح به مدت ۱۸ روز نزد پدیدآوران بوده است.

نشریه علمی | رتبه بین‌المللی  
پژوهشگاه علوم و فناوری اطلاعات ایران  
(ایرانداک)

شاپا (چاپی) ۸۲۲۳-۲۲۵۱

شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱

نمایه در SCOPUS، ISI، و LISTA

jipm.irandoc.ac.ir

دوره ۳۹ | شماره ۴ | صص ۱۴۷۷-۱۵۰۶

تابستان ۱۴۰۳

<https://doi.org/10.22034/jipm.2024.2022598.1545>

**چکیده:** ارتباط علمی از انواع و گونه‌های ارتباطات است که از طریق به کارگیری روش‌ها و ابزارهای ارتباطی و با هدف تبادل دانش و اطلاعات علمی انجام می‌شود. به‌منظور احراز دیدگاهی جامع درباره ارتباطات علمی و پژوهشی و تقویت آن باید اصطلاحات و مفاهیم آن شناسایی گردد. از این‌رو، هدف اصلی پژوهش حاضر شناسایی و خوشه‌بندی مفهومی اصطلاحات و مفاهیم کلیدی در حوزه ارتباطات علمی با استفاده از تکنیک‌های متن‌کاوی است.

روش پژوهش حاضر از لحاظ رویکرد، کمی و از لحاظ هدف، کاربردی است. برای شناسایی و خوشه‌بندی اصطلاحات کلیدی در حوزه ارتباطات علمی از تکنیک‌های مختلف متن‌کاوی استفاده گردید. جامعه آماری پژوهش حاضر، مشتمل بر چکیده مقالات مرتبط با حوزه ارتباطات علمی مستخرج از پایگاه‌های اطلاعاتی «وب‌آوساینس» و «اسکوپوس» به تعداد ۵۵۸ مقاله و روش نمونه‌گیری، سرشماری بود. ابتدا تمامی اصطلاحات



و عبارات اسمی با استفاده از کتابخانه‌های موجود با کدنویسی به زبان «پایتون» استخراج، و فراوانی واژگان محاسبه گردید. سپس هر عبارت مرکب به واژگان تشکیل دهنده آن تجزیه شده و بر مبنای واژه‌نامه «گلاو» و با محاسبه میانگین بردارهای آن کلمات، یک بردار عددی تخصیص یافت. برای اصطلاحات ناشناخته نیز - که در واژه‌نامه «گلاو» وجود نداشت - یک عبارت معادل با استفاده از واژگان موجود جهت توصیف آن، جایگزین و بردار عددی عبارت ساخته شد و خوشه‌بندی (به روش کا-مینز) بر روی آن واژگان انجام شد. یافته‌ها نشان داد که از ۱۷۹۳۰ کلیدواژه مستخرج، تعداد ۱۳۶۵۱ واژه، عبارت اسمی بود. همچنین، ۱۶ درصد از اصطلاحات حوزه ارتباطات علمی، تک‌واژه‌ای و ۸۴ درصد آن مرکب بودند. پس از ایجاد بردارهای اصطلاحات مرکب و انجام خوشه‌بندی، از ۷۹۲ عبارت یا اصطلاح در حوزه ارتباطات علمی، تعداد ۴۰ خوشه مفهومی ایجاد گردید. پس از تعدیل و حذف خوشه‌های ضعیف، سرانجام تعداد ۲۲ خوشه در حوزه ارتباطات علمی شناسایی گردید. شناسایی مفاهیم اصلی در ارتباطات علمی در قالب خوشه‌های مفهومی و عناصر آن از نتایج پژوهش حاضر قلمداد می‌شود. از مهم‌ترین یافته‌های دیگر پژوهش، تخصیص بردارهای عددی به عبارات ترکیبی از روی بردار واژگان تشکیل دهنده آن‌ها و سپس استفاده از این بردارها برای خوشه‌بندی و دسته‌بندی عبارات و همچنین قابلیت بهبود و تصحیح بعضی از خوشه‌ها بود. این روش در خوشه‌بندی و دسته‌بندی مفاهیم امکان توجه به جنبه‌های معناشناسی و یادگیری را مورد توجه قرار می‌دهد و به تحلیل دقیق اصطلاحات و عبارات کلیدی در حوزه‌های مختلف کمک خواهد کرد.

**کلیدواژه‌ها:** ارتباطات علمی، متن کاوی، خوشه‌بندی مفهومی، عبارت اسمی، واژه‌نامه گلاو، رویکردهای معنایی

## ۱. مقدمه

توسعه و نشر علم بدون ارتباطات قابل تصور نیست. ارتباطات علمی به صورت رسمی با قدمتی بیش از سه‌ونیم قرن، همواره در طی زمان شاهد تغییر و تحولات عمده‌ای بوده است. این شیوه از ارتباط از عوامل سیاسی، اقتصادی، اجتماعی، و فناوری متأثر بوده و بر توسعه علمی و پیشرفت اندیشه در جامعه بشری تأثیر گذاشته است (د سیلوا و ونس<sup>۱</sup> ۲۰۱۷). فعالیت علمی در خلاء اتفاق نمی‌افتد و فرد نمی‌تواند جدای از دیگران به یافته‌های علمی دست یابد. ارتباطات و ارتباطات علمی ناگزیر، جزء لاینفک تأمین نیاز اطلاعاتی در هر کار علمی است. درک و گسترش کاربرد مفهوم ارتباطات علمی به طور قطع به توسعه علمی سرعت خواهد بخشید (منصوری ۱۳۹۵). ارتباطات علمی، زمینه‌ساز ایجاد یک جامعه علمی پویا و دانش‌مدار بوده و فعالیت علمی دانشمندان، زمانی هویت و اهمیت می‌یابد که

1. De Silva & Vance

به جامعه علمی عرضه شود، در آثار و مطالعات دیگران متجلی شود، و با تحلیل‌های فردی دیگران در آمیزد تا سرانجام بر ذخیره‌های دانش افزوده شود.

واژه «ارتباطات»، مفهومی به‌ظاهر بدیهی و قابل درک است، اما تعاریف گوناگون این مفهوم نشان می‌دهد که بیش از حد پیچیدگی دارد. «دنس و لارسن»<sup>۱</sup> برای واژه «ارتباطات» بیش از ۱۲۶ تعریف یافتند (پرهام‌نیا ۱۴۰۰). از دوره جنگ جهانی دوم بود که به مفهوم ارتباطات برای رسیدن به اهداف متنوع توجه شد، اصطلاح «ارتباطات علمی»<sup>۲</sup> در متون و اسناد، بیشتر مشابه با «ارتباطات پژوهشی»<sup>۳</sup> یا «ارتباطات علمی-پژوهشی» و بیشتر متفاوت از «همکاری‌های علمی»<sup>۴</sup> بوده، ولی گاهی مرتبط با یکدیگر به کار رفته است. از نظر «راجرز»<sup>۵</sup>، ارتباط علمی فرایندی است که به موجب آن پیام‌های علمی از منابع به گیرنده انتقال یافته و عناصر اصلی آن مانند هر گونه اشکال دیگر ارتباط، عبارت است از: منبع پیام، کانال ارتباطی، و گیرنده پیام. نوشتن و انتشار، مجرای اصلی ارتباط علمی است (داورپناه ۱۳۹۵). انجمن کتابداری آمریکا همچنین ارتباطات علمی را نظامی دانسته که در آن کیفیت پژوهش و سایر تحقیقات علمی تولیدی ارزیابی شده، در اختیار اجتماع علمی قرار گرفته و از آن‌ها برای کاربردهای آینده، حفاظت و نگهداری می‌شود (ALA نقل در Xia 2006 و همچنین در علیدوستی، خسروجردی و دوران ۱۳۸۸). ارتباط علمی از انواع ارتباطات است که از طریق به کارگیری روش‌ها و ابزارهای ارتباطی و با هدف تبادل دانش و اطلاعات علمی انجام می‌شود. ارتباط علمی شکل مفرد دارد و جمع آن عبارت «ارتباطات علمی» و شامل مجموعه متنوعی از ارتباط علمی به‌طور گسترده و با ابزارهای متنوع است.

داده کاوی و متن کاوی از روش‌ها و تکنیک‌های مهمی است که بارها در پژوهش‌های متعدد برای استخراج مفاهیم، کلیدواژه‌ها و الگوها از متون و اسناد مرتبط استفاده شده است. هدف از این تکنیک‌ها، سپردن بخشی از وظیفه انسان به ماشین برای خودکارسازی، سرعت بخشی و قابلیت تکرار وظیفه مذکور با شمار و در مقیاس بالاست. داده کاوی، اکتشاف و تجزیه و تحلیل داده‌ها برای کسب و آشکارسازی قوانین و الگوهای جالب، روابط و اطلاعات پنهان و مفید موجود در حجم زیادی از داده‌هاست. متن کاوی نیز نوعی داده کاوی است که مبتنی بر داده‌های متنی انجام می‌پذیرد. تکنیک‌های متن کاوی یا

1. Dance & Larson

2. scientific communication or scholarly communication

3. research communication

4. scientific collaboration or scientific cooperation

5. Rogers

تحلیل متن این امکان را فراهم آورده که استفاده‌کننده، بدون خواندن متن‌های طولانی و فاقد ساختار، بتواند اطلاعات مفیدی از متن، استخراج کرده و بر روی آن‌ها عملیاتی نظیر کشف روابط، خلاصه‌سازی، خوشه‌بندی و دسته‌بندی انجام دهد.

برای هر نوع داده‌کاوی متون و پردازش زبان طبیعی لازم است ابتدا واژگان به شکل عددی یا به بردارهای عددی تبدیل شوند. بدین منظور روش‌ها و مدل‌های مختلفی توسعه یافته است. یکی از این موارد، واژه‌نامه / مدل برداری «گلاو»<sup>۱</sup> است. در این مدل که در سال ۲۰۱۴ توسط «جفری پنینگتون»<sup>۲</sup> در دانشگاه استنفورد ارائه شد، بردارهایی برای نمایش واژگان با تعداد اعداد مختلف استفاده شد. این مدل که بر اساس داده‌های کمی موجود در متون بزرگ و معنای هر واژه در متن، قبلاً آموزش داده شده، با استفاده از ترکیب معنای واژگان مجاور، بردارهای n تایی را برای نمایش واژگان فراهم می‌آورد. از این رو، امکان استفاده از فضای برداری تولیدی، برای یافتن روابط بین واژگان و انجام هر نوع پردازش زبان طبیعی وجود دارد. به‌طور معمول، این مدل دارای ۴۰۰۰۰۰ واژه است. نسخه‌های مختلف بردارهای این واژه‌نامه قابل استفاده بوده و از منابع مختلفی نظیر سامانه رسمی پروژه «گلاو» در دانشگاه استنفورد<sup>۳</sup>، مخزن گیت‌هاب<sup>۴</sup> پروژه گلاو، وب‌سایت کگل<sup>۵</sup>، و وب‌سایت هاگینگ فیس<sup>۶</sup> قابل دسترس است. ویژگی بردارهای «گلاو» توجه به ملاحظات معنایی است که در خوشه‌بندی واژگان بسیار مفید و مؤثر خواهد بود. به گفته دیگر، اگر دو کلمه از نظر معنایی، مشابه و به هم نزدیک باشند، به همان نسبت در بردارهای «گلاو»، بردار عددی متناظر آن‌ها به هم نزدیک بوده و مشابهت بیشتری نشان خواهد داد.

به‌منظور تقویت ارتباطات علمی در جوامع علمی و دانشگاهی بایسته است که عوامل و مؤلفه‌های تأثیرگذار بر آن شناسایی گردد. پژوهش‌های چندی نظیر «حسن‌زاده، کلینی مقدم و اجاق» (۱۴۰۲)، «ریاحی و موسوی چلک» (۱۳۹۹)، «مهرآور گیگلو» و همکاران (۱۳۹۹)، «رضادوست، نواح و ادیب‌زاده» (۱۳۹۶)، و «احمدی، عصاره و حیدری» (۱۳۹۴) به‌منظور شناسایی این عوامل و مؤلفه‌ها صورت گرفته است. در این میان، برخی عوامل نیز خارج از این فرایند تأثیر گذاشته که ارتباطات علمی را دستخوش تغییر کرده است. طی سال‌های اخیر با شیوع و همه‌گیری بیماری کرونا و لزوم پیشگیری، رعایت فاصله افراد، لغو بسیاری

1. Glove Vectors

2. Jeffrey Pennington

3. <https://nlp.stanford.edu/projects/glove>

4. GitHub

5. <https://github.com/stanfordnlp/GloVe>6. <https://www.kaggle.com>

7. Hugging Face

از جلسات آموزشی و پژوهشی دانشگاهی یا برگزاری آن‌ها به شکل پیوسته و محدودیت ارتباطات انسانی این سؤال مطرح می‌شود که در نتیجه این همه‌گیری، چه عوامل دیگری بر ارتباطات علمی مؤثر بوده یا خواهد بود. به نظر می‌رسد خلأ پژوهشی در این حوزه، مطالعه جامع و کامل عوامل مؤثر بر ارتباطات علمی در متون و مقالات این حوزه در پایگاه‌های اطلاعاتی معتبر است.

بهره‌جویی از تکنیک‌های داده‌کاوی و متن‌کاوی نظیر خوشه‌بندی اصطلاحات برای شناسایی عوامل مؤثر بر ارتباطات علمی سودمند خواهد بود. نکته مهم این است که در بسیاری از روش‌های خوشه‌بندی، اصطلاحات مرکب در روش‌های مبتنی بر مجموعه واژگان، مورد پردازش قرار نمی‌گیرند. لازم است بر روی این دسته از اصطلاحات که تعداد آن در حوزه‌های مختلف موضوعی و زبانی زیاد است، تبدیلی صورت گیرد تا امکان خوشه‌بندی و دسته‌بندی بر روی آن‌ها نیز وجود داشته باشد. با توجه به ویژگی معنایی بردار «گلاو»، در پژوهش حاضر برای استخراج و خوشه‌بندی مفهومی حوزه ارتباطات علمی از آن استفاده گردید. از این‌رو با توجه به مطالب پیشگفته، مسئله اصلی پژوهش حاضر، خوشه‌بندی مفهومی اصطلاحات کلیدی (مفرد و مرکب) در حوزه ارتباطات علمی در پایگاه‌های استنادی «اسکوپوس» و «وب‌آوساینس» به‌منظور شناسایی عوامل مؤثر بر آن است. پژوهش حاضر بر آن است که با استفاده از تکنیک خوشه‌بندی، مفاهیم مهم را با توجه به روابط معنایی در اسناد مرتبط در پایگاه‌های استنادی «اسکوپوس» و «وب‌آوساینس» استخراج کند. از این‌رو برای دستیابی به اهداف بالا، سؤالات زیر مطرح گردید:

۱. چگونه می‌توان اصطلاحات مرکب (چند کلمه‌ای) مستخرج از متون حوزه ارتباطات علمی در پایگاه «اسکوپوس» و «وب‌آوساینس» را خوشه‌بندی کرد؟
۲. خوشه‌بندی مفهومی واژگان مستخرج از متون حوزه ارتباطات علمی در پایگاه «اسکوپوس» و «وب‌آوساینس»، با استفاده از تکنیک متن‌کاوی چگونه است؟

## ۲. پیشینه پژوهش

با توجه به هدف پژوهش حاضر، می‌توان دو دسته از پیشینه‌ها را از یکدیگر متمایز ساخت. دسته اول، پژوهش‌های مرتبط با ارتباطات علمی و دسته دیگر پژوهش‌های داده‌کاوی، متن‌کاوی و تحلیل متون است. مرور پیشینه‌ها در حوزه ارتباطات علمی در پایگاه‌های اطلاعاتی داخلی و خارجی نظیر «گوگل اسکولار»، «اسکوپوس»، «وب‌آوساینس»،

«مرکز اطلاعات علمی جهاد دانشگاهی» و «پرتال جامع علوم انسانی» نشان داد که امکان دسته‌بندی آن‌ها در قالب چند دسته کلی وجود دارد. برخی پژوهش‌ها به اهداف، نتایج، مصادیق، ابزارها و بررسی ارتباطات علمی در کشورهای مختلف و کاربردهای آن نظیر تولید دانش و ترویج علم پرداخته‌اند. پژوهش‌های «حسن‌زاده، گلینی مقدم و اجاق» (۱۴۰۲)، Shehata and Eldakar (2023)، Khoiriah et al. (2023)، «کفاشان کاخکی و خلیلی» (۱۳۹۸)، Kappel and Holmen (2019)، Kuzminska (2021)، Gherab Martin (2021) و «باقری بنجار» و همکاران (۱۳۹۴) در این دسته هستند. تعداد محدودی پژوهش نیز به مفهوم‌پردازی در ارتباطات علمی نظیر توجه به آرای «لومان»<sup>۱</sup> و دیدگاه مکتب شیکاگو توجه کرده‌اند. پژوهش‌های «پرهام‌نیا» (۱۴۰۰)، Fährnich (2021)، «سهرابی و غفاری» (۱۳۹۸)، «پلویی و نقشینه» (۱۳۹۵) و Khosrowjerdi (2011) در این دسته هستند. بیشتر پژوهش‌ها نیز بر نحوه ارتباط، عناصر و فرایندهای ارتباطات علمی نظیر کمیت همکاری علمی و سهولت دسترسی متمرکز بودند. برخی پژوهش‌ها نظیر «ریاحی و موسوی چلک» (۱۳۹۹) و «احمدی» و همکاران (۱۳۹۶) نیز به بررسی عوامل ترغیب‌کننده و بازدارنده همکاری‌های علمی اعضای هیئت علمی دانشگاه‌ها پرداختند. از نظر بافت نیز بیشتر به مطالعه ارتباطات علمی در سطوح مختلف دانشگاه‌ها و تعداد اندکی نیز به مطالعه از روی مجلات، اسناد و پایگاه‌ها پرداخته‌اند. پژوهش‌های «سهرابی و غفاری» (۱۳۹۸)، «رضی» و همکاران (۱۳۹۵)، «سهیلی، دانش و فتاحی» (۱۳۸۹)، «برهمند» (۱۳۸۶) و «خسروجردی» (۱۳۸۵) از این نمونه هستند. در ادامه، به چند پژوهش مرتبط با حوزه ارتباطات علمی اشاره می‌گردد.

«حسن‌زاده، گلینی مقدم و اجاق» پژوهشی را درباره میزان آشنایی دانشجویان رشته علم اطلاعات و دانش‌شناسی دانشگاه علامه طباطبائی درباره مصادیق ارتباطات علمی رسمی و غیررسمی، عوامل فردی، عوامل محیطی، شاخص‌های اعتبار، کانال‌ها و ابزارهای گوناگون برقراری ارتباطات علمی و میزان استفاده از این ابزارها انجام دادند. نتایج حاکی از آن بود که کانال‌ها و ابزارهایی که انتقال اطلاعات علمی از طریق آن‌ها صورت گرفته و برای برقراری ارتباط از آن‌ها استفاده می‌شود، از مهم‌ترین عوامل تأثیرگذار بر ارتباطات علمی است. همچنین این ابزارها به سه دسته شفاهی، کتبی، و الکترونیکی قابل تقسیم است (۱۴۰۲). «شهاتا و الداکار» پژوهشی در کشورهای در حال توسعه به بررسی شیوه‌های ارتباطی علمی

1. Luhmann



محققان تازه کار در مصر پرداختند. یافته‌های آنان نشان داد که دانش پژوهان جوان بیشتر به روش‌های علمی مرسوم در فعالیت‌های تحقیقاتی خود پایبند هستند. این انطباق تا حدی به سیستم تبلیغاتی در مصر مرتبط است که تأکید قابل توجهی بر رویه‌های مرسوم نظیر انتشار تک‌نویسنده دارد؛ در حالی که استفاده از مجراهای غیررسمی برای کسب شناخت و ترویج را منع می‌کند (Shehata and Eldakar 2023). «خوریا» و همکاران نیز به مطالعه و ساخت ابزار برای اندازه‌گیری مهارت‌های ارتباط علمی دانش‌آموزان اندونزیایی پرداختند. یافته‌های آن‌ها نشان داد که مهم‌ترین مهارت‌های ارتباط علمی عبارت‌اند از: بازیابی اطلاعات، علمی خواندن، گوش دادن، مشاهده تأملی، و نگارش علمی (Khoiriah et al. 2023). بررسی عملکرد دو گانه ارتباطات علمی دسترسی آزاد در پایگاه arXiv پژوهش دیگری بود که توسط «غراب مارتین» انجام پذیرفت. وی به این نتیجه رسید که سیستم‌های ارتباطات و انتشارات علمی باید با واقعیت‌های جدید فناوری (اینترنت) و اجتماعی (دسترسی آزاد) طراحی گردد. مجلات دانشگاهی به‌منظور رفع نیازهای جدید علم و جامعه نیازمند بازنگری و تطبیق هستند. در این پژوهش، تمایز بین کارکرد تحقیقاتی و کارکرد نهادی و چگونگی انجام این دو کارکرد در مخازن موضوع‌محور و مجلات علمی arXiv نشان داده شده است (Gherab Martin 2021). «کوزمیناسکا» در پژوهشی ابزارهایی را برای تقویت ارتباطات علمی در طول چرخه حیات تحقیقات علمی پژوهشگران معرفی و شناسایی کرد. دیجیتالی‌سازی ارتباطات علمی به دلیل همه‌گیری کووید ۱۹ به سرعت پیش رفته است. در عین حال، انتقال اجباری به ارتباطات علمی دیجیتالی از طریق همه‌گیری کووید ۱۹ می‌تواند به ادغام دانشمندان جوان در فضای علمی بین‌المللی کمک کند. این پژوهش تجزیه و تحلیل ابزارهایی را برای حمایت از ارتباطات علمی توسعه یافته در پروژه «۱۰۱ نوآوری در ارتباطات علمی» ارائه می‌دهد (Kuzminska 2021). «سهرابی و غفاری» پژوهشی را با هدف شناسایی موضوعات پرکاربرد تولیدات علمی حوزه ارتباطات علمی در پایگاه اطلاعاتی «وب‌آوساینس» با استفاده از روش تحلیل محتوای کمی و تحلیل هم‌واژگانی انجام دادند. یافته‌های آنان نشان داد که از نظر هم‌رخدادی واژگان، کلیدواژه‌های «دسترسی آزاد» و «ارتباطات علمی» دارای بیشترین فراوانی بودند. نتایج خوشه‌بندی موضوعات نیز نشان داد که خوشه‌های «علم ارتباطات»، «دسترس پذیری علم» و «سنجش علمی» از مهم‌ترین آن‌هاست (۱۳۹۸).

دسته دوم پیشینه‌ها، پژوهش‌های مرتبط با حوزه داده‌کاوی، متن‌کاوی، خوشه‌بندی و استخراج خودکار مفاهیم از متون است. برخی از پژوهش‌ها بر کاربرد خوشه‌بندی معنایی

در بافت‌های مختلف نظیر تشخیص تصویر و رده‌بندی، سیستم‌های پیشنهاددهنده، تحلیل مدولارسازی یا ماژولارسازی<sup>۱</sup> مجدد نرم‌افزارهای کاربردی، و شناسایی موضوعات در کد منبع متمرکز بودند. پژوهش‌های (Abbasi-Moud, Vahdat-Nejad and, Ma et al. (2022)، (Sadri (2021)، (Santos, Valente and Anquetil (2014)، و (Kuhn, Ducasse and Girba (2007) از این دسته‌اند. دسته دیگر پژوهش‌ها نظیر (Santhiappan, Sharavan and Ravindran (2023)، (Kapellas and Kapidakis (2023)، «غیائی، نظافتی و شکوهار» (۱۳۹۴) و «ابویی اردکان، عابدی جعفری و آفازاده» (۱۳۸۹) به کاربرد خوشه‌بندی در رده‌بندی منابع، استخراج رویدادهای خبری، داده‌های حوزه علوم دریایی و همچنین ترسیم نقشه‌های علمی پرداختند. همچنین برخی پژوهش‌ها نظیر «بلوه و نوروزی» (۱۴۰۰) و «سلیمانی نژاد، سلاجقه و طیبی» (۱۳۹۷) بر الگوریتم‌های ویژه‌ای نظیر کا-مینز برای خوشه‌بندی متمرکز بودند. شماری دیگر از پژوهش‌ها با استفاده از رویکردهای داده‌کاوی و یادگیری ماشین به کشف و استخراج کلیدواژه‌ها و روابط موضوعی مبادرت کردند. «زرمهر، منصوری و کارشناس نجف‌آبادی» (۱۴۰۲)، «شکوهیان» و همکاران (۱۳۹۸)، و (Salloum et al. (2018) در این دسته قرار دارند. در ادامه، به برخی از پژوهش‌های مرتبط اشاره می‌گردد.

«زرمهر، منصوری و کارشناس نجف‌آبادی» با استفاده از تکنیک‌های متن‌کاوی و یادگیری ماشین، به مقایسه عملکرد رویکردهای کشف و استخراج کلیدواژه‌های موضوعی کتاب‌های الکترونیک پرداختند. سه رویکرد آزمایشی شامل اجرای متوالی فرایند خوشه‌بندی، ارتقای کیفیت خوشه‌ها از نظر معنایی و غنی‌سازی واژگان غیرمجاز حوزه خاص؛ استفاده از الگوی کلیدواژه‌های تخصصی؛ و استفاده از بخش‌های مهم متن در کشف و استخراج واژگان کلیدی و موضوعات مهم متن معرفی و مقایسه شده‌اند. نتایج تحلیل‌ها نشان داد که کیفیت و انسجام معنایی خوشه‌های موضوعی حاصل از رویکرد سوم، یعنی استفاده از بخش‌های مهم متن در کشف و استخراج موضوع، از دو رویکرد دیگر عملکرد بهتری داشت (۱۴۰۲). «شکوهیان» و همکاران به دسته‌بندی موضوعی متون علمی حوزه سلامت با روش متن‌کاوی پرداختند. نتایج پژوهش آن‌ها نشان داد که رویکرد ترکیبی استفاده از دسته‌بندی موضوعی و خوشه‌بندی با استحکام بیشتری به زمینه‌های موضوعی در حوزه اطلاعات سلامت مساعدت می‌کند (۱۳۹۸). «سالوم» و همکاران در

پژوهشی با عنوان «استفاده از تکنیک‌های متن کاوی برای استخراج اطلاعات از مقالات پژوهشی»، تعداد ۳۰۰ مقاله با موضوع «یادگیری سیار» را از پایگاه‌هایی نظیر Science, IEEE, Wiley, Springer, Direct استخراج کردند. بخش عمده تحلیل‌ها بر روی واژه‌های این حوزه در سطح تک‌واژه‌ای بوده است. تجزیه و تحلیل متنی متون این حوزه نشان داد که واژه‌های «یادگیری»، «بیماران» و «دانشجویان» به ترتیب، پر تکرارترین واژه‌ها و پایگاه «اشپرینگر»، منبع اصلی مقالات پژوهشی در زمینه یادگیری سیار در حوزه پزشکی بود. به دلیل روابط متقابل یا ابهام در معانی موضوعات، امکان تشخیص شباهت میان آن‌ها وجود نداشت (Salloum et al. 2018). از سال ۲۰۰۹، افزایش قابل توجهی در تعداد مقالات ایجاد گردید و مقالات منتشر شده در طول سال‌های ۲۰۱۵ تا ۲۰۱۶ افزایش چشمگیری داشته است.

پژوهش‌هایی نظیر «ژئائویان، راگا و ژوئمی» نیز از مدل برداری «گلاو» برای بردارسازی کلمات استفاده کردند. در این پژوهش که مدل تحلیل احساسات مبتنی بر GloVe-CNN-BiLSTM بهینه پیشنهاد می‌شود، از مدل «گلاو» برای بردارسازی کلمات به کاررفته در نظرات «توییتر» در مورد کووید ۱۹ به‌عنوان مجموعه داده آزمایشی استفاده گردید. نتایج، حاکی از آن بود که این روش می‌تواند به‌طوری مؤثر تمایلات احساساتی نظرات برخط کاربران را شناسایی کند. دقت طبقه‌بندی احساسات در متن کامل، بلند و کوتاه مقدار قابل قبولی بود (Xiaoyan, Raga, and Xuemei 2022). همچنین (Aburass, Dorgham and Al-Shaqsi (2023). از این مدل برداری در یک مدل یادگیری ماشینی برای جاسازی کلمات در طبقه‌بندی جهش‌های ژنی در سرطان بهره‌جویی کردند که نتایج، حاکی از سودمندی آن بود.

بخش عمده پیشینه‌های حوزه داده کاوی، متن کاوی و استخراج خودکار مفاهیم نظیر (Salloum et al. (2018), (Xiaoyan, Raga, and Xuemei (2022), (Aburass, Dorgham and Al-Shaqsi (2023) بر روی مفاهیم تک‌واژه‌ای متمرکز بود. این پیشینه‌ها به‌طور مستقیم از مدل برداری «گلاو» برای نمایش برداری عبارات‌های تک‌واژه‌ای استفاده کرده‌اند؛ در حالی که در پژوهش حاضر، از مدل برداری «گلاو» برای برداری کردن اصطلاحات مرکب بهره‌جویی شده است. به‌واقع، می‌توان بیان کرد که استفاده از عبارات‌های مرکب در استخراج مفاهیم از متون ارتباطات علمی در این تحقیق مورد توجه بوده است. پژوهش حاضر گامی ارزنده به‌سوی تعمیق مبحث مذکور در بازایی اطلاعات و دانش با بهره‌مندی از مفاهیم معنایی و اصطلاحات مرکب تلقی می‌شود.

## ۳. روش پژوهش

پژوهش حاضر از لحاظ هدف، کاربردی و از لحاظ رویکرد، کمی است، زیرا از تکنیک‌های داده‌کاوی برای خوشه‌بندی مفهومی اسناد استفاده کرده است. جامعه پژوهش مشتمل بر چکیده مقالات ارتباطات علمی در پایگاه‌های استنادی «اسکوپوس» و «وب‌آوساینس» تا سال ۲۰۲۲ میلادی بود. با جست‌وجو در پایگاه‌های فوق داده‌های اولیه در اسفند ماه ۱۴۰۱ (مارس ۲۰۲۳) بازیابی شدند. تعداد ۵۵۸ مقاله از دو پایگاه «وب‌آوساینس» (۱۸۶) و «اسکوپوس» (۳۷۲) بازیابی شد که پس از ادغام مقالات مشابه، حذف مقالات تکراری و فاقد چکیده، و پاکسازی اولیه، تعداد ۴۹۱ مقاله برای مرحله بعدی و انجام تکنیک‌های متن‌کاوی به‌عنوان نمونه پژوهش در نظر گرفته شد.

هر متن مشتمل بر جملاتی است که عبارت اسمی موضوع جمله مربوط را نشان می‌دهد. از این‌رو در پژوهش حاضر، عبارات اسمی به‌عنوان واژگان و اصطلاحات کلیدی در نظر گرفته شد. همچنین به‌منظور آماده‌سازی داده‌های پژوهش برای اجرای تکنیک‌های داده‌کاوی پردازش‌هایی برای پاک‌سازی داده‌ها بر روی آن‌ها اعمال گردید. داده‌های دو ستون «کلیدواژه مؤلف» و «کلیدواژه اضافی» توسط نرم‌افزار خواننده و از محل جداکننده ویرگول (,) یا نقطه ویرگول (;) تفکیک شده و فاصله‌های خالی اضافی طرفین واژگان نیز با استفاده از تابع strip() حذف گردید و در یک فهرست قرار داده شدند. همچنین از دو ستون «عنوان مقاله» و «چکیده مقالات» با استفاده از تابع (noun\_phrases) از کتابخانه TextBlob، عبارات اسمی به‌عنوان عبارات کلیدی به‌دست آمد و به فهرست واژگان قبلی اضافه گردید. تمامی اصطلاحات و عبارات اسمی با کدنویسی به زبان «پایتون» با استفاده از کتابخانه‌هایی نظیر «پانداس»<sup>۱</sup>، «نامپای»<sup>۲</sup>، «سایکیت لرن»<sup>۳</sup>، پردازش زبان طبیعی (ان‌ال‌تی‌کا)<sup>۴</sup> استخراج شد. در این فهرست، واژگان تکراری زیادی وجود داشت. از این‌رو، یک واژه‌نامه شامل واژگان یکتا به همراه فراوانی آن‌ها تشکیل گردید. برای این کار تابع (Counter) از کتابخانه collections استفاده شد. اصطلاحات در این واژه‌نامه نیز در دو سطح نیاز به بررسی و پاکسازی داشت که برای انجام این مهم اقدامات زیر صورت پذیرفت:

الف. پاکسازی در سطح حروف و علائم هر واژه: واژگان خروجی، دارای بخش‌های زائد و

1. Pandas

2. Numpy

3. Scikit-learn

4. natural language toolkit (NLTK)

بی‌معنا نظیر پراتتز و اعشار بوده که در این مرحله با تابع (`replace()`) حذف گردید. همچنین یکسان‌سازی تبدیل حروف به حروف کوچک نیز مورد توجه بود. قبل از محاسبه فراوانی‌ها، عبارات تکراری، مشابه هم می‌شوند تا هنگام محاسبه فراوانی عبارت‌ها، با هم ادغام شوند. این کار نیز با تابع (`lower()`) انجام پذیرفت.

**ب. پاکسازی کلی با حذف و افزودن واژگان:** در این سطح، بررسی یک واژه و مقایسه یک واژه با سایر واژگان انجام گردید و بیشتر، واژگان اضافی حذف شدند. در این مرحله واژگان تک-حرفی و دو-حرفی از واژه‌نامه حذف شدند. از جمله عملیات ضروری در این مرحله، ادغام عبارات جمع و مفرد و تجمیع فراوانی آن‌هاست؛ زیرا نیاز به استخراج مفاهیم است، نه اصطلاحات. توابع (`pluralize()`) و (`singularize()`) از کتابخانه `Textblob` برای به‌دست آوردن شکل‌های جمع و مفرد از یک عبارت بهره‌جویی شد. افزون بر این، در مجموعه فهرست، جمع هر عبارت جست‌وجو شده است که آیا شکل جمع در واژگان وجود دارد یا خیر؛ اگر وجود داشته باشد، یکی از دو شکل جمع یا مفرد را حذف و فراوانی دو عبارت را با هم جمع می‌کند. حذف اسامی خاص نظیر نام کشورها، ملیت‌ها، ماه، و سال نیز در این مرحله انجام گردید. با دقت در عبارات حاصل ملاحظه گردید که تعدادی از عبارت‌ها و واژگان، اصطلاحات ویژه چکیده هستند؛ نظیر `finding`، `case study`، `paper_shows`، `present_study_aims`. بنابراین، از این نوع عبارت‌ها هم صرف نظر شد. همچنین دو عبارت `scientific_communication` و `scholarly_communication` که موضوع اصلی این پژوهش‌ها بودند، به‌رغم اینکه به لحاظ فراوانی رتبه‌های اول و دوم را داشتند، نیز کنار گذاشته شدند.

به‌طور کلی، به‌منظور اجرای پردازش‌های داده‌کاوی نظیر خوشه‌بندی، باید واژگان و اصطلاحات مستخرج، به شکل بردارهای عددی نشان داده شوند. بدین منظور، از مدل برداری «گلاو» به‌عنوان مدل پایه استفاده و هر یک از اصطلاحات مستخرج در این مرحله به شکل بردارهای ۳۰۰ عددی نمایش داده شد. مشکلی که برای نمایش عددی عبارت‌ها وجود داشت، اینکه مدل «گلاو» به‌صورت یک واژه‌نامه از اصطلاحات تک‌واژه‌ای بود و لازم بود برای عبارات دو یا چند واژه‌ای مستخرج نیز، که قسمت عمده خروجی را تشکیل داده، بردارهای متناسب ساخته شود. این فرایند با اجرای سه مرحله شکستن عبارت به واژگان تشکیل‌دهنده، یافتن بردار عددی هر واژه از روی مدل «گلاو»، و محاسبه بردار میانگین از بردارهای واژگان انجام گرفت. بدین ترتیب، این عبارات نیز جایگاهی را در بین

بردارها و واژگان به دست آورده و در ادامه، امکان انجام خوشه‌بندی، دسته‌بندی و استخراج مفاهیم را خواهند داشت.

به منظور استقلال این مرحله، مدل برداری ساخته شده برای مجموعه واژگان و عبارات مستخرج ذخیره گردید تا برای مراحل بعدی بارها و بارها از آن استفاده شود. به این ترتیب، در مراحل بعدی نیازی به استفاده از مدل اصلی (لیست شامل ۴۰۰۰۰۰ بردار عبارت تک‌واژه‌ای) «گلاو» نیست؛ زیرا مدل برداری حاصل و سفارشی شده برای عبارت‌ها، به نسبت، حجم کمتر و سرعت بیشتری داشته و دیگر بارگذاری و استفاده از مدل اصلی اولیه «گلاو» مورد نیاز نخواهد بود. اکنون از روی مدل حاصل، که به شکل یک واژه‌نامه کوچک ذخیره شده، می‌توان برای هر کلمه یا عبارت از خروجی مرحله قبل، بردار ۳۰۰ عددی متناظر را به دست آورد و در ادامه، عبارت‌های دو و چند واژه‌ای را نیز همراه با عبارات تک‌واژه‌ای مورد خوشه‌بندی قرار داد.

پس از اینکه برای واژگان و عبارات باقی مانده در مرحله قبل، از روی واژه‌نامه «گلاو» بردارهای عددی تعیین شد، به منظور خوشه‌بندی آن‌ها از الگوریتم کا-مینز<sup>۱</sup> استفاده شد و متغیرها و شاخص‌های دیگری هم برای استفاده‌های بعدی از خوشه‌ها به دست آمد؛ از جمله: تعداد عناصر، مجموع فراوانی عناصر، نماینده خوشه (سرخوشه) و فراوانی.

در ادامه، نسبت به حذف خوشه‌های ضعیف بر اساس شاخص‌های محاسبه شده برای هر خوشه اقدام گردید. یعنی خوشه‌هایی که با توجه به تعداد عنصر (اندازه خوشه)، مجموع فراوانی عناصر (ارزش جمعی) یا فراوانی نماینده آن (ارزش سرخوشه) نسبت به سایر خوشه‌ها امتیاز کمتری به دست آورده‌اند، حذف شدند. شاخص‌های الف) تعداد عناصر خوشه کمتر از ۶، ب) مجموع فراوانی عناصر خوشه، کمتر از ۲۰، و ج) بیشترین فراوانی عناصر خوشه، کمتر از ۵، برای حذف خوشه‌های ضعیف و ادغام عناصر آن‌ها در سایر خوشه‌ها مد نظر قرار گرفت. پس از تعیین و حذف خوشه‌های ضعیف، نمایندگان خوشه‌های باقی مانده به عنوان نمایندگان دسته‌ها برای دسته‌بندی انتخاب شد. کل عبارت‌ها به طور مجدد دسته‌بندی شده و هر یک از عبارت‌ها بر اساس میزان شباهت با نمایندگان در نزدیک‌ترین دسته قرار گرفت. با انجام دسته‌بندی، افزون بر عناصر و نمایندگان اولیه هر دسته، مهم‌ترین اصطلاحات در هر یک از دسته‌ها تعیین گردید (عبارات با بیشترین فراوانی). توجه ویژه به عبارت‌های مزبور در گام آخر می‌تواند در شناسایی مفهوم مستخرج از آن دسته کمک کند.

1. K-means

#### ۴. تجزیه و تحلیل یافته‌ها

به‌منظور پاسخگویی به پرسش اول، ابتدا پردازش بر روی داده‌های مستخرج صورت گرفت. استخراج واژگان کلیدی، موضوعات و عبارات‌های اسمی از متون با تکنیک‌های متن کاوی در زبان «پایتون» مطابق جدول ۱، به‌دست آمده است.

جدول ۱. تعداد آیت‌های مستخرج از داده‌های اولیه

ردیف	نوع آیت‌های مستخرج	تعداد	توضیحات
۱	کلید واژه	۴۵۵۸	مستخرج از دو ستون کلیدواژه مؤلف و اضافی
۲	عبارت اسمی	۱۳۶۵۱	مستخرج از دو ستون عنوان و چکیده
۳	واژگان کوتاه	۲۷۹	واژگان دو-حرفی و تک-حرفی (طول > ۳)
۴	واژگان اولیه مورد بررسی	۱۷۹۳۰	مجموع دو سطر ۱ و ۲ منهای تعداد واژگان کوتاه
۵	واژگان مانده منحصربه‌فرد	۷۴۴۲	اصطلاحات تک‌کلمه‌ای و بیشتر
۶	واژگان با تکرار ۲ و بیشتر	۳۱۹۶	پس از حذف کلماتی که فقط یک بار ظاهر شدند.
۷	واژگان با تکرار ۳ و بیشتر	۹۷۹	پس از حذف کلماتی که فقط دو بار ظاهر شدند.
۸	عبارات پس از ادغام جمع	۸۹۶	باقی‌مانده پس از ادغام جمع و مفرد
۹	عبارات نهایی	۷۹۲	پس از حذف اصطلاحات عمومی ویژه چکیده

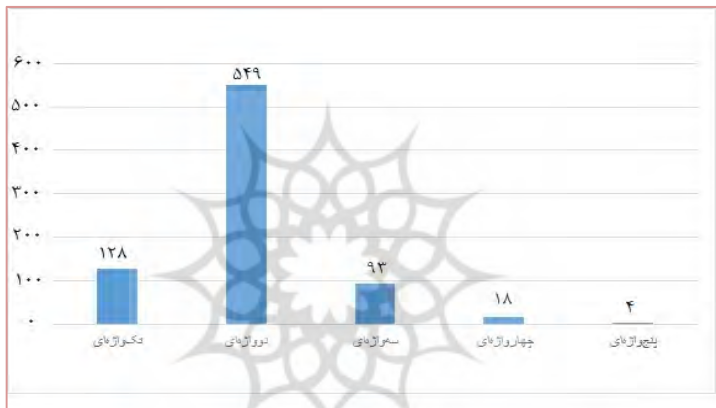
تعداد کلیدواژه‌ها بالغ بر ۱۷۹۳۰ مورد بود که ۱۳۶۵۱ عبارت آن، عبارت اسمی است. در ادامه، توزیع فراوانی عبارات با بیشترین تکرار در جدول ۲، ارائه شده است. در این جدول، تعداد ۱۰ عبارت پر تکرار ذکر شده است.

جدول ۲. توزیع فراوانی ۱۰ عبارت پر تکرار حوزه ارتباطات علمی

ردیف	عبارت/ واژگان	فراوانی	ردیف	عبارت/ واژگان	فراوانی
۱	scientific_communication	۵۶۱	۶	institutional_repositories	۸۷
۲	scholarly_communication	۴۸۰	۷	academic_libraries	۷۸
۳	open_access	۲۴۰	۸	access	۷۵
۴	social_media	۱۰۷	۹	Web	۷۱
۵	scientific_communication_skills	۹۹	۱۰	open_science	۶۳

پس از آشکارسازی عبارت‌ها، واژگانی را که از نظر معنایی دارای قرابت معنایی بود در یک دسته گنج‌نموده و پس از آنکه فهرست عبارت‌ها مشخص شد، مشاهده گردید که کلماتی مانند "Web" و "Internet" از نظر معنایی قرابت و شباهت زیادی با یکدیگر داشته و می‌توان این نوع عبارات مشابه را در یک دسته قرار داد. به گفته دیگر، می‌توان عبارات را از روی معنا و مفهوم خوشه‌بندی کرد تا به‌جای بررسی و کار با فهرستی گسترده، تعداد آن‌ها محدود شده و مفاهیم برآورد شوند.

از سوی دیگر، وضعیت اصطلاحات حوزه ارتباطات علمی نیز نیازمند بررسی است که یافته‌های آن از منظر چند واژه‌ای در شکل ۱، به تصویر کشیده شده است.



شکل ۱. فراوانی عبارات از منظر تعداد واژه

همان‌طور که در نمودار مشخص است، تنها ۱۶ درصد از اصطلاحات حوزه ارتباطات علمی، تک‌واژه‌ای و ۸۴ درصد آن‌ها عبارت مرکب هستند. از این‌رو باید در فرایند خوشه‌بندی نیز مد نظر قرار گیرند. برای پاسخ به پرسش اول مبنی بر دخیل کردن اصطلاحات مرکب (چند کلمه‌ای) در فرایند خوشه‌بندی، یک مدل برداری یا واژه‌نامه مشابه مدل «گلاو» که هر کلمه را به یک بردار عددی ۳۰۰ عنصری متناظر می‌کند، ساخته شد. بردارهای عددی که بر مبنای بردارهای «گلاو» بنا شده، به‌صورت نهفته شباهت معنایی و قرابت عبارات مشابه را در خود داشته است. از این‌رو، این بردارها می‌توانند به‌جای عبارت‌های مربوط در خوشه‌بندی بهره‌جویی شده و سپس عبارت معادل را به‌عنوان نتیجه خوشه‌بندی به‌دست آورد. برای مثال، اگر برای عبارت publish or perish culture از مدل «گلاو» استفاده کنیم، با خطای "Key 'publish or perish culture' not present in vocabulary"



مواجه خواهیم شد. برای رفع این مشکل، ابتدا عبارت مورد نظر به واژه‌های or, publish, perish و culture شکسته می‌شود. این واژه‌ها در مدل «گلاو» وجود دارند و می‌توان بردار عددی هر یک را از روی مدل «گلاو» به صورت زیر بازیابی کرد:

publish: [ 0.054996 -0.29319 -0.26528 -0.057689 -0.056067 ... ]

or: [-0.26234 0.12633 -0.084185 -0.077551 -0.08509 ... ]

perish: [ 0.12621 -0.2245 -0.22975 0.04251 -0.15619 ... ]

culture: [ 0.15588 0.011506 -0.16631 -0.16458 0.41527 ... ]

در این مرحله، از بردارهای عددی بالا میانگین گرفته شد تا بردار عددی معادل (میانگین آرایه‌ها) به صورت زیر محاسبه گردد:

publish\_or\_perish\_culture: [ 0.0186865 -0.0949635 -0.18638127 -0.0643275 0.02948075 ... ]

نکته دیگر، واژه‌هایی نظیر Covid-19 بوده که اگرچه مرکب نیست، ولی در مدل «گلاو» نیست و به جای آن عبارت global\_pandemic جایگزین و سپس به طور مشابه از بردارهای واژگان آن میانگین گرفته شد. بر این قرار، عبارت‌های publish\_or\_perish\_culture و global\_pandemic همراه با بردارهای عددی به دست آمده به مجموعه واژگانی اضافه می‌شود و سرانجام، با یک مجموعه واژگانی متشکل از ۷۹۲ بردار عددی (مطابق جدول ۱) محدود و بسیار کمتر نسبت به مدل اصلی «گلاو» (با ۴۰۰۰۰۰ واژه) برای پاسخ به پرسش دوم سروکار خواهیم داشت که هم ساده و کوچک بوده و هم عبارات مرکب مورد نیاز را در برمی‌گیرد.

سرانجام، به منظور پاسخ به پرسش دوم، پس از آنکه بردار عددی معادل برای تمامی اصطلاحات به دست آمد، فرایند خوشه‌بندی با تعداد ۴۰ خوشه بر روی این بردارها انجام گردید که مشخصات خوشه‌ها در جدول ۳، ذکر شده است. تعداد عناصر خوشه‌ها از ۱ تا ۸۰ عبارت متغیر است. در یافته‌های اولیه مشاهده گردید که خوشه [speed] یک خوشه منفرد (تک عنصری)، خوشه [serials, serials\_crisis] دو عضوی، و غنی‌ترین خوشه‌ها، خوشه {smart\_contracts, effect\_sizes, impact, ...} دارای ۸۰ عضو و خوشه {scientific\_communication\_skills, communications, academic\_communication, ...} ۶۷ عضو داشت. همچنین در این جدول، برای هر خوشه، افزون بر مرکز، سه متغیر (شاخص) دیگر خوشه نیز ذکر شده است که عبارت‌اند از: تعداد عناصر، بیشترین عدد بین فراوانی عناصر، و مجموع فراوانی‌های کل عناصر آن خوشه. مهم‌ترین اصطلاحات (دو واژه دارای بیشترین فراوانی) نیز برای هر خوشه تعیین و در جدول ۳، ارائه شده است.

جدول ۳. خوشه‌بندی اولیه مفاهیم مستخرج از متون حوزه ارتباطات علمی در پایگاه‌های «اسکوپوس» و «وب‌آوساینس»

ردیف	مرکز خوشه	دو واژه پرتکرار	تعداد	بیشینه	جمع
۱	current_problems	[smart_contracts, effect_sizes]	۸۰	۱۳	۳۳۳
۲	direct_scientific_communication	[scientific_communication_skills, communications]	۶۷	۹۹	۵۳۳
۳	scientific_research	[research_work, research_evaluation]	۶۰	۱۹	۳۶۳
۴	concept_analysis	[ambient_intelligence, latent_profile_analysis]	۴۲	۶	۱۵۵
۵	scientific_information_management	[information_literacy, scientific_information]	۴۱	۳۲	۲۹۱
۶	theory_approach	[practical, social_cognitive_career_theory]	۳۴	۱۴	۱۶۳
۷	scientific_knowledge	[scientific_literature, scientific_production]	۳۳	۲۸	۳۰۷
۸	academic_institutions	[academic_libraries, universities]	۳۲	۷۸	۲۸۲
۹	online_video_journal	[web, internet]	۳۰	۷۱	۳۰۰
۱۰	science_education	[open_science, social_sciences]	۲۷	۶۳	۲۱۹
۱۱	social_media	[social_media, social_networks]	۲۳	۱۰۷	۲۴۰
۱۲	communication_technologies	[communication_technologies, electronic_resour...]	۲۱	۲۹	۱۳۹
۱۳	open_access	[open_access, access]	۱۸	۲۴۰	۴۰۱
۱۴	new_information	[new_media, new_models]	۱۸	۱۰	۸۳
۱۵	citation_index	[citation_analysis, citation_data]	۱۷	۳۷	۱۳۰
۱۶	scholarly_journals	[scholarly_publishing, electronic_journals]	۱۷	۳۶	۲۰۲
۱۷	doctoral_students	[doctoral_students, faculty_members]	۱۷	۳۲	۱۱۵
۱۸	main_objective	[main_objective, ground_truth]	۱۷	۸	۶۷
۱۹	data	[data, data_management]	۱۶	۱۵	۱۰۲
۲۰	researchers	[early_career_researchers, researchers]	۱۵	۲۷	۹۰
۲۱	learning_model	[problem_based_learning, technology_acceptance...]	۱۵	۹	۷۱
۲۲	digital	[digital_libraries, digital_scholarship]	۱۴	۳۱	۱۱۷
۲۳	university_level	[conference_papers, state_university]	۱۴	۱۸	۶۹

ردیف	مرکز خوشه	دو واژه تکرار	تعداد	بیشینه	جمع
۲۴	organizational_repositories	[institutional_repositories, repositories]	۱۳	۸۷	۱۷۴
۲۵	college_library	[university_libraries, libraries]	۱۳	۴۵	۱۰۶
۲۶	academic_research_video	[google_scholar, collaboration]	۱۳	۱۵	۷۱
۲۷	skills	[communication_skills, training]	۱۱	۱۷	۷۲
۲۸	peer_review	[peer_review, literature_review]	۱۰	۳۰	۷۲
۲۹	liaison_librarians	[librarians, mentoring]	۹	۱۳	۵۷
۳۰	scientometric_assessment	[evaluation, impact_evaluation]	۹	۱۰	۴۱
۳۱	empirical_studies	[bibliometric_analyses, bibliometrics_analysis]	۸	۷	۴۰
۳۲	preprints	[preprints, scientometrics]	۶	۱۶	۵۷
۳۳	knowledge_graphs	[information_graphs, knowledge_graphs]	۶	۸	۳۵
۳۴	language	[language, language_distribution]	۶	۶	۳۰
۳۵	global_pandemic	[global_survey, global_networks]	۵	۶	۲۱
۳۶	research_productivity	[research_productivity, utilization]	۴	۸	۲۴
۳۷	instruction	[skills_instruction, direct_instruction]	۴	۶	۲۰
۳۸	group_investigation	[experimental_group, group_investigation]	۴	۴	۱۴
۳۹	serials	[serials, serials_crisis]	۲	۴	۷
۴۰	first-citation_speed	[first-citation_speed]	۱	۴	۴

در این مرحله با داشتن مقادیر متغیرها از مرحله قبل، می‌توان به حذف خوشه‌های ضعیف پرداخت. مطابق داده‌های جدول ۳، خوشه‌های سطور ۳۵ تا ۴۰، مطابق با معیارهای مدنظر خوشه‌های ضعیف محسوب شده و حذف می‌شوند. پس از حذف این خوشه‌ها، به دسته‌بندی پرداخته شد. در دسته‌بندی مفهومی هر یک از سرخوشه‌های باقی‌مانده از مرحله قبل را به‌عنوان نمایندگان دسته‌ها، در یک دسته (سبد) قرار داده و با استفاده از شباهت سایر کلمات با هر یک از این نمایندگان، در این دسته‌ها قرار می‌گیرند. نتیجه نهایی دسته‌بندی مفهومی با رویکرد معنایی در جدول ۴، ارائه شده است:

جدول ۴. دسته‌های حوزه ارتباطات علمی با رویکرد معنایی

ردیف	نام دسته	ردیف	نام دسته	ردیف	نام دسته
۱	social_media	۹	scientific_production	۱۶	scientific_communication_skills
۲	academic_libraries	۱۰	communication_ technologies	۱۷	scientific_literature
۳	training	۱۱	data	۱۸	scholarly_publishing
۴	peer_review	۱۲	open_access	۱۹	information_literacy
۵	web	۱۳	early_career_researchers	۲۰	conference_papers
۶	doctoral_students	۱۴	research_evaluation	۲۱	university_researchers
۷	librarians	۱۵	institutional_repositories	۲۲	bibliometric_analysis
۸	preprints				

ستون «نام دسته» در جدول ۴، عنوان یا برچسبی انتخابی است و با توجه به مهم‌ترین واژه‌هایی انتخاب شده که در آن دسته قرار گرفته است. برای مثال، واژه‌های bibliometric\_ empirical\_، empirical\_، study\_analyses، qualitative، forensic، bibliometrics\_analysis، analyses، evidence، و empirical\_studies که پیش‌تر در خوشه‌ای با نماینده empirical\_studies بوده، اکنون با برچسب bibliometric\_analyses نشان داده شده است. همچنین به جهت درک بهتر، خوشه‌ها در فضایی دو بعدی با استفاده از الگوریتم کاهش ابعاد<sup>۱</sup> در شکل ۲، ارائه شده است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
رتال جامع علوم انسانی

1. principal component analysis (PCA)



عناصر دسته	مرکز دسته
['information_literacy', 'information_sources', 'information_behavior', 'scientific_literacy', 'information_exchange', 'information_management', 'information_professionals', 'information_infrastructure', 'information_resources', 'critical_information_literacy', 'electronic_information', 'electronic_information_environment', 'environmental_literacy', 'information_needs', 'modern_information', 'digital_literacy', ...]	information_literacy
['peer_review', 'evaluation', 'assessment', 'continuous_peer_review', 'peer', 'peer_review_innovation_activities', 'review', 'scientometric_assessment', 'bibliometric_evaluation', 'critical_review', 'peer_review_journals', 'recommendations', 'review_alternatives']	peer_review

## ۵. بحث و نتیجه‌گیری

خوشه‌بندی، روشی بدون نظارت در یادگیری ماشین است. بر این قرار که هنگام خوشه‌بندی جدا از همان داده‌های مسئله، که باید به خوشه‌های مختلف تقسیم شوند، هیچ داده دیگری برای آموزش به ماشین داده نمی‌شود. الگوریتم خوشه‌بندی تنها می‌تواند داده‌ها را با هم مقایسه کرده، بر اساس شباهت و عدم شباهت بین داده‌ها (و نه مقایسه با نمونه آموزشی) عمل کند تا داده‌ها در یک خوشه قرار گرفته یا به خوشه‌های مختلف تقسیم شوند. تکنیک مزبور می‌تواند در شناسایی الگوهای مخفی و رؤیت‌پذیری تنوع داده‌ها مفید باشد. در انجام پژوهش حاضر، از روش معروف و متداول خوشه‌بندی «کی-مینز» استفاده شده است که در آن هر مورد بر اساس فاصله با میانگین خوشه برای عضویت یا عدم عضویت در یک خوشه تصمیم می‌گیرد. برای پیاده‌سازی الگوریتم خوشه‌بندی «کی-مینز» در زبان «پایتون» می‌توان از کتابخانه‌های مختلفی نظیر NumPy، Pandas و Scikit-learn استفاده کرد.

هدف از پژوهش حاضر، استخراج و خوشه‌بندی مفهومی متون حوزه ارتباطات علمی در پایگاه‌های استنادی «اسکوپوس» و «وب‌آوساینس» با استفاده از تکنیک‌های داده‌کاوی بود. بدین منظور از روش‌های مختلفی از جمله استخراج واژگان کلیدی و ایجاد خوشه‌های معنایی برای تجزیه و تحلیل داده‌های متنی بهره‌جویی گردید. در خوشه‌بندی که از روش‌های متداول در داده‌کاوی است، برای واژگانی که زمینه‌های مفهومی را انتقال می‌دهند، بایستی از بردارهای عددی بهره جست تا دسته‌بندی بر روی داده‌ها مبتنی بر بردارهای معادل صورت پذیرد. همان‌طور که در بخش روش‌شناسی نیز بیان گردید، مدل برداری «گلاو» به صورت یک واژه‌نامه کمی، واژگان را فهرست کرده و به آن‌ها بردار

عددی نسبت می‌دهد. اما بسیاری از مفاهیم مورد استخراج پژوهش حاضر، اصطلاحات چند کلمه‌ای بوده که در بردارهای مزبور نمی‌گنجد و برای استفاده از این نوع اصطلاحات در خوشه‌بندی و دسته‌بندی باید معادل بُرداری مناسب به آن‌ها نسبت داده شود. یکی از نوآوری‌های پژوهش حاضر گذر از خوشه‌بندی بر روی واژه‌های تک کلمه‌ای به عبارتی و توجه به ارتباطات معنایی میان واژگان بود.

یافته‌های پژوهش حاضر نشان داد که متون حوزه ارتباطات علمی را می‌توان در ۲۲ خوشه با توجه به رویکردهای معنایی جای داد. برخی از این خوشه‌ها مشتمل بر «مخازن سازمانی»، «سواد اطلاعاتی»، «شبکه‌های اجتماعی»، «دسترسی آزاد» و «فناوری‌های ارتباطی» است. به‌طور کلی و با توجه به مفاهیم مرتبط با این حوزه که در خوشه‌بندی مفهومی پژوهش حاضر نیز نشان داده شد، ارتباطات علمی به‌عنوان حوزه‌ای مهم در تحقیقات علمی و انتقال دانش مورد توجه است. بررسی خوشه‌های مستخرج از متون حوزه ارتباطات علمی نشان می‌دهد که برخی از آن‌ها امکان ادغام یا قرار گرفتن به‌عنوان یک عنصر در آن خوشه را دارند. به‌طور مثال، خوشه‌های «دسترسی آزاد» با «مخازن سازمانی»، و «مهارت‌های ارتباط علمی» با «سواد اطلاعاتی» با یکدیگر همپوشانی معنایی دارند. در ادامه، به تبیین برخی از خوشه‌های مستخرج با رویکرد معنایی مبادرت خواهد شد.

خوشه «مخازن سازمانی» به‌عنوان یک مفهوم در حوزه ارتباطات علمی مطرح است. عناصر این خوشه همان‌گونه که در جدول ۵، مشخص است، مشتمل بر مفاهیمی نظیر کتابخانه‌های دانشگاهی، اساتید دانشگاهی، بایگانی‌ها و منابع دانشگاهی است. این امر نشان می‌دهد که روش پیشنهادی در این خوشه توانسته است مفاهیم مرتبطی را در خود جای دهد. «حسن‌زاده، گلینی‌مقدم و اجاق» (۱۴۰۲) نیز به استفاده از مخازن سازمانی و به‌نوعی واسپارگاه‌های دانشگاهی به‌عنوان یک عامل مؤثر در ارتباطات علمی اشاره داشته‌اند. «باقری بنجار» و همکاران (۱۳۹۴) از دانشگاه‌ها، کتابخانه‌ها و فضای مجازی به‌عنوان محیط‌هایی که امکان ارتباط و تعامل علمی را فراهم می‌کنند، یاد کرده است. همچنین پژوهش‌های چندی نظیر Martin-Yeboah, Alemna & Adjei (2018) و Gozetti (2006) نیز به تأثیر مخازن سازمانی بر ارتباطات علمی و بهره‌وری علمی دانشگاه‌ها اشاره کرده‌اند که به نوعی نتایج خوشه‌بندی حاضر را در این خوشه تأیید می‌کند.

خوشه دیگر «سواد اطلاعاتی» بود که عناصر این خوشه مطابق با جدول ۵، مشتمل بر مفاهیمی نظیر منابع اطلاعاتی، رفتار اطلاعاتی، سواد علمی و مدیریت اطلاعات است. سواد

اطلاعاتی به‌عنوان مفهومی کلی بر ارتباطات علمی تأثیرگذار است. (Buitrago Ciro (2022). معتقد است که سواد اطلاعاتی، توانایی ارزیابی انتقادی از اطلاعات و منابع آن و دسترسی به آن به شیوه‌ای مؤثر و کارآمد است. (Xiao (2024) نیز با اشاره به مفهوم سواد دیجیتال، اظهار می‌دارد که فناوری‌های دیجیتال به طرز فوق‌العاده‌ای در حال تغییر نحوه ارتباط و بازیابی اطلاعات و دانش هستند و پژوهشگران می‌توانند با استفاده از فناوری‌های دیجیتال بیاموزند و ارتباط برقرار کنند. برقراری ارتباط در این فضا با داشتن و کسب مهارت‌های سواد اطلاعاتی موجبات تقویت ارتباطات علمی را به‌دنبال خواهد داشت. در پژوهش «حسن‌زاده، گلینی مقدم و اجاق» (۱۴۰۲) بالا بودن سواد اطلاعاتی دیجیتال از عوامل فردی مؤثر بر ارتباطات علمی بیان شده است. (Gullbekk (2016) نیز به مزیت‌های «سواد اطلاعاتی» می‌پردازد که به‌عنوان یک زمینه اجتماعی برای ارتباطات علمی میان‌رشته‌ای بوده و می‌تواند تغییرات و نوآوری‌ها را در شیوه‌های اجتماعی از جمله در ارتباطات علمی تبیین کند.

شبکه‌ها و رسانه‌های اجتماعی، مفهومی تأثیرگذار در حوزه ارتباطات علمی است که یکی از خوشه‌های اصلی در پژوهش حاضر بود و مطابق با داده‌های جدول ۶، مفاهیمی نظیر وب اجتماعی، اینترنت، شبکه‌های اجتماعی، و ارتباطات برخط را در خوشه خود جای داده بود. موضوع اهمیت رسانه‌های اجتماعی و تأثیرگذاری به‌عنوان بستر یا ابزار در ارتباطات علمی در پژوهش‌های بی‌شماری نظیر پژوهش‌های «دهقانپور، صمدی و نوروزی چاکلی» (۱۴۰۲)، (Al-Daihani, Al-Qallaf & AISaheeb (2018)، «دستیار، رزاقی و فقیه عبداللهی» (۱۳۹۹)، «گلینی مقدم، سمیعی و کمالی سروسستانی» (۱۳۹۷)، و «شمسی و سلیمانی» (۱۳۹۴) مورد تأیید و تأکید قرار گرفته است. «مرادویسی، پناهی و صدقی» (۱۴۰۰) به نقش شبکه‌های اجتماعی در ارتباطات علمی دانشجویان و اعضای هیئت علمی در «دانشگاه علوم پزشکی ایران» توجه کرده‌اند. با افزایش محبوبیت رسانه‌های اجتماعی، اغلب پژوهشگران با استفاده از خدمات ارائه‌شده در این شبکه‌ها به تبادلات علمی می‌پردازند که منجر به افزایش ارتباطات علمی پژوهشگران می‌شود. همچنین مهارت‌های ارتباطی نیز جزو عناصر این خوشه قرار دارد که در پژوهش «پرهام‌نیا» (۱۴۰۰) نیز در بخش عوامل فردی اجتماعی به مهارت‌های ارتباطی به‌عنوان یکی از عوامل مؤثر بر ارتباطات علمی پرداخته شده است.

خوشه «فناوری‌های ارتباطی» به ارتباطات و فناوری‌های ارتباطی اشاره دارد و شامل مفاهیمی نظیر فرایندهای ارتباطی، سیستم‌های ارتباطی، فناوری‌های دیجیتال، فناوری‌های



اطلاعاتی و ارتباطاتی و ابزارهای ارتباطی است. کانال‌ها و ابزارهای برقراری ارتباط به‌عنوان یکی از زیرساخت‌های هر نوع ارتباط علمی همیشه مورد توجه بوده و در پژوهش‌های بی‌شماری نظیر «حسن‌زاده، گلینی مقدم و اجاق» (۱۴۰۲) و «سهیلی، دانش، و فتاحی» (۱۳۸۹) مورد تأکید قرار گرفته است. «مهرآور گیگلو» و همکاران (۱۳۹۹) موانع مالی و تجهیزات و زیرساخت‌های فیزیکی و ارتباطی را از مهم‌ترین موانع همکاری‌های علمی بین‌المللی ذکر می‌کنند. بنابراین، توجه یا عدم توجه به این مفهوم، موفقیت یا شکست در ارتباطات علمی را به‌دنبال خواهد داشت.

«تولید علمی» خوشه دیگری بود که به تولید محصولات علمی و نتایج پژوهش‌ها اشاره دارد. این محصولات شامل مقالات علمی، کتب، پایان‌نامه‌ها، گزارش‌های پژوهشی، نرم‌افزارهای علمی و غیره هستند. این مفهوم می‌تواند زاویه‌ها و جنبه‌های مختلف از ارتباطات علمی را دربرداشته باشد؛ از جمله تولید علمی، همکاری، گزارشات، سیاست‌های عمومی، و انتشارات پژوهشی. پژوهش‌های متعددی نیز به‌صورت مجزا به نقش هر کدام از تولیدات علمی در ارتباطات علمی پرداخته‌اند. «برهمنند» (۱۳۸۶) به جایگاه مجلات علمی در چرخه ارتباطات علمی و به نقش مجلات الکترونیکی و ارزیابی همکاران پرداخته است. «خسروجردی» (۱۳۸۵) بیان کرده است که نشریات، کتاب‌ها، گزارش‌ها، شرکت در همایش‌های علمی، ارسال پیش‌چاپ‌ها و غیره کانال‌هایی هستند که اطلاعات را به اعضای دانشکده نامرتبی منتقل می‌کنند. نتایج پژوهش «رضادوست، نواح و ادیب‌زاده» (۱۳۹۶) نیز نشان می‌دهد بین متغیرهای مستقل ارتباطات علمی و متغیر وابسته میزان تولید علمی رابطه مثبت و معناداری وجود دارد.

خوشه دیگر، خوشه «داوری همتا» بود که مشتمل بر مفاهیمی نظیر مرور، ارزیابی و بازبینی همتایان یا همکاران، سنجش، نقد، بحث و بررسی، مرور و داوری در نشریات علمی، ارزیابی و اندازه‌گیری‌های علم‌سنجی و توصیه‌هاست. این موارد می‌توانند به بهبود کیفیت تحقیقات و اعتبار علمی نتایج کمک کرده و نقش مهمی در ارتباطات علمی و تقویت آن دارند. «برهمنند» (۱۳۸۶) و (Nosek and Bar-Anan (2012) به ارزیابی همتایان، و Khosrowjerdi (2011) نیز به ممیزی و کنترل کیفیت محتوای علمی تأکید داشته‌اند. همچنین «مهرابی» و همکاران (۱۴۰۱) ارزیابی مقاله را از مهم‌ترین عوامل اثرگذار برای کیفیت بخشی به پژوهش‌های حوزه علوم انسانی و اجتماعی دانستند. این کیفیت بخشی سرانجام، به‌صورت غیرمستقیم بر ارتباطات علمی و کیفیت آن تأثیر خواهد گذاشت.

خوشه «دسترسی آزاد» شامل مفاهیمی نظیر انتشارات دسترسی آزاد، سیاست‌های دسترسی آزاد، اثر دسترسی، زیرساخت‌های دسترسی آزاد، کانال‌ها و سکوها‌ی غیررسمی، اطلاعات دسترسی، گسترش و استفاده از دسترسی آزاد، دسترسی رایگان، جنبه‌های عمده دسترسی آزاد، اشتراک‌گذاری و دسترسی عمومی بود. دسترسی آزاد به اطلاعات علمی یکی از تسهیل‌کننده‌های اصلی ارتباطات علمی است. با دسترسی آزاد، اطلاعات علمی به‌سرعت و به‌سادگی در دسترس جامعه علمی و عموم خواهد بود. دسترسی آزاد به بهبود اشتراک دانش، افزایش تفاهم و تبادل اطلاعات بین محققان و دانشگاه‌ها نیز کمک می‌کند. «پاکدامن» (۱۳۸۴)، «زره‌ساز» (۱۳۹۰)، «روزبهرانی و ریاحی‌نیا» (۱۳۹۵) و «سهرابی و غفاری» (۱۳۹۸) دسترسی آزاد به منابع را در پژوهش‌های خود مطرح و مطالعه کردند. همان‌گونه که ملاحظه می‌شود، عمده خوشه‌های استخراج‌شده با رویکرد معنایی از مؤلفه‌های مؤثر در حوزه ارتباطات علمی قلمداد می‌شوند که نشان از دقت خوشه‌بندی انجام‌شده دارد.

یافته‌های پژوهش حاضر نشان داد که مفاهیم و مؤلفه‌های تأثیرگذار و متعددی در حوزه ارتباطات علمی وجود دارند و این مفاهیم می‌توانند در تحقیقات و پژوهش‌های آتی مورد توجه قرار گیرند. همچنین، این پژوهش نشان می‌دهد که تحلیل متون علمی و استفاده از روش‌های خوشه‌بندی می‌تواند به تفهیم بهتر مفاهیم و موضوعات در حوزه ارتباطات علمی کمک کند. این ابزارها می‌توانند در تحقیقات بیشتر و همچنین در تدوین راهبردها و سیاست‌های مرتبط با علم و ارتباطات علمی مفید باشند. به‌طور کلی، پژوهش حاضر به توسعه دانش و فهم عمیق‌تر در زمینه ارتباطات علمی کمک خواهد کرد و می‌تواند به سازمان‌ها، پژوهشگران، و تصمیم‌گیرندگان این حوزه کمک کند تا منابع و راهبردهای خود را بهبود بخشند و در ارتقای علم و تحقیقات موفق‌تر باشند.

انجام پژوهش حاضر و سایر پژوهش‌های مشابه در راستای آشکارسازی جایگاه عوامل مختلف مؤثر بر ارتباطات علمی به‌عنوان یکی از مهم‌ترین مسائل تأثیرگذار بر توسعه علمی و رشد علم است. بدین منظور پیشنهاد می‌گردد که پژوهش‌هایی در ارتباط با هر یک از مؤلفه‌های تأثیرگذار یا خوشه‌های اصلی پژوهش حاضر اجرا گردد. همچنین توجه به ارتباطات علمی و عوامل مؤثر بر آن در حوزه‌های علوم انسانی و به‌ویژه در دانشگاه‌ها و مراکز آموزش عالی توصیه می‌گردد.

## فهرست منابع

- ابویی اردکان، محمد، حسن عابدی جعفری، و فتاح آقازاده. ۱۳۸۹. کاربرد روش‌های خوشه‌بندی در ترسیم نقشه‌های علم: مورد کاوی نقشه علم مدیریت شهری. *پژوهشنامه پردازش و مدیریت اطلاعات* ۲۵ (۳): ۳۴۷-۳۷۱.
- احمدی، الهام، فریده عصاره، و غلامرضا حیدری. ۱۳۹۶. شناسایی و تحلیل عوامل ترغیب‌کننده و بازدارنده همکاری علمی اعضای هیئت علمی در سطوح محلی، ملی و بین‌المللی در دو دانشگاه شهید چمران و علوم پزشکی جندی شاپور اهواز. *مدیریت اطلاعات سلامت* ۱۲ (۲): ۱۸۳-۱۹۳.
- باقری بنجار، عبدالرضا، نرگس مصلحی جنابیان، بانو بیگی ملک‌آبادی، و میثم محمدی. ۱۳۹۴. بررسی رابطه ارتباطات علمی و خوداثربخشی دانشگاهی دانشجویان. *توسعه اجتماعی (توسعه انسانی سابق)* ۱۰ (۲): ۷۹-۱۰۰.
- برهمند، نیلوفر. ۱۳۸۶. مجلات الکترونیکی و ارتباطات علمی. *فصلنامه کتاب* ۷۲: ۲۰۱-۲۱۲.
- پاکدامن، نشانه. ۱۳۸۴. ارتباطات علمی در دانشگاه‌ها با تأکید بر مشکلات دسترسی به مجلات علمی. *تحقیقات کتابداری و اطلاع‌رسانی دانشگاهی* ۳۹ (۴۳): ۴۱-۵۶.
- پرهام‌نیا، فرشاد. ۱۴۰۰. واکاوی مفهوم ارتباطات علمی در مکتب شیکاگو با تأکید بر کنش متقابل نمادین. *بازیابی دانش و نظام‌های معنایی* ۷ (۲۶): ۱۴۳-۱۶۳.
- پلویی، آرزو، و نادر نقشینه. ۱۳۹۵. ارتباطات علمی از منظر آرای نیکلاس لومان. *مطالعات ملی کتابداری و سازماندهی اطلاعات* ۲۷ (۲): ۷-۲۰.
- حسن‌زاده، آرزو، گلنسا گلینی‌مقدم، و زهرا اجاق. ۱۴۰۲. بررسی سطح آگاهی دانشجویان رشته علم اطلاعات و دانش‌شناسی دانشگاه علامه طباطبائی درباره ارتباطات علمی. *مطالعات دانش‌پژوهی* ۲ (۱): ۳۳-۵۳.
- خسروجردی، محمود. ۱۳۸۵. ترسیم شبکه ارتباطات علمی میان دانشمندان با استفاده از رویکرد شبکه استنادی. *کتابداری* ۴۰: ۹۷-۱۱۲.
- داورپناه، محمدرضا. ۱۳۹۵. *ارتباط علمی: نیاز اطلاعاتی و رفتار اطلاعاتی*. تهران: دبیرش؛ چاپار.
- د سلووا، پالی یو. کی.، و کنديس کی. ونس. ۲۰۱۷. *ارتباطات علمی: تغییر چشم‌انداز*. ترجمه صالح رحیمی. ۱۳۹۸. کرمانشاه: دانشگاه رازی.
- دستیار، وحید، نادر رازقی، و مهدی فقیه‌عبداللهی. ۱۳۹۹. نقش رسانه‌های اجتماعی در ارتباطات علمی اعضای هیئت علمی دانشگاه مازندران و تأثیر آن بر تولیدات علمی آنان. *کتابداری و اطلاع‌رسانی* ۲۳ (۴): ۱۷۳-۱۹۶.
- دهقانپور، زهرا، لاله صمدی، و عبدالرضا نوروزی چاکلی. ۱۴۰۲. تحلیل رویکردهای کارکردی و ساختاری ارتباطات علمی در تولید دانش توسط پژوهشگران حوزه علم‌سنجی در ایران. *پژوهشنامه علم‌سنجی* ۹ (۱): ۲۸۶-۲۵۹.

رضادوست، کریم، عبدالرضا نواح و الهام ادیب‌زاده. ۱۳۹۶. بررسی عوامل هنجاری و سازمانی مؤثر بر میزان تولید علمی اعضای هیئت علمی: مورد مطالعه اعضای هیئت علمی دانشگاه شهید چمران اهواز. *مطالعات کتابداری و علم اطلاعات* ۹ (۲۰): ۲۳-۴۲.

رضی، مصطفی، جواد هاشم‌زاده، فریده عصاره، و مرتضی محمدی استانی. ۱۳۹۵. تبیین الگوی رفتار استنادی و تعیین میزان همکاری علمی پژوهشگران در مجله‌های علمی-پژوهشی زمین‌شناسی ایران. *مطالعات کتابداری و علم اطلاعات* ۸ (۱۷): ۶۱-۸۲.

ریاحی، عارف، و افشین موسوی چلک. ۱۳۹۹. بررسی عوامل مؤثر و موانع پیش روی همکاری‌های علمی بین‌المللی از دیدگاه اعضای هیئت علمی حوزه پرستاری. *مجله بالینی پرستاری و مامایی* ۹ (۳): ۷۵۳-۷۶۸. روزبهبانی، معصومه، و نصرت ریاحی‌نیا. ۱۳۹۵. شناسایی شاخص‌های اعتبار در ارتباطات علمی (مطالعه و استناد) از دیدگاه اعضای هیئت علمی دانشگاه خوارزمی. *تعامل انسان و اطلاعات* ۳ (۲): ۱-۱۱.

زرمهر، فاطمه، علی منصوری، و حسین کارشناس نجف‌آبادی. ۱۴۰۲. مقایسه عملکرد رویکردهای کشف و استخراج موضوعات کتاب‌های الکترونیک. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۸ (۴): ۱۳۶۹-۱۳۹۴. زره‌ساز، محمد. ۱۳۹۰. ارتباطات علمی به‌منزله یک نظام اطلاعاتی: جنبش دسترسی آزاد و تغییرات پارادایمی. *شمسه* ۳ (۱۲-۱۳): ۱-۱۷.

سلیمانی‌نژاد، عادل، مزده سلاجقه، و الهام طیبی. ۱۳۹۷. خوشه‌بندی مقالات علمی بر پایه الگوریتم k\_means مطالعه موردی: پایگاه پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک). *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۴ (۲): ۸۷۱-۸۹۶.

سهرابی، طیبه، و سعید غفاری. ۱۳۹۸. شناسایی موضوعات پرکاربرد تولیدات علمی در حوزه «ارتباطات علمی» با استفاده از روش تحلیل هم‌رخدادی واژگان. *پژوهشنامه علم‌سنجی* ۵ (۲): ۴۵-۶۲. سهیلی، فرامرز، فرشید دانش، و رحمت‌الله فتاحی. ۱۳۸۹. کاربردهای وب‌سنجی در سنجش ارتباطات علمی. *مطالعات ملی کتابداری و سازماندهی اطلاعات* ۲۳ (۱): ۱۴۸-۱۶۳.

شکوهیان، محبوبه، عاصفه عاصمی، احمد شعبانی، و مظفر چشمه‌سهرابی. ۱۳۹۸. ارائه مدل دسته‌بندی موضوعی تولیدات علمی حوزه سلامت با استفاده از روش‌های متن‌کاوی. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۵ (۲): ۵۵۳-۵۷۴.

شمسی، مهناز، و مریم سلیمانی. ۱۳۹۴. بررسی رابطه ارتباطات علمی و رسانه‌ها با تأکید بر رسانه‌های اجتماعی. *کتاب مهر* (۱۷ - ۱۸): ۶۲ - ۷۹.

علیدوستی، سیروس، محمود خسروجردی، و بهزاد دوران. ۱۳۸۸. *مدیریت ارتباطات علمی*. تهران: پژوهشگاه اطلاعات و مدارک علمی ایران؛ چاپار.

گیانی، فرناز، نوید نظافتی، و سجاد شکوهیار. ۱۳۹۴. خوشه‌بندی کاربران داده‌های دریایی با استفاده از تکنیک داده‌کاوی. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۰ (۴): ۱۰۲۵-۱۰۴۹.

- کفاشان کاخکی، مجتبی، و محسن خلیلی. ۱۳۹۸. نقش ارتباط علمی در فرایند تولید دانش در علوم انسانی. *پژوهشنامه کتابداری و اطلاع‌رسانی* ۹ (۱): ۱۰۷-۱۲۳.
- گلینی مقدم، گلنسا، میترا صمعی، و مریم کمالی سروستانی. ۱۳۹۷. نقش رسانه‌های اجتماعی در ارتباطات علمی اعضای هیئت علمی دانشگاه علامه طباطبائی. *مطالعات رسانه‌های نوین* ۴ (۱۵): ۳۲۲-۳۵۱.
- مرادویسی، ژاله، سیروس پناهی، و شهرام صدقی. ۱۴۰۰. نقش شبکه‌های اجتماعی در ارتباطات علمی دانشجویان و اعضای هیئت علمی دانشگاه علوم پزشکی ایران. *مدیریت سلامت* ۲۴ (۴): ۴۷-۵۷.
- منصوری، رضا. ۱۳۹۵. ارتباطات علم و ترویج علم. *ترویج علم* ۷ (۱): ۵-۶.
- مهرآور گیگلو، شهرام، علی خورسندی طاسکوه، سعید غیائی ندوشن، و عباس عباسی. ۱۳۹۹. شناسایی عوامل بازدارنده همکاری‌های علمی بین‌المللی در نظام آموزش عالی کشور و ارائه راهکار: مطالعه‌ای کیفی. *پژوهش در آموزش علوم پزشکی* ۱۲ (۳): ۶۸-۷۸.
- مهرابی، گلپر، عباس عباس‌پور، علی دلاور، و علی خورسندی طاسکوه. ۱۴۰۱. شناسایی عوامل مؤثر بر ارتقای کیفیت پژوهش دانشگاهی در حوزه علوم انسانی و اجتماعی. *پژوهش در نظام‌های آموزشی* ۱۶ (۵۶): ۵-۲۱.
- یلوه، الهام، و یعقوب نوروزی. ۱۴۰۰. مروری نظام‌مند بر پژوهش‌های بهبود الگوریتم کایانه برای خوشه‌بندی داده‌ها. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۷ (۲): ۵۲۷-۵۵۶.

## References

- Abbasi-Moud, Z., H. Vahdat-Nejad, and J. Sadri. 2021. Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Systems with Applications* 167, Article 114324. <https://doi.org/10.1016/j.eswa.2020.114324>
- Aburass, S., O. Dorgham, and J. Al-Shaqsi. 2023. A hybrid machine learning model for classifying gene mutations in cancer using LSTM, BiLSTM, CNN, GRU, and GloVe. <https://arxiv.org/abs/2307.14361> (accessed Jan. 3, 2024)
- Al-Daihani, S. M., J. S. Al-Qallaf, & S. A. Al-Saheeb. 2018. Use of social media by social science academics for scholarly communication. *Library Review* 67 (6-7): 412-424.
- Buitrago Ciro, J. 2022. How are academic libraries in Spanish-speaking Latin America responding to new models of scholarly communication and predatory publishing? *Journal of Librarianship and Information Science* 54 (3): 373-388. <https://doi.org/10.1177/09610006211016533>
- Fährnich, B. 2021. Conceptualizing science communication in flux: A framework for analyzing science communication in a digital media environment. *JCOM* 20 (03): Y02. <https://doi.org/10.22323/2.20030402>
- Gherab Martín, K. J. 2021. The dual function of open access scholarly communication: An arXiv case study. *Techno Review. International Technology, Science and Society Review* 10 (2): 199-211. <https://doi.org/10.37467/gkarevtechno.v10.3196>
- Gozetti, P. 2006. Institutional repositories in scholarly communication: A literature review on models, issues and current trends. MSc. in International Information Studies, University of Parma. <https://core.ac.uk/download/pdf/41181177.pdf> (accessed Jan. 21, 2024)
- Gullbekk, E. 2016. Apt information literacy? A case of interdisciplinary scholarly communication. *Journal of Documentation* 72 (4): 716-736. <https://doi.org/10.1108/JDOC-08-2015-0101>

- Kapellas, N. and S. Kapidakis. 2023. Event detection in news articles: A hybrid approach combining topic modeling, clustering, and named entity recognition. In *Proceedings of the 15<sup>th</sup> International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2023)* – V. 2: KEOD, 272-279, Science and Technology Publications. doi:10.5220/0012234300003598
- Kappel, K. and S. J. Holmen. 2019. Why science communication, and does it work? A taxonomy of science communication aims and a survey of the empirical evidence. *Frontiers in Communication* 4 (55). doi: 10.3389/fcomm.2019.00055
- Khoiriah, Kh., A. Suyatna, A. Abdurrahman, and T. Jalmo, T. 2023. Reviewing of Indonesian students' scientific communication skills: A structural equation modeling analysis. *International Journal of Evaluation and Research in Education (IJERE)* 12 (1): 292-301.
- Khosrowjerdi, M. 2011. Designing a viable scientific communication model: VSM approach. *Library Hi Tech* 29 (2): 359-372.
- Kuhn, A., S. Ducasse and T. Girba. 2007. Semantic clustering: Identifying topics in source code. *Information and Software Technology* 49 (3): 230-243. <https://doi.org/10.1016/j.infsof.2006.10.017>
- Kuzminska, O. 2021. Selecting tools to enhance scholarly communication through the life cycle of scientific research. *Educational Technology Quarterly* 2021 (3): 402–414.
- Ma, W., X. Tu, B. Luo, and G. Wang. 2022. Semantic clustering-based deduction learning for image recognition and classification. *Pattern Recognition* 124, Article 108440. <https://doi.org/10.1016/j.patcog.2021.108440>
- Martin-Yeboah, E., A. A. Alemna, & E. Adjei. 2018. Scholarly Communication via Institutional Repositories: A Ghanaian Perspective. *Library Philosophy and Practice*. 2024. <https://digitalcommons.unl.edu/libphilprac/2024/> (accessed Jan. 18, 2024)
- Nosek, B. A., & Y. Bar-Anan. 2012. Scientific utopia: I. Opening scientific communication. *Psychological Inquiry* 23 (3): 217-243.
- Salloum, S.A., M. Al-Emran, A. A. Monem, and K. Shaalan. 2018. Using text mining techniques for extracting information from research articles. In *Intelligent Natural Language Processing: Trends and Applications. Studies in Computational Intelligence*, K. Shaalan, A. Hassanien, F. Tolba. (eds), 373-397, Springer, Cham. [https://doi.org/10.1007/978-3-319-67056-0\\_18](https://doi.org/10.1007/978-3-319-67056-0_18)
- Santhiappan, S., N. Shravan, and B. Ravindran. 2023. CIAMS: Clustering indices-based automatic classification model selection. *International Journal of Data Science and Analytics*, 1-23. <https://doi.org/10.1007/s41060-023-00441-5>
- Santos, G., M. T. Valente, and N. Anquetil. 2014. Remodularization analysis using semantic clustering, In *2014 Software Evolution Week - IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering (CSMR-WCRE)*, Antwerp, Belgium, 224-233, IEEE. doi: 10.1109/CSMR-WCRE.2014.6747174
- Shehata, A. M.K. and M. A. M. Eldakar. 2023. Scholarly communication practices and attitudes of Egyptian junior researchers: An exploratory study. *The Journal of Academic Librarianship* 49 (6). <https://doi.org/10.1016/j.acalib.2023.102811> (accessed 6 Jan 2024).
- Xia, J. 2006. Scholarly communication in east and Southeast Asia: Traditions and challenges. *IFLA Journal* 32 (2): 104-112.
- Xiao, J. 2024. Integrating digital literacies and scientific communication in a multimedia anatomy group assignment to advance contextual learning. *Anatomical Sciences Education* 17 (1): 55–65. <https://doi.org/10.1002/ase.2331>
- Xiaoyan, L., R. C. Raga, R. C., and S. Xuemei. 2022. GloVe-CNN-BiLSTM model for sentiment analysis on Text Reviews. *Journal of Sensors*, Article ID 7212366. <https://doi.org/10.1155/2022/7212366> (accessed 2 Jan 2024).

### رجب کیانی شاهوندی

متولد ۱۳۵۳، دانشجوی دکتری رشته علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است.

بازیابی اطلاعات، هستی‌شناسی، وب معنایی، داده‌کاوی و یادگیری ماشین، الگوریتم‌های هوش مصنوعی، برنامه‌نویسی رایانه، مدیریت و طراحی پایگاه داده و یادگیری الکترونیکی از جمله علایق پژوهشی وی است.



### احمد شعبانی

متولد ۱۳۳۵، دارای مدرک دکتری علم اطلاعات و دانش‌شناسی است. ایشان هم‌اکنون استاد گروه علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است. مدیریت اطلاعات و دانش، روش تحقیق و حوزه کتابخانه‌های عمومی از جمله علایق پژوهشی وی است.



### عاصفه عاصمی

متولد ۱۳۴۹، دارای مدرک دکتری علم اطلاعات و دانش‌شناسی از دانشگاه پونا هندوستان و دکتری بیزینس اینفورماتیک از دانشگاه کورویونس بوداپست است. ایشان هم‌اکنون عضو هیئت علمی و محقق دانشگاه کورویونس بوداپست است. سیستم‌های اطلاعاتی هوشمند، تجارت هوشمند، کریپتوکارسی، سیستم‌های پیشنهاددهنده سرمایه‌گذاری و هوش مصنوعی مولد در همکاری با بشر از جمله علایق پژوهشی وی است.



### مرتضی محمدی استانی

متولد ۱۳۶۴، دارای مدرک دکتری علم اطلاعات و دانش‌شناسی از دانشگاه اصفهان است. ایشان هم‌اکنون استادیار گروه علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است.



سازماندهی اطلاعات و دانش، هستی‌شناسی، وب معنایی، فراداده، مدیریت دانش و کتابخانه‌های عمومی از جمله علایق پژوهشی وی است.

پژوهش نامه  
پژدازش و  
مدیریت  
اطلاعات

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
رتال جامع علوم انسانی