

The Impact of Data Lack and Data Sparsity on the Effectiveness of the Results of the RICeST Journal Finder Results: A Case Study in the Field of Engineering

Narjes Vara

PhD Candidate in Knowledge and Information Science
Department; School of Education & Psychology; Shiraz University;
Shiraz, Iran; Faculty Member RICeST Email: vara@ricest.ac.ir

Mahdieh Mirzabeigi*

PhD in Knowledge and Information Science; Associate Professor;
School of Education & Psychology; Shiraz University; Shiraz, Iran;
Email: mmirzabeigi@gmail.com

Hajar Sotudeh

PhD in Knowledge and Information Science; Professor;
School of Education & Psychology; Shiraz University; Shiraz, Iran;
Email: sotudeh@shirazu.ac.ir

Seyed Mostafa Fakhrahmad

PhD in Computer Engineering; Associate Professor; Department of
Computer Science and Engineering & IT; Shiraz University; Shiraz,
Iran Email: fakhrahmad@shirazu.ac.ir

Niloofar Mozafari

PhD in Artificial Intelligence; Assistant Professor; Regional
Information Center for Science and Technology; Islamic World
Science Citation Center; Shiraz, Iran Email: mozafari@ricest.ac.ir

Received: 21, Jun. 2021 Accepted: 03, Sep. 2021

Abstract: Several factors are involved in the production and presentation of recommender systems. The aim of this study was to investigate the effect of the two challenges lack and sparsity of data on the effectiveness of the proposed results of the RICeST Journal Finder. The corpus includes more than 15,000 articles from technical and engineering publications in the period 2013 to 2017, which have been collected from their website. Textual elements of these articles were extracted, normalized and processed, and a research body database was created. Based on the number of collected articles, by using Cochran's formula, 400 basic articles that previously published

* Corresponding Author

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. 37 | No. 4 | pp. 1293-1318

Summer 2022

<https://doi.org/10.35050/JIPM010.2022.011>



in related to the topic of each journal were selected in a random-proportional method. Title and abstract of these articles as a query, in order to receive the system journals suggested, to print the article in two stages of before and after improving the two challenges of lack and sparsity of data in the test corpus. The suggested results in each step were saved in Excel. Finally, the effectiveness of the system results in each stage was determined by Leave-one-out Cross-Validation method and based on the accuracy criterion in k. The relative abundance of categories showed that, in the current situation, the target journal was suggested in only 26% of searches in the first 3 ranks. After enriching, normalizing and processing the data and thus improving the lack of data challenge, although 30% of the results were still rated above 10; but the accuracy of the results in the first 3 ranks increased by 15%. Also, after thematically categorizing the data with the aim of improving the sparsity challenge, 30% increase in the accuracy of the system results in the first 3 ranks compared to the previous step was achieved. The results of this study showed that enriching the database, improving the processing process and thematic classification of data in RICeST journal finder can reduce the two challenges lack and sparsity of data and increase the effectiveness of the proposed results of this systems.

Keywords: Efficiency, Journal Finder, Lack of Data, Data Sparsity, RICeST Journal Finder

تأثیر کمبود و پراکندگی داده بر اثربخشی نتایج سامانه ژورنال یاب رایست؛ مطالعه موردی حوزه فنی و مهندسی

نرجس ورع

دانشجوی دکتری علم اطلاعات و دانش‌شناسی؛
مربی؛ مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری؛
شیراز، ایران vara@ricest.ac.ir

مهديه ميرزاييگي

دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛
گروه علم اطلاعات و دانش‌شناسی؛ دانشگاه شیراز؛
شیراز، ایران؛
mmirzabeigi@gmail.com پدیدآور رابط

هاجر ستوده

دکتری علم اطلاعات و دانش‌شناسی؛ استاد؛ گروه علم
اطلاعات و دانش‌شناسی؛ دانشگاه شیراز؛ شیراز، ایران؛
sotudeh@shirazu.ac.ir

سید مصطفی فخر احمد

دکتری مهندسی کامپیوتر؛ دانشیار؛ گروه مهندسی و
علوم کامپیوتر و فناوری اطلاعات؛ دانشگاه شیراز؛
شیراز، ایران fakhrmahmad@shirazu.ac.ir

نیلوفر مظفری

دکتری هوش مصنوعی؛ استادیار؛ مرکز منطقه‌ای
اطلاع‌رسانی علوم و فناوری؛ شیراز، ایران؛
mozafari@ricest.ac.ir



دریافت: ۱۴۰۰/۰۳/۳۱ پذیرش: ۱۴۰۰/۰۶/۱۲ مقاله برای اصلاح به مدت ۱۱ روز نزد پدیدآوران بوده است.

چکیده: عوامل متعددی از مجموعه عناصر تشکیل‌دهنده سامانه‌های پیشنهاددهنده در تولید و ارائه پیشنهاد دخیل هستند. مطالعه حاضر، با هدف شناخت تأثیر دو چالش کمبود و پراکندگی داده بر اثربخشی نتایج پیشنهادی سامانه ژورنال‌یاب «رایست» انجام شده است. بدین منظور، بیش از ۱۵۰۰۰ مقاله از نشریه‌های فنی و مهندسی در بازه زمانی ۱۳۹۲ تا ۱۳۹۶ از وب‌سایت نشریه‌ها گردآوری شد. در مرحله بعد عناصر متنی این مقاله‌ها شامل عنوان، چکیده و واژه‌های کلیدی استخراج، نرمال‌سازی و پردازش شد و پایگاه داده پیکره پژوهش ایجاد گردید. بر اساس تعداد مقاله‌های گردآوری‌شده، با استفاده از فرمول «کوکران» تعداد ۴۰۰ مقاله پایه، که پیش از این در نشریه‌های مرتبط با موضوع منتشر شده بودند،

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

شاپا (چاپی) ۲۲۵۱-۸۲۲۳

شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱

نمایه در SCOPUS، ISI، LISTA، و

jipm.irandoc.ac.ir

دوره ۳۷ | شماره ۴ | صص ۱۲۹۳-۱۳۱۸

تابستان ۱۴۰۱

<https://doi.org/10.35050/JIPM010.2022.011>



به روش تصادفی-تناسبی انتخاب شد. عنوان و چکیده این مقاله‌ها، به‌منظور دریافت نشریه‌های پیشنهادی سامانه، جهت چاپ مقاله در دو مرحله پیش و پس از بهبود دو چالش کمبود و پراکندگی داده به‌عنوان پرمسمن وارد سامانه شد. سپس، نتایج پیشنهادی در هر مرحله در قالب فایل «اکسل» ذخیره گردید. سرانجام، میزان اثربخشی نتایج سامانه در هر مرحله به روش اعتبارسنجی یک‌طرفه و بر اساس معیار دقت در k تعیین شد. فراوانی نسبی رده‌ها نشان داد که در وضعیت موجود، نشریه هدف تنها در ۲۶ درصد از پرمسمن‌ها در ۳ رتبه نخست پیشنهاد شده است. در راستای بهبود چالش کمبود داده با غنی‌سازی، نرمال‌سازی و پردازش داده‌ها اثربخشی نتایج در ۳ رتبه نخست به میزان ۱۵ درصد افزایش یافت. اما همچنان در بیش از ۳۰ درصد پرمسمن‌ها، نشریه هدف در رتبه ۱۰ و بالاتر پیشنهاد شده بود. بنابراین، در مرحله بعد، به‌منظور بهبود چالش پراکندگی، دسته‌بندی موضوعی داده‌ها انجام و افزایش ۳۰ درصدی اثربخشی نتایج نسبت به مرحله پیشین در ۳ رتبه نخست حاصل گردید. بر این اساس، یکی از عواملی که منجر به کاهش اثربخشی نتایج پیشنهادی سامانه ژورنال‌یاب «رایست» می‌شود، کمبود و پراکندگی داده‌هاست که با غنی‌سازی پایگاه داده، بهبود فرایند پردازش، و دسته‌بندی موضوعی داده‌ها می‌توان با این دو چالش به میزان قابل توجهی مقابله کرد و اثربخشی نتایج پیشنهادی سامانه را بهبود بخشید.

کلیدواژه‌ها: اثربخشی، سامانه پیشنهاددهنده نشریه، کمبود داده، پراکندگی داده، سامانه ژورنال‌یاب رایست

۱. مقدمه

انتشار یافته‌های پژوهشی برای به اشتراک‌گذاری دانش و همچنین اعتبار حرفه‌ای پژوهشگران حائز اهمیت است (Wang and Hou 2015). اما در تصمیم‌نهایی، عوامل انتخاب یک نشریه مناسب جهت نشر، متأثر از عوامل گوناگون بوده و در سه دسته قابل بررسی است: دسته اول، شامل ویژگی‌های نشریه^۱ مانند اعتبار علمی^۲، مدل انتشار^۳، نرخ پذیرش^۴، دامنه و قلمرو موضوعی^۵ و مخاطبان^۶ است، دسته دوم، مشخصات مقاله^۷ مانند ارزش ذاتی به معنای تازگی تحقیق^۸، صحت روش^۹، تأثیر بالقوه در قلمرو موضوعی نشریه و پیامدهای آن^{۱۰} را دربرمی‌گیرد، و دسته سوم، حاوی اولویت‌ها و محدودیت‌های نویسنده^{۱۱} مانند زمان، شرایط ارتقای شغلی، حمایت‌های سازمانی و هزینه انتشار است (Bahadoran et al. 2021). انتخاب درست نشریه یکی از مهم‌ترین و دشوارترین جنبه‌های انتشار نتایج تحقیقات است،

- | | | |
|--|---|-----------------------------|
| 1. journal's characteristics | 2. scientific prestige | 3. publishing models |
| 4. acceptance rate | 5. specialty | 6. audience |
| 7. characteristics of the manuscript | 8. novelty | 9. soundness of methodology |
| 10. potential impact and its alignment with the journal's aims and scope | 11. authors' priorities and limitations | |

زیرا ارسال مقاله‌های حاصل از یک پژوهش به نشریه‌هایی که ربط کمتری با محتوا و نیاز خوانندگان دارند، باعث اتلاف وقت پژوهشگر و همچنین، حوزه سردبیری می‌شود. نشریه مدت‌زمانی را صرف بررسی و به احتمال، ارسال مقاله به داوری نموده و اگر طی هر یک از فرایندهای مزبور مقاله رد شود، این فرایند باید مجدد برای نشریه‌ای دیگر تکرار گردد. همچنین، این تأخیر افزون بر اقتصاد نشریه، بر ضریب تأثیر^۱ و به عبارتی، اعتبار نشریه نیز اثر منفی دارد (Guo, Li and Yu 2021). نتایج مطالعات متعدد نشان می‌دهد که متوسط زمان تصمیم‌گیری و اعلام نظر اولیه به نویسنده، فرایندی طولانی و به‌طور متوسط ۴۱ روز است (Huisman and Smits 2017; Nguyen et al. 2018) و این مدت در ۳۱ درصد موارد به بالای شش ماه نیز می‌رسد (Mulligan, Hall & Raphael 2013). در بخش بررسی اولیه توسط سردبیران و داوران، عدم تطابق و تناسب موضوعی مقاله با دامنه و قلمرو موضوعی نشریه، دلیل اصلی رد مقاله در بسیاری از زمینه‌های موضوعی است و نرخ رد اولیه حتی پیش از ارسال به داوری تا ۸۸ درصد اعلام شده است (Anderson 2012). بنابراین، ارسال مقاله به نشریه‌های غیرمرتبط، افزون بر این که مزایای کمتری به جهت ارتقای فردی و سازمانی برای پژوهشگر به همراه دارد، منجر به از دست رفتن زمان قابل ملاحظه، تازگی، اهمیت و تأثیرگذاری یافته‌های تحقیقاتی خواهد شد (Lin, Hou and Wu 2016). این نتایج محرک اصلی ایجاد و توسعه سامانه‌های پیشنهاددهنده^۲ در زمینه‌های علمی، از جمله پیشنهاد شبکه همکاری نویسندگان، پیشنهاد استناد، پیشنهاد مقاله و پیشنهاد محل انتشار مانند نشریه‌ها و سمینارها بود. در این میان از اواسط سال ۲۰۱۴، پایگاه‌های اطلاعاتی تجاری بزرگ مانند «الزویر»^۳، «وایلی»^۴ و «اشپرینگر»^۵، اقدام به ایجاد سامانه‌های پیشنهاددهنده نشریه برای چاپ مقاله کردند. اما، هیچ‌یک از سامانه‌های پیشنهاددهنده نشریه در سطح بین‌المللی، نشریه‌های فارسی زبان را پوشش نمی‌دهند و فارغ از مبحث زبان نیز، این سامانه‌ها فقط به پیشنهاد نشریه‌های موجود در پایگاه‌های خاص خود می‌پردازند. بدین ترتیب، نویسندگان ممکن است، بسیاری از نشریه‌هایی را که به‌صورت بالقوه می‌تواند محمل انتشار یافته‌های پژوهشی آن‌ها باشد، از دست بدهند. شواهد، حاکی از آن است که در ایران سالانه بالغ بر ۶۰۰۰۰ مقاله علمی - تحقیقاتی در بیش از ۱۷۰۰ نشریه علمی معتبر (وزارت علوم، تحقیقات و فناوری) و «وزارت بهداشت، درمان و آموزش پزشکی» منتشر

1. impact factor

2. recommender system

3. Elsevier

4. Wiley

5. Springer

می‌گردد^۱. از این رو، مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری^۲ در راستای رسالت فراهم‌آوری خدمات مورد نیاز کاربران، سامانه پیشنهاددهنده نشریه جهت چاپ مقاله را مشتمل بر نشریه‌های داخل کشور، تحت عنوان «سامانه ژورنال‌یاب رایست»^۳، طراحی و پیاده‌سازی نمود.

غالب مطالعات در زمینه سامانه‌های پیشنهاددهنده از به کارگیری چهار رویکرد اصلی پالایشی مشارکتی^۴، مبتنی بر محتوا^۵، جمعیت‌شناختی^۶ و ترکیبی^۷ در این سامانه‌ها نام برده‌اند (Mohamed, Khafagy and Ibrahim 2019; Fayyaz et al. 2020). در رویکرد مشارکتی، از رفتار گذشته کاربر برای تولید پیشنهاد استفاده و اقلامی توصیه می‌شود که توسط وی یا سایر کاربران با سلاقی مشابه انتخاب شده بودند (Das, Sahoo and Datta 2017). در رویکرد محتوای محور، ویژگی‌های توصیفی اقلام^۸ برای تولید و ارائه پیشنهاد مورد استفاده قرار می‌گیرد (Aggarwal 2016). این رویکرد بیشتر برای پیشنهاد محتوایی به کار می‌رود که حاوی اطلاعات آگاهی‌بخش و آموزنده نظیر کتاب، صفحات وب، انتشارات و اخبار باشند (Park et al. 2011; Isinkaye, Folajimi and Ojokoh 2015) و به منظور یافتن شباهت میان مدارک برای تولید پیشنهاد، از مدل‌های مختلفی مانند مدل فضای برداری از جمله فراوانی عبارت ضربدر معکوس فراوانی مدرک^۹ یا مدل‌های احتمالاتی نظیر مدل طبقه‌بندی و خوشه‌بندی^{۱۰} استفاده می‌شود. سامانه‌های پیشنهاددهنده جمعیت‌شناختی همان‌گونه که از نامشان برمی‌آید، بر پایه اطلاعات جمعیت‌شناختی کاربران مانند سن، جنسیت، تحصیلات و منطقه زندگی برای یافتن شباهت و ارائه پیشنهاد استفاده می‌کنند (Sharma and Singh 2016). سه سامانه محتوای محور، مشارکتی و جمعیت‌شناختی در بسترهای مختلف به صورت متفاوتی عمل می‌کنند. بنابراین، در بسیاری از موارد که تنوع ورودی‌های در دسترس زیاد است، پتانسیل ترکیب رویکردهای سامانه‌های پیشنهاددهنده برای یک کار واحد وجود دارد. به همین دلیل، رویکرد ترکیبی معرفی شد که در آن، دو یا چند روش مختلف با یکدیگر ترکیب می‌شود تا دقت^{۱۱} و عملکرد سامانه‌های پیشنهاددهنده را بهبود

1. <https://jcr.isc.ac/main.aspx>

۲. در ادامه متن، از مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری با عنوان «مرکز منطقه‌ای» نام برده خواهد شد.

3. Ricest

4. collaborative

5. content-based

6. demographic

7. hybrid

8. item

9. Term Frequency - Inverse Document Frequency (TF-IDF)

10. probabilistic classification and clustering

11. precision

بخشد (Göksedef and Gündüz-Öğüdücü 2010). به‌رغم کاربردها و مزایای سامانه‌های پیشنهاددهنده، این سامانه‌ها با توجه به رویکردهای مورد استفاده با چالش‌هایی از جمله شروع سرد، کمبود داده، پراکندگی داده، تنوع^۴ و مقیاس‌پذیری^۵ مواجه هستند (Ricci, 2020; Fayyaz et al. 2016; Khusro, Ali and Ullah 2016; Rokach & Shapira 2015). از این رو، پژوهشگران می‌کوشند در مواجهه با این چالش‌ها، عملکرد سامانه‌ها ارتقا یافته و دقت نتایج ارائه‌شده بهبود یابد.

از بین رویکردهای پیش گفته، سامانه‌های پیشنهاددهنده نشریه به‌طور معمول، با رویکرد محتوامحور پیاده‌سازی شده و از ویژگی‌های مقاله مانند عنوان، چکیده و واژه‌های کلیدی در تولید و ارائه پیشنهاد استفاده می‌کنند (Aggarwal 2016)، زیرا اثربخشی آن‌ها در بازنمایی متن تأیید شده است (Jenuwine & Floyd 2004). چالش‌های اساسی کمبود و پراکندگی داده در تولید و ارائه پیشنهاد در این رویکرد، همواره مطرح است (Ricci, Rokach and Shapira 2015; Khusro, Ali & Ullah 2016; Mohammad et al. 2019; Fayyaz et al. 2020). در این سامانه‌ها که عملکرد الگوریتم وابسته به نحوه بازنمایی^۶ اقلامی است که در پایگاه داده رابطه‌ای ذخیره می‌شوند، اگر اطلاعات کافی در مورد محتوا برای تولید پیشنهاد وجود نداشته باشد، پایگاه با کمبود داده مواجه شده و پیشنهادها دقیق نخواهد بود. همچنین، چنانچه داده‌ها به‌درستی سازماندهی نشوند، دچار پراکندگی شده و مطابقت به‌درستی انجام نخواهد شد (Beel et al. 2016).

با توجه به این که سامانه ژورنال‌یاب «رایست» نیز از رویکرد محتوامحور و شباهت‌سنجی عنوان، واژه‌های کلیدی و چکیده مقاله با منابع موجود در سامانه، جهت ارائه پیشنهاد استفاده می‌کند، همانند بسیاری از سامانه‌های پیشنهاددهنده نشریه با دو چالش پیش گفته روبه‌روست. نتایج پژوهش‌هایی که در جهت بهبود دو چالش مذکور گام برداشته‌اند، نشان داد که کمیت و سازماندهی داده‌هایی که یک سامانه پیشنهاددهنده مبتنی بر محتوا بر آن‌ها استوار است، تأثیر مهمی بر دقت نتایج پیشنهادی دارد (Forrester, Björk and Tenopir 2017; Wang, Wang and Zhang 2018; Heinrich et al. 2021; Fayyaz et al. 2020). هرچند تجربیات جهانی در این زمینه در برخی مطالعات به کار رفته است، اما به دلیل محرمانه بودن الگوریتم‌ها، اطلاعات به‌گونه‌ای ارائه نشده که از آن طریق

1. cold start
4. diversity

2. lack of data
5. scalability

3. data sparsity
6. represent

بتوان برای زبان فارسی و به صورت خاص برای سامانه ژورنال‌یاب «رایسست» که بیشتر محتوای آن به زبان فارسی است، بهبود ایجاد کرد. بر این اساس، ارزیابی این سامانه از منظر دو چالش پیش گفته به منظور سنجش میزان اثربخشی و همچنین، هزینه-سودمندی را می‌توان یکی از ضروریات اجتناب‌ناپذیر تحقیقات این زمینه دانست. در مجموع نیز بهبود عملکرد سامانه‌های اطلاعات علمی، به خصوص در زبان فارسی و سایر زبان‌های دارای دبیره مشابه، به معنای بهبود دسترسی به محتوای تولیدشده برای مخاطبان و گسترش علم در سطح منطقه خواهد بود. همچنین، شناسایی راهکارهایی در راستای بهبود اثربخشی نتایج این سامانه می‌تواند به ارائه الگو جهت طراحی سامانه‌های مشابه، به منظور سهولت دسترسی به منابع مرتبط بیانجامد. از این رو، پژوهش حاضر اثربخشی پیشنهادی سامانه ژورنال‌یاب «رایسست» را در دو مرحله مورد آزمون قرار داده است:

۱. اثربخشی نتایج پیشنهادی ارائه شده توسط سامانه ژورنال‌یاب «رایسست»
۲. اثربخشی نتایج پیشنهادی ارائه شده توسط سامانه ژورنال‌یاب «رایسست» پس از بهبود چالش‌های کمبود و پراکندگی داده‌ها

۲. مرور پیشینه‌های پژوهش

۲-۱. چالش کمبود داده در سامانه‌های پیشنهاددهنده

از آنجا که تولید و ارائه پیشنهاد در سامانه‌های پیشنهاددهنده محتوای محور با محتوای زبانی سروکار دارد، اطلاعات توصیفی اقلام باید با استفاده از پردازش زبان طبیعی تجزیه و تحلیل شوند تا استفاده از این عناصر برای ماشین امکان‌پذیر گردد (Balyan, McCarthy and McNamara 2020). از این رو، نتایج زمانی قابل اعتماد و مفید خواهند بود که داده‌های کافی در اختیار باشد. چنانچه اطلاعات کافی در مورد محتوا برای تولید پیشنهاد وجود نداشته باشد، یا داده‌ها دارای نواقصی همچون ناسازگاری، خطا و مقادیر گمشده باشد، ارائه پیشنهاد با دشواری مواجه شده و توصیه دقیق نخواهد بود (Fayyaz et al. 2020). پژوهش‌های انجام شده در زمینه‌های گوناگون، مانند شبکه‌های اجتماعی (Camacho and Zhang, 2019; Alves-Souza 2018; Eirinaki et al. 2018; Zoetekouw 2019) تجارت الکترونیک (Kang, Doornenbal and Schijvenaars 2019; Zhao 2019; Abbas and Sun 2019) و باز یابی اطلاعات (Kanakia et al. 2015) نشان داده است که کمبود داده ساخت یافته نتایج سامانه را تحت تأثیر قرار می‌دهد. بدیهی است که در زبان فارسی به دلیل پیچیدگی‌ها و ویژگی‌های

خاص آن، مانند راست به چپ بودن، عدم مشخص بودن مرز واژه، پیوسته و جدانویسی و پیچیدگی‌های بی‌شمار موجود در واژگان بیشتر مستعد خطا بوده و این چالش در محیط دیجیتال بر تأثیر پردازش خودکار متن و نتایج جست‌وجو اثرگذار است (QasemiZadeh, Rahimi and Ghalati 2014). پژوهش‌های انجام‌گرفته هر یک به‌نوعی مسایل زبان فارسی را در تعامل انسان با رایانه، پایگاه‌های اطلاعاتی و موتورهای کاوش بررسی نموده و در بیشتر موارد به ویژگی‌های نوشتاری زبان فارسی معطوف بوده است. «حری» (۱۳۷۲) و نشاط (۱۳۷۹) به تبیین مشکلات رسم‌الخط فارسی در رویارویی با فناوری‌های اطلاعاتی پرداخته و استفاده از راهکارهایی مانند سیاست‌گذاری واحد برای یکسان‌سازی رسم‌الخط را پیشنهاد داده‌اند. چنین راهکارهایی نیازمند مشارکت فعالانه و آموزش نویسندگان متون (ماشین‌نویس‌ها و کاربران) است و از روندی بلندمدت و هزینه‌بر برخوردار است. از این رو، حرکت به‌سوی راهکارهای خودکارسازی پردازش متن و نمایه‌سازی ضروری است (ستوده و هنرجویان ۱۳۹۱). شناسایی چالش‌های ریخت‌شناسی و توجه به هنجارسازی چنددستی نگارشی آن‌ها می‌تواند چالش کمبود داده‌های ساخت‌یافته در زبان فارسی را بهبود بخشیده و بر اثربخشی نتایج سامانه‌های ذخیره و بازیابی اطلاعات تأثیرگذار باشد (آخشیک و فتاحی ۱۳۹۱؛ ستوده و هنرجویان ۱۳۹۱، رنجبر و عباس‌پور ۱۳۹۷؛ کامیابی گل و همکاران ۱۳۹۷؛ هماوندی، نوروزی و حسینی بهشتی ۱۳۹۷).

۲-۲. چالش پراکندگی داده در سامانه‌های پیشنهاددهنده

یک نقطه ضعف بحرانی برای سیستم‌های پیشنهاددهنده امروزی، چالش پراکندگی داده‌هاست. این مشکل به‌طور مستقیم بر دقت نتایج سامانه تأثیر می‌گذارد و باعث می‌شود پیشنهادها به‌درستی ارائه نشده و قابل اعتماد نباشد (Schedl et al. 2018). چالش پراکندگی از دو بعد مبتنی بر کاربر و مبتنی بر اقلام قابل بررسی است. در سامانه‌های پیشنهاددهنده مبتنی بر کاربر، ابتدا ماتریسی از کاربران/اقلام ایجاد می‌شود و نبود رتبه‌بندی اقلام توسط کاربران منجر به ایجاد ماتریس تُنک و خلوتی می‌گردد که شباهت‌سنجی و ارائه پیشنهاد را با دشواری مواجه خواهد کرد (Sachan and Richariya 2013; Thorat, Goudar and Barve 2015; Schedl et al. 2018; Wang, Wang & Zhang 2018). در سامانه‌های پیشنهاددهنده مبتنی بر اقلام مانند سامانه‌های پیشنهاددهنده نشریه که از رتبه‌بندی کاربران استفاده نمی‌شود، این پراکندگی ماتریس به اطلاعات توصیفی و محتوایی اقلام برمی‌گردد (Sarwar et al. 2000).

در مجموع، نتایج پژوهش‌های Wang, Wang and Yeung (2015); Wang, Wang and Zhang (2018) نشان داده است که هرچه کاربران/ اقلام بیشتری درگیر محاسبات شباهت‌سنجی شوند، پراکندگی کمتری در مجموعه داده ایجاد خواهد شد. همچنین، نتایج پژوهش Dror et al. (2011) نشان داده است که مقادیر پراکنده داده در سامانه‌های پیشنهاددهنده از جمله موسیقی یا هو، فیلم نتفلیکس^۱ و مووی‌لنز^۲ چالش‌هایی را برای ارائه نتایج اثربخش ایجاد کرده است.

۲-۳. استفاده از فنون داده‌کاوی جهت بهبود چالش کمبود داده

اگرچه سامانه‌های پیشنهاددهنده بر استفاده گسترده‌ای از تکنولوژی‌ها، مانند پردازش زبان طبیعی و یادگیری ماشین استوار است، اما آنچه نقش اصلی را در عملکرد صحیح الگوریتم‌ها برای پیش‌بینی و ارائه پیشنهاد دارد، یک پایگاه دانش غنی است که از روش‌های مختلف و با استفاده از مدارک موجود در مجموعه ساخته می‌شود. با آنکه واژگان طبیعی به کار رفته در مدارک علمی می‌تواند به بازنمون محتوای آن‌ها کمک کند، اما به دلیل ماهیت ساخت‌نیافته، نیاز به پردازش‌هایی دارد که بتواند منجر به داده‌های قابل استفاده برای ماشین‌گردد. در نتایج مطالعات مربوط به چالش کمبود داده، استفاده از اطلاعات اضافی دامنه و تولید داده با کیفیت بالا به‌عنوان راه‌حلی برای بهبود این چالش پیشنهاد شده است (Sarwar et al. 2001; Sachan and Richariya 2013). همچنین، استفاده از روش‌های مختلف یادگیری ماشین جهت استخراج واژه‌های کلیدی مناسب می‌تواند به افزایش قابل توجه عملکرد سامانه و به‌طور برجسته در سه پیشنهاد اول عملکرد خوبی داشته باشد (Hong et al. 2020). در پژوهشی مبتنی بر داده‌های «پابمد»^۳ نیز استفاده از بردار کلمات در کنار یادگیری عمیق، دقت نتایج سامانه را افزایش داده است (Wu et al. 2021).

در محدود پژوهش‌های انجام‌شده در قلمرو سامانه‌های پیشنهاددهنده نشریه، نتایج بررسی سامانه پیشنهاددهنده «الزویر» نشان داد که پس از نرمال‌سازی عبارت‌ها و واژه‌های کلیدی اثربخشی نتایج بیش از ۱۰ درصد و با بهینه‌سازی الگوریتم به همراه نرمال‌سازی واژه‌ها، و اثربخشی نتایج پیشنهادی سامانه بیش از ۲۰ درصد بهبود یافت (Kang, Doornenbal and Schijvenaars 2015).

1. netflix

2. movielens

3. PubMed

۲-۴. استفاده از فنون داده‌کاوی جهت بهبود چالش پراکندگی داده

جمع‌آوری داده‌ها در یک مجموعه به خودی خود ارزش بسیار بالایی ندارد. ارزش این داده‌ها زمانی به حداکثر خود می‌رسد که با اطلاعات جانبی با عنوان فراداده غنی شده و نیز مدیریت شوند. از این رو، شناسایی و استفاده از ویژگی‌های منتخب در تکنیک‌های برچسب‌زنی و دسته‌بندی داده‌ها می‌تواند در سرعت و کیفیت نتایج سامانه‌های پیشنهاددهنده اثرگذار باشد (Lucas, Segre and Moreno 2012). نتایج پژوهش‌هایی در زمینه بهبود چالش پراکندگی داده در سامانه‌های پیشنهاددهنده نشان داده است که برچسب‌گذاری موضوعی برای دسته‌بندی یا طبقه‌بندی متون مبتنی بر خصیصه ساختاری یا معنایی (Lampert 2009)، یکپارچه‌سازی اطلاعات اقلام (Liu et al. 2015)، به‌کارگیری مدل‌سازی موضوعی (Yao et al. 2015)، حذف اقلامی که اهمیت چندانی ندارند (Kant and Mahara 2018)، استفاده از هستی‌شناسی‌ها (Nilashi, Ibrahim and Bagherifard 2018) و دسته‌بندی داده‌ها (Wei et al. 2012; Yin et al. 2020) می‌تواند منجر به کاهش پراکندگی داده‌ها و در نتیجه، افزایش دقت در نتایج پیشنهادی سامانه گردد.

۲-۵. جمع‌بندی

انواع داده‌ها در سامانه‌های پیشنهاددهنده به دلیل وجود فناوری‌های پیشرفته به سرعت در حال رشد هستند و چگونگی مقابله با این چالش‌ها در جهتی که بتواند علایق کاربران در زمینه‌های گوناگون را با مجموعه‌ای از روش‌های یادگیری ماشینی تأمین کند، ضروری است (Su and Chiu 2016). آنچه از نتایج پژوهش‌ها حاصل می‌شود، استفاده از روش‌های مختلف جهت کشف یا استخراج دانش از مستندات متنی است که بنا به هدف، نوع، و سطح پژوهش نتایج متفاوتی در پی داشته است.

جدول ۱. خلاصه مرور پیشنهادها

چالش مورد مطالعه	نام نویسنده / نویسندگان	عنوان مقاله	راهکار پیشنهادی برای بهبود چالش
کمبود داده	Hong et al. (2020)	فرصت‌ها و چالش‌های روش‌های یادگیری عمیق برای داده‌های الکتروکاردیوگرام: مرور سیستماتیک	استفاده از روش‌های مختلف یادگیری ماشین جهت استخراج واژه‌های کلیدی
	Kang, Doornenbal and Schijvenaars (2015)	پیشنهاددهنده‌ی نشریه الزویر: پیشنهاد نشریه برای شما	نرمال‌سازی عبارت‌ها، واژه‌های کلیدی و بهینه‌سازی الگوریتم
	وو و همکاران	الگوریتم پیشنهاد مشارکتی مبتنی بر قلم با ترکیب دسته‌بندی اقلام و علائق	استفاده از بردار کلمات در کنار یادگیری عمیق
پراکندگی داده	Yin et al. (2020)	استفاده از روش‌های پردازش زبان طبیعی و روش‌های یادگیری سلسله‌مراتبی ماشینی برای طبقه‌بندی متون	دسته‌بندی داده‌ها
	Kant and Mahara (2018)	ترکیب رویکرد مشترک مبتنی بر کاربر و قلم برای کاهش پراکندگی داده‌ها	حذف اقلام زاید
	Nilashi, Ibrahim & Bagherifard (2018)	سیستم پیشنهاددهنده‌ی مشارکتی با استفاده از روش‌های هستی‌شناسی و کاهش ابعاد	استفاده از هستی‌شناسی‌ها
	Liu et al. (2015)	طبقه‌بندی اقلام مبتنی بر ماشین بولتزمن	یکپارچه‌سازی اطلاعات اقلام
	Yao et al. (2015)	رتبه‌بندی موضوعی: استفاده از فراداده برای کاهش پراکندگی	مدل‌سازی موضوعی

کشف روش‌های جدید، به کارگیری فنون متن‌کاوی در قلمروهای مختلف مطالعاتی، دسته‌بندی موضوع‌های مشابه و مرتبط به هم، صرفه‌جویی در وقت و هزینه‌های جست‌وجو و رضایت کاربران از نتایج به‌دست‌آمده از این مطالعات هستند. نتایج پژوهش‌ها نشان داد که استفاده از فنون داده‌کاوی در دسترسی و بازیابی اطلاعات و منابع مرتبط با یکدیگر می‌تواند مفید و مؤثر واقع شود. اما همان‌گونه که مشاهده شد، با وجود افزایش قابل توجه تعداد رویدادها و نشریه‌های علمی و اهمیت این‌گونه سامانه‌ها در تسریع فرایند انتشار و کاهش اشتباهات احتمالی، پژوهشی که به‌طور مستقیم به ارزیابی سامانه‌های پیشنهاددهنده‌ی نشریه پرداخته باشد، یافت نشد. از این رو، به‌دلیل نوپا بودن این سامانه‌ها

شناسایی عواملی که بتواند افزایش اثربخشی نتایج پیشنهادی و رضایت کاربر را به دنبال داشته باشد، به‌عنوان عاملی برای بقا یا زوال سامانه ضروری است.

۳. روش پژوهش

۳-۱. معماری سامانه ژورنال‌یاب «رایست»

در این سامانه که از رویکرد محتوامحور و شباهت‌سنجی مشارکتی بهره می‌برد، اطلاعات متنی تمامی مدارک پیکره طی یک فرایند برون‌خطی استخراج و پردازش شده است. بدین ترتیب، که ترکیب‌های واژه‌های تکی و دوبه‌دو از عناصر متنی مقالات استخراج و به شکل بردار برای هر مدرک محاسبه و ذخیره شده است. پس از وارد کردن عنوان و چکیده مقاله کاربر به سامانه، این اطلاعات مورد پردازش قرار گرفته و بسامد واژه‌های کلیدی محاسبه و بر اساس آن، برداری از وزن واژه‌های کلیدی تشکیل می‌گردد. برای تبدیل هر مدرک به بردار از ویژگی فراوانی عبارت ضربدر معکوس فراوانی مدرک، و برای پیاده‌سازی مدرک در فضای برداری از تک‌واژه و دو واژه‌های استخراج‌شده از مدرک استفاده می‌شود. بردار به‌دست آمده به‌عنوان نماینده مقاله است. سرانجام، با استفاده از معیار شباهت کسینوسی و با مقایسه زاویه بین بردار پرسش و بردار مدارک موجود در پیکره، میزان شباهت بین بردار پرسش و بردارهای مدارک پیکره، محاسبه و نشریه هدف تعیین می‌گردد. شایان ذکر است که در لایه پایگاه داده این سامانه، نشریه‌ای که مقاله در آن منتشر شده، به‌عنوان مرتبط‌ترین نشریه جهت چاپ مقاله برچسب خورده و بر این اساس، قضاوت‌هایی برای ارزیابی ربط مدارک بازیابی شده وجود دارد.

۳-۲. تشکیل پیکره

با توجه به پراکندگی و تعدد موضوعی داده‌های موجود در پایگاه داده سامانه ژورنال‌یاب «رایست»، به‌منظور آزمون تأثیر دو چالش کمبود و پراکندگی داده، نیاز به پیکره آزمون وجود داشت. به‌دلیل نبود پیکره جامع مقاله‌های نشریه‌های علمی داخل کشور که بیشتر محتوای آن به زبان فارسی است، فرایندی جهت تولید پیکره و ایجاد امکان ارزیابی نتایج ضروری بود. نخستین مرحله در ایجاد پیکره، استخراج و تشکیل مجموعه‌ای از متون مربوط بود. برای این منظور، نشریه‌های فنی و مهندسی انتخاب شد. دلیل انتخاب این قلمرو موضوعی وجود تعداد قابل قبولی نشریه بود که دایره وسیعی از

کلمات مربوط به موضوع را پوشش می‌داد و بدین ترتیب، قابلیت ارزیابی و تعمیم نتایج وجود داشت.

فراهم آوردن امکان مقایسه یافته‌ها پیش و پس از بهبود چالش‌های مورد بررسی مستلزم وجود پایگاه داده مشترک میان سامانه ژورنال‌یاب «رایست» و پیکره آزمون بود. از این رو، ابتدا فهرستی از نشریه‌های فنی و مهندسی موجود در سامانه و منطبق با فهرست نشریه‌های معتبر «وزارت علوم، تحقیقات و فناوری» استخراج شد. از مجموع فهرست نشریه‌های معتبر فنی و مهندسی، داده‌های ۱۱۰ عنوان نشریه در سامانه وجود داشت. از این رو، برای تشکیل پیکره آزمون نیز این ۱۱۰ عنوان نشریه انتخاب شد. از آنجا که به‌منظور عملکرد صحیح الگوریتم‌ها برای پیش‌بینی، تولید و ارائه پیشنهادها وجود حداقل ۱۰۰ مقاله برای هر عنوان نشریه ضروری است (Kang, Doornenbal and Schijvenaars 2015)، با فرض بر وجود ۲۴ مقاله برای هر نشریه در طول هر سال، بازه زمانی پنج‌ساله ۱۳۹۲ الی ۱۳۹۶ انتخاب گردید. کلیه مقاله‌های منتشرشده در بازه زمانی مورد نظر، از طریق وبسایت نشریه‌ها گردآوری و ذخیره شد. بدین ترتیب، در نهایت، برای تشکیل پیکره مجموعه‌ای بالغ بر ۱۵۰۰۰ مقاله جمع‌آوری گردید.

تجزیه و تحلیل داده‌ها با استفاده از نرم‌افزار «اس‌پی‌اس‌اس»^۱ نسخه ۲۳ انجام شد. به‌منظور مقایسه دوه‌دو رتبه‌ها در هر یک از مراحل بهینه‌سازی، از آزمون «ویلکا کسون» استفاده شد. همچنین، با توجه به نرمال نبودن توزیع داده‌ها، جهت مقایسه رتبه‌ها در سه مرحله و به‌دلیل وابستگی نمونه‌ای (رتبه تخصیص داده‌شده به هر نشریه)، از آزمون ناپارامتری «فریدمن» استفاده شد.

۳-۳. غنی‌سازی و پردازش داده‌ها

عناصر متنی مقاله‌های پیکره باید به داده‌هایی تبدیل شوند که به‌وسیله ماشین قابل خواندن باشند. مطالعات نشان می‌دهد که ابزارهای استاندارد پیش‌پردازش و نرمال‌سازی ایجادشده برای متون زبان فارسی به‌صورت رایگان منتشر نشده یا دقت مناسب را ندارند (کامیابی و همکاران ۱۳۹۷). از این رو، مراحل با کدنویسی در برنامه «پایتون» نسخه ۳/۶ انجام شد. این فرایند با هدف کاهش ابعاد فضای نمایش واژه‌های موجود در سند انجام

1. SPSS

شد. نتایج پژوهش‌ها نشان می‌دهد که هرچه دقت بیشتر در این مرحله به کار رود، اثربخشی بیشتر نتایج سامانه را به دنبال خواهد داشت (Berbatova 2019; Wang et al. 2020). (Gupta and Lehal 2009:). بر این اساس، دو محور اصلی پیش‌پردازش و نرمال‌سازی داده‌ها مد نظر قرار گرفت.

ابتدا جهت استانداردسازی متن، پیش‌پردازش‌هایی روی آن‌ها انجام شد. به دلیل نزدیکی دبی‌ره زبان فارسی با عربی، همواره در نگارش تعدادی از حرف‌ها مشکل کاراکترهای عربی معادل وجود دارد. از جمله آن‌ها می‌توان به حروف «ک»، «ی» و همزه اشاره کرد. در گام نخست، مشکلات مربوط به این حروف با یکسان‌سازی آن‌ها برطرف شد. در این مرحله همه نویسه‌ها (حروف) متن با معادل استاندارد آن جایگزین و یکسان‌سازی شد. افزون بر این، اصلاح نویسه نیم‌فاصله و فاصله در کاربردهای مختلف آن و همچنین، حذف نویسه «ا»، تشدید، تنوین و موارد مشابه (کامیابی و همکاران ۱۳۹۷؛ Seraji 2013) از جمله اقدام‌های لازم پیش از شروع پردازش متن بود. سپس، تنوع ریخت‌شناسی واژه‌ها در زبان فارسی طبق آخرین یافته‌ها از متون استخراج و مورد استفاده قرار گرفت (هماوندی، نوروزی و حسینی بهشتی ۱۳۹۷؛ پرئی و حمیدی ۱۳۹۶؛ ستوده و هنرجویان ۱۳۹۱). این بخش شامل حذف ایست‌واژه‌ها از جمله حروف اضافه، بسیاری از قیده‌ها، حروف ربط و فعل‌ها بود. بدین ترتیب، پیش‌پردازش انجام شد.

حذف مواردی مانند حروف اضافه، ربط، قیود، افعال که اهمیت چندانی در شباهت‌سنجی متن ندارند (Kant and Mahara 2018) و همچنین یکپارچه‌سازی واژه‌ها (Liu et al. 2015) به‌عنوان راه‌حل‌هایی برای بهبود چالش کمبود داده در چنین سامانه‌هایی پیشنهاد شده است. بنابراین، رویکرد این مطالعه در بخش پیش‌پردازش، استفاده از روش نیمه‌خودکار بود. بدین ترتیب که بررسی دستی نیز توسط پژوهشگر انجام شد و تا جای ممکن نواقص و اشتباه‌های نگارشی که ناشی از خطاهای تایپی در هنگام ورود یا مشکلاتی در انتقال داده بود، برطرف گردید. با انجام مراحل بالا، پایگاه داده تا حد ممکن پردازش و پاک‌سازی گردید. در این گام، چالش کمبود داده بهبود یافت.

۴-۳. دسته‌بندی داده‌ها

نشریات معتبر «وزارت عتف» در شش گروه موضوعی اصلی (سطح کلان) و بیش ۶۵ زیرگروه موضوعی (سطح خرد) تقسیم‌بندی شده‌اند. به‌منظور کاهش پراکندگی داده‌ها از دسته‌بندی موضوعی سطح خرد که از پیش توسط «وزارت عتف» برای نشریه‌های تعیین شده است، استفاده شد.

با توجه به این که مفروض است هر مقاله در نشریه‌ای مرتبط با موضوع مقاله منتشر شده است، برای وزندهی به کلیدواژه‌ها می‌توان از قلمروهای موضوعی نشریه استفاده نمود. از طرفی، هر کلیدواژه ممکن است در چندین مقاله وجود داشته باشد و از آنجا که هر مقاله در یک نشریه به چاپ می‌رسد، از این رو، یک کلیدواژه ممکن است در نشریه‌های متفاوتی وجود داشته باشد. وجود رابطه تعدی^۱ با فرمول $xRy \& yRz \rightarrow xRz$ (Robinson 1964) میان کلیدواژه، مقاله و نشریه موجب می‌شود که یک کلیدواژه موضوع‌های متفاوتی را دربرداشته باشد. اگر x با y در ارتباط باشد و y با z در نتیجه، x با z در ارتباط خواهد بود. فرض کنید x نشانگر کلیدواژه، y نشانگر مقاله و z نشانگر نشریه باشد. از آنجا که کلیدواژه با مقاله و مقاله با نشریه در ارتباط است، در نتیجه، کلیدواژه با نشریه نیز در ارتباط خواهد بود. بنابراین، هر کلیدواژه می‌تواند در چندین نشریه با موضوع‌های متفاوت به کار گرفته شده باشد. بدین ترتیب، بر اساس موضوع‌های از پیش تعیین شده برای هر عنوان نشریه، جدولی با عنوان دسته موضوعی در پایگاه داده رابطه‌ای تعریف، و پیوند آن با جدول‌های واژه‌های کلیدی و عنوان‌های نشریه‌ها ایجاد گردید.

۳-۵. ارزیابی اثربخشی نتایج سامانه

به‌منظور ارزیابی اثربخشی نتایج پیشنهادی سامانه از روش اعتبارسنجی یک‌طرفه و معیار دقت در نقطه k^2 استفاده شد. در این روش تعدادی مقاله که در نشریه‌های مرتبط منتشر شده‌اند، به‌صورت تصادفی انتخاب و از پایگاه داده حذف می‌شوند. سپس، عناصر متنی این مقاله‌ها به‌عنوان پرسمان به سامانه وارد می‌شود. چنانچه ۳ یا ۱۰ پیشنهاد برتر دربردارنده نشریه‌ای بود که مقاله ورودی در آن منتشر شده است، به‌عنوان یک پیشنهاد درست، و در غیر این صورت به‌عنوان یک توصیه اشتباه محاسبه می‌گردد (Kang, Doornenbal and Schijvenaars 2015; Feng et al. 2019). بدین منظور، با توجه به تعداد

1. transitive

2. p@k

کل مقاله‌های گردآوری شده (۱۵۰۰۰ مقاله)، با استفاده از فرمول (Cochran 1977)، تعداد ۴۰۰ مقاله که در نشریه‌های مورد بررسی منتشر شده بودند، به روش تصادفی-تناسبی (دیانی ۱۳۹۲، ۷۶) انتخاب و از پیکره آزمون خارج گردید. سپس، عناصر متنی مقاله‌های انتخاب‌شده به‌عنوان یک پرسمان جدید، توسط پژوهشگر به سامانه وارد و نتایج پیشنهادی ذخیره و مورد ارزیابی قرار گرفت.

۴. یافته‌ها

۴-۱. اثربخشی نتایج پیشنهادی سامانه ژورنال‌یاب «رایست» پیش و پس از بهبود کمبود داده

در این مرحله، نتایج در وضعیت پیش و پس از بهبود کمبود داده آزمون شد. فراوانی نسبی رده‌ها نشان می‌دهد که در وضعیت فعلی، نشریه هدف در ۵۵ درصد پرسمان‌ها در رتبه‌های بالاتر از ۱۰ ($k > 10$) و تنها در ۲۶ درصد پرسمان‌ها در ۳ رتبه نخست ($k=1$ و $k=3$) توسط سامانه پیشنهاد شده است (جدول ۲).

جدول ۲. اثربخشی نتایج پیش و پس از بهبود کمبود داده

نقطه دقت	بهبود اثربخشی	
	پیش از بهبود (درصد)	پس از بهبود (درصد)
۱	۱۴	۱۸
۳	۱۲	۲۳
۶	۱۲	۲۰
۹	۷	۸
۱۰ و بالاتر	۵۵	۳۱

پس از بهبود چالش کمبود داده، در ۴۱ درصد پرسمان‌ها نشریه هدف در سه رتبه نخست ($k=1$ و $k=3$) قرار داشت. در این مرحله، اثربخشی نتایج سامانه نسبت به وضعیت پیش، به میزان ۱۵ درصد بهبود داشت (جدول ۱). در ادامه، مقایسه دو به دو رتبه‌ها با استفاده از آزمون «ویلکاکسون» انجام شد (جدول ۳).

جدول ۳. نتایج آزمون «ویلکاکسون» پیش و پس از بهبود کمبود داده

Z	-۵/۵۸۶
سطح معناداری	۰/۰۰۰

نتایج حاکی از آن است که رتبه‌های مرحله اول ارزیابی ژورنال یاب «رایست» (با میانۀ ۱۰) با رتبه‌های مرحله دوم و پس از بهبود چالش کمبود داده (با میانۀ ۶) در سطح یک درصد تفاوت آماری معناداری دارند. به بیان دیگر، بهبود این چالش منجر به افزایش اثربخشی در نتایج سامانه گردید.

۴-۲. اثربخشی نتایج پیشنهادی سامانه ژورنال یاب «رایست» پیش و پس از بهبود چالش پراکندگی داده

در این مرحله نتایج، پیش و پس از بهبود پراکندگی داده آزمون شد. فراوانی نسبی رده‌ها نشان می‌دهد که پس از دسته‌بندی داده‌ها، اثربخشی نتایج در سه رتبه نخست ($k=1$) و ($k=3$)، نسبت به مرحله پیش ۳۰ درصد بهبود یافت و تنها در ۹ درصد پرسمان‌ها، نشریه هدف در رتبه‌های بالاتر از ۱۰ ($k > 10$)، پیشنهاد شد (جدول ۴).

جدول ۴. اثربخشی نتایج پیش و پس از بهبود پراکندگی داده

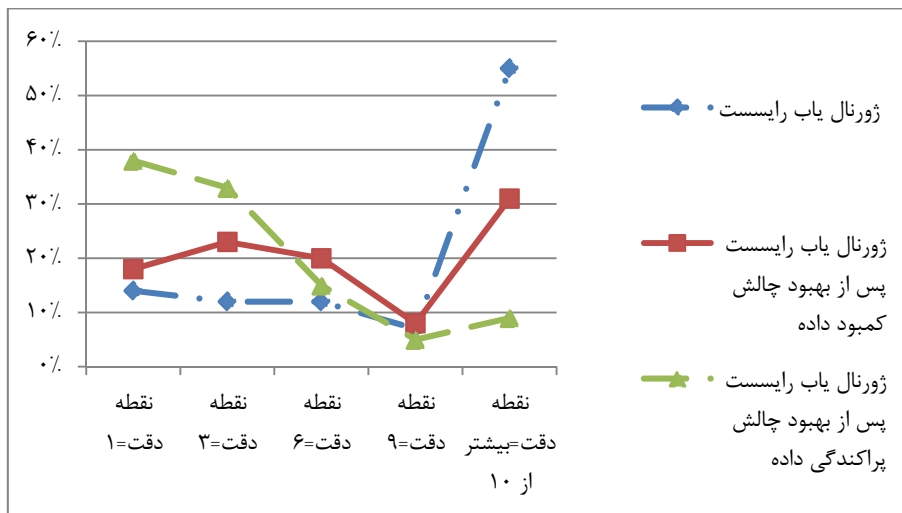
نقطه دقت	بهبود اثربخشی	
	پیش از بهبود (درصد)	پس از بهبود (درصد)
۱	۱۸	۳۸
۳	۲۳	۳۳
۶	۲۰	۱۵
۹	۸	۵
۱۰ و بالاتر	۳۱	۹

در ادامه، مقایسه دو به دو رتبه‌ها با استفاده از آزمون «ویلکاکسون» انجام شد.

جدول ۵. نتایج آزمون ناپارامتری «ویلکاکسون» پیش و پس از بهبود پراکندگی داده

Z	-۹/۰۷۷
سطح معناداری	۰/۰۰۰

نتایج نشان می‌دهد که رتبه مرحله دوم با مرحله سوم نیز تفاوت معناداری در سطح یک درصد دارند (جدول ۵). به بیان دیگر، با کاهش پراکندگی داده‌ها دقت نتایج پیشنهادی سامانه افزایش یافت.



نمودار ۱. اثربخشی نتایج پیشنهادی سامانه پیش و پس از بهبود دو چالش کمبود و پراکندگی داده

سرانجام، نتایج نشان می‌دهد که پیش از بهبود چالش‌های کمبود و پراکندگی داده، بیشترین فراوانی پیشنهاد نشریه هدف توسط سامانه ژورنال‌یاب «رایست» در رتبه‌های ۱۰ و بالاتر قرار داشت. در حالی که پس از بهینه‌سازی دو چالش پیش‌گفته، در بیش از ۷۰ درصد از پرسمان‌ها، نشریه هدف در رتبه نخست ($K=1$ و $K=3$) پیشنهاد شده است (نمودار ۱).

۳-۴. اثربخشی نتایج پیشنهادی سامانه ژورنال‌یاب «رایست»، پیش و پس از بهبود کمبود و پراکندگی داده

با توجه به نرمال نبودن توزیع داده‌ها، جهت مقایسه رتبه‌ها، در سه مرحله و به دلیل وابستگی نمونه‌ای (رتبه تخصیص داده شده به هر نشریه)، به منظور بررسی تفاوت اثربخشی نتایج پیشنهادی سامانه، پیش و پس از بهبود از آزمون ناپارامتری «فریدمن» استفاده شد.

جدول ۶. نتایج آزمون «فریدمن» پیش و پس از بهبود کمبود و پراکندگی داده

میانگین رتبه	مرحله ۱ تا ۳
۲/۴۷	ژورنال یاب رایسست (مرحله ۱)
۲/۱۱	پس از بهبود چالش کمبود داده (مرحله ۲)
۱/۴۲	پس از بهبود چالش پراکندگی داده (مرحله ۳)
۱۷۹/۰۴۲	مربع کای
** ۰/۰۰۰	سطح معناداری

** معناداری در سطح ۱ درصد

نتایج نشان می‌دهد که تفاوت معناداری بین اثربخشی نتایج در ژورنال‌یاب فعلی و پس از دو مرحله بهینه‌سازی وجود دارد (جدول ۶). بدین معنا که بهبود دو چالش کمبود و پراکندگی داده، منجر به افزایش اثربخشی نتایج سامانه ژورنال‌یاب «رایسست» شده است.

۵. بحث و نتیجه‌گیری

در این پژوهش اثربخشی نتایج پیشنهادی سامانه ژورنال‌یاب «رایسست» که نشریه‌های مرتبط جهت چاپ مقاله در نشریه‌های داخلی را پیشنهاد می‌دهد، پیش و پس از بهبود دو چالش کمبود و پراکندگی داده مورد آزمون قرار گرفت. با توجه به این که این سامانه از رویکرد محتوای محور و شباهت‌سنجی عنوان، چکیده و کلیدواژه‌های مقاله نویسنده با منابع موجود در سامانه، جهت ارائه پیشنهاد استفاده می‌کند، همانند بسیاری از سامانه‌های پیشنهاددهنده نشریه، همواره با دو چالش پیش‌گفته روبه‌روست. نتایج این پژوهش نشان داد که غنی‌سازی پایگاه داده، بهینه‌سازی فرایند پردازش، و دسته‌بندی داده‌ها می‌تواند تا حد قابل توجهی مشکل کمبود و پراکندگی داده را در سامانه‌های پیشنهاددهنده محتوای محور فارسی و سایر زبان‌های دارای دبیره مشابه مانند اردو، افغان و عربی جهت تسهیل دسترسی به نشریه‌ها، مقاله‌ها و سایر اقلام اطلاعاتی بهبود بخشد.

نتایج پژوهش حاضر با تأکید بر زبان فارسی همراستا با نتایج پژوهش‌های پیشین در قلمرو بهبود چالش کمبود داده در سامانه‌های بین‌المللی پیشنهاددهنده نشریه با پایگاه داده به زبان انگلیسی است (Kang, Doornenbal and Schijvenaars 2015). به بیان دیگر، غنی‌سازی پایگاه داده و بهبود فرایند پردازش می‌تواند فارغ از مبحث زبان منجر به بهبود

اثربخشی در نتایج سامانه‌های پیشنهاددهنده نشریه گردد. همچنین، نتایج پژوهش حاضر حاکی از آن است که دسته‌بندی یا طبقه‌بندی داده‌ها می‌تواند در کاهش پراکندگی داده و افزایش اثربخشی نتایج پیشنهادی مؤثر باشد. گرچه پیش از این، پژوهشی که از دسته‌بندی موضوعی داده‌ها جهت چالش پراکندگی داده و بهبود اثربخشی سامانه‌های پیشنهاددهنده نشریه بهره برده باشد، مشاهده نشد. اما نتایج به‌دست آمده با نتایج بسیاری از پژوهش‌ها در قلمرو بهبود چالش پراکندگی داده مانند «آمازون» همراستاست (Lampert 2009; Wei et al. 2012; Liu et al. 2015; Kant and Mahara 2018; Yin et al. 2020).

همان‌گونه که Rollins et al. (2017) نیز بیان می‌کنند، پژوهش‌های انجام شده در قلمرو سامانه‌های پیشنهاددهنده نشریه بسیار انگشت‌شمار است. همچنین، نتایج پژوهش «حسین‌زاده» (۱۴۰۰) نیز نشان داد که با وجود احساس نیاز پژوهشگران به استفاده از سامانه‌های پیشنهاددهنده نشریه، اثربخشی نتایج سامانه‌های پیشنهاددهنده نشریه از نگاه آن‌ها چندان مطلوب نیست. این دو نکته می‌تواند ضرورت انجام پژوهش‌های بیشتر و غنی‌تر در قلمرو سامانه‌های پیشنهاددهنده نشریه را نشان دهد. هرچند نبود پیکره مقاله‌های نشریه‌های علمی داخل کشور و همچنین، ابزار استاندارد پیش‌پردازش در زبان فارسی از محدودیت‌های این پژوهش بود، اما نتایج این پژوهش نشان می‌دهد که غنی‌سازی پایگاه داده، بهبود فرایند پردازش، و دسته‌بندی موضوعی داده‌ها می‌تواند مشکل کمبود و پراکندگی داده‌ها در سامانه ژورنال‌یاب «رایست» را کاهش و اثربخشی نتایج پیشنهادی این سامانه را افزایش دهد. بر این اساس، پیشنهاد می‌شود، نتایج این پژوهش در سامانه عملیاتی گردد. همچنین، انجام پژوهش‌های بیشتر در قلمرو موضوعی سامانه‌های پیشنهاددهنده نشریه نیز ضروری به نظر می‌رسد. بر این اساس، توصیه می‌شود که در پژوهش‌های آینده مقاله‌های علمی سایر قلمروهای موضوعی سطح کلان از جمله علوم انسانی، علوم پایه، کشاورزی و ... نیز به‌عنوان پیکره و نمونه پژوهش انتخاب شود تا امکان مقایسه نتایج فراهم آمده و با اطمینان بیشتری قابل تعمیم باشد. همچنین، پتانسیل استفاده از تکنیک‌های کاهش بعد داده مانند الگوریتم‌های خوشه‌بندی، استفاده از شبکه‌های اجتماعی، شبکه همکاران، مدل‌سازی موضوعی و استنادی با رویکردهای شبکه عصبی، مدل‌های معنایی، شناسایی و استخراج به‌روز چالش‌های زبان‌شناختی به‌خصوص در زبان فارسی نیز به‌منظور بهبود چالش کمبود و پراکندگی داده‌ها مورد مطالعه قرار گیرد. همچنین، با توجه به این که در این سامانه‌ها در کنار چالش کمبود و پراکندگی داده‌ها،

مقیاس‌پذیری نیز یکی از چالش‌های اساسی است، پیشنهاد می‌شود در پژوهشی مستقل این مسئله بررسی گردد و الگوریتم‌ها و شیوه‌های به‌کاررفته در سایر سامانه‌های پیشنهاددهنده، بر روی سامانه‌های پیشنهاددهنده نشریه و به‌صورت خاص بر تأثیر آن در بهبود اثربخشی نتایج پیشنهادی سامانه ژورنال‌یاب «رایسست» مورد آزمون قرار گیرد. بررسی و مطالعه تأثیر سایر مجموعه‌عناصر تشکیل‌دهنده سامانه‌های پیشنهاددهنده مانند انواع الگوریتم‌های شباهت‌سنجی و رتبه‌بندی نیز در سامانه‌های پیشنهاددهنده نشریه، به‌خصوص در زبان فارسی پیشنهاد می‌شود.

۶. قدردانی

بدین وسیله از حمایت‌های جناب آقای دکتر محمدجواد دهقانی، ریاست محترم مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری و پایگاه ISC جهت در اختیار قرار دادن امکانات پژوهش صمیمانه سپاسگزارم.

فهرست منابع

- آخسیک، سمیه‌سادات، و رحمت‌الله فتاحی. ۱۳۹۱. تحلیل چالش‌های پیوسته‌نویسی و جدانویسی واژگان فارسی در ذخیره و بازیابی اطلاعات در پایگاه‌های اطلاعاتی. *کتابداری و اطلاع‌رسانی* ۳ (۵۹): ۹-۳۰.
- پرنی، اعظم‌السادات، و حجت‌اله حمیدی. ۱۳۹۶. ارائه رویکردی برای مدیریت و سازمان‌دهی اسناد متنی با استفاده از تجزیه و تحلیل هوشمند متن. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۲ (۴): ۱۱۷۱-۱۲۰۲.
- حرری، عباس. ۱۳۷۲. کامپیوتر و رسم‌الخط فارسی. *فصلنامه تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی (پیام کتابخانه سابق)* ۳ (۱): ۶-۱۱.
- حسین‌زاده، پریسا. ۱۴۰۰. اثربخشی مجله‌های پیشنهادی سامانه توصیه‌گر مجلات الزویر. پایان‌نامه کارشناسی ارشد علم اطلاعات و دانش‌شناسی. دانشگاه شیراز.
- دیانی، محمدحسین. ۱۳۹۲. *روش‌های تحقیق در کتابداری*. مشهد: کتابخانه رایانه‌ای.
- رنجبر، ایوب، و جواد عباس‌پور. ۱۳۹۷. گسترش‌پذیری جست‌وجو و بازیابی مدارک در پایگاه‌های اطلاعات علمی فارسی: مورد پژوهشی پیوسته‌نویسی و جدانویسی. *کتابداری و اطلاع‌رسانی* ۲۱ (۳): ۵۷-۹۰.
- ستوده، هاجر، و زهره هنرجویان. ۱۳۹۱. مروری بر دشواری‌های زبان فارسی در محیط دیجیتال و تأثیرات آن‌ها بر اثربخشی پردازش خودکار متن و بازیابی اطلاعات. *کتابداری و اطلاع‌رسانی* ۱۵ (۴): ۵۹-۹۲.
- ضیائی‌بیده، علیرضا، و سید یعقوب حسینی. ۱۳۹۵. *آمار ناپارامتریک و روش پژوهش با کاربرد نرم‌افزار SPSS*. تهران: دانشگاه علامه طباطبایی.

کامیابی گل، عطیه، الهام اخلاقی باقوجری، احسان عسگریان، و هانیه حبیبی. ۱۳۹۷. استخراج اطلاعات از پیکره زبانی: معرفی پیکره مقاله‌های علمی پژوهشی دانشگاه فردوسی مشهد. *کتابداری و اطلاع‌رسانی* ۲۱ (۲): ۳-۲۵.

نشاط، نرگس. ۱۳۷۹. مسائل رسم الخط فارسی در رویارویی با فناوری نوین اطلاعاتی در فهرست‌های رایانه‌ای؛ کاربرد و توسعه. مجموعه مقالات همایش کاربرد و توسعه فهرست‌های رایانه‌ای کتابخانه‌های ایران. آبان ۲۷-۲۸، (۴۰۸-۴۰۱). مشهد: دانشگاه فردوسی مشهد.

هماوندی، هدی، یعقوب نوروزی، و ملوک‌السادات حسینی بهشتی. ۱۳۹۷. بررسی مشکلات جست‌وجو و بازیابی اطلاعات در پایگاه‌های اطلاعاتی از جنبه ویژگی‌های نگارشی زبان فارسی. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۳ (۳): ۱۰۸۷-۱۱۱۰.

References

- Aggarwal, C. C. 2016. An introduction to recommender systems. In: *Recommender systems: The Textbook*. Springer, Cham. https://doi.org/10.1007/978-3-319-29659-3_1
- Anderson, K. 2012. *Editorial Rejection — Increasingly Important, Yet Often Overlooked Or Dismissed*, in *The Scholarly Kitchen*. <https://scholarlykitchen.sspnet.org/2012/04/19/post-publication-peer-review-what-value-do-usage-based-metrics-offer/> (accessed Dec. 22, 2020)
- Bahadoran, Z., P. Mirmiran, K. Kashfi, and A. Ghasemi. 2021. Scientific Publishing in Biomedicine: How to Choose a Journal? *International Journal of Endocrinology and Metabolism* 19 (1): e108417.
- Balyan, R., K. S. McCarthy, and D. S. McNamara. 2020. Applying natural language processing and hierarchical machine learning approaches to text difficulty classification. *International Journal of Artificial Intelligence in Education* 30 (3): 337-370.
- Beel, J., B. Gipp, S. Langer, & C. Breitinger. 2016. Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries* 17 (4): 305-338.
- Berbatova, M. 2019. Overview on NLP Techniques for Content-Based Recommender Systems for Books. In *Proceedings of the Student Research Workshop Associated with RANLP 2019* (pp. 55-61). Varna, Bulgaria. INCOMA Lt.d
- Camacho, L. A. G., and S. N. Alves-Souza. 2018. Social network data to alleviate cold-start in recommender system: A systematic review. *Information Processing & Management* 54 (4): 529-544.
- Cochran, W. G. 1977. *Sampling techniques* (3rd ed.). NewYork: Wiley.
- Das, D., L. Sahoo, and S. Datta. 2017. A survey on recommendation system. *International Journal of Computer Applications* 160 (7): 6-10.
- Dror G, N. Koenigstein, Y. Koren, & M. Weimer. 2012. The Yahoo! music dataset and kdd-cup'11. *Journal of Machine Learning Research Workshop and Conference Proceedings: Proceedings of KDD Cup 18*: 3-18.
- Eirinaki, M., J. Gao, I. Varlamis, and K. Tserpes, K. 2018. Recommender systems for large-scale social networks: A review of challenges and solutions. *Future Generation Computer Systems* 78 (Part 1): 413-318.
- Fayyaz, Z., M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef. 2020. Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. *Applied Sciences* 10 (21): 7748. <https://doi.org/10.3390/app10217748>.

- Feng, X., H. Zhang, Y. Ren, P. Shang, Y. Zhu, Y. Liang, and D. Xu. 2019. The Deep Learning–Based Recommender System “Pubmender” for Choosing a Biomedical Publication Venue: Development and Validation Study. *Journal of medical Internet research* 21 (5): e12957.
- Forrester, A., B. C. Björk, and C. Tenopir. 2017. New web services that help authors choose journals. *Learned Publishing* 30 (4): 281–287.
- Göksedef, M., and S. Gündüz-Öğüdücü. 2010. Combination of Web page recommender systems. *Expert Systems with Applications* 37 (4): 2911-2922.
- Guo, X., X. Li, and Y. Yu. 2021. Publication delay adjusted impact factor: The effect of publication delay of articles on journal impact factor. *Journal of Informetrics* 15 (1): 101100.
- Gupta, V., & G. S. Lehal. 2009. A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence* 1 (1)60-76 .
- Heinrich, B., M. Hopf, D. Lohninger, A. Schiller, and M. Szubartowicz. 2021. Data quality in recommender systems: the impact of completeness of item content data on prediction accuracy of recommender systems. *Electronic Markets* 31 (2): 389-409
- Hong, S., Y. Zhou, J. Shang, C. Xiao, & J. Sun. 2020. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine* 122: 103801.
- Huisman, J., and J. Smits. 2017. Duration and quality of the peer review process: the author's perspective. *Scientometrics* 113 (1): 633-650.
- Isinkaye, F. O., Y. O. Folajimi, and B. A. Ojokoh. 2015. Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal* 16 (3): 261-273.
- Jenuwine, E. S., & J. A. Floyd. 2004. Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. *Journal of the Medical Library Association* 92 (3): 349.
- Kanakia, A., Z. Shen, D. Eide, & K. Wang. 2019. A scalable hybrid research paper recommender system for Microsoft academic. In *The World Wide Web conference* (pp. 2893-2899). NewYork, NY, USA.
- Kang, N., M. Doornenbal, & B. Schijvenaars. 2015. Elsevier Journal Finder: Recommending Journals for your Paper. *RecSys '15*, September 16 - 20, Vienna, Austria.
- Kant, Mahara2018 .. Merging user and item based collaborative filtering to alleviate data sparsity. *International Journal of System Assurance Engineering* 9 (1): 173-179.
- Khusro, S., Z. Ali, & I. Ullah. 2016. Recommender systems: issues, challenges, and research opportunities. In *Information Science and Applications (ICISA) 2016* (pp. 1179-1189). Singapore: Springer.
- Lampert, C. H. 2009. Learning to detect unseen object classes by between-class attribute transfer. In *IEEE Conference on Computer Vision and Pattern Recognitio. CVPR*. Miami, FL, USA.
- Liu, X., Y. Ouyang, W. Rong, & Z. Xiong. 2015. Item category aware conditional restricted boltzmann machine based recommendation. In *International Conference on Neural Information Processing* (pp. 609-616). Cham: Springer.
- Lin, Z., S. Hou, and J. Wu. 2016. The correlation between editorial delay and the ratio of highly cited papers in Nature, Science and Physical Review Letters. *Scientometrics* 107 (3): 1457-1464.
- Lucas, J. P., S. Segrera, & M. N. Moreno. 2012. Making use of associative classifiers in order to alleviate typical drawbacks in recommender systems. *Expert Systems with Applications* 39 (1): 1273-1283.
- Mohamed, M. H., M. H. Khafagy, & M. H. Ibrahim. 2019. Recommender systems challenges and solutions survey. In *2019 International Conference on Innovative Trends in Computer Engineering (ITCE)* (pp. 149-155). Aswan, Egypt. doi: 10.1109/ ITCE.2019.8646645.

- Mulligan, A., L. Hall, & E. Raphael. 2013. Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology* 64 (1): 132-161.
- Nguyen, T. T., F. Maxwell Harper, L. Terveen, et al. 2018. User Personality and User Satisfaction with Recommender Systems. *Information Systems Frontiers* 20 (6): 1173-1189.
- Nilashi, M., O. Ibrahim, & K. Bagherifard. 2018. A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. *Expert Systems with Applications* 92 (February): 507-520.
- Park, D. H., H. K. Kim, J. K. Kim, I. Y. Choi, & J. K. Kim. 2011. A review and classification of recommender systems research. *International Proceedings of Economics Development & Research* 5 (1): 290-294.
- QasemiZadeh, B., S. Rahimi, & M. S. Ghalati. 2014. Challenges in Persian Electronic Text Analysis. *arXiv preprint arXiv:1404.4740*. <https://arxiv.org/ftp/arxiv/papers/1404/1404.4740.pdf> (accessed Nov. 10, 2020)
- Ricci, F., L. Rokach, & B. Shapira. 2015. Recommender systems: introduction and challenges. In *Recommender systems handbook* (pp. 1-34). Boston, MA: Springer.
- Robinson. 1964. Groups in which normality is a transitive relation. In *Mathematical Proceedings of the Cambridge Philosophical Society* 60 (1): 21-38. Cambridge University Press.
- Rollins, J., M. McCusker, J. Carlson, & J. Stroll. 2017. Manuscript Matcher: A Content and Bibliometrics-based Scholarly Journal Recommendation System. Proceedings of the Fifth Workshop on Bibliometric-enhanced Information Retrieval (BIR) co-located with the 39th European Conference on Information Retrieval (ECIR 2017), Aberdeen, UK, April 9th, 2017. (pp. 18-29).
- Sachan, A., & V. Richhariya. 2013. Reduction of data sparsity in collaborative filtering based on fuzzy inference rules. *International Journal of Advanced Computer Research* 3 (2): 101.
- Sarwar B., G. Karypis, J. Konstan, & J. Riedl. 2000. Application of dimensionality reduction in recommender system – a case study. In: ACM WebKDD Workshop, 2000b, pp. 264–272. University of Minnesota, Minneapolis.
- Sarwar B., G. Karypis, J. Konstan, and J. Riedl. 2001. Item-Based Collaborative Filtering Recommendation Algorithms In *Proceedings of the 10th international conference on World Wide*. Hong Kong.
- Schedl, M., H. Zamani, C. W. Chen, Y. Deldjoo, and M. Elahi. 2018. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval* 7 (2): 95-116.
- Seraji, M. 2013. PrePer: A Pre-processor for Persian, Proceedings of the Fifth International Conference on Iranian Linguistics (ICIL5), Bamberg, Germany.
- Sharma, R., & R. Singh. 2016. Evolution of recommender systems from ancient times to modern era: a survey. *Indian Journal of Science and Technology* 9 (20): 1-12.
- Su, J. H., & T. W. Chiu. 2016. An item-based music recommender system using music content similarity. In *Asian Conference on Intelligent Information and Database Systems* (pp. 179-190). Springer, Berlin, Heidelberg.
- Thorat, P. B., R. M. Goudar, & S. Barve. 2015. Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications* 110 (4): 31-36.
- Wang, W. T., and Y. P. Hou. 2015. Motivations of employees' knowledge sharing behaviors: A self-determination perspective. *Information and Organization* 25 (1): 1-26.
- Wang, Z., H. Huang, L. Cui, J. Chen, J. An, H. Duan, & N. Deng. 2020. Using Natural Language Processing Techniques to Provide Personalized Educational Materials for Chronic Disease Patients in China: Development and Assessment of a Knowledge-Based Health Recommender System. *JMIR medical informatics* 8 (4): e17642.

- Wang, H., Z. Wang, & W. Zhang. 2018. Quantitative analysis of Matthew effect and sparsity problem of recommender systems. In *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 78-82). Chengdu, China.
- Wei, S., N. Ye, S. Zhang, X. Huang, & J. Zhu. 2012, August. Item-based collaborative filtering recommendation algorithm combining item category with interestingness measure. In *2012 International Conference on Computer Science and Service System* (pp. 2038-2041). Nanjing, China.
- Wu, F., Yang, R., Zhang, C., & Zhang, L. 2021. A deep learning framework combined with word embedding to identify DNA replication origins. *Scientific reports* 11 (1): 1-19.
- Yao, W., J. He, H. Wang, Y. Zhang, & J. Cao. 2015. Collaborative topic ranking: Leveraging item meta-data for sparsity reduction. In *Proceedings of the AAAI Conference on Artificial Intelligence* 29 (1): 374-380.
- Yin, H., Q. Wang, K. Zheng, Z. Li, & X. Zhou. 2020. Overcoming Data Sparsity in Group Recommendation. *IEEE Transactions on Knowledge and Data Engineering*. doi: 10.1109/TKDE.2020.3023787.
- Zhao, X. 2019. A study on e-commerce recommender system based on big data. In *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 222-226). Chengdu, China.
- Zhang, Y., H. Abbas, & Y. Sun. 2019. Smart e-commerce integration with recommender systems. *Electronic Markets* 29 (2): 219-220
- Zoetekouw, K. F. A. 2019. A critical analysis of the negative consequences caused by recommender systems used on social media platforms. Bachelor's thesis. University of Twente.

نرجس وردع

متولد ۱۳۵۸، دانشجوی دکتری علم اطلاعات و دانش‌شناسی دانشگاه شیراز است. ایشان هم‌اکنون مربی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری است.

پردازش زبان طبیعی، داده‌کاوی، سامانه‌های بازیابی اطلاعات، تعامل انسان و رایانه، تعامل انسان و اطلاعات، توسعه و ارزیابی منابع از جمله علایق پژوهشی وی است.



مهدیه میرزاییگی

متولد سال ۱۳۶۰، دارای مدرک تحصیلی دکتری رشته علم اطلاعات و دانش‌شناسی از دانشگاه فردوسی مشهد است. ایشان هم‌اکنون دانشیار دانشگاه شیراز است.

مطالعات کاربرمدار در فرایند بازیابی، وب معنایی و هستی‌شناسی‌ها از جمله علایق پژوهشی وی است.



هاجر ستوده

متولد سال ۱۳۵۰، دارای مدرک تحصیلی دکتری رشته علم اطلاعات و دانش‌شناسی است. ایشان هم‌اکنون استاد دانشگاه شیراز است. علم‌سنجی و دگرسنجی، دسترسی آزاد و علم آزاد، سازماندهی و مدیریت دانش، و بازیابی اطلاعات از جمله علایق پژوهشی وی است.



سید مصطفی فخر احمد

متولد سال ۱۳۵۹، دارای مدرک تحصیلی دکتری در رشته مهندسی کامپیوتر از دانشگاه شیراز است. ایشان هم‌اکنون دانشیار دانشگاه شیراز است. داده‌کاوی، متن‌کاوی، پردازش زبان طبیعی و طراحی سیستم‌های خبره از جمله علایق پژوهشی وی است.



نیلوفر مظفری

متولد ۱۳۶۴، دارای مدرک دکتری در رشته هوش مصنوعی از دانشگاه شیراز است. ایشان هم‌اکنون استادیار مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری است. داده‌کاوی، یادگیری ماشین، تحلیل شبکه‌های اجتماعی و پردازش زبان‌های طبیعی از جمله علایق پژوهشی وی است.

