


Comparison of the Textual Similarity of Representation Elements (Title, Abstract, and Keyword) of Articles in the Citation Network of References with a Research Proposal


Sanaz Rahrovani¹

 1. Ph.D Candidate in Knowledge & Information Science, Shiraz University, Shiraz, Iran.
Email: Sanaz.rahrovani@gmail.com


Javad Abbaspour^{2*}

 2. Associate Professor, Department of Knowledge & Information Science, Faculty of Education & Psychology, Shiraz University, Shiraz, Iran. (Corresponding Author)

Mahdieh Mirzabeigi³

 3. Associate Professor, Department of Knowledge & Information Science, Faculty of Education & Psychology, Shiraz University, Shiraz, Iran.
Email: mmirzabeigi@gmail.com

Hamid R. Jamali⁴

 4. Associate Professor, School of Information Studies, Charles Sturt University, Wagga Wagga, Australia.
Email: h.jamali@gmail.com

Email: javad.abbaspour@gmail.com

Abstract

Date of Reception:
07/05/2023

Date of Acceptation:
01/09/2023

Purpose: The current research compares the representative elements (title, abstract, and keywords) of the articles that existed in the proposal references' citation network with the proposals' elements. The other goal of this research is to calculate representative elements' weighted average (title, abstract, and keywords) from a textual similarity perspective.



Methodology: This is an applied and quantitative research that uses citation analysis and content analysis. The research sample is 3019 articles extracted from the citation network of 31 graduated students' proposals (M.Sc. and Ph.D.) in Chemistry at Shiraz University. All English articles' titles in the proposals' references were searched on the Web of Science database, and each article's file and all articles' files in its citation network were saved in Excel format. All retrieved files were merged into one file and sorted based on citation count to have the unit citation network for each user's proposal. Because some of the proposals had an extended citation network with more than a thousand articles, 100 articles with the greatest citation count of each network were analyzed to create uniformity and balance among the proposals' citation networks. Next, the scale of textual similarity of 100 articles' representative elements with the greatest citation count in the citation network, was calculated with the proposal's title, the proposal's text, and the titles of the proposal's references. The scale of textual similarity was checked using designed software based on the Python programming language and measuring the cosine similarity.

*Sanaz Rahrovani*¹

Javad Abbaspour^{2*}

*Mahdieh Mirzabeigi*³

*Hamid R. Jamali*⁴

Date of Reception:
07/05/2023

Date of Acceptation:
01/09/2023



Findings: The results of the Kruskal-Wallis test showed that there was a significant difference between the articles' representative elements and the title, text, and references' titles of the proposals from a textual similarity viewpoint; and in all three cases articles' abstracts had the most textual similarity with the proposal elements, then, the title and keywords of the articles' citation network were in the second and third ranks; In addition, the representative elements' weighted average was calculated. The obtained value was 0.62 for the abstract, 0.5 for the title, and 0.22 for the keywords, respectively.

Conclusion: Despite the use of different platforms to measure the similarity between the documents searched and the documents desired by the user, there is still a distance to reach the ideal level. Until now, no research had used the representative elements of the articles that existed in the proposal references' citation network to measure the textual similarity with the proposal elements and had not evaluated their capability. The confirmation of textual similarity among the representative elements of the articles that existed in the proposal references' citation network with proposals' elements, indicates that the student's proposal can be used as a platform for recommending related articles. Hence, the designers of scientific recommender systems, scientific information retrieval systems, digital libraries, and scientific social networks such as LinkedIn, Academia, and ResearchGate can use the elements of articles' citation networks to recommend related articles. In addition, considering the articles' representative elements as independent units is important not only for similarity measurement but also for keyword expansion and suggesting the appropriate journal to the authors for publishing their articles. According to the determined weight of representative elements and to increase the efficiency of information systems, it is suggested that designers of such systems use the abstracts and the titles of the articles to measure the similarity and avoid calculating the similarity of the texts as a whole unit. This saves time, resources, and energy, presents better results, and users can reach their target and desired information more easily and faster than before. In addition, for indexing articles in databases and search engines, the articles' abstracts and titles can be prioritized to save financial resources and energy.

Keywords: citation network, proposal, dissertation, thesis, related articles, research references, textual similarity, cosine similarity, representative elements, title, abstract, keywords, weighted average of representative elements.

مقایسه عناصر بازنمونی (عنوان، چکیده و کلیدواژه‌ها) مقاله‌های شبکه استنادی فهرست منابع از نظر شباهت متنی با پیشنهاد پژوهش

ساناز رهروانی^۱

۱. دانشجوی دکتری علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه شیراز، شیراز، ایران.

Email: Sanaz.rahrovani@gmail.com

جواد عباس‌پور^{*۲}

۲. دانشیار علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه شیراز، شیراز، ایران. (نویسنده مسئول)

مهديه میرزابیگی^۳

۳. دانشیار علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه شیراز، شیراز، ایران.

Email: mmirzabeigi@gmail.com

حمیدرضا جمالی مهموئی^۴

۴. دانشیار علم اطلاعات و دانش‌شناسی، دانشگاه چارلز استوارت، واگاوگا، استرالیا.

Email: h.jamali@gmail.com

Email: javad.abbaspour@gmail.com

چکیده

هدف: پژوهش حاضر عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقاله‌های موجود در شبکه استنادی فهرست منابع پیشنهاد پژوهش (پروپوزال) را از نظر شباهت متنی با پیشنهاد پژوهش مقایسه می‌کند.

روش شناسی: این پژوهش از روش‌های تحلیل استنادی و تحلیل محتوایی استفاده می‌کند. نمونه پژوهش ۳۰۱۹ مقاله مستخرج از شبکه استنادی ۳۱ پیشنهاد پژوهش دانشجویان تحصیلات تکمیلی رشته شیمی دانشگاه شیراز است. میزان شباهت متنی عناصر بازنمونی ۱۰۰ مقاله دارای بیشترین میزان استناد که در شبکه استنادی بودند، با عنوان و متن اصلی و عنوان مقالات موجود در فهرست منابع پیشنهاد محاسبه شد. میزان شباهت متنی با کمک نرم‌افزاری که بر اساس زبان برنامه‌نویسی پایتون طراحی شده بود و شباهت کسینوسی را اندازه می‌گرفت، بررسی شد.

یافته‌ها: نتایج آزمون کروسکال والیس نشان داد میان عناصر بازنمونی مقالات شبکه استنادی با عنوان و متن اصلی و عنوان مقالات فهرست منابع پیشنهاد تفاوت معناداری وجود دارد و در هر سه مورد، چکیده مقالات شبکه استنادی بیشترین شباهت متنی را با عناصر پیشنهاد پژوهش دارد. به علاوه، میانگین وزنی شباهت عناصر بازنمونی شبکه استنادی با عناصر پیشنهاد پژوهش به ترتیب برای چکیده ۰.۶۲، عنوان ۰.۵ و کلیدواژه‌ها ۰.۲۲ به دست آمده است.

نتیجه‌گیری: تأیید وجود شباهت متنی میان عناصر بازنمونی مقالات موجود در شبکه استنادی فهرست منابع پیشنهاد پژوهش با پیشنهاد پژوهش، در کل حاکی از آن است که می‌توان از پیشنهاد پژوهش دانشجویان به‌عنوان بستی برای پیشنهاد مقالات مرتبط به آنها استفاده کرد. **واژگان کلیدی:** شبکه استنادی، پیشنهاد پژوهش (پروپوزال)، شباهت متنی، شباهت کسینوسی، فهرست منابع، عناصر بازنمونی.

صفحه ۴۴-۲۳

دریافت: ۱۴۰۲/۰۲/۱۷

پذیرش: ۱۴۰۲/۰۶/۱۰



مقدمه و بیان مسئله

در اغلب دانشگاه‌های جهان بخشی از دوره تحصیلات تکمیلی دانشجویان به پژوهش در قالب پایان‌نامه اختصاص می‌یابد. پایان‌نامه نخستین گام نظام‌مند در بعد پژوهش است که طی آن دانشجو یکی از موضوع‌های تخصصی را زیر نظر استاد (ان) راهنما و مشاور انتخاب، اجرا و ارائه می‌کند. پایان‌نامه اهداف مختلفی نظیر؛ پرکردن شکاف دانش، ارزیابی عملکرد، یافتن و مرتفع کردن چالش‌ها و پیشنهاد راهکارهای مؤثر برای بهبود امور را دنبال می‌کند.

معمولاً پیش از نگارش پایان‌نامه، پیشنهاد پژوهش^۱ (پروپوزال) نگاشته می‌شود. پیشنهاد پژوهش، شاکله و ساختار اولیه یک پژوهش است و اغلب شامل بخش‌های مقدمه، بیان مسئله، اهداف، اهمیت و ضرورت، مبانی نظری، پیشینه پژوهش و ابزار گردآوری و تحلیل داده‌ها و فهرست منابع است (حری، ۱۳۸۸). پژوهشگران به طور معمول و پیش از نگارش پیشنهاد پژوهش متون متعددی را مطالعه می‌کنند تا به اشراف نسبی در ارتباط با موضوع مورد پژوهش برسند و پس از نگارش پیشنهاد پژوهش، با کلیت پژوهشی که قرار است انجام دهند آشنا می‌شوند و در قالب فهرست منابع هم به برخی از متون مورد مطالعه و استفاده استناد می‌کنند (حری، ۱۳۸۸). بنابراین، در فهرست منابع پیشنهاد پژوهش مجموعه‌ای از منابع گرد آمده‌اند که حداقل بخشی از آنها با موضوع مورد پژوهش مرتبط است. از این رو، فهرست منابع پیشنهاد پژوهش می‌تواند تا حدودی نیازها و علائق پژوهشی پژوهشگر را با شرح و بسط نسبی منعکس کند. در فهرست منابع مدارک متنوعی نظیر مقاله مجلات علمی یا ارائه شده در کنفرانس‌ها و همایش‌ها، کتاب، پایان‌نامه، طرح پژوهشی و جز آن درج می‌شود و برای هر مدرک، به فراخور، اطلاعاتی چون عنوان، نام نویسنده/ نویسندگان، تاریخ نشر، محل نشر، نام ناشر و شماره صفحات و مواردی از این قبیل آورده می‌شود.

در صورتی که فهرست منابع پیشنهاد پژوهش را نماینده‌ای از علایق و نیازهای پژوهشی پژوهشگر، سابقه‌ای از روند مطالعات علمی او و در نهایت حاوی برخی از منابع مرتبط با کار پژوهش بدانیم (Arif, 2016; Sun et al., 2018) ، انتظار می‌رود شبکه استنادی که بر اساس منابع مندرج در فهرست منابع و پیوندهای میان آنان تشکیل می‌شود نیز این قابلیت را داشته باشد که برای ردیابی پیشینه پژوهش، تصویرسازی روند پژوهش و شناسایی سایر مقاله‌های مرتبط به کار گرفته شود^۲ (Alfraidi et al., 2015). البته متخصصان، برخی از محدودیت‌ها را هم برای استفاده از شبکه استنادی برای شناخت ارتباط میان مدارک برشمرده‌اند؛ برای نمونه، گفته می‌شود برخی از منابع مهم و مرتبط که به تازگی منتشر شده‌اند و هنوز فرصت کافی برای دریافت استناد پیدا نکرده‌اند و یا به هر دلیلی مورد استناد قرار نگرفته‌اند در شبکه استنادی قرار نخواهند گرفت (Kusumastuti et al., 2016; Bornmann & Mutz, 2015). اما، اینکه منابع مرتبطی با موضوع پژوهشگر در این شبکه استنادی یافت شود که پژوهشگر به هر دلیلی آنها را ندیده است یا از آنها غفلت کرده باشد محتمل به نظر می‌رسد.

در تأیید این گفته باید اشاره کنیم که دانشجویان دوره تحصیلات تکمیلی، به ویژه دانشجویان مقطع کارشناسی ارشد، تجربه پژوهش رسمی ندارند. آنها علی‌رغم آگاهی نسبی از نیاز اطلاعاتی خود و تمایل به استفاده از مدارک مرتبط با منابع پیشنهاد پژوهش، به دلایلی چون موانع زبانی، عدم تشخیص کامل و جامع نیاز اطلاعاتی خویش،

۱. قابل ذکر است که در این مقاله، کلمات «پروپوزال»، «پیشنهاد پژوهش» و «پیشنهاد» به جای یکدیگر به کار رفته‌اند و هر سه آنها به یک مفهوم واحد اشاره می‌کنند.

۲. منظور از شبکه استنادی فهرست منابع، مجموعه مدارکی است که به منابع موجود در فهرست منابع استناد کرده‌اند و یا از آنها استناد دریافت کرده‌اند.

اشراف‌نداشتن به موضوع در زمان نگارش پیشنهاد پژوهش، ناتوانی در استفاده از عبارت جستجوی مناسب برای بازیابی مدارک و ناآشنایی با جستجوی پیشرفته و امکانات دیگر جستجو، ممکن است نتوانند به همه مدارک مرتبط دست یابند (رحمان پور و همکاران، ۱۳۹۶؛ Luu & Kim, 2012; Fonseca et al., 2005). علاوه بر این، دانشجویان پس از نگارش پیشنهاد پژوهش، فرصت چندانی برای نگارش پایان‌نامه و یافتن اطلاعات مرتبط با موضوع پژوهش ندارند. ضمن اینکه، ناآگاهی آنها از هر پژوهش مرتبط انجام شده با موضوع مشابه، می‌تواند تبعات و آثار نامطلوبی از جمله انجام پژوهش تکراری، عدم اطلاع از یافته‌ها و نتایج سایر پژوهش‌ها و روش‌ها و ابزارهای جدید به همراه داشته باشد؛ از این رو، ضروری است که دانشجویان بتوانند به کلیه مقالات مرتبط با موضوع پیشنهاد پژوهش با سهولت بیشتر و در زمان کمتر دسترسی داشته باشند.

در همین راستا، پژوهش‌هایی به منظور بازیابی بهتر منابع اطلاعاتی و پیشنهاد مقالات مرتبط به کاربران انجام شده است. برخی پژوهشگران به معرفی و ارائه ابزارها، فنون و سنج‌های جدید برای سنجش شباهت میان مدارک با یکدیگر و یا عبارت جستجوی کاربران با هدف شباهت‌سنجی بهتر مدارک و یافتن منابع مرتبط با نیاز اطلاعاتی کاربران پرداختند (Peng et al., 2008; Renu & Mocko, 2016; Kerzendorf, 2019; Farouk, 2019; Farouk, 2020). در این میان، پژوهشگران بسترهای متعددی نظیر مقالات و عناصر موجود در آن، جستجوهای قبلی کاربران، نظرات و علایق آنها و کلیدواژه‌های مورد جستجو را برای سنجش شباهت میان مدارک مورد نظر با مفاهیم مورد جستجو در نظر گرفتند (Porcel et al., 2012; Beel et al., 2017)؛ اما بر اساس بررسی‌های انجام شده پیشنهاد پژوهش و به طور خاص فهرست منابع آن چندان مورد توجه قرار نگرفته است و تنها پابرجا و همکاران (۱۴۰۰) با بررسی شبکه استنادی منابع موجود در پیشنهاد پژوهش دانشجویان تحصیلات تکمیلی دریافتند که کاربران، مقاله‌های مرتبط پیشنهادی از طریق شبکه استنادی منابع پیشنهاد پژوهش را اثربخش دانسته‌اند.

اهمیت این پژوهش از چند بعد قابل بررسی است: نخست آنکه علی‌رغم به کارگیری بسترهای مختلف برای سنجش شباهت میان مدارک مورد جستجو و مدارک مورد نظر کاربر، هنوز تا رسیدن به حد آرمانی فاصله وجود دارد؛ بدین معنی که می‌بایست بسترهای مختلف کنکاش شود تا بتوان دریافت کدام بستر دارای قابلیت بهتری برای سنجش شباهت و ارائه مقاله‌های مرتبط به کاربر دارد؛ در همین راستا، در این پژوهش، عناصر بازنمونی موجود در شبکه استنادی فهرست منابع، بستری است که برای سنجش شباهت مورد توجه قرار گرفت. با تحلیل عناصر بازنمونی موجود در شبکه استنادی فهرست منابع و تأیید شباهت این عناصر با پیشنهاد پژوهش کاربران، می‌توان به بازیابی مقاله‌های مرتبط با پیشنهاد پژوهش دانشجویان و پژوهشگران کمک کرد و مقالات مرتبطی را از این طریق به آنها پیشنهاد داد. افزون بر این، با تعیین وزن هر یک از عناصر بازنمونی موجود در شبکه استنادی فهرست منابع پیشنهاد پژوهش دانشجویان، می‌توان از میانگین وزنی این عناصر برای تحلیل مقاله‌ها در سامانه‌های بازیابی اطلاعات استفاده کرد و با هدف صرفه‌جویی در زمان و انرژی، عناصر با میانگین وزنی بیشتر را با وزن بیشتری در الگوریتم سامانه‌ها لحاظ کرد.

با توجه به اهمیت این پژوهش و قابلیت شبکه استنادی منابع موجود در پیشنهاد پژوهش برای پیشنهاد مقاله‌های مرتبط، سؤال اصلی پژوهش حاضر بدین صورت است که از میان عناصر مختلف مقاله‌های موجود در شبکه استنادی، کدام یک پتانسیل بیشتری برای به کارگیری در سامانه‌های بازیابی اطلاعات برای سنجش شباهت و ارائه مقاله‌های مرتبط به دانشجویان دارد؟

پرسش‌های پژوهش

هدف پژوهش حاضر مقایسه عناصر بازنمونی (عنوان، چکیده و کلیدواژه‌ها) مقاله‌های موجود در شبکه استنادی فهرست منابع پیشنهاد پژوهش (پروپوزال) از نظر شباهت متنی با پیشنهاد پژوهش است؛ بنابراین، پرسش‌های زیر برای این پژوهش مفروض است:

۱. آیا اختلاف معناداری در شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقاله‌های شبکه استنادی با عنوان پیشنهاد پژوهش وجود دارد؟
۲. آیا اختلاف معناداری در شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقاله‌های شبکه استنادی با عنوان منابع پیشنهاد پژوهش وجود دارد؟
۳. آیا اختلاف معناداری در شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقاله‌های شبکه استنادی با متن پیشنهاد پژوهش وجود دارد؟
۴. میانگین وزنی شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقالات موجود در شبکه استنادی به چه میزان است؟

چارچوب نظری

برای سنجش شباهت میان مقالات پیشنهادی با مقالات مدنظر کاربران و یا عبارت جستجوی آنها، از روش‌ها و رویکردهای مختلفی در سامانه‌های بازیابی اطلاعات استفاده می‌شود. برخی از آنها برای شباهت‌سنجی از اطلاعات متنی نظیر عنوان، چکیده، اطلاعات کتابشناختی و کلیدواژه‌ها و برخی از اطلاعات استنادی موجود در مقالات استفاده می‌کنند. سنجش شباهت میان مقاله‌ها بر اساس منابع اطلاعاتی مختلف دارای مزایا و معایبی است؛ برای مثال، شباهت‌سنجی بر اساس اطلاعات متنی دارای مزایایی نظیر سهولت دسترسی به این نوع اطلاعات است؛ در مقابل، استفاده از این نوع اطلاعات، مسائلی را در پی خواهد داشت؛ برای نمونه، یکی از چالش‌های استفاده از شباهت‌سنجی مقاله بر اساس عناوین مشابه، این است که ممکن است مقاله‌هایی بازیابی شوند که کلمات مشترکی با مقاله هدف داشته باشند؛ ولی از نظر معنایی به آن مقاله نزدیک نباشند و یا مقاله‌هایی که از نظر محتوایی، با مقاله هدف نزدیک‌اند ولی عنوان مشابهی با آن مقاله ندارند، بازیابی نشوند (Küçüktunç et al., 2012). چالش دیگر این است که تفسیر متون برای کامپیوتر جهت سنجش شباهت دشوار است و درنهایت، امکان فیلترکردن و اضافه و کم کردن مواردی نظیر کیفیت، نوع و یا نقطه نظر خاصی برای بهبود جستجو در این نوع شباهت‌سنجی وجود ندارد (Komkhao et al., 2013; Li et al., 2015). برای سنجش شباهت متنی میان اجزای مختلف متن از سنج‌ها و معیارهای متفاوتی استفاده می‌شود؛ یکی از این سنج‌ها سنج کسینوسی است؛ سنج‌ای که در مدل برداری برای تعیین میزان شباهت پرسش‌ها و مدارک استفاده می‌شود. این سنج میزان شباهت را با اندازه‌گیری کسینوس زاویه بین بردار پرسش و مدارک و یا دو مدارک با هم محاسبه می‌کند (گراسمن و فریدر، ۱۳۸۴). بر اساس این سنج، میزان شباهت متنی عددی بین صفر و یک تعیین می‌شود که هرچه این عدد به یک نزدیک‌تر باشد، میزان شباهت میان عناصر مختلف بیشتر است.

از طرفی، استفاده از سنج‌های شبکه استنادی دارای برتری‌هایی چون مستقل بودن آنها از زبان است؛ بدین معنی که مشکلات هم‌نویسه‌ها و ابهام معنایی را ندارد، پویا هستند، نیاز به قضاوت متخصصان موضوعی ندارند و معمولاً، منابع دارای پیوند استنادی نسبت به منابعی که به صورت معمول نمایه‌سازی شده‌اند، آسان‌تر بازیابی می‌شوند (Eto, 2013).

درواقع، یکی از پیش فرض های اصلی تحلیل استنادی که بنیان اصلی استناد قرار گرفته این است که میان متن (استنادکننده) و سند (استنادشده) نوعی رابطه وجود دارد؛ و متن کم و بیش به همان موضوعی می پردازد که سند به آن پرداخته است (Harter & Nisonger, 1993). هر فردی که در مورد موضوعی جست و جو می کند، فقط با دانستن یکی از مقاله هایی که به آن استناد شده است، می تواند یک یا چند مقاله دیگر را بیابد و هر مقاله به نوبه خود فهرستی از استنادهای جدید به دست می دهد که می توان جست و جو را با آن پی گرفت (موئند، ۱۳۸۷). کاربرد روزافزون نمایه های استنادی شاهدهی بر مفیدبودن پیوندهای استنادی به عنوان نشانه ای از ارتباط مدارک با یکدیگر است (Bichteler & Eaton, 1980)؛ اما درعین حال دارای معایبی نظیر داشتن حالت دودویی است؛ بدین معنی که نمی توان بین پیوندهای نزدیک و دور تمایز قائل شد. همچنین این سنج، بافتار استنادی را در نظر نمی گیرد و فقط به شمارش استنادهای منابع بسنده می کند (Eto, 2013).

یکی از سنج های مورد استفاده در سنجش شباهت استنادی شمار استناد است. شمار استناد به تعداد استنادهایی اشاره دارد که چند منبع به یک منبع خاص داده اند. با در نظر گرفتن اینکه اساس اندیشه تحلیل استنادی بر این فرضیه استوار است که بین مدرک استنادکننده و مدرک مورد استناد، نوعی رابطه محتوایی وجود دارد؛ می توان نتیجه گرفت که منابع با شمار استناد بیشتر در یک موضوع خاص، با آن موضوع مرتبط ترند و دارای اعتبار بیشتری هستند، پس شمار استناد می تواند یکی از سنج های مهم شباهت استنادی تلقی شود.

با توجه به مزایای استفاده از اطلاعات متنی و شبکه استنادی منابع بهتر است سنج های شباهت سنجی هر دو مورد به صورت هم زمان با هم به کار گرفته شوند تا چالش های همدیگر را جبران کنند. با توجه به اینکه مسائل مربوط به شبکه استنادی در حوزه علم سنجی نیز مطرح می شود و تاکنون نقش مهمی در شباهت سنجی میان مدارک داشته است (Xia et al., 2016)؛ لذا استفاده از شبکه استنادی فهرست منابع پیشنهاد و مقایسه عناصر بازنمونی موجود در آن با پیشنهاد پژوهش از نظر سنجش شباهت متنی می تواند در مجموعه کاربردهای حوزه علم سنجی قرار بگیرد.

پیشینه پژوهش

پژوهش صانعی نژاد و همکاران (۱۳۹۵) با هدف افزایش دقت در انتخاب اقلام پیشنهادی، دو روش مبتنی بر محتوا و پالایش مشارکتی در گراف داده ای را با یکدیگر ترکیب کردند و الگویی جدید جهت سنجش شباهت ارائه کردند. این الگو میزان شباهت میان موجودیت ها را با ترکیب الگوریتم های ساختاری و غیرساختاری محاسبه می کرد و پیشنهادها را بر اساس بیشترین میزان مشابهت به کاربر ارائه می داد. الگوی پیشنهادی بر روی مجموعه داده فروشگاه اینترنتی آی هدیه ارزیابی شد. حمیدزاده و صادق زاده (۱۳۹۶) نیز روش جدیدی ارائه کردند که در آن رویکرد پالایش مشارکتی و مجموعه فازی ناهموار با یکدیگر ترکیب و شباهت بر اساس فاصله اقلیدسی محاسبه می شد. این روش ترکیبی، با استفاده از مجموعه های فازی ناهموار، صحت پیش بینی را افزایش داد. آزمایش ها نشان داد که روش پیشنهادی، میزان صحت را نسبت به روش های قبلی افزایش داده بود. پژوهش پابرجا و همکاران (۱۴۰۰) نیز به امکان سنجی استفاده از شبکه استنادی منابع موجود در پیشنهاد پژوهش دانشجویان تحصیلات تکمیلی برای پیشنهاد مقاله های مرتبط پرداختند و دریافتند که کاربران، مقاله های مرتبط پیشنهادی از طریق شبکه استنادی منابع پیشنهاد پژوهش را اثربخش دانسته اند.

برخی پژوهش ها نیز به تحلیل علم سنجی مقالات و برون داده های علمی پرداختند و از پایگاه هایی نظیر وب آو

ساینس^۱ و اسکوپوس استفاده کردند؛ برای مثال، ملکی و همکاران (۱۳۹۸) رابطه میان رتبه آموزشی کتاب‌های درسی دانشگاه‌های برتر جهان را با شاخص‌های استناد، پیچ رنگ و هیتس با کمک تحلیل استنادی تعیین کردند. یافته‌ها نشان داد بین رتبه آموزشی کتاب‌ها با سنجه استناد رابطه ضعیفی وجود دارد و با رتبه پیچ ارتباط معنادار متوسطی وجود دارد. صراطی شیرازی (۱۴۰۰) ارتباط میان فناوری و علم بر اساس استناد به پروانه‌های ثبت اختراع در هم‌انتشاری دانشگاه و صنعت در نمایه گسترش یافته علوم پایگاه وب آو ساینس در سال ۲۰۱۸ را مطالعه کرد و از روش مطالعه استنادی معکوس بهره برد. وی دریافت ارتباطی که باید میان فناوری و علم بر اساس استناد به پروانه‌های ثبت اختراع در مدارک علمی برقرار باشد، ارتباط قوی نبوده است. محمدزاده و همکاران (۱۴۰۰) به توصیف مقاله‌های پراستناد ایرانی و ترسیم و تحلیل شبکه‌های اجتماعی و کشف الگوهای آنها در مقالات پراستناد ایرانی با رویکرد علم‌سنجی پرداختند. یافته‌ها نشان داد روند تولیدات علمی مقالات پراستناد روند صعودی داشته است و حوزه‌های مهندسی، شیمی، مکانیک و همچنین دانشگاه آزاد اسلامی و دانشگاه تهران بیشترین سهم را در تولید مقالات پراستناد داشته‌اند. اکبری نیسانی و همکاران (۱۴۰۰) جایگاه وزنی مقالات علمی تولیدشده در انستیتوکانسر دانشگاه علوم پزشکی تهران را با استفاده از نرمال‌سازی مقالات بر اساس معیار سنجش استناد، تعیین کردند و دریافتند مقالات انستیتوکانسر روند رو به رشدی طی سال‌های ۱۹۶۵ تا ۲۰۱۸ داشته است و بیشترین پژوهش‌ها در حوزه سرطان پستان، مری، معده، جراحی سرطان و رادیوتراپی منتشر شده است. خلیلی و محمدی (۱۴۰۰) دو مجله ایرانی حوزه علم اطلاعات و دانش‌شناسی (مجله علوم اطلاع‌رسانی و مدیریت اطلاعات و وب‌لوژی) را که به زبان انگلیسی منتشر شده‌اند، بر اساس شاخص‌های هم‌تألفی و رخداد واژگانی تحلیل کردند. نتایج نشان داد اگرچه بیشتر مقالات دو مجله توسط پژوهشگران ایرانی نوشته شده بود، اما تعداد استناد دریافتی مقالات پژوهشگران خارجی بیشتر بود و الگوی نویسندگی مقالات به ترتیب دو، یک و سه نویسنده بود. عبدالمهی و نقشینه (۱۴۰۱) به تحلیل موضوعی استنادات و ارجاعات به پروانه‌های ثبت اختراع بین‌المللی ایران با استفاده از روش تحلیل استنادی و تحلیل روابط استنادی معکوس پرداختند. نتایج نشان داد بیشتر پروانه‌های ثبت اختراع بین‌المللی ایران متعلق به رده ضروریات انسانی بودند و بیشترین ارجاعات را به رده فیزیک و حوزه‌های بیوشیمی، ژنتیک و زیست‌شناسی مولکولی داشتند. نجفی برازجانی و همکاران (۱۴۰۱) به بررسی تطبیقی مقاله‌های نمایه‌شده پژوهشگران کشورهای خاورمیانه در حوزه دیابت در پایگاه وب آو ساینس با استفاده از شاخص‌های نفوذ اندیشه‌ای و نفوذ اجتماعی پرداختند و دریافتند بر اساس مجموع امتیازهای شاخص‌های نفوذ اندیشه‌ای و نفوذ اجتماعی، کشورهای رژیم صهیونیستی، عربستان سعودی، ترکیه، ایران و مصر رتبه نخست تا پنجم را کسب کردند. علی‌نژاد چمازکتی و میرحق‌جو لنگرودی (۱۴۰۱) به بررسی برون‌دادهای علمی در نشریات ایران و ترکیه در پایگاه وب آو ساینس پرداختند و الگوی رفتار استنادی نویسندگان تبیین و نقشه هم‌نویسندگی آنان را ترسیم کردند. یافته‌ها نشان داد در نشریات ایران ۵۰ درصد مدارک توسط نویسندگان ایرانی و در نشریات ترکیه ۶۰ درصد مدارک توسط نویسندگان ترک تولید شده است.

پژوهش‌های انجام‌شده در خارج به چند دسته تقسیم می‌شوند: برخی پژوهشگران در پژوهش‌های خود به ارائه روش‌های نوین شباهت‌سنجی میان مدارک با هم و عبارات مورد جستجوی کاربران با مدارک موجود در سامانه‌ها پرداخته‌اند؛ در مواردی سنج‌های شباهت‌سنجی را با یکدیگر ترکیب کرده‌اند و یا در پی ارائه ابزارهای جدید شباهت‌سنجی بوده‌اند؛ برخی دیگر، با در نظر گرفتن عناصر بازنمونی مدارک به‌عنوان عناصر مستقل، به شباهت‌سنجی

میان مدارک پرداخته اند و برخی از سنجه‌های شباهت کسینوسی در سامانه‌های پیشنهاددهنده و طبقه‌بندی متون استفاده کردند.

از جمله پژوهش‌هایی که با استفاده از روش‌های نوین به شباهت‌سنجی پرداخته‌اند می‌توان به پنگ و همکاران (Peng et al., 2008) اشاره کرد. آنها تلاش کردند روش جدیدی را برای محاسبه شباهت متنی در متون چینی پیشنهاد کنند؛ به این منظور، ابتدا متن را به فضای برداری کلمات تبدیل کرده، و سپس کلمات را به مجموعه‌ای از مفاهیم تقسیم کردند؛ در نهایت، با کمک ضرب برداری میان مفاهیم، شباهت را به دست آوردند. یافته‌ها نشان داد که میزان مانعیت افزایش یافته است؛ همچنین، رنو و موکو (Renu & Mocko, 2016) از پتانسیل الگوریتم‌های شباهت متنی به منظور بازیابی و اشتراک دانش برنامه‌های مبتنی بر متن استفاده کردند. نتایج حاکی از آن بود که ضریب ژاکارد مشابه قضاوت انسانی در حوزه سنجش شباهت متن عمل می‌کند و نتایج دقیق تری نسبت به روش هم‌پوشانی کلمات، TF-IDF و تحلیل معنایی پنهان به دست می‌دهد.

برخی دیگر از پژوهشگران رویکردها و سنجه‌های شباهت را با هدف دستیابی به سنجه‌های بهینه با یکدیگر ترکیب کردند؛ برای مثال، منزر (Menczer, 2004) مقیاس‌های شباهت متنی، پیوندی و معنایی را در صفحات وب بررسی کرد. او از هر یک از ۱۵ شاخه فرعی راهنمای Open Directory (OD) تعداد ده هزار آدرس وب انتخاب و با استفاده از مقیاس‌های شباهت با یکدیگر مقایسه کرد. نتایج بیانگر این نکته بود که هرچند همبستگی میان مقیاس‌های مختلف اندک بود ولی با توجه به حجم بسیار زیاد داده‌های مورد استفاده، قابل اعتنا بود. وی همچنین نتیجه گرفت که ترکیب سنجه‌های شباهت محتوایی با پیوندی برای بازیابی و رتبه‌بندی صفحات وب می‌تواند حائز اهمیت باشد گرچه ممکن است با گزارش‌های کاذب منفی و مثبتی نیز همراه باشد. چیخی و همکاران (Chikhi et al., 2008) اطلاعات متنی و استنادی را با هدف کشف موضوعات علمی ترکیب کردند و معتقد بودند استفاده از استنادها به غنی‌سازی محتوای متنی مدارک کمک می‌کند. نتایج آزمایش بر روی دو مجموعه واقعی نشان داد عملکرد این روش نسبت به سایر روش‌های ترکیب اطلاعات متنی و استنادی بهتر است. همدانی و همکاران (Hamedani et al., 2013) نیز برای بررسی شباهت میان مقالات علمی از ترکیب معیارهای متنی و پیوندی استفاده کردند. در روش پیشنهادی آنها محتوا و استنادهای علمی به صورت هم‌زمان در نظر گرفته می‌شد. نتایج آزمایش روی داده‌های واقعی دقت را نسبت به روش‌های قبلی ۲۰ درصد بهبود داده بود. فاروک (Farouk, 2019) نیز ابتدا رویکردهای مختلف اندازه‌گیری شباهت جملات را بررسی و با هدف یافتن بهترین روش، آنها را به سه گروه شباهت کلمه به کلمه، شباهت مبتنی بر ساختار و شباهت مبتنی بر بردار تقسیم کرد. هر روش، شباهت میان متون کوتاه را از منظر خاصی می‌سنجید. یافته‌ها نشان داد ترکیب رویکردها نتایج بهتری داشت. او براساس همین نتایج، در سال ۲۰۲۰ به منظور محاسبه شباهت میان جملات با یکدیگر معیار شباهت کلمه به کلمه را با تکنیک ساختار معنایی جمله ترکیب کرد و به این نتیجه رسید که لحاظ کردن ساختار جملات و مقاله‌ها هنگام محاسبه شباهت امری مطلوب است. شباهت به دست آمده در نتیجه اعمال ساختار جملات و مقاله‌ها حدود ۸۸ درصد شبیه به قضاوت انسان بود (Farouk, 2020).

برخی از پژوهشگران سامانه و ابزار جدیدی را با هدف سنجش شباهت میان مدارک ارائه کردند؛ برای مثال، بورگس و همکاران (Borges et al., 2012) سامانه بازیابی اطلاعات به نام آرگوسرچ^۱ را ارائه کردند که کیفیت جستجو را در سامانه‌های مدیریت کتابخانه بهبود می‌بخشید. رتبه‌بندی نتایج این سامانه بر اساس معیارهای ربط بود که

1. ARGOSearch

این کار با استفاده از شباهت متنی و اطلاعات زمینه‌ای نظیر پروفایل کاربر و اطلاعات آماری صورت می‌گرفت. در این سامانه، جامعیت نتایج بازیابی شده افزایش یافته بود. همدانی و همکاران (Hamedani et al., 2016) برای محاسبه شباهت میان مقالات علمی روش جدیدی به نام سیم‌سی‌سی^۱ پیشنهاد کردند که محتوا و استنادهای مقالات را به صورت هم‌زمان دربرمی‌گرفت. آنها میزان شباهت متنی و استنادی را با استفاده از رتبه ماشینی بردار پشتیبانی^۲ با یکدیگر ترکیب می‌کردند. یافته‌های آنها روی داده‌های واقعی مقالات علمی نشان داد این روش نسبت به روش‌های پیشین بیش از ۲۰ درصد مانعیت را بهبود داده است. گوما و فامی (Gomaa & Fahmy, 2017) نیز با توجه به قابلیت سنج‌های متنی در بازیابی اطلاعات، پردازش زبان طبیعی، خلاصه‌سازی متن و طبقه‌بندی متون، فنون سنجش شباهت متنی نظیر رشته، پیکره و رویکرد تشابه دانش را در ابزار واحدی جهت سنجش شباهت ترکیب کردند و آن را سیم‌آل^۳ نامیدند. این ابزار شامل انواع مختلف ماژول‌های پیش‌پردازش^۴ و سطوح شباهت بود و به منظور سنجش شباهت متنی به وجود آمد که از زبان عربی و انگلیسی پشتیبانی می‌کرد. کرزندورف (Kerzendorf, 2019) ابزاری برای پیشنهاد مقالات مرتبط در حوزه نجوم بر اساس محتوای متن مقاله ورودی ارائه کرد. با استفاده از این تکنیک مقالاتی بازیابی می‌شدند که از طریق استنادات و سامانه‌های اطلاعاتی تخصصی و مدارک موجود در فهرست منابع یافت نشده بودند.

برخی دیگر از پژوهشگران برای محاسبه شباهت میان مقالات، هرکدام از عناصر بازنمونی مقالات را به صورت یک واحد مستقل در نظر گرفتند و برای هرکدام یک وزن قائل شدند. در این روش برای محاسبه میزان شباهت میان دو مقاله، عناصر بازنمونی همسان با یکدیگر مقایسه می‌شدند؛ یعنی، عناوین با هم، چکیده‌ها با هم و متن مقالات هم با هم. پژوهشگران دریافته‌اند که میزان شباهت حاصل از مقایسه جفت به جفت عناصر بازنمونی با یکدیگر، در مقایسه با حالتی که دو مدرک صرف‌نظر از عناصر بازنمونی و به صورت یک واحد کل در نظر گرفته می‌شود، تفاوت زیادی دارد؛ بنابراین، به زعم آنها استفاده از هر عنصر بازنمونی به صورت مجزا برای سنجش شباهت بسیار حائز اهمیت است (Yoon et al., 2011)؛ در همین راستا، یانگ و لیو (Yang & Liu, 2006) به محاسبه شباهت مدارک نیمه ساختاریافته بر اساس وزن‌های زمینه‌با هدف بازیابی بهتر در محیط XML پرداختند. آنها با کمک مدل فضای برداری توسعه‌یافته و با در نظر گرفتن محتوا و ساختار مدرک شباهت میان مدارک را محاسبه کردند به این نتیجه رسیدند که روش آنها بالاترین عملکرد را نسبت به روش‌های دیگر سنجش شباهت داشته است. یون و همکاران (Yoon et al., 2011) نیز شباهت میان مقالات را با استفاده از سنج‌های متنی و با هدف ارائه طرحی برای بسط کلیدواژه‌ها بررسی کردند. آنها شباهت عناصر بازنمونی مقالات را به صورت جداگانه محاسبه کردند و روشی برای بسط کلیدواژه پیشنهاد کردند که در مواقعی که اطلاعات متنی ناقص بود، کارایی خوبی داشت. آنها به هر یک از عناصر بازنمونی مقالات، یک وزن اختصاص دادند و در سال ۲۰۱۲ با گسترش طرح اولیه، طرح جدیدی برای محاسبه شباهت ارائه کردند که اطلاعات متنی و پیوندی مقالات را ترکیب می‌کرد. یافته‌ها نشان داد استفاده از عنوان و چکیده برای محاسبه شباهت متنی بهترین نتایج را داشت (Yoon et al., 2011). ریستانتی و همکاران (Ristanti et al., 2019) نیز با هدف پیشنهاد مجله مناسب به نویسندگان جهت انتشار مقاله، مجلات تخصصی حوزه اقتصاد را طبقه‌بندی

1. SimCC
 2. SVM^{rank} (support vector machine)
 3. SimAll
 4. preprocessing modules

کردند. به این منظور، عناوین و چکیده‌های مقالات مجلات تخصصی حوزه اقتصاد را استخراج کردند و برای محاسبه شباهت، سنجه کسینوسی را به کار بردند. آزمون عملکرد، مانعیت سامانه را ۵۷.۷۹ درصد و جامعیت آن را ۶۲.۹۶ درصد برآورد کرد.

برخی پژوهش‌ها از سنجه شباهت کسینوسی در سامانه‌های پیشنهاددهنده و طبقه‌بندی متون استفاده کردند؛ برای مثال، حسن و همکاران (Hassan et al., 2019) در سامانه پیشنهاددهنده مقالات پژوهشی که بر اساس عنوان مقالات، پیشنهادهای را به کاربران ارائه می‌کرد، از سنجه شباهت کسینوسی استفاده کردند. حبیبی و کایو (Habibi & Cahyo, 2020) مدلی به منظور طبقه‌بندی مجلات به صورت خودکار با استفاده از الگوریتم شباهت کسینوسی و ماشین بردار پشتیبان^۱ ایجاد کرد. هدف پژوهش طبقه‌بندی مجلات علمی بر اساس چکیده‌های آنان بود. نتایج نشان داد ماشین بردار پشتیبان دقت عملکرد بهتری نسبت به شیوه شباهت کسینوسی دارد. پارک و همکاران (Park et al., 2020) روش مبتنی بر شباهت کسینوسی با هدف بهبود طبقه‌بندی متون ارائه کردند. آنها شباهت کسینوسی را با طبقه‌بندی کننده‌های معمول نظیر ماشین بردار پشتیبان، نایو بیس چندجمله‌ای و شبکه عصبی ترکیب کردند تا عملکرد آنها را بهبود ببخشند. یافته‌ها نشان داد دقت طبقه‌بندی کننده‌ها به میزان قابل توجهی افزایش یافته است. استرلینگ و مونته‌مور (Sterling & Montemore, 2021) سامانه پیشنهاددهنده مقالات پژوهش طراحی کردند که در آن برای سنجش شباهت میان مقالات بازیابی شده با مقالات مدنظر کاربران، از ترکیب شباهت متنی و ارتباط شبکه استنادی استفاده می‌شد. ابتدا آنها مقالاتی را که در شبکه استنادی مقاله مدنظر کاربر دارای بیشترین میزان استناد بود، جستجو کردند؛ سپس با استفاده از سنجه شباهت کسینوسی میزان شباهت متنی عناوین و چکیده‌ها را مقایسه کردند. یافته‌های آنان اثربخشی سامانه مذکور را نشان داد. همچنین، رنوکا و همکاران (Renuka et al., 2021) سامانه پیشنهاددهنده مقالات طراحی کردند که از رویکرد مبتنی بر محتوا استفاده می‌کرد. مجموعه داده‌های آنها را ۲۳۰ مقاله تشکیل می‌داد و برای اندازه‌گیری شباهت از سنجه کسینوسی و برای خوشه‌بندی مقالات از K-means استفاده کردند. یافته‌ها نشان داد این شیوه برای پیشنهاد مقالات به‌خوبی عمل کرده است.

مرور پژوهش‌ها نشان داد برخی پژوهش‌ها به تحلیل علم‌سنجی مقالات و برون‌دادهای علمی پرداختند و از پایگاه‌های مختلفی نظیر وب آو ساینس و اسکوپوس استفاده کردند (اکبری نیسیانی و همکاران، ۱۴۰۰؛ خلیلی و محمدی، ۱۴۰۰؛ نجفی برازجانی و همکاران، ۱۴۰۱؛ علی‌نژاد چمازکتی و میرحق‌جو لنگرودی، ۱۴۰۱)؛ برخی دیگر از پژوهشگران در حوزه سنجش شباهت میان مدارک و عبارت جستجوی کاربر با مدارک به ارائه روش‌های نوین شباهت‌سنجی پرداختند (Peng et al., 2008; Renu & Mocko, 2016) و برخی دیگر رویکردها و سنجه‌های شباهت‌سنجی را با هدف بهبود سنجش شباهت با یکدیگر ترکیب کردند (Chikhi et al., 2008; Hamedani et al., 2013; Farouk, 2019; Farouk, 2020). گروهی هم ابزارهای مختلفی جهت ارزیابی متون با استفاده از سنجه‌های مختلف شباهت‌سنجی ایجاد کردند (Borges et al., 2012; Hamedani et al., 2016; Gomma & Fuhmy, 2017; Kerzendorf, 2019). در مواردی دیگر، پژوهشگران با در نظر گرفتن عناصر بازنمونی مقالات به‌عنوان عناصر مستقل، به شباهت‌سنجی مدارک پرداختند (Yang & Liu, 2006; Yoon et al., 2012; Yoon et al., 2011; Ristani et al., 2019) و تنها در یک پژوهش قابلیت استفاده از شبکه استنادی منابع پیشنهاد پژوهش برای پیشنهاد مدارک مرتبط مورد مطالعه قرار گرفته بود (پابرجا و همکاران، ۱۴۰۰). درنهایت، برخی

1. Support Vector Machine

پژوهشگران از سنجه شباهت کسینوسی در سامانه‌های پیشنهاددهنده و طبقه‌بندی متون استفاده کردند (Hassan et al., 2019; Habibi & Cahyo, 2020; Park et al., 2020; Sterling & Montemore, 2021; Renuka 2021).

روش شناسی پژوهش

پژوهش حاضر از نظر هدف، کاربردی است؛ از نظر روش شناختی رویکرد کمی دارد و از روش‌های تحلیل استنادی و تحلیل محتوایی بهره برده است.

جامعه آماری این پژوهش، ۳۰۱۹ مقاله موجود در شبکه استنادی فهرست منابع پیشنهاددهنده‌های پژوهش دانشجویان تحصیلات تکمیلی شامل دانشجویان کارشناسی ارشد و دکتری رشته شیمی دانشگاه شیراز است. برای شناسایی و انتخاب آزمودنی‌های واجد شرایط، روش‌های معمول نمونه‌گیری تصادفی امکان‌پذیر نبود؛ زیرا، تمایل آزمودنی‌ها برای در دسترس قراردادن پیشنهاد پژوهش از اهمیت زیادی برخوردار بود؛ بنابراین، از روش نمونه‌گیری داوطلبانه استفاده شد.

با توجه به رویکرد پژوهش که از روش‌های تحلیل استنادی و تحلیل محتوایی بهره می‌برد، از ۳۱ نفر از دانشجویان تحصیلات تکمیلی حوزه شیمی دانشگاه شیراز خواسته شد پیشنهاد پژوهش خود را که به زبان انگلیسی نگاشته می‌شوند در اختیار پژوهشگر قرار دهند تا بتوان به وسیله فهرست منابع آنها، شبکه استنادی فهرست منابع پیشنهاددهنده آنها را ترسیم کرد؛ بدین منظور، عنوان تمامی مقالات انگلیسی موجود در فهرست منابع پیشنهاددهنده پژوهش آنها در پایگاه وب آو ساینس جستجو شد و پس از تطبیق اطلاعات کتاب‌شناختی آنها، فایل شبکه استنادی مقاله و تمامی مقالات موجود در شبکه استنادی آن شامل تمام مقالاتی که به مقالات فهرست منابع استناد داده و یا از آن استناد دریافت کرده بودند، در قالب اکسل ذخیره شد؛ سپس، تمامی فایل‌های به دست آمده با هدف دستیابی به شبکه واحد استنادی برای تمامی منابع موجود در پیشنهاد پژوهش هر یک از کاربران، در یک فایل ادغام شد و بر اساس تعداد استنادها مرتب شد. برخی از پیشنهاددهنده‌های پژوهش دارای شبکه استنادی بسیار بزرگی با بیش از هزار مقاله بودند که باعث می‌شد میان شبکه‌های استنادی پیشنهاددهنده‌های پژوهش توازن وجود نداشته باشد؛ از این رو، با هدف ایجاد یکدستی، ۱۰۰ مقاله دارای بیشترین میزان استناد در هر شبکه انتخاب و مورد تحلیل قرار گرفت.

در مرحله بعد، با استفاده از نرم‌افزاری که توسط یکی از متخصصان علوم رایانه با استفاده از زبان برنامه‌نویسی پایتون و فرمول سنجه شباهت کسینوسی طراحی شده بود، میزان شباهت متنی عنوان پیشنهاددهنده، متن پیشنهاددهنده و عنوان مقالات موجود در فهرست منابع پیشنهاددهنده با عناصر بازنمونی (عنوان، چکیده، واژه‌های کلیدی) ۱۰۰ مقاله دارای بیشترین میزان استناد در شبکه استنادی بر اساس سنجه کسینوسی به دست آمد. نحوه کارکرد این برنامه نرم‌افزاری توسط کمیته‌ای متشکل از دو نفر دکتری رشته کامپیوتر و دو نفر دکتری رشته علم اطلاعات و دانش‌شناسی مورد سنجش، اصلاح و نهایتاً تأیید قرار گرفت. ساز و کار این نرم‌افزار بدین گونه بود که ابتدا پوشه‌ای توسط پژوهشگر ایجاد می‌شد که حاوی ۵ فایل بود؛ فایل اول، حاوی عنوان پیشنهاددهنده پژوهش در قالب تکست بود؛ فایل دوم، عنوان مقالات موجود در پیشنهاددهنده پژوهش وی را که در پایگاه وب آو ساینس نمایه شده بود، در قالب تکست دربرمی‌گرفت و فایل سوم، دربردارنده متن پیشنهاددهنده پژوهش در قالب تکست بود. فایل چهارم، به فرمت اکسل حاوی اطلاعات کتاب‌شناختی مقالات شبکه استنادی هر کاربر بود که در یک فایل واحد ادغام شده بود. درنهایت، فایل اجرایی برنامه نرم‌افزاری طراحی شده نیز همراه با چهار فایل قبلی در این پوشه قرار گرفت. با کلیک‌کردن روی برنامه نرم‌افزاری،

فایل اکسل حاوی نتایج ساخته می‌شد و شباهت کسینوسی عنوان، عنوان مقالات و متن پیشنهاد پژوهش با عناصر بازنمونی ۱۰۰ مقاله با بیشترین میزان استناد موجود در شبکه استنادی هر کاربر تعیین می‌شد.

برای ۳۱ پیشنهاد پژوهش که در این پژوهش به کار گرفته شد، تعداد کل مقالات موجود در شبکه استنادی ۳۱ پیشنهاد پژوهش، ۳۱۰۰ مقاله می‌شد؛ اما به دلیل اینکه تعداد مقالات موجود در شبکه استنادی منابع برخی از پیشنهاد‌های پژوهش کمتر از ۱۰۰ مورد بود، در نهایت ۳۰۱۹ مقاله حاصل شد؛ در مرحله آخر، میزان شباهت متنی عناصر بازنمونی با عناصر پیشنهاد پژوهش که با استفاده از نرم افزار طراحی شده به دست آمده بود به نرم افزار اس پی اس نسخه ۲۳ منتقل و تحلیل شد.

این توضیح لازم است که در پژوهش حاضر منظور از عناصر بازنمونی موجود در شبکه استنادی فهرست منابع، اطلاعاتی نظیر عنوان، چکیده و کلیدواژه‌های کلیه مقالات پژوهشی بازیابی شده از پایگاه وب آو ساینس است که از مقالات موجود در فهرست منابع استناد دریافت کرده‌اند و یا به مقاله‌های فهرست منابع استناد داده‌اند؛ در واقع، مجموعه این منابع با هم یک شبکه را تشکیل می‌دهند که به آن شبکه استنادی گفته می‌شود؛ همچنین، عناصر پیشنهاد پژوهش که در این مقاله از نظر شباهت متنی با عناصر بازنمونی شبکه استنادی سنجیده می‌شود شامل سه بخش عنوان پیشنهاد پژوهش، عنوان منابعی که در انتهای پیشنهاد پژوهش و در قالب فهرست منابع آمده‌اند و متن پیشنهاد پژوهش است.

یافته‌های پژوهش

پاسخ به پرسش نخست پژوهش. آیا اختلاف معنی‌داری در شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقاله‌های شبکه استنادی با عنوان پیشنهاد پژوهش وجود دارد؟

با توجه به میانگین رتبه‌های به دست آمده برای عنوان (۵۰۵۶.۲۳)، چکیده (۶۰۱۰.۲۰) و واژه‌های کلیدی (۲۵۲۰.۵۷)، به منظور مقایسه شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقاله‌های شبکه استنادی با عنوان پیشنهاد پژوهش به دلیل نرمال نبودن توزیع داده‌ها و مقایسه سه عنصر شامل عنوان، چکیده و واژه‌های کلیدی، از آزمون ناپارامتریک کروسکال والیس استفاده شد. متغیرها شامل عنوان پیشنهاد پژوهش به عنوان متغیر مستقل و عناصر بازنمونی موجود در شبکه استنادی به عنوان متغیرهای وابسته بود. نتایج این آزمون نشان داد سطح معناداری به دست آمده برای این سنج از ۰/۰۵ کمتر است و فرض صفر رد می‌شود؛ به عبارتی میان عنوان، چکیده و واژه‌های کلیدی مقالات موجود در شبکه استنادی از نظر سنجش شباهت عنوان پیشنهاد پژوهش اختلاف معناداری وجود دارد ($P=0.001$).

جدول ۱. نتایج آزمون کروسکال والیس برای بررسی اختلاف میان عناصر بازنمونی از نظر سنجش شباهت عنوان

پیشنهاد پژوهش

متغیر	میانگین رتبه‌ها			سطح معناداری
	عنوان	چکیده	کلیدواژه	
میزان	۵۰۵۶.۲۳	۶۰۱۰.۲۰	۲۵۲۰.۵۷	۰.۰۰۱
درجه آزادی	۲	۲	۲	
خی دو	۳۲۰۱.۶۳۱	۳۲۰۱.۶۳۱	۳۲۰۱.۶۳۱	

مقایسه عناصر بازنمونی (عنوان، چکیده و کلیدواژه‌ها) مقاله‌های شبکه استنادی فهرست ...

برای بررسی دقیق‌تر تفاوت‌های مشاهده شده در آزمون کروسکال والیس، از مقایسه میانگین‌ها استفاده شد. میانگین رتبه چکیده ۶۰۱۰.۲۰، میانگین رتبه عنوان ۵۰۵۶.۲۳ و میانگین رتبه واژه‌های کلیدی ۲۵۲۰.۵۷ است؛ بنابراین، می‌توان گفت ابتدا چکیده، سپس عنوان و در نهایت واژه‌های کلیدی مقالات موجود در شبکه استنادی بیشترین شباهت را با عنوان پیشنهاد پژوهش دارند (جدول ۱).

پاسخ به پرسش دوم پژوهش. آیا اختلاف معناداری در شباهت متنی عناصر بازنمونی (عنوان، چکیده و کلیدواژه‌ها) مقاله‌های شبکه استنادی با عنوان منابع پیشنهاد پژوهش وجود دارد؟

مقایسه شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقاله‌های شبکه استنادی با عنوان منابع پیشنهاد پژوهش با استفاده از آزمون کروسکال والیس نشان داد که از نظر آماری اختلاف معناداری میان عنوان، چکیده و واژه‌های کلیدی مقالات موجود در شبکه استنادی از نظر سنجش شباهت عنوان منابع پیشنهاد پژوهش وجود دارد ($P=0.001$) (جدول ۲).

جدول ۲. نتایج آزمون کروسکال والیس برای بررسی اختلاف میان عناصر بازنمونی از نظر سنجش شباهت عنوان منابع پیشنهاد پژوهش

متغیر	میانگین رتبه‌ها			سطح معناداری
	عنوان	چکیده	کلیدواژه	
میزان	۵۱۴۴.۴۶	۶۳۱۰.۴۳	۲۱۳۲.۱۱	۰.۰۰۱

همان‌گونه که در جدول ۲ آمده است میانگین رتبه چکیده ۶۳۱۰.۴۳، میانگین رتبه عنوان ۵۱۴۴.۴۶ و میانگین رتبه کلیدواژه‌ها ۲۱۳۲.۱۱ است. مقایسه میانگین‌ها نشان می‌دهد ابتدا چکیده، سپس عنوان و در نهایت واژه‌های کلیدی بیشترین شباهت را با عنوان منابع پیشنهاد پژوهش داشتند.

پاسخ به پرسش سوم پژوهش. آیا اختلاف معناداری در شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقاله‌های شبکه استنادی با متن پیشنهاد پژوهش وجود دارد؟

مقایسه شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه‌های کلیدی) مقاله‌های شبکه استنادی با متن پیشنهاد پژوهش با استفاده از آزمون کروسکال والیس صورت پذیرفت. نتایج جدول ۳ بیان‌کننده این حقیقت است که از نظر آماری اختلاف معناداری میان عنوان، چکیده و واژه‌های کلیدی مقالات موجود در شبکه استنادی از نظر سنجش شباهت متن پیشنهاد پژوهش وجود دارد ($P=0.001$).

جدول ۳. نتایج آزمون کروسکال والیس برای بررسی اختلاف میان عناصر بازنمونی از نظر سنجش شباهت متن پیشنهاد پژوهش

متغیر	میانگین رتبه‌ها			سطح معناداری
	عنوان	چکیده	کلیدواژه	
میزان	۴۹۳۰.۱۰	۶۵۴۷.۹۰	۲۱۰۹.۰۰	۰.۰۰۱

افزون بر این، همان‌گونه که یافته‌های مندرج در جدول ۳ حکایت می‌کند میانگین رتبه چکیده ۶۵۴۷.۹۰، میانگین رتبه عنوان ۴۹۳۰.۱۰ و میانگین رتبه واژه‌های کلیدی ۲۱۰۹.۰۰ است. مقایسه میانگین‌ها نشان می‌دهد که ابتدا چکیده، سپس عنوان و در نهایت واژه‌های کلیدی بیشترین شباهت متنی را با متن پیشنهاد پژوهش داشتند.

پاسخ به پرسش چهارم پژوهش. میانگین وزنی شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه های کلیدی) مقالات موجود در شبکه استنادی به چه میزان است؟

به منظور تعیین وزن کلی شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه های کلیدی) مقاله های شبکه استنادی از میانگین های وزنی عناصر بازنمونی موجود در جداول ۱، ۲ و ۳ استفاده شد. نتایج نشان داد که در هر سه حالت مقایسه میزان شباهت متنی با عنوان، عنوان منابع و متن پیشنهاد پژوهش چکیده دارای بیشترین وزن و واژه های کلیدی دارای کمترین وزن هستند و عنوان در میان این دو قرار می گیرد (جدول ۴). همان گونه که از یافته ها برمی آید، چکیده و عنوان بسیار قدرتمندتر و با فاصله بسیار زیاد با کلیدواژه می توانند شباهت متنی را با عناصر پیشنهاد پژوهش تبیین کنند. در ستون آخر جدول ۴، میانگین وزنی کل عناصر بازنمونی بر عدد ۱۰۰۰۰ تقسیم شده است تا عددی میان صفر و یک به دست آید و وزن صحیح عناصر بازنمونی تعیین شود.

جدول ۴. میانگین وزنی شباهت متنی عناصر بازنمونی (عنوان، چکیده و واژه های کلیدی) مقالات شبکه استنادی با

پیشنهاد پژوهش

عناصر بازنمونی شبکه استنادی	میزان شباهت متنی با عنوان پیشنهاد	میزان شباهت متنی با عنوان منابع پیشنهاد	میزان شباهت متنی با متن پیشنهاد	میانگین وزنی کل بر اساس سنجه کسینوسی	میانگین وزنی کل
عنوان	۵۰۵۶.۲۳	۵۱۴۴.۴۶	۴۹۳۰.۱۰	۰.۵۰	۵۰۴۳.۵۹
چکیده	۶۰۱۰.۲۰	۶۳۱۰.۴۳	۶۵۴۷.۹۰	۰.۶۲	۶۲۸۹.۵۱
کلیدواژه ها	۲۵۲۰.۵۷	۲۱۳۲.۱۱	۲۱۰۹.۰۰	۰.۲۲	۲۲۵۳.۸۹

بحث و نتیجه گیری

همان گونه که پیش تر گفته شد، پیشنهاد پژوهش، متنی است که نویسنده حین نگارش آن از اطلاعات کافی و جامع درباره موضوع پژوهش برخوردار نیست و پس از نگارش پیشنهاد پژوهش و حین گردآوری اطلاعات و نگارش پایان نامه به اشراف بهتری نسبت به موضوع می رسد. پیشنهاد به نوعی طرح کلی پژوهش است که قرار است پژوهش بر مبنای آن صورت گیرد. در نتیجه، کسب دانش و اطلاعات کافی پس از نگارش پیشنهاد پژوهش از اهمیت خاصی برخوردار است؛ برای مثال، نویسنده پس از نگارش پیشنهاد پژوهش می بایست بتواند تمام پیشینه های مرتبط با کار خود را که از نظر وی دور افتاده است، بازیابی کند، راه های تجزیه و تحلیل یافته های خود را بشناسد و برای تبیین یافته ها و توجیه روش شناسی کار از منابع مرتبط بهره ببرد. برای مرتفع کردن همه نیازهای اطلاعاتی گفته شده، پژوهشگر می تواند از منابع مرتبط با موضوع پیشنهاد پژوهش استفاده کند تا فرایند پژوهش خود را به نحو مطلوب به انجام رساند. یافته ها نشان داد تفاوت معناداری میان عنوان، چکیده و واژه های کلیدی مقالات موجود در شبکه استنادی از نظر شباهت آنها با متن، عنوان و عنوان منابع پیشنهاد پژوهش وجود دارد. تأیید وجود شباهت متنی میان عناصر بازنمونی مقالات موجود در شبکه استنادی فهرست منابع پیشنهاد پژوهش با پیشنهاد پژوهش، در کل حاکی از آن است که می توان از پیشنهاد پژوهش دانشجویان به عنوان بستری برای پیشنهاد مقالات مرتبط به آنها استفاده کرد که پابرجا و همکاران (۱۴۰۰) نیز به نتایجی مشابه دست یافته بودند. آنها اذعان داشتند که کاربران، مقاله های مرتبط پیشنهادی از طریق شبکه استنادی منابع پیشنهاد پژوهش را مرتبط و اثربخش قلمداد کردند.

علاوه بر این، مقایسه میانگین ها حاکی از آن بود که در هر سه جزء مورد بررسی پیشنهاد (متن، عنوان منابع و عنوان پیشنهاد پژوهش)، ابتدا چکیده، سپس عنوان و در نهایت واژه های کلیدی بیشترین شباهت متنی را با آنها

داشت؛ به گونه‌ای که میانگین وزنی چکیده ۰.۶۲، عنوان ۰.۵ و کلیدواژه‌ها ۰.۲۲ بود. این یافته و میانگین وزنی به‌دست‌آمده برای هریک از این عناصر بسیار نزدیک به یافته‌های یون و همکاران (Yoon et al., 2012) بود. بدین‌صورت که آنها ضریب وزنی چکیده را ۰.۷ و عنوان را ۰.۳ برآورد کردند. آنها به محاسبه ضریب وزنی واژه‌های کلیدی پرداخته بودند. این یافته به‌طور ضمنی به یافته‌های یون و همکاران (Yoon et al., 2011) اشاره دارد؛ بدین‌صورت که برای محاسبه میزان شباهت میان دو مقاله بهتر است هرکدام از عناصر بازنمونی مقالات را به‌عنوان واحدهای مستقل در نظر گرفت. به عبارت دیگر، بهتر است چکیده‌های دو مقاله با یکدیگر، عناوین آنها با هم و واژه‌های کلیدی هم با یکدیگر مقایسه شوند؛ چراکه میزان شباهت به‌دست‌آمده در این حالت با زمانی که شباهت دو متن به‌عنوان یک واحد کل در نظر گرفته شود تفاوت زیادی دارد (Yoon et al., 2011). این یافته با آنچه سایر پژوهشگران به آن دست یافته‌اند همخوانی دارد. علاوه‌براین، در نظر گرفتن عناصر بازنمونی مقالات به‌عنوان واحدهای مستقل نه تنها برای شباهت‌سنجی، بلکه برای بسط واژه‌های کلیدی و پیشنهاد مجله مناسب به نویسندگان جهت انتشار مقاله نیز اهمیت دارد (Yang & Liu, 2006; Yoon et al., 2012; Ristani et al., 2019).

علاوه‌براین، نتایج این پژوهش هم‌راستا با نتایج پژوهش یوسفی و همکاران (Yousefi et al., 2019) است که نشان دادند چکیده‌ها قابلیت‌های معنایی بیشتری نسبت به بدنه و متن کامل دارند. چکیده قسمتی از مدرک است که کاربران برای دستیابی به اطلاعات مهم به آن مراجعه می‌کنند و معمولاً بیشتر از قسمت‌های دیگر مورد مطالعه قرار می‌گیرد؛ چراکه حاوی اطلاعات مهم درباره مدرک (نظیر هدف، روش‌شناسی، یافته‌ها و نتیجه‌گیری) است و در کمترین فضا و زمان، اطلاعات بسیار مهمی را در اختیار مخاطب قرار می‌دهد و می‌تواند نقش به‌سزایی در بازیابی اطلاعات مرتبط ایفا کند. علاوه‌براین، چکیده خلاصه‌ای از ایده‌های اصلی پژوهش است و طرح و یافته‌های پژوهش را بیان می‌کند و اطلاعات مهمی را در کمترین زمان به کاربر منتقل می‌کند. چکیده به کاربران کمک می‌کند تا به آسانی درباره مرتبط یا نامرتب بودن اثری قضاوت کنند (Gazni, 2011). ضمن اینکه چکیده بیش از سایر بخش‌های یک مقاله مورد توجه کاربران قرار می‌گیرد (Nicholas et al., 2003)؛ از این‌رو، می‌توان آن را مهم‌ترین عنصر یک مقاله برای سنجش ارتباط با سایر منابع و از جمله پیشنهاد پژوهش قلمداد کرد.

در ارتباط با عنوان هم باید توجه داشته باشیم که در برخی از موارد اولین و گاهی تنها منبعی است که خوانندگان می‌توانند اطلاعاتی درباره یک مدرک به دست آورند (Garcia et al., 2019). در واقع، عنوان مهم‌ترین عنصر هر مقاله علمی و نشان‌دهنده موضوع اصلی مقاله است که کاربران از آن به‌عنوان منبع اصلی اطلاعات برای قضاوت و یافتن متون مرتبط هنگام جستجو استفاده می‌کنند و نقش مهمی در بازیابی مقاله ایفا می‌کند (Jamali & Nikzad, 2011). افزون بر این، عنوان خوانندگان را از محتوای مقاله آگاه و آنها را تشویق به مطالعه می‌کند و برای نمایه‌سازی در پایگاه‌های داده و اهداف ارجاعی کاربرد دارد (Subotic & Mukherjee, 2014). پژوهشگران به این نتیجه رسیده‌اند که حدود ۷۰ تا ۹۰ درصد استنادات علمی بدون مطالعه کافی یا خواندن کل مدرک، از عنوان فهرست منابع مورد استفاده در مقالات دیگر بوده است (Simkin & Roychowdhury, 2002). بایرن (Byrne, 1975) نیز تأکید داشت استفاده از عنوان در کنار واژه‌های کلیدی می‌تواند بازیابی اثربخش را به دنبال داشته باشد.

پیشنهاد‌های اجرایی پژوهش

یافته‌های پژوهش حاضر مبنی بر تأیید وجود شباهت میان عناصر بازنمونی مقالات موجود در شبکه استنادی

فهرست منابع با پیشنهاد پژوهش نشان می‌دهد که پیشنهاد پژوهش دانشجویان پتانسیل استفاده به عنوان بستر مناسبی جهت بازیابی و پیشنهاد مقالات مرتبط را دارد؛ از این رو:

- طراحان سامانه‌های پیشنهاددهنده مقالات علمی، سامانه‌های بازیابی اطلاعات علمی، کتابخانه‌های دیجیتال و شبکه‌های اجتماعی علمی نظیر لینکدین و ریسرچ گیت می‌توانند از عناصر بازنمونی مقالات شبکه استنادی منابع آن به منظور پیشنهاد مقالات مرتبط استفاده کنند؛

- با توجه به وزن تعیین شده عناصر بازنمونی و در راستای افزایش بهره‌وری سامانه‌های اطلاعاتی، به طراحان این گونه سامانه‌ها پیشنهاد می‌شود که برای محاسبه شباهت، به ویژه از چکیده و عنوان مقالات برای سنجش شباهت استفاده کنند و از محاسبه شباهت متون به صورت یک واحد کل بپرهیزند. این امر موجب صرفه‌جویی در زمان، منابع و انرژی شده، نتایج مطلوب‌تری به دست می‌دهد و کاربران را سریع‌تر به مطالب مورد نظرشان می‌رساند؛

- برای نمایه‌سازی مقالات در پایگاه‌ها و موتورهای جستجو می‌توان چکیده و عنوان مقالات را با هدف صرفه‌جویی در منابع مالی و انرژی اولویت قرار داد.

پیشنهاد برای پژوهش‌های آتی

- این پژوهش عناصر بازنمونی مقالات موجود در شبکه استنادی منابع پیشنهاد پژوهش حوزه شیمی را مورد تحلیل و بررسی قرارداد؛ بنابراین، شایسته است این پژوهش در حوزه‌های دیگر مانند علوم انسانی و اجتماعی و حوزه فنی و مهندسی نیز تکرار شود و نتایج آنها با پژوهش حاضر مقایسه شود تا از این رهگذر پتانسیل شبکه استنادی فهرست منابع پیشنهاد پژوهش سایر حوزه‌ها برای به‌کارگیری در سامانه‌های بازیابی اطلاعات نمایان شود.

تقدیر و تشکر

پژوهش حاضر برگرفته از رساله دکتری با عنوان «مطالعه قابلیت سنج‌های شبکه استنادی و اطلاعات متنی منابع پیشنهاد پژوهش دانشجویان تحصیلات تکمیلی به منظور پیشنهاد مقاله‌های مرتبط» است. در پایان بر خود لازم می‌دانیم از تمامی کسانی که ما را یاری کردند؛ به ویژه دانشجویانی که در این پژوهش مشارکت داشته‌اند و آقای دکتر مسعود نبوی‌زاده، استاد بخش شیمی دانشگاه شیراز و همکارانشان که از هیچ‌گونه کمکی مضایقه نکردند، تشکر و قدردانی کنیم.

فهرست منابع

اکبری نیسیانی، س.، احتشام، ح.، تقی‌زاد، ح.، و دانشور، ح. (۱۴۰۰). جایگاه وزنی مقالات علمی تولیدشده انستیتوکانسر دانشگاه علوم پزشکی تهران: یک مطالعه علم‌سنجی. پژوهش‌نامه علم‌سنجی، ۷ (۱۳)، ۲۱۷-۲۳۴.

<https://www.sid.ir/paper/401315/fa>

پابرجا، ا.، عباس پور، ج.، و نبوی، س. م. (۱۴۰۰). امکان‌سنجی استفاده از شبکه استنادی فهرست منابع موجود در پیشنهاد پژوهش (پروپوزال) دانشجویان تحصیلات تکمیلی برای پیشنهاد مقاله‌های مرتبط به آنها. پژوهش‌نامه

کتابداری و اطلاع‌رسانی، ۱ (۲۱)، ۳۲۲-۳۳۴. <https://doi.org/10.22067/INFOSCI.2021.24102.0>

مقایسه عناصر بازنمونی (عنوان، چکیده و کلیدواژه‌ها) مقاله‌های شبکه استنادی فهرست ...

حری، ع. (۱۳۸۸). آئین نگارش علمی. تهران: نهاد کتابخانه‌های عمومی کشور.

حمیدزاده، ج.، و صادق‌زاد، م. (۱۳۹۶). فیلترکننده مشارکتی فازی ناهموار مبتنی بر کاربر در سیستم‌های پیشنهاددهنده.

مجله مهندسی برق دانشگاه تبریز، ۴۷ (۲)، ۴۹۱-۵۰۰. <https://civilica.com/doc/722526>

خلیلی، ل.، و محمدی، ف. (۱۴۰۰). تحلیل علم‌سنجی مجلات منتشرشده به زبان انگلیسی در حوزه علم اطلاعات و

دانش‌شناسی ایران بر اساس داده‌های اسکوپوس. پژوهش‌نامه علم‌سنجی، ۷ (۱۴)، ۱۹۷-۲۲۰.

<https://doi.org/10.22070/RSCI.2020.5329.1368>

رحمان‌پور، م.، لیاقت‌دار، م. ج.، و افشار، ا. (۱۳۹۶). بررسی چالش‌های فرهنگی-اجتماعی و منابع انسانی فراروی

توسعه فناوری اطلاعات در آموزش عالی ایران از دیدگاه دانشجویان تحصیلات تکمیلی. فصلنامه تحقیقات

فرهنگی ایران، ۱۰ (۲)، ۱۵۱-۱۸۱. <http://dx.doi.org/10.22631/jicr.2017.1460.2172>

صانعی‌نژاد، ا.، خسروی فارسانی، ه.، و خیام‌باشی، م. ر. (۱۳۹۵). بهبود کارایی در سامانه‌های پیشنهاددهنده مبتنی بر

گراف. کنفرانس بین‌المللی چشم‌اندازهای نو در مهندسی برق و کامپیوتر. تهران.

صراطی شیرازی، م. (۱۴۰۰). مطالعه ارتباط استنادی معکوس در هم‌انتشاری‌های صنعت و دانشگاه بر اساس پایگاه

وب آو ساینس در سال ۲۰۱۸. پژوهش‌نامه علم‌سنجی، ۷ (۱۴)، ۱-۲۲.

<https://doi.org/10.22070/rsci.2020.5303.1361>

عبداللهی، ف.، و نقشینه، ن. (۱۴۰۱). تحلیل موضوعی استنادات و ارجاعات به پروانه‌های ثبت اختراع بین‌المللی ایران.

پژوهش‌نامه علم‌سنجی، ۸ (۱۵)، ۱۶۷-۱۹۰. <https://doi.org/10.22070/rsci.2020.5617.1410>

علی‌نژاد چمازکتی، ف.، و میرحق‌جو لنگرودی، س. (۱۴۰۱). تحلیل علم‌سنجی و الگوهای هم‌نویسندگی و استناد

مقالات منتشرشده در نشریات وب آو ساینس کشورهای ایران و ترکیه. پژوهش‌نامه علم‌سنجی، (زودآیند)،

<https://doi.org/10.22070/rsci.2022.14765.1513>

گراسمن، د. ا.، و فریدر، اف. (۱۳۸۴). بازیابی اطلاعات: الگوریتم‌ها و روش‌های اکتشافی. ترجمه جعفر مهرداد و سارا

کلینی. مشهد: کتابخانه رایانه‌ای.

محمدزاده، ف.، فهیمی‌فر، س.، و حسن‌زاده، م. (۱۴۰۰). بررسی مقاله‌های پراستناد پژوهشگران ایرانی در پایگاه وب

آو ساینس بر اساس الگوی همکاری‌ها در سال‌های ۲۰۰۷ تا ۲۰۱۷. پژوهش‌نامه علم‌سنجی، ۷ (۱۴)، ۱۷-۹۱.

<https://doi.org/10.22070/rsci.2020.3850.1241>

ملکی، ا.، عباس‌پور، ج.، جوکار، ع.، و ستوده، ه. (۱۳۹۸). رابطه رتبه آموزشی کتاب‌های درسی دانشگاه‌های برتر جهان

با شاخص‌های استناد، پیچ‌رنگ و هیتس. پژوهش‌نامه علم‌سنجی، ۵ (۱۰)، ۲۲۱-۲۴۰.

<https://doi.org/10.22070/rsci.2019.4125.1267>

موئد، ه. (۱۳۸۷). تحلیل استنادی در ارزیابی پژوهش. ترجمه عباس میرزایی و حیدر مختاری، تهران: چاپار.

نجفی برازجانی، ا.، بصیریان جهرمی، ر. و.، حمیدی، ع. (۱۴۰۱). بررسی تطبیقی مقاله‌های نمایه‌شده پژوهشگران کشورهای خاورمیانه در حوزه دیابت در پایگاه وب آو ساینس با استفاده از شاخص‌های نفوذ اندیشه‌ای و نفوذ اجتماعی. *پژوهش‌نامه علم‌سنجی*، ۹ (۱)، (بهار و تابستان).
<https://doi.org/10.22070/RSCI.2022.15411.1540>

Abdollahi, F., & Naghshineh, N. (2022). Topical Analysis of References from and Citations to Iranian International Patents. *Scientometrics Research Journal*, 8 (1), (Spring & Summer), 167-190. <https://doi.org/10.22070/rsci.2020.5617.1410> [In Persian].

Akbari Neisiani, S., Ehtesham, H., Taghizad, H., & Daneshvar, H. (2021). Position of scientific articles produced by the Cancer Institute of Tehran University of Medical Sciences in terms of weight: a scientometric study. *Scientometrics Research Journal*, 7 (1), (Spring & Summer), 217-234. <https://doi.org/10.22070/rsci.2020.5124.1348> [In Persian].

Alfraidi, H., Lee, W. S., & Sankoff, D. (2015). *Literature visualization and similarity measurement based on citation relations*. [In 2015 19th International Conference on Information Visualisation], 217-222. IEEE. <https://doi.org/10.1109/iV.2015.47>

Alinezhad Chamazkoti, F., & mirhaghjoo langerudi, S. (2022). Scientometric Analysis and Co-authorship Patterns and Citation of the Articles Published in Iranian and Turkish Journals in WoS. *Scientometrics Research Journal*, (Online Published, 6 June).
<https://doi.org/10.22070/rsci.2022.14765.1513> [In Persian].

Arif, M. A. (2016). *Content aware citation recommendation system*. [In 2016 International Conference on Emerging Technological Trends (ICETT)], 1-6. IEEE.
<https://doi.org/10.1109/ICETT.2016.7873690>

Beel, J., Aizawa, A., Breitingner, C., & Gipp, B. (2017). *Mr. DLib: recommendations-as-a-service (RaaS) for academia*. [In 2017 ACM/IEEE Joint Conference on Digital Libraries], 1-2. <https://ieeexplore.ieee.org/document/7991606>

Bichteler, J., & Eaton, E. A. (1980). The Combined Use of Bibliographic Coupling and Cocitation for Document Retrieval. *Journal of the American Society for Information Science*, 31(4), 278-282. <https://doi.org/10.1002/asi.4630310408>

Borges, E. N., Pereira, I. A., Tomasini, C., & Vargas, A. P. (2012). *ARGOsearch: An Information Retrieval System based on text similarity and extensible relevance criteria*. [In 2012 31st International Conference of the Chilean Computer Science Society], 133-141. <https://doi.org/10.1109/SCCC.2012.23>

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215-2222. <https://doi.org/10.1002/asi.23329>

Byrne, J. R. (1975). Relative effectiveness of titles, abstracts, and subject headings for machine retrieval from the COMPENDEX services. *Journal of the American Society for Information Science*, 26(4), 223-229. <https://doi.org/10.1002/asi.4630260405>

- Chikhi, N. F., Rothenburger, B., & Aussenac-Gilles, N. (2008). *Combining link and content information for scientific topics discovery*. [In 2008 20th IEEE International Conference on Tools with Artificial Intelligence], 2, 211-214. IEEE.
<https://doi.org/10.1109/ICTAI.2008.136>
- Eto, M. (2013). Evaluations of context-based co-citation searching. *Scientometrics*, 94(2), 651-673. <https://doi.org/10.1007/s11192-012-0756-z>
- Farouk, M. (2019). Measuring sentences similarity: a survey. *Indian Journal of Science and Technology*, 12(25). <https://doi.org/10.17485/ijst/2019/v12i25/143977>
- Farouk, M. (2020). Measuring text similarity based on structure and word embedding. *Cognitive Systems Research*, 63, 1-10. <https://doi.org/10.1016/j.cogsys.2020.04.002>.
- Fonseca, B. M., Golgher, P., Pôssas, B., Ribeiro-Neto, B., & Ziviani, N. (2005). *Concept-based interactive query expansion*. [In Proceedings of the 14th ACM international conference on Information and knowledge management], 696-703.
<https://doi.org/10.1145/1099554.1099726>
- Garcia, D. C. F., Gattaz, C. C., & Gattaz, N. C. (2019). The relevance of title, abstract and keywords for scientific paper writing. *Revista de Administração Contemporânea*, 23, 1-9.
<https://doi.org/10.1590/1982-7849rac2019190178>
- Gazni, A. (2011). Are the abstracts of high impact articles more readable? Investigating the evidence from top research institutions in the world. *Journal of Information Science*, 37(3), 273-281. <https://doi.org/10.1177/0165551511401658>
- Gomaa, W. H., & Fahmy, A. A. (2017). *SimAll: A flexible tool for text similarity*. [In the Seventeenth Conference on Language Engineering ESOLEC], 17, 122-127.
https://www.academia.edu/35381793/SimAll_A_flexible_tool_for_text_similarity
- Grossman, D. A., & Frieder, F. (2004). *Information retrieval: Algorithms and heuristic methods*. Translated by Jafar Mehrad and Sara Koleini. Mashhad: Computer Library Publications. [In Persian].
- Habibi, M., & Cahyo, P. W. (2020). Journal Classification Based on Abstract Using Cosine Similarity and Support Vector Machine. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 4(3), 185-192.
https://www.researchgate.net/publication/342867323_Journal_Classification_Based_on_Abstract_Using_Cosine_Similarity_and_Support_Vector_Machine
- Hamedani, M. R., Kim, S. W., & Kim, D. J. (2016). SimCC: A novel method to consider both content and citations for computing similarity of scientific papers. *Information Sciences*, 334, 273-292. <https://doi.org/10.1016/j.ins.2015.12.001>
- Hamedani, M. R., Lee, S. C., & Kim, S. W. (2013). On combining text-based and link-based similarity measures for scientific papers. In *Proceedings of the 2013 Research in Adaptive and Convergent Systems*, 111-115. <https://doi.org/10.1145/2513228.2513321>

- Hamidzadeh, J., & Sadeghzadeh, M. (2017). A User Based Fuzzy Rough Collaborative Filtering in Recommender Systems. *Tabriz Journal of Electrical Engineering*, 47(2), 491-500. <https://civilica.com/doc/722526> [In Persian].
- Harter, S. P., & Nisonger, T. E. (1993). Semantic Relationships between Cited and Citing Articles in Library and Information Science Journals. *Journal of the American Society for Information Science*, 44(9), 543-552. [https://doi.org/10.1002/\(SICI\)1097-4571\(199310\)44:9<543::AID-ASI4>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-4571(199310)44:9<543::AID-ASI4>3.0.CO;2-F)
- Hassan, H. A. M., Sansonetti, G., Gasparetti, F., Micarelli, A., & Beel, J. (2019). Bert, elmo, use and infersent sentence encoders: The panacea for research-paper recommendation? In *RecSys (Late-Breaking Results)*, 6-10. https://www.researchgate.net/publication/335555312_BERT_ELMo_USE_and_InferSent_Sentence_Encoders_The_Panacea_for_Research-Paper_Recommendation
- Horri, A. (2009). *Scientific writing*. Tehran: Iran Public Libraries. [In Persian].
- Jamali, H. R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, 88(2), 653-661. <https://doi.org/10.1007/s11192-011-0412-z>
- Kerzendorf, W. E. (2019). Knowledge discovery through text-based similarity searches for astronomy literature. *Journal of Astrophysics and Astronomy*, 40(3), 1-7. <https://doi.org/10.1007/s12036-019-9590-5>
- Khalili, L., & Mohammadi, F. (2021). Scientometric Analysis of English-language Journals in the Field of Knowledge and Information Science in Iran Based on Scopus Data. *Scientometrics Research Journal*, 7 (2), (Autumn & Winter), 197-220. <https://doi.org/10.22070/rsci.2020.5329.1368> [In Persian].
- Kinley, K., Tjondronegoro, D., Partridge, H., & Edwards, S. (2014). Modeling users' web search behavior and their cognitive styles. *Journal of the Association for Information Science and Technology*, 65(6), 1107-1123. <https://doi.org/10.1002/asi.23053>
- Komkhao, M., Lu, J., Li, Z., & Halang, W. A. (2013). Incremental collaborative filtering based on Mahalanobis distance and fuzzy membership for recommender systems. *International Journal of General Systems*, 42(1), 41-66. <https://doi.org/10.1080/03081079.2012.710437>
- Küçükünç, O., Saule, E., Kaya, K., & Çatalyürek, Ü. V. (2012). Recommendation on academic networks using direction aware citation analysis. *arXiv preprint arXiv*, 1205.1143. <https://doi.org/10.48550/arXiv.1205.1143>
- Kusumastuti, S., Derks, M. G., Tellier, S., Di Nucci, E., Lund, R., Mortensen, E. L., & Westendorp, R. G. (2016). Successful ageing: A study of the literature using citation network analysis. *Maturitas*, 93, 4-12. <https://doi.org/10.1016/j.maturitas.2016.04.010>
- Li, X., Wang, H., & Yan, X. (2015). Accurate recommendation based on opinion mining. In *Genetic and Evolutionary Computing*, 399-408. https://doi.org/10.1007/978-3-319-12286-1_41

- Luu, L. A., & Kim, J. J. (2012). Automatic suggestion for PubMed query reformulation. *Journal of Computing Science and Engineering*, 6(2), 161-167.
<https://doi.org/10.5626/JCSE.2012.6.2.161>
- Maleki, A., Abbaspour, G., Jowkar, A., & Sotudeh, H. (2019). The Relationship Between Text-books' Teaching Ranks in World Top Universities and Citation, PageRank and HITS Indicators. *Scientometrics Research Journal*, 5 (2), (Autumn & Winter), 221-240.
<https://doi.org/10.22070/rsci.2019.4125.1267> [In Persian].
- Menczer, F. (2004). *Combining Link and Content Analysis to Estimate Semantic Similarity*. [In Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters], May 19-21, 2004, New York, NY, USA, 452-453.
<https://doi.org/10.1145/1013367.1013521>
- Moed, Hank. (2008). *Citation analysis in research evaluation*. Translated by Abbas Mirzaei and Haider Mokhtari. Tehran: Chapar. [In Persian].
- Mohammadzadeh, F., Fahimifar, S., & Hasanzadeh, M. (2021). Investigating and Visualization of Iranian highly cited papers in order to discover the most effective at the international level in the period of ten years 2007-2017. *Scientometrics Research Journal*, 7(2), (Autumn & Winter), 77-98. <https://doi.org/10.22070/rsci.2020.3850.1241> [In Persian].
- Najafi Borazjani, A., Basirian Jahromi, R., & Hamidi, A. (2023). A Comparative Study of Indexed Articles on Diabetes by Middle Eastern Scholars in the Web of Science from 2010 to 2019 using the Indicators of Ideational Influence and Social Influence. *Scientometrics Research Journal*, 9 (1), (Spring & Summer), 447-468.
<https://doi.org/10.22070/rsci.2022.15411.1540> [In Persian].
- Nicholas, D., Huntington, P., & Watkinson, A. (2003). Digital journals, Big Deals and online searching behavior: A pilot study. In *Aslib Proceedings: New Information Perspectives*, 55(1/2), 84-109. <https://doi.org/10.1108/00012530310462742>
- Pabarja, E., Abbaspour, J., & Nabavi, S. (2021). Feasibility Study of Using References Citation Network in Postgraduate Students' Proposal for Suggesting Related Articles. *Library and Information Science Research*, 11(1), 322-334.
<https://doi.org/10.22067/infosci.2021.24102.0> [In Persian].
- Park, K., Hong, J. S., & Kim, W. (2020). A methodology combining cosine similarity with classifier for text classification. *Applied Artificial Intelligence*, 34(5), 396-411.
<https://doi.org/10.1080/08839514.2020.1723868>
- Peng, J., Yang, D., Tang, S., Wang, T., & Gao, J. (2008). A new similarity computing method based on concept similarity in Chinese text processing. *Science in China Series F: Information Sciences*, 51(9), 1215-1230. <https://doi.org/10.1007/s11432-008-0103-4>
- Porcel, C., Tejada-Lorente, A., Martínez, M. A., & Herrera-Viedma, E. (2012). A hybrid recommender system for the selective dissemination of research resources in a technology transfer office. *Information Sciences*, 184(1), 1-19.
<https://doi.org/10.1016/j.ins.2011.08.026>

- Rahmanpoor, M., Liaghatdar, M., & Afshar, E. (2017). Cultural-social and human resource challenges facing development of information technology in Iran's higher education in viewpoint of graduate students. *Journal of Iranian Cultural Research*, 10(2), 151-181. <https://doi.org/10.22631/jicr.2017.1460.2172> [In Persian].
- Renu, R. S., & Mocko, G. (2016). Computing similarity of text-based assembly processes for knowledge retrieval and reuse. *Journal of Manufacturing Systems*, 39, 101-110. <https://doi.org/10.1016/j.jmsy.2016.03.004>
- Renuka, S., Raj Kiran, G. S. S., & Rohit, P. (2021). An unsupervised content-based article recommendation system using natural language processing. In *Data Intelligence and Cognitive Informatics*, 165-180. https://doi.org/10.1007/978-981-15-8530-2_13
- Ristani, P. Y., Wibawa, A. P., & Pujianto, U. (2019). *Cosine similarity for title and abstract of economic journal classification*. [In 2019 5th International Conference on Science in Information Technology (ICSITech)], 123-127. <https://doi.org/10.1109/ICSITech46713.2019.8987547>
- Saneinejad, E., Khosravi F., Hadi, K., M. R. (2015). *Improving efficiency in graph-based recommender systems*, [In The first international conference on new perspectives in electrical and computer engineering], Tehran. [In Persian].
- Serati Shirazi, M. (2021). Studying the Reverse citation relations in university and industry co-publications based on Web of Science database in 2018. *Scientometrics Research Journal*, 7(2), (Autumn & Winter), 1-22. <https://doi.org/10.22070/rsci.2020.5303.1361> [In Persian].
- Simkin MV & Roychowdhury VP. (2002). Read before you cite! *Complex Systems*, 14, 269–274. <https://doi.org/10.48550/arXiv.cond-mat/0212043>
- Spink, A., Park, M., & Koshman, S. (2006). Factors affecting assigned information problem ordering during Web search: An exploratory study. *Information Processing & Management*, 42(5), 1366-1378. <https://doi.org/10.1016/j.ipm.2006.01.007>
- Sterling, J. A., & Montemore, M. M. (2021). Combining Citation Network Information and Text Similarity for Research Article Recommender Systems. *IEEE Access*, 10, 16-23. <https://doi.org/10.1109/ACCESS.2021.3137960>
- Subotic, S., & Mukherjee, B. (2014). Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of information science*, 40(1), 115-124. <https://doi.org/10.1177/0165551513511393>
- Sun, J., Jiang, Y., Cheng, X., Du, W., Liu, Y., & Ma, J. (2018). A hybrid approach for article recommendation in research social networks. *Journal of Information Science*, 44(5), 696-711. <https://doi.org/10.1177/0165551517728449>
- Xia, F., Liu, H., Lee, I., & Cao, L. (2016). Scientific article recommendation: Exploiting common author relations and historical preferences. *IEEE Transactions on Big Data*, 2(2), 101-112. Scientific Article Recommendation: Exploiting Common Author Relations and Historical Preferences | IEEE Journals & Magazine | IEEE Xplore

- Yang, C. C., & Liu, N. (2006). *Measuring similarity of semi-structured documents with context weights*. [In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval], 719-720.
<https://doi.org/10.1145/1148170.1148334>
- Yoon, S. H., Kim, J. S., Kim, S. W., & Lee, C. (2012). TL-Rank: A Blend of Text and Link Information for Measuring Similarity in Scientific Literature Databases. *IEICE TRANSACTIONS on Information and Systems*, 95(10), 2556-2559.
https://search.ieice.org/bin/summary.php?id=e95-d_10_2556
- Yoon, S. H., Kim, S. W., Kim, J. S., & Hwang, W. S. (2011). *On computing text-based similarity in scientific literature*. [In Proceedings of the 20th international conference companion on World wide web], 169-170. <https://doi.org/10.1145/1963192.1963278>
- Yousefi, Z., Sotudeh, H., Mirzabeigi, M., Fakhrahmad, S. M., Nikseresht, A., & Mohammadi, M. (2019). Investigating text power in predicting semantic similarity. *International Journal of Information Science and Management (IJISM)*, 17(1), 17.
https://ijism.ricest.ac.ir/article_698288.html