A Context-Aware System for Selective Dissemination of Information in Electronic Theses and Dissertations

Abbas Sadeghi Pouryani
PhD Candidate in Information Technology Management; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran Email: sadeghi@students.irandoc.ac.ir

Marzieh Zarinbal*
PhD in Industrial Engineering; Assistant Professor; Amirkabir University of Technology; Faculty of Industrial Engineering and Management Systems; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran;
Email: mzarinbal@aut.ac.ir

Iranian Journal of Information Processing and Management

Received: 01, Dec. 2024 **Accepted: 15, Feb. 2025**

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 40 | No. 3 | pp. 691-722
Spring 2025
https://doi.org/10.22034/jipm.2025.2047228.1865

Abstract: In an electronic dissertation and thesis (ETDs) system, selective dissemination of information is the delivery of ETD documents to users which are more likely to be useful to those users. This research aims to present a model for selective dissemination of information using user and ETD contexts in an ETD database. The core of this model is a context-aware content-based recommender system in which the ETD contexts are compared with the user contexts, and the target users with a higher likelihood of being useful are selected. The research method is a design science study in which an artifact (software system) is built and the research contribution is the user and ETD context model for the ETDs system. In the selective dissemination of information, user contexts and ETD contexts were weighted and compared based on the proposed method of this paper, CF-ICF, to select the users closest to the ETD. From the published ETDs of the Iranian academic year 1401-1402, 10,860 ETDs were randomly selected. The 30 ETDs with the highest similarity between the ETD context and the user context were sent to each user for feedback. The results indicate that the first recommendations were highly beneficial to users in at least 56% of cases. For these recommendations, the precision is 0.73 out of 1, and the Normalized Discount Cumulative Gain index is 0.69 out of 1. This method provides a model selective dissemination of information based on the context of users and the context of ETDs, achieving optimal precision in the first five recommendations. It can be highly beneficial for developers and system administrators of ETDs.

Keywords: Context-Aware Recommender, Dissertation and Thesis Database, Selective Dissemination of Information

* Corresponding Author

اشاعه گزینشی اطلاعات بافت‌آگاه برای پایان‌نامه‌ها و رساله‌های الکترونیک*

عباس صادقی پوریانی
دانشجوی دکتری مدیریت فناوری اطلاعات؛ پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛ تهران، ایران sadeghi@students.irandoc.ac.ir

مرضیه زرین‌بال
دکتری تخصصی مهندسی صنایع؛ استادیار؛ دانشگاه صنعتی امیرکبیر؛ دانشکده مهندسی صنایع و سیستم‌های مدیریت؛ پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛ تهران، ایران؛ پدیدآور رابط mzarinbal@aut.ac.ir

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۴۰۳/۰۹/۱۱ پذیرش: ۱۴۰۳/۱۱/۲۷ مقاله برای اصلاح به مدت ۳ روز نزد پدیدآوران بوده است.

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۴۰ | شمارهٔ ۳ | ص‌ص ۶۹۱-۷۲۲
بهار ۱۴۰۴
https://doi.org/10.22034/jipm.2025.2047228.1865

چکیده: در سامانهٔ پایان‌نامه‌ها و رساله‌های الکترونیک (پارسا)، اشاعه گزینشی اطلاعات رساندن اسناد «پارسا» به کاربرانی است که آن اسناد احتمال سودمندی بیشتری برای کاربران داشته باشند. هدف این پژوهش ارائه مدلی برای اشاعه گزینشی اطلاعات با بهره‌گیری از بافت‌های کاربران و بافت‌های «پارسا» در یک پایگاه اطلاعات «پارسا»هاست. هسته این مدل یک سیستم پیشنهادهنده بافت‌آگاه محتوامحور است که در آن بافت «پارسا» با بافت کاربر مقایسه و کاربران هدف با احتمال سودمندی بیشتر برگزیده می‌شوند. روش پژوهش، مطالعه به شیوهٔ علم طراحی است که در آن یک مصنوع (سامانه نرم‌افزاری) دارای نوآوری ساخته می‌شود. نوآوری این پژوهش مدل بافت کاربران، مدل بافت «پارسا»ها، و روش وزن‌دهی به بافت‌های کاربر و «پارسا» برای سامانهٔ «پارسا»ست. در اشاعه گزینشی اطلاعات، بافت‌های کاربر و «پارسا» بر اساس روش پیشنهادی مقاله حاضر، «سی‌اف-آی‌سی‌اف» وزن‌دهی و مقایسه شدند تا نزدیک‌ترین کاربران به «پارسا»های قابل انتشار گزینش شوند. برای ارزیابی مدل از پارسا»های منتشرشده در سال تحصیلی ۱۴۰۱-۱۴۰۲ تعداد ۱۰۸۶۰ «پارسا» به‌صورت تصادفی انتخاب و تعداد ۳۰ «پارسا»ی دارای بیشترین شباهت بین بافت «پارسا» با بافت کاربر برای ۱۶۳ کاربر بر اساس مدل یادشده ارسال شد. «پارسا»های بازگشتی از آن‌ها رتبه‌بندی و بازخورد دریافت شد.

* این مقاله از رساله دکتری نویسنده اول با عنوان «ارائه مدل بازیابی و اشاعه گزینشی اطلاعات با رویکرد سیستم‌های بافت‌آگاه؛ مورد مطالعه: پایگاه اطلاعات علمی ایران» استخراج شده است

یافته‌ها نشان می‌دهد که پنج پیشنهاد اول کمینه به میزان ۵۶ درصد برای کاربران خیلی مفید بوده‌اند. برای همین پیشنهادها، شاخص تمرکز برابر ۰/۷۳ از ۱ و کمینه شاخص سود تجمعی نرمال‌شده برابر ۰/۶۹ از ۱ است. این روش مدلی را برای اشاعه گزینشی اطلاعات مبتنی بر محتوا و بافت کاربران و «پارسا»ها ارائه می‌دهد که در پنج گزینه پیشنهادی اول بهترین تمرکز را داراست و می‌تواند برای توسعه‌دهندگان و مدیران سامانه «پارسا» مفید باشد

کلیدواژه‌ها: پیشنهاددهنده بافت‌آگاه، پایگاه اطلاعاتی پایان‌نامه‌ها و رساله‌ها، اشاعه گزینشی اطلاعات

۱. مقدمه

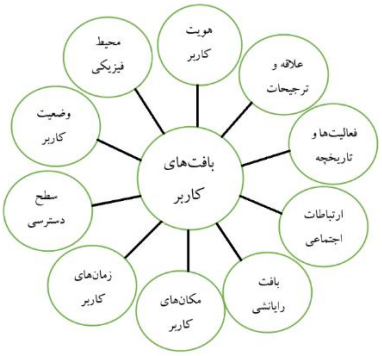
مفهوم اشاعه گزینشی اطلاعات[1] را نخستین‌بار «لوهن» در سال ۱۹۵۸ ارائه داد و آن خدمتی برای هدایت عناصر جدید اطلاعات از هر منبع به نقطه‌ای از سازمان است که احتمال سودمندی در ارتباط با فعالیت‌های جاری یا مورد علاقه آن نقطه وجود دارد (Uzohue and Yaya 2016; Anjaian 2013; Luhn 1958). اشاعه گزینشی اطلاعات یک فناوری است که برای تطبیق نیازهای کاربران و افزایش استفاده مؤثر از منابع در کتابخانه‌ها و مراکز اطلاعاتی به‌کار می‌رود (Soleimani, Elango and Hassanzadeh 2024). سامانهٔ «پایان‌نامه‌ها و رساله‌های الکترونیک»[2] که در این مقاله سامانهٔ «پارسا»» گفته می‌شود (علیدوستی و صابری ۱۳۸۶)، منبعی غنی از اطلاعات و دانش است. سامانه «گنج ایرانداک» در ایران و «اندی‌ال‌تی‌دی»[3] و «اُای‌تی‌دی»[4] نمونه بین‌المللی از این نوع پایگاه‌های اطلاعاتی هستند که این نوع اسناد را برای مجموعه‌ای از مؤسسات آموزشی به شکل یک درگاه واحد عرضه می‌کنند.

«پارسا»های گزینش‌شده را می‌توان با استفاده از مدل‌هایی کارآمد، به سمت کاربرانی هدایت کرد که بیشترین سود را از داشتن آن اطلاعات به‌دست آورند. در این مطالعه، هدفِ اشاعه گزینشی اطلاعات ارائه روشی است که «پارسا»ها را به‌صورت هدفمند برای آن بخش از کاربران ارسال کند که اسناد بر اساس بافت، علاقه و فعالیت‌های جاری یا قبلی آن‌ها احتمال سودمندی بیشتری برای آن‌ها دارد.

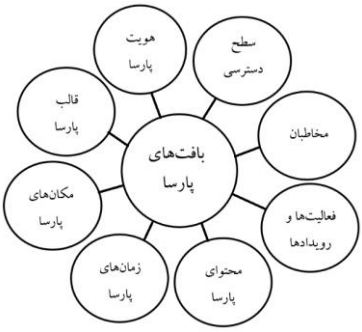
1. selective dissemination of information (SDI)
2. system of electronic theses and dissertationsn (ETDs)
3. http://www.ndltd.org/resources/find-etds
4. OATD.org

در جامعه اطلاعاتی توانایی انتشار محتوای مرتبط با افراد از اهمیت فزاینده‌ای برخوردار شده است (Saeidnia 2023). با پیشرفت فناوری اطلاعات داده‌های پروفایل کاربران، محتوای مورد درخواست، فعالیت‌ها، و رفتار آن‌ها در سامانه‌های «پارسا» ذخیره شده و افزون بر آن، رویدادهای مرتبط با «پارسا»ها (چون: دانلودهای «پارسا»، مکان مورد مطالعه و زمان انتشار تمام‌متن «پارسا») را نیز نگهداری می‌کنند. با وجود این داده‌های بزرگ، اشاعه گزینشی اطلاعات (پیشنهاددهی پارسا به کاربر با احتمال سودمندی بیشتر) راه‌حلی برای نیازهای روبه گسترش کاربران در جست‌وجوی اطلاعات بین داده‌های فراوان است (Mezzi and Benblidia 2017; Gollagi, Math and Kulkarni 2019)

از نظر «دی» بافت «هر اطلاعاتی است که می‌تواند برای مشخص کردن وضعیت یک موجودیت و تفسیر خاص-دامنهٔ آن مورد استفاده قرار گیرد» (Dey 2001). «صادقی پوریانی و زرین‌بال ماسوله» بافت را «ویژگی‌هایی از موجودیت‌های سامانه می‌دانند که در ایجاد، توسعه یا افزایش کیفیت رابطه بین موجودیت‌های کاربر و «پارسا» مؤثر هستند» (۱۴۰۰). در مطالعه آن‌ها بافت‌های کاربر در ۱۰ دسته و بافت‌های «پارسا» در ۸ دسته به‌شرح شکل ۱، جای گرفتند



شکل ۱. الف. بافت‌های کاربر



شکل ۱. ب. بافت‌های «پارسا»

منبع: صادقی پوریانی و زرین‌بال (۱۴۰۰)

هدف پژوهش حاضر، ارائه مدلی برای ارسال «پارسا»های گزینش‌شده به کاربران بر اساس بافت‌های ذخیره‌شده از هر دو موجودیت کاربران و «پارسا»هاست. سؤال اصلی پژوهش این است که چگونه می‌توان پارسا»های گزینش‌شده را به سمت کاربرانی ارسال کرد که بیشترین سودمندی را برای ایشان داشته باشد. در مثالی فرض کنید «پارسا»هایی

گزینش شده‌اند که بافت مکانی مورد مطالعه «پارسا» در آن‌ها «شهر سنندج» و بافت فرهنگی مورد مطالعه آن‌ها «زبان کردی» است. حال کدام کاربران برای ارسال «پارسا» مناسب‌تر هستند؟ حتی می‌توان کاربران را در یک پالایش اولیه قرار داد. یعنی کاربرانی انتخاب شوند که در بافت مکانی خود «شهر سنندج» را در بین پارسا»های دانلودشده قبلی یا محل سکونت خود داشته باشند. در این حالت، پالایش اولیه کمک می‌کند تا فرایند اشاعه گزینشی با دقت بیشتر و سریع‌تر انجام شود

نوآوری پژوهش حاضر، ارائه مدلی مبتنی بر انطباق بافت‌های «پارسا» با بافت‌های کاربر برای پیشنهاددهی است. در این مدل فقط انطباق یک اصطلاح در محتوای «پارسا» با همان اصطلاح از پروفایل کاربر مد نظر نیست، و افزون بر آن، انطباق بافتی از سوی هر دو موجودیت «پارسا» و «کاربر» ملاک تصمیم است. نوآوری دیگر، انجام اشاعه گزینشی اطلاعات مبتنی بر رساله‌ها و پایان‌نامه‌هاست. در پیشنهاددهی، بافت‌ها وزن‌دهی شده‌اند و این وزن در دو رده اعمال شده است: اولین رده، وزن بافت در بافت‌های «پارسا» و دومین رده، وزن بافت در بافت‌های کاربر است. وزن‌دهی بافت‌ها به شیوه پیشنهادشده این مقاله با نام «سی‌اِف-آی‌سی‌اِف»[1] است که توسعه‌ای در روش وزن‌دهی «تی‌اف-آی‌دی اف»[2] است. بر اساس آنچه گفته شد، یک سامانهٔ بافت‌آگاه توسعه داده شد و نتایج حاصل از اِعمال مدل پیشنهاددهی برای کاربران ارسال شد. کاربران سودمندی پارسا»های ارسال‌شده برای آنان را بر اساس نمره‌ای بین ۰ تا ۱۰ مشخص کردند.

در بخش دوم این مقاله، پیشینه پژوهش برای هدایت خواننده به مدل بافت کاربر و بافت «پارسا» در یک سامانه «پارسا» با قابلیت پیشنهاددهی آمده است. در بخش سوم، روش پژوهش که رویکرد آن پژوهش به شیوه «علم طراحی»[3] است، بیان شده و در بخش چهارم، مدل اشاعه گزینشی بافت‌آگاه پیشنهاد شده و نتایج آزمون آمده است. بخش پنجم، نتیجه‌گیری برای ارائه یک سامانه پیشنهاددهنده محتوامحور به‌منظور اشاعه گزینشی اطلاعات است

1. CF-icf: context frequency-inverse context frequency
2. TF-idf: term frequency-inverse term frequency
3. design science

۲. ادبیات و پیشینه پژوهش

در دو دهه اخیر، استفاده از بافت به موضوع مهمی در سامانه‌های پیشنهاددهنده تبدیل شده است (Bauman, Vasilev and Tuzhilin 2022). سیستمی بافت‌آگاه خوانده می‌شود که از بافت برای خدمات‌دهی به کاربر به‌طور مناسب استفاده کند (Dey 2001). این سیستم‌ها نیازمند مدل‌سازی بافت هستند. روش‌های مطرح‌شده در مدل‌سازی بافت مبتنی بر دسته‌بندی Strang, Linnhoff-Popien (2004) و Baldauf, Dustdar, and Rosenberg (2007) به شش دسته تقسیم می‌شوند: دسته اول، زوج ویژگی/ ارزش[1] که انتساب یک ارزش به هر ویژگی در بافت است. دسته دوم، رویکردهای مبتنی بر XML است. دسته سوم، مدل‌های مبتنی بر منطق است که در آن فرایند استنتاج (و استدلال) می‌تواند برای به‌دست آوردن حقایق جدید بر اساس قوانین موجود در سیستم استفاده شود. دسته چهارم، مدل‌های گرافیکی است که جنبه‌های بافتی با استفاده از UML مدل‌سازی شده‌اند. دسته دیگر، مدل‌های شیءگراست که رویکردهای موجود از اشیای متنوع برای نشان دادن انواع مختلف بافت (مانند درجه حرارت، مکان، و غیره) استفاده می‌کنند (Strang and Linnhoff Popien 2004) و جزئیات پردازش و نمایش بافت را ارائه می‌دهند، و سرانجام مدل‌های مبتنی بر هستی‌شناسی (Chaari, Zouari and Laforest 2009) است. محققان بسیاری بر اهمیت مدل‌سازی بافت برای ساختن سیستم بافت‌آگاه تمرکز کرده‌اند. یکی از چالش‌های اساسی سال‌های اخیر، ساخت استاندارد احتمالی مدل‌های بافت بوده است (Mezzi and Benblidia 2017).

رشد محتوای دیجیتال یک مشکل بزرگ است. در این زمینه، روش‌هایی برای مهیا کردن اطلاعات با ایجاد دسترسی برای کاربران به‌کار گرفته می‌شوند (Porcel et al. 2017). «پرسل» و همکاران، چندین روش که فرایندهای اشاعه اطلاعات در کتابخانه‌های دیجیتال دانشگاه را بهبود می‌بخشد، بررسی و تحلیل نمودند. این روش‌ها با توجه به نیازها و ترجیحات کاربر، دسترسی کاربر را به اطلاعات فراهم می‌ساخت. یکی از این روش‌ها با نتایج خوب، استفاده از سیستم‌های پیشنهاددهنده بود. آن‌ها به‌طور خاص بر روی تجزیه‌وتحلیل سیستم پیشنهاددهنده زبانی فازی متمرکز شدندکه از ابزارهای مدل‌سازی زبانی فازی برای مدیریت تنظیمات کاربر و عدم اطمینان سیستم به‌صورت کیفی استفاده

1. attribute/ value

می‌کند (ibid). سیستم‌های پیشنهاددهنده در دو نوع اصلی طبقه‌بندی می‌شوند: سیستم‌های محتوامحور[1] که پیشنهاد بر اساس محتوای مشابه در پروفایل و ویژگی‌های کاربران ارائه می‌شود و سیستم‌های پالایش مبتنی بر همکاری[2] که پیشنهاد بر اساس ترجیحات کاربران مشابه و تاریخچه تعاملات انجام می‌شود (Mobasher 2013). این سیستم‌ها نقش مهمی در ارائه پیشنهادهای شخصی، توسعه کسب‌وکار و ساده‌سازی فرایند تصمیم‌گیری دارند (Hasan et al. 2024)

مطالعات داخل کشور با موضوع بافت‌آگاهی در کتابخانه الکترونیک یا پایان‌نامه‌ها و رساله‌ها بسیار کم است. «رجبی، معین‌آزاد طهرانی و درخوش» در مطالعه خود به بافت‌آگاهی در کتابخانه دیجیتال توجه کرده‌اند. آن‌ها در این مطالعه پس از مرور مفاهیم بافت‌آگاهی، شش دسته فعالیت مبتنی بر بافت را در کتابخانه پیشنهاد دادند (۱۳۹۳). در مطالعه یادشده تکیه بر فضای فیزیکی کتابخانه و دیجیتال‌سازی فیزیکی کتابخانه با بافت‌آگاهی کاربر بیشتر مورد توجه بوده است. یکی از پژوهش‌های انجام‌شده در داخل کشور پایان‌نامهٔ کارشناسی ارشد «دهقانی» (۱۳۸۹) است. وی به شناسایی عوامل بافتی (زمینه‌ای) کاربر با نظریهٔ داده‌بنیاد در کتابخانهٔ دیجیتال پرداخته و برای استفاده در سامانهٔ بافت‌آگاه الگوی کلی چهارسطحی را پس از شناسایی بافت‌های کاربر مورد استفاده قرار می‌دهد. اگرچه در این اثر به شناسایی بافت‌های کاربر توجه شده، اما راه‌حلی برای نحوهٔ به‌کارگیری و ارتباطات بین آن‌ها و چگونگی تأثیر عوامل بافتی برای پیشنهاددهنده ارائه نمی‌شود. «سعیدنیا و حسن‌زاده» به طراحی مدل تعاملی برای اشاعه گزینشی اطلاعات در کتابخانه‌های دانشگاهی پرداختند. روش پژوهش آن‌ها، پیمایشی و با توزیع پرسشنامه محقق‌ساخته در بین اساتید و دانشجویان دانشگاه تربیت مدرس بود که ۴۳ خدمت را در قالب ۷ ویژگی (گنجینه دانش، بازخورد، ارتباط دوسویه، آموزش، جست‌وجو، سفارشات، و مدیریت دانش) دسته‌بندی و نمونه اولیه نرم‌افزاری را برای این موضوع طراحی کردند (۱۴۰۱).

«سلیمانی، بخت‌اواچلام، و حسن‌زاده» در یک مقاله مروری به بررسی سامانه‌های تحت وب اشاعه گزینشی اطلاعات پرداختند. مطالعه آن‌ها به‌طور خاص مزایا و محدودیت‌های

1. content-based systems
2. collaborative filtering systems

مرتبط با انتشار اطلاعات شخصی‌سازی‌شده را بررسی کرد، و سه جزء اصلیِ تحویل اطلاعات مبتنی بر وب، چالش‌ها، و ویژگی‌های کلیدی اشاعه گزینشی اطلاعات را بحث کردند. در پژوهش آن‌ها همچنین به مشکلات خاصی که کشورهای در حال توسعه با آن مواجه هستند، توجه شده است (Soleimani, Bakthavachalam and Hassanzadeh 2024)

استفاده از پیشنهاددهنده‌های بافت‌آگاه را می‌توان در نمایش فیلم (Wu et al. 2017)، موسیقی (Selvi and Sivasankar 2019; Hansen et al. 2020)، رزرو هتل و رستوران‌داری (Chen and Chen 2015, 2014) و در بانکداری با بهره‌گیری از دنباله بافت‌ها (Bauman, Vasilev, and Tuzhilin 2022) مشاهده کرد. «رضا و دینگ» در یک مقاله مروری به سیستم‌های پیشنهاددهنده بافت‌آگاه پرداختند. در این مقاله ابتدا الگوریتم‌های پیشنهاددهنده در سیستم‌های بافت‌آگاه تجزیه‌وتحلیل شده و سپس چگونگی مواجه شدن این الگوریتم‌ها با مشکل ابعاد مختلف (یعنی، ماتریس به‌دست‌آمده از کاربران، گزینه‌ها، مکان‌ها، زمان‌ها و غیره) بحث شده است. از ویژگی مطالعه آن‌ها، عدم تمرکز بر کاربرد خاص سیستم‌های پیشنهاددهنده بافت‌آگاه و تمرکز بر الگوریتم‌های مهم و اساسی، پایگاه داده، مدل‌سازی کاربر و مدل‌های ارزیابی است (Raza and Ding 2019). «دِگویستی» و همکاران بر اشاعه گزینشی اطلاعات تمرکز کردند و با ارائه تعریف عمومی‌تر از بافت، به‌صورت «هر اطلاعاتی که برای تعریف محیط کاربر استفاده می‌شود»، شامل علایق، پرس‌وجوها، گروه‌های کاری و انتخاب‌ها، سرویسی برای اشاعه گزینشی اطلاعات توصیف کردند که یک سیستم بافت‌آگاه مبتنی بر هستی‌شناسی برای تعریف بافت کاربر است. این سیستم نیازمند یک بازیابی بافتی برای به‌دست آوردن اطلاعات و رفتار کاربر و یک موتور استنتاج برای پشتیبانی استنتاج بافتی از منابع فراوان اطلاعات (اسناد انتخابی و پرس‌وجوهای کاربران) است (De Giusti et al. 2010)

بررسی پژوهش‌های پیشین نشان می‌دهد که نمی‌توان پژوهش‌های زیادی با تمرکز بر سامانه «پارسا» در بافت‌آگاهی یافت. خاص آنکه در مطالعات قبلی به بافت کاربر با اهدف انجام یک تراکنش توجه شده است؛ حال آنکه توجه به بافت «پارسا» و مدل‌های ممکن برای اشاعه اطلاعات نیز در این سامانه مهم است. به‌طور خلاصه، این پژوهش سعی دارد مدل و روابطی را بیابد که بتواند نقش و مفهوم بافت را در اشاعه گزینشی اطلاعات به نمایش بگذارد. در مرحله اول، مدل بازنمایی بافت‌ها برای موجودیت‌های کاربر و «پارسا» بررسی می‌شود، در مرحله دوم، مدلی پیشنهاد می‌شود که بتوان اشاعه

گزینشی اطلاعات را با سیستم‌های پیشنهادهنده محتوامحور بافت‌آگاه انجام داد و سرانجام، ارزیابی مدل و اعتباربخشی به آن تکمیل‌کننده این پژوهش است

۳. روش پژوهش

نوع تحقیق از نظر جهت‌گیری اصلی کاربردی-توسعه‌ای است و ماهیت تحلیل، تحقیق به شیوه علم طراحی است. علم طراحی به‌عنوان یک روش پژوهشی پذیرفته شده است (حسان، شریف‌زاده و صدیقی ۱۳۹۹). در این شیوه از مطالعه، دستاورد پژوهش یک مصنوع[1] است (Recker 2013). نوآوری این پژوهش، پیشنهاد مدل‌سازی بافت شامل روش کسب داده‌ها، پالایش داده‌ها و ذخیره‌سازی بافت‌های کاربر و سامانه برای یک سامانه «پارسا»ست. همچنین در نمایه‌سازی بافت‌ها، وزن‌دهی به انواع بافت (کاربر و «پارسا») انجام شده است که دارای دو بخش وزن‌دهی به بافت‌های کاربر و وزن‌دهی به بافت‌های «پارسا» با روش «سی‌اف-آی‌سی‌اف» منطبق با معادله (۴) و معادله (۵) است.

در استفاده از علم طراحی گام‌های تکراری شش‌گانه (Peffers et al. 2007) شاملِ شناخت مسئله، تعریف اهداف راه‌حل، طراحی و توسعه، اثبات، ارزیابی، و انتشار است (حسان، شریف‌زاده و صدیقی ۱۳۹۹). این گام‌ها در زیر توضیح داده می‌شود

شناخت مسئله: مسئله عدم اطلاع کاربران سامانهٔ «پارسا» از پارسا»هایی است که در سامانه ذخیره شده یا تولید می‌شوند؛ مگر اینکه با ورود مداوم و جست‌وجو در سامانه از آن‌ها مطلع شوند. از سوی دیگر، حتی با جست‌وجوی کاربر، اطلاعات کسب‌شده قبلی از کاربر در نتیجهٔ پرس‌وجوی فعلی او نقشی ندارد و بافت‌های کاربر در پیشنهاددهی مهم نیستند

تعریف اهداف راه‌حل: راه‌حل، توسعه یک سامانهٔ پیشنهاددهنده مبتنی بر داده‌های ذخیره‌شده قبلی از کاربر و «پارسا»ست. این راه‌حل، کاربر را از مراجعه مداوم به سامانه جست‌وجو بی‌نیاز می‌کند و ابزاری برای او در فرایند کار علمی‌خواهد بود تا بتواند از تولید «پارسا»های مرتبط با بافت خود مطلع شود.

طراحی و توسعه: در پژوهش حاضر، مصنوع یک نرم‌افزار تحت وب، توسعه داده‌شده در سکوی «لینوکس» نسخه اوبونتو ۲۲/۰۴ و دارای پایگاه داده «مای‌اس‌کیوال»[2] است که

1. artifact
2. MYSQL

از نرم‌افزار «پایتون» برای سمت میزبان[1] و از «اچ‌تی‌ام‌ال»، «جاوااسکریپت» و «جی‌کوئری»[2] برای سمت کاربر[3] استفاده کرده است. در تولید نرم افزار، فرایند تکراری (شامل بازبینی مکرر کدها، روش‌ها و مراحل پیشین با هدف تولید نسخه جدید و کارآمدتر) برای رفع نیازمندی‌ها در نظر گرفته شده است

اثبات و ارزیابی: مدل بافت با یک روش اکتشافی در شکل ۲، بیان شده است. با استفاده از این مدل، داده‌هایِ بافت کاربر u_i $(1 \le i \le n)$ از لاگ‌ها، ورود دستی، خوانش خودکار، پروفایل کاربر، دانلودها و غیره کسب و پس از ذخیره‌سازی اولیه، استنتاج و به‌صورت سه‌تایی رابطه (۳) ذخیره می‌شود. فرض کنیم مجموعه T، انواع بافت (مانند: مکان، جنسیت، رشته تحصیلی و علاقه‌ها)، به‌شرح رابطه (۱) باشد؛ بافت‌های کاربر i به‌صورت مجموعه $Conu_i$ در رابطه (۲) تعریف می‌شود

$$T = \{ t_j | \ 1 \le j \le m \ , t_j \ is \ a \ type \ of \ context \} \quad (۱)$$

$$Conu_i = \{ (t_j, c_{ijk}) | \ t_j \in T \ , 1 \le k \le p \} \quad (۲)$$

و بافت‌های وزن‌دار ذخیره‌شده کاربر i به‌صورت مجموعه $ConwU_i$ در رابطه (۳) است.

$$Conwu_i = \{ (t_j, c_{ijk}, w_{ijk}) | \ t_j \in T \ , 1 \le k \le p \} \quad (۳)$$

در این رابطه‌ها اندیس i برای کاربر i–ام و اندیس j برای انواع بافت‌ها و اندیس k برای مفاهیم موجود در آن نوع از بافت است. t_j نوع بافت (مانند: مکان)، c_{ijk} مفهومی در بافت همین کاربر (مانند: شهر رشت) است. حال، تابع وزن را برای هر بافت (t_j, c_{ijk}) عدد حقیقی w_{ijk} به شکل رابطه (۴) تعریف می‌کنیم.

در روش وزن‌دهی به اصطلاحات در پروفایل کاربر از روش «تی‌اف–آی‌یواِف»[4] (Wu et al. 2015) و به اصطلاحات در سند از روش «تی‌اف–آی‌دی‌اف» (Lops, de Gemmis and Semeraro 2011a) استفاده می‌شود، که در هر دوی آن‌ها وزن، از حاصل‌ضرب فراوانی یک اصطلاح (یک بعدی) در لگاریتم کاهنده بر اساس زیادی اصطلاح در همه پیکره

1. backend
2. jQuery
3. frontend
4. Tf-Iuf

به‌دست می‌آید. با الگو گرفتن از این دو روش، برای یافتن وزن هر بافتِ دو بعدی (رابطه ۲) در بافت کاربر از روش «سی‌اف-آی‌سی‌اف» استفاده شد.

$$W_{ijk}(u_i, \mathrm{t}_j, \mathrm{c}_{ijk}) = \mathrm{Cf}(u_i, \mathrm{t}_j, \mathrm{c}_{ijk}) \times \mathrm{ICf}(u_i, \mathrm{t}_j, \mathrm{c}_{ijk}) \quad (۴)$$

$$\mathrm{Cf}(u_i, \mathrm{t}_j, \mathrm{c}_{ijk}) = number\ of\ context\ (t_j, C_{ijk})\ \text{appers}\ in\ \text{the contexts set of}\ u_i$$

$$ICf(u_i, \mathrm{t}_j, \mathrm{c}_{ijk}) = log \frac{total\ number\ of\ users}{\text{number of}\ users\ \text{in whose contexts}(t_j, C_{ijk})\ \text{appears}}$$

اگر d_i «پارسا»یی متعلق به پیکرهٔ در حال مطالعه باشد، وزن هر بافت دو بعدی مانند (مکان مورد مطالعه «پارسا»، سد زاینده‌رود) برای آن «پارسا» نیز با روش «سی‌اف-آی‌سی‌اف» وزن‌دهی می‌شود.آن حاصل‌ضرب تعداد تکرار یک مفهوم در لگاریتم طبیعی تعداد کل «پارسا»ها بر تعداد «پارسا»های دارای همان بافت با رابطه (۵) تقسیم شد.

$$W_{ijk}(d_i, \mathrm{t}_j, \mathrm{c}_{ijk}) = \mathrm{Cf}(d_i, \mathrm{t}_j, \mathrm{c}_{ijk}) \times \mathrm{ICf}(d_i, \mathrm{t}_j, \mathrm{c}_{ijk}) \quad (۵)$$

$$\mathrm{Cf}(d_i, \mathrm{t}_j, \mathrm{c}_{ijk}) = number\ of\ context\ (t_j, C_{ijk})\ \text{appers}\ in\ \text{the contexts set of}\ d_i$$

$$ICf(d_i, \mathrm{t}_j, \mathrm{c}_{ijk}) = log \frac{total\ number\ of\ documents}{\text{number of}\ documents\ \text{in whose contexts}\ (t_j, C_{ijk})\ \text{appears}}$$

سرانجام، وزن برای هر بافت کاربر [یا بافت «پارسا»] برای یافتن اثر نسبی مفاهیم مورد توجه کاربر [«پارسا»] به همه مفاهیم مورد توجه همان کاربر [«پارسا»] نرمال می‌شود؛ یعنی بر مجموع اوزان بافت‌های همان کاربر [یا «پارسا»] تقسیم می‌شود

دو بافت $X_i = (t_i, c_i)\ (1 \le i \le n)$ از کاربر و $Y_j = (t_j, c_j)\ (1 \le j \le m)$ از «پارسا» برابر هستند هرگاه هر دو نوع بافت t_i و t_j (بر اساس جدول ۱) قابل انطباق (همگن) بود و $c_i = c_j$

جدول ۱. انطباق بافت کاربر با بافت «پارسا»

بافت پارسا	بافت کاربر
رشته تحصیلی تولیدکننده پارسا	رشته تحصیلی کاربر
گرایش تحصیلی تولیدکننده پارسا	گرایش تحصیلی کاربر
جغرافیای مورد مطالعه در پارسا	جغرافیای مورد علاقه یا موقعیت حضور کاربر
مکان مورد مطالعه در پارسا	مکان‌های کاربر (تحصیل، مورد علاقه، تولد، کار، حضور کنونی و مورد مطالعه)
فرهنگ مورد مطالعه در پارسا	فرهنگ مورد علاقه کاربر

بافت پارسا	بافت کاربر
زمان مورد مطالعه در پارسا	زمان مورد علاقه کاربر
شخصیت مورد مطالعه در پارسا	شخصیت مورد علاقه کاربر
محتوای پارسا (عنوان، کلیدواژه، چکیده، متن)	فعالیت‌ها و تاریخچه کاربر (کارهای پیشین، پرس‌وجوها، دانلودها، علاقه ثبت‌شده و ...)
زبان تولید پارسا	زبان کاربر

برای مثال، دو بافت (شهر محل مطالعه «پارسا»، رشت) و (شهر محل سکونت کاربر، رشت) دو بافت قابل انطباق هستند. در انطباق زبان کاربر با زبان تولید «پارسا»، «پارسا»هایی به زبان غیرفارسی نیز در سامانه «گنج» وجود دارند و تعدادی از کاربران نیز فارسی‌زبان نیستند

پس از یافتن وزن مفاهیم در بافت هر کاربر و هر «پارسا»، برای مشابهت‌یابی کاربر u و «پارسا»ی d، مجموعه بافت‌های متعلق به آن‌ها را در نظر گرفته و بردارهای V_u و V_d را برداری بر این مجموعه به‌ترتیب برای کاربر و «پارسا» تعریف می‌کنیم که برای هر عضو در اجتماع دو مجموعه، اگر در بافت کاربر [یا «پارسا»] نباشد، مساوی عدد صفر، و اگر باشد، مساوی وزن بافت است.

به‌صورت دقیق‌تر، اگر $E = \{u, d\}$ و مجموعه $C = Con_u \cup Con_d$ را اجتماع بافت‌های کاربر و «پارسا» در نظر بگیریم و $k = |C|$. دو بردار V_u و V_d بر مجموعه $E \times C$ به‌صورت رابطه (۶) تعریف می‌شود

$$w(e,x) = \begin{cases} w(u,t,c) & e = u\ , x = (t,c) \in con_u \\ w(d,t,c) & e = d\ , x = (t,c) \in con_d \\ 0 & o.w \end{cases}$$

$$V_u = (w(u,x_1), w(u,x_2), \ldots, w(u,x_k))\ , x_i \in C\ , 1 \le i \le k \qquad (۶)$$

$$V_d = (w(d,x_1), w(d,x_2), \ldots, w(d,x_k))\ , x_i \in C\ , 1 \le i \le k$$

رابطه شباهت‌یابی بافتی برای کاربر u و «پارسا»ی d به‌صورت رابطه کسینوسی برداری تعریف شده است: اگر بردار بافت‌های دو کاربر را بر مجموعه متشکل از همه بافت‌های دو کاربر به‌ترتیب V_u و V_d در نظر بگیریم، شباهت کاربر و «پارسا» از رابطه (۷) به‌دست می‌آید (Bobadilla et al. 2013).

$$sim(u,d) = \frac{(V_u . V_d)}{\sqrt{{V_u}^2}\sqrt{{V_d}^2}} \qquad (۷)$$

پس از شباهت‌یابی مجموعه کاربران و «پارسا»ها بر اساس n-امین «پارسا»ی دارای بیشترین شباهت یا «پارسا»های دارای شباهت بیشتر از یک عدد مشخص به کاربران پیشنهاد می‌شوند. شبه‌کد[1] مربوط به اشاعه گزینشی اطلاعات در ادامه آمده است

START
GET a corps of documents $D1$
GET a set of users $U1$
FOR $d \in D1$ **DO**
FOR $u \in U1$ **DO**
$sim(u,d) \leftarrow$ **FUNCTION** $cosine\ (u,d)$
IF $sim(u,d) > 0$, **THEN**
$DIS(u) \leftarrow DIS(u) \cup \{(u,d,(sim(u,d))\}$
ENDIF
ENDFOR
ENDFOR
FOR $u \in U1$ **DO**
$DISO(u) \leftarrow list\ of\ documents\ in\ DIS(u)\ sorted\ by$ sim(u. d)
SHOW $DISO(u,n) \leftarrow\ Sending\ the\ first\ n\ document\ of\ DISO(u)\ to\ user\ u$
ENDFOR
END
FUNCTION $cosine\ (u,d)$:
$U_{context} \leftarrow is\ a\ set\ of\ contexts\ of\ the\ user\ u\ contains\ \{(t_1,c_1,w_1),\dots,(t_n,c_n,w_n)\}$
$d_{context} \leftarrow is\ a\ set\ of\ contexts\ of\ the\ document\ d\ contains\ \{(t'_1,c'_1,w'_1),\dots,(t'_k,c'_k,w'_k)\}$
FOR $(t,c) \in \{(t_1,c_1),\dots,(t_n,c_n)\} \cap \{(t'_1,c'_1),\dots,(t'_k,c'_k)\}$ **Do**
$numerator \leftarrow\ numerator + w \times w'\ \boldsymbol{s.t.}\ (\mathrm{t,c,w}) \in \mathrm{U_{context}}\ , (\mathrm{t,c,w'}) \in \mathrm{d_{context}}$
ENDFOR
RETURN $sim(u,d) \leftarrow \dfrac{numerator}{\sqrt{\sum_1^n w_i} \times \sqrt{\sum_1^k w'_j}}$
END FUNCTION

جامعه آماری در پیشنهاددهندهٔ محتوامحور بافت‌آگاه

از مجموع رساله‌های تولیدشده در سال ۱۴۰۱-۱۴۰۲ که در «سامانه ثبت» پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) ثبت شده‌اند، حدود ۱۰۸۶۰ رساله به‌صورت تصادفی انتخاب و برای اشاعه گزینشی به ۱۶۳ کاربر دارای اطلاعات که پژوهشگران رساله حاضر،

1. pseudocode

امکان ارتباط با ایشان را داشته و برای آزمون مدل در دسترس بودند، ارسال شد. ۵۳ نفر از بین آن‌ها «پارسا»های ارسال‌شده را رتبه‌بندی کردند. از بین ۵۳ نفر پاسخ‌دهنده ۲۹ نفر در مقطع دکتری تخصصی، ۲۱ نفر در مقطع کارشناسی ارشد و ۳ نفر پایین‌تر از کارشناسی ارشد بودند. در اشاعه گزینشی اطلاعات محتوامحور چون نیاز به تعامل با کاربر وجود داشت و از سویی، قواعد «ایرانداک» اجازه دسترسی پژوهشگران به داده‌های ارتباطی کاربران (ایمیل یا شماره تماس) را نمی‌داد، به‌ناچار کاربران در دسترس که امکان ارتباط با آن‌ها فراهم بود، برگزیده و داده‌های آن‌ها (رشته، تحصیلات، گرایش، مطالعات قبلی، مکان‌ها و ...) در سامانه توسعه‌یافته این مطالعه ذخیره شده و برای پیشنهاددهنده استفاده شد.

ارزیابی اشاعه گزینشی محتوامحور بافت‌آگاه

اندازه‌گیری ادراکات و تجربیات کاربران (مانند میزان مفید بودن یک پیشنهاد) بر خلاف ویژگی‌های عینی (مانند سن) کار آسانی نیست. مقیاس‌های رتبه‌بندی، وزن ویژگی‌ها، انتقادات و رفتار ضمنی، متداول‌ترین روش‌هایی هستند که سامانه‌های پیشنهاددهنده از آن‌ها استفاده می‌کنند. در مقیاس‌های رتبه‌بندی از دوتایی (پسند/ ناپسند؛ بالا/ پایین)، یا رتبه‌بندی ستاره‌ها (۵ ستاره یا ۱۰ نیم‌ستاره)، تا لغزنده (هر تعداد مرحله) بهره می‌برند (Knijnenburg and Willemsen 2015). در این بخش از ارزیابی، تعامل کاربران مد نظر قرار گرفته است. ابتدا اسناد به کاربران منتخب پیشنهاد شد و از ایشان خواسته شد تا عددی بین ۰ تا ۱۰ برای مرتبط یا سودمند بودن اسناد با فعالیت‌ها، علایق و بافت‌های خود ارائه دهند. برای ارزیابی مفید بودن n–پیشنهاد اول در پیشنهاددهنده آن‌ها را با روش‌هایی چون تمرکز[1]، فراخوانی[2]، معیار–اف، سود تجمعی کاهنده نرمال‌شده (NDCG)[3] ارزیابی می‌کنند (Hasan et al. 2024).

$$\text{precision}=\frac{\text{تعداد پیشنهاد مورد قبول کاربر در } n \text{ پیشنهاد اول}}{n} \qquad (۸)$$

$$\text{recall}=\frac{\text{تعداد پیشنهاد مورد قبول کاربر در } \boldsymbol{n} \text{ پیشنهاد اول}}{\text{تعداد پیشنهاد مورد قبول کاربر}}$$

$$\text{Fmeasure}=2\times\frac{persicion\times recall}{persicion+recall}$$

1. precision
2. recall
3. normalized discounted cumulative gain (NDCG)

رابطه‌های (۸) میزان اهمیت در امتیازدهی کاربر به گزینه‌ها را نشان نمی‌دهد و فقط قابل قبول بودن گزینه از نظر کاربر مد نظر است. در این رابطه، گزینه‌ای که از کاربر امتیاز ۶ دریافت کرده، میزان اثرش در رابطه (۸) به اندازه گزینه‌ای است که امتیاز بازخورد ۱۰ را دریافت کرده است. برای رفع این مشکل و اندازه‌گیری همراستایی ترتیب پیشنهادها با امتیاز بازخورد کاربران روش NDCG پیشنهاد شده است (Burges et al. 2005; Ahlqvist 2015; Hasan et al. 2024). در این روش، اگر rel_i امتیاز بازخورد مرتبط بودن i-امین سند از n سند برای کاربر باشد، رابطه (۹) کیفیت بازیابی را نشان می‌دهد که می‌توان آن را با تقسیم بر بهترین نتیجه (یعنی $IDCG$[1]) نرمال کرد. برای محاسبه IDCG، بهترین ارزیابی‌های کاربر را در اولویت بالاتر بازمرتب کرده و برای آن DCG را به‌دست می‌آوریم

$$DCG = \sum_{i=1}^{n} \frac{rel_i}{log_2(i+1)} \quad (۹)$$

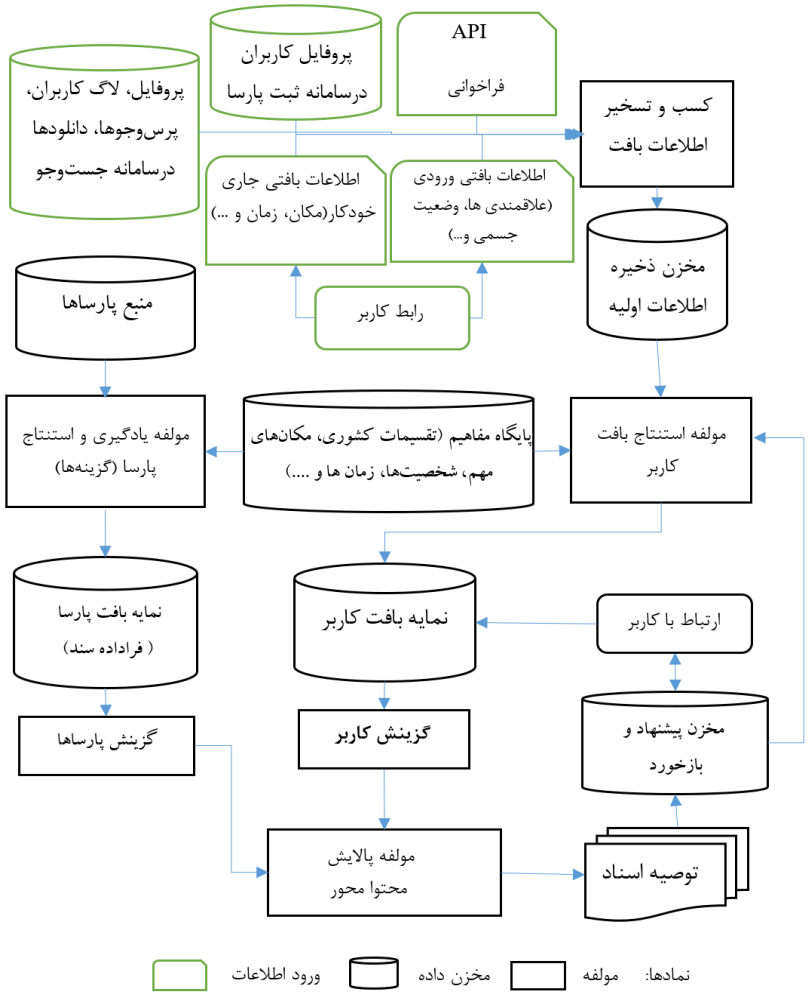
$$NDCG = \frac{DCG}{IDCG}$$

بخش پایانی روش تحقیق، «انتشار» است که سامانه نرم‌افزاری توسعه‌داده‌شده در یک آدرس اینترنتی، در دسترس کاربران قرار گرفت

٤. تجزیه‌وتحلیل یافته‌ها

بیشتر سامانه‌های بافت‌آگاه در یک معماری پنج-لایه‌ای کلان مشترک هستند که شامل لایه‌های تسخیر بافت، تفسیر بافت، ذخیره بافت، انتشار، و کاربرد است (Chaari, Zouari, and Laforest 2009). این معماری کلی در مقاله حاضر پذیرفته شده، اما به‌گونه‌ای خاص برای اشاعه گزینشی سامانهٔ «پارسا» (شکل ۲) بیان شده است.

1. ideal discounted cumulative gain (IDCG)



شکل۲. مدل اشاعه گزینشی بافت‌آگاه در سامانه «پارسا»

٤-١. مؤلفه کسب و تسخیر اطلاعات بافت‌ها در سامانهٔ «پارسا»: یکی از مهم‌ترین بخش‌های سامانه بافت‌آگاه «پارسا»، مؤلفه کسب و تسخیر بافت‌هاست. این مؤلفه با سامانه‌های ثبت «پارسا» و سامانه جست‌وجوهای قبلی کاربر ارتباط دارد. همچنین اطلاعات را از رابط کاربری (نرم‌افزار سامانه بافت‌آگاه و تجهیزات کاربر مانند تلفن همراه هوشمند) می‌توان به‌دست آورد. از سوی دیگر، با داشتن یک رابط (API) می‌توان از سامانه‌های دیگر، اطلاعات مرتبط با کاربر را دریافت کرد. به‌طور کلی، روش‌های تسخیر اطلاعات در جدول ۲، آمده است

جدول ۲. روش‌های کسب بافت کاربر/ بافت «پارسا»

روش تسخیر	بافت‌ها
فراخوانی از سامانه‌های دیگر به خواست مدیر سامانه	◇ اطلاعات هویتی، بافت‌های محتوایی کاربر (مانند لاگ کاربر، دانلودها، مطالعات قبلی) و بافت ارتباطات اجتماعی از سامانه‌های ثبت «پارسا»، سامانهٔ جست‌وجو و سامانه‌های دیگر با API و وب سرویس، بافت‌های محتوایی «پارسا» (دریافت عنوان، چکیده، کلمات کلیدی، پدیدآور، راهنما و مشاور و غیره)
در زمان ثبت نام در سامانهٔ	◇ اطلاعات هویتی توسط کاربر
ورود علاقه‌مندی‌ها با فرمت‌های مشخص	◇ در این بخش فهرست فعالیت‌های علمی فرد با فرمت‌های BibTeX، EndNote، RefMan، RefWorks قابل دریافت است.
ورود دستی پس از ورود به سامانه	◇ هویت، مکان تولد، مکان تحصیل و علایق کاربر، ناتوانی کاربر ◇ اطلاعاتی درباره «پارسا»
در زمان ورود از دریافت‌کننده‌های فیزیکی یا منطقی	◇ اطلاعات رایانشی کاربر (مانند نرم‌افزار، سخت‌افزار، مرورگر، آدرس شبکه) ◇ بافت‌های محیطی کاربر (در حین حرکت، روشنایی)
اقدام کاربر	◇ کسب بافت دانلودها، علایق در زمان اقدام به دانلود یا پرس و جو ◇ کسب بافت علایق از بازخوردهای پیشنهاددهی یا انتخاب نتایج پرس‌وجو
در زمان رویداد	◇ کشف تقلب برای بافت «پارسا»

۴-۲. ذخیره اطلاعات اولیه بافت: پس از کسب اطلاعات بافت، بافت می‌تواند به روش‌های گوناگون نمایه و ذخیره شود. یک روش استفاده از پایگاه داده رابطه‌ای است. در این روش «مجموعه‌ای از داده‌های ذخیره‌شده پایا، به‌هم مرتبط و مجتمع، تا حد ممکن فاقد افزونگی و دارای معماری خاص مبتنی بر یک مدل داده‌ای تحت مدیریت متمرکز و مورد استفاده یک یا چند کاربر قرار می‌گیرد» (روحانی رانکوهی ۱۳۹۶). برای نگهداری اطلاعات اولیه بافت کاربر [پارسا] مانند ابزار معنایی[1] (موضوع، محمول، مفعول) از سه‌تایی (کاربر، نوع بافت، مقداربافت) [(پارسا، نوع بافت و مقداربافت)] استفاده شده است (Cyganiak, Wood and Lanthaler 2014). جدول ۳، بافت اولیه کاربران شامل سه فیلد افزوده، زمان ایجاد بافت، نشست کاربری و نحوه یادگیری بافت است.

1. subject, predicate, object (SPO)

جدول ۳. نمونه ثبت اولیه بافت کاربران

کاربر	نوع بافت	مقدار بافت	زمان ایجاد	شماره نشست	نحوه یادگیری
U1	رشته تحصیلی	مهندسی کامپیوتر	۱۳۹۹/۰۵ /۰۱	-	زمان ثبت نام
U1	زبان کاربر	فارسی	۱۳۹۹/۰۵ /۰۱	-	از «پارسا»های دانلودشده
U101	مکان-وابستگی سازمانی	دانشگاه فردوسی	۱۳۹۹/۰۹ /۰۱	S1	ورود دستی توسط کاربر
U50	دانلود تمام‌متن	کاوش‌های باستان‌شناسی در قلعه الموت	۱۴۰۲/۰۵ /۰۱	S1	از پرس‌وجوی کاربر
U50	پرس‌وجو	شهر+ قزوین+ تاریخچه	۱۴۰۲/۰۵ /۰۱	S2	از پرس‌وجوی کاربر
U1	شبکه	۷۸,۳۹,۲۰۷,۹	۱۴۰۲/۰۷ /۱۴	S22	از اطلاعات نشست

۴-۳. مؤلفه استنتاج بافت‌ها در سامانهٔ «پارسا»: برخی اطلاعات کسب‌شده از موجودیت‌ها (مانند زبان کاربر) به‌صورت مستقیم (یا با اصلاح جزئی در صورت اشتباه کاربر) در پایگاه داده بافت ذخیره می‌شوند. اما اطلاعات بسیاری از کاربران یا «پارسا»ها پس از عملیات پالایش یا تبدیل و با اعمال روش‌های استنتاج (مانند جدول ۴) از مخزن اولیه بافت به محل ذخیره اصلی منتقل می‌شوند. اطلاعات اولیه استنتاج در این مرحله بعد از مدت زمانی مشخص و بر اساس سیاست‌های سامانه، حذف یا بایگانی می‌شوند. یک نمونه از این داده‌ها، اطلاعات شبکه و آدرس منطقی کاربر است که پس از استخراج مکان یا مالک آدرس از آن به مرحله بعدی راه پیدا نمی‌کنند (چون نیازی به آدرس منطقی وجود ندارد و مکان کاربر مورد نظر است) و پس از مدتی از حافظه موقت پاک می‌شوند. پاک شدن این داده‌ها به معنای پاک شدن بافت‌های یافت‌شده توسط آن‌ها نیست.

جدول ٤. روش‌های استنتاج بافت

روش استنتاج	بافت‌ها
استنتاج توسط مؤلفه یادگیر	◇ مکان مورد علاقه کاربر (یا مورد مطالعه «پارسا») از تطبیق مطالعات قبلی فرد (یا اطلاعات در «پارسا») با داده‌های مکانی موجود در بانک اطلاعاتی مفاهیم ◇ فرهنگ‌ها، زمان‌ها و شخصیت‌های مورد علاقه کاربر (یا مورد مطالعه «پارسا») از سازگاری داده‌های فرهنگی، زمانی و شخصیت‌ها در بانک اطلاعاتی مفاهیم؛ ◇ اطلاعات مکان کنونی از روی آدرس منطقی با وب سرویسی مانند Geogoogle؛ ◇ اطلاعات وضعیت کاربر از تعداد کلیک‌ها یا پرس‌وجوی سریع، یا ورود اطلاعات مانند کم‌بینایی توسط خود کاربر؛ ◇ بافت ارتباطات اجتماعی از فعالیت‌های علمی قبلی؛ ◇ اطلاعات محتوایی کاربر (مانند پرس‌وجو، دانلودها و برچسب‌ها) پس از پردازش متن، بافت محتوایی را برای ذخیره‌سازی آماده می‌کند؛ ◇ استخراج اصطلاحات از عنوان «پارسا»های دانلودشده می‌تواند بافت علاقه‌مندی‌ها را به‌دست دهد؛ ◇ استخراج زمانی از شبانه‌روز در کار با سامانه از نشست‌های کاربر.

٤-٤. نمایه‌سازی بافت‌ها در سامانهٔ «پارسا»: پس از کسب داده‌های اولیه و استنتاج بافت از آن در مرحله پایانی مؤلفه استنتاج، دو کار بر روی داده‌های بافت انجام شد. اول) مکان کاربر، جنسیت، سن، رشته و گرایش در صورت تکرار، در بافت نهایی تکرار نشده است. درباره اصطلاحات در بافت پرس‌وجو، در صورت تکرار اصطلاح برای هر نشست کاربر یک بار منظور شده است و در صورت تکرار در نشست‌های متفاوت به تعداد نشست در نظر گرفته شده است. در اصطلاحات به‌دست‌آمده از دانلودها برای یادگیری بافت کاربر، تکرار اصطلاح در دانلودهای متفاوت در نظر گرفته شده است و در پایان برای هر اصطلاح در بافت کاربر وزن‌دهی برابر رابطه «سی‌اف-آی‌سی‌اف» (تعداد حضور اصطلاح در بافت کاربر در ضریب کاهنده لگاریتمی برای کاهش اثر تکرار بیش از اندازه) گفته‌شده در رابطه‌های (۴) و (۵) اعمال شده است. نمونه‌ای از رکوردهای درج‌شده در ذخیره نهایی بافت کاربر (نمایه بافت) به‌شرح جدول ۵، است.

جدول ۵. نمونه نمایه‌سازی اطلاعات بافت کاربران پس از استنتاج

کد کاربر	نوع بافت	مقدار بافت	وزن بافت Cf-Icf
U101	مکان -وابستگی سازمانی	دانشگاه فردوسی	۰/۲
U1	مکان-مورد علاقه	شهر ساوه	۰/۰۹
U50	دانلودهای قبلی	داده‌کاوی	۰/۱
U50	پرس‌وجو	باستان‌شناسی	۰/۳۴
U50	علاقه‌مندی‌ها	قلعه الموت	۰/۲۲
U1	رشته تحصیلی	مهندسی کامپیوتر	۰/۰۵
U120	زمان-مورد علاقه	عید فطر	۰/۱

٤-۵. گزینش «پارسا»ها: «پارسا»ها بر اساس بافت‌های سند، زمان خاص در تولید یا محتوای سند یا رویدادی خاص گزینش می‌شوند (شکل ۳). به‌طور نمونه، این گزینش می‌تواند اسنادی باشد که در هفته جاری به سامانه افزوده شده‌اند یا می‌تواند «پارسا»هایی باشند که زمان انتشار تمام‌متن آن‌ها فرارسیده است.

شکل ۳. گزینش «پارسا»ها بر اساس بافت آن‌ها

مثال دوم، در زمان یک رویداد در دنیای واقعی است. در شرایط زمانی که سیل خوزستان جاری است، «پارسا»های با بافت مکان جغرافیایی خوزستان و کلیدواژه‌های مرتبط با سیل، گزینش شده و به جامعه مخاطب ارسال شود

پس از گزینش «پارسا»ها، نوبت به گزینش کاربران می‌رسد. می‌توان این بخش را بدون هیچ محدودیتی برای کاربران ادامه داد تا خود سامانه بر اساس مؤلفه پالایش،

کاربران نزدیک به اسناد گزینش‌شده را بیابد یا آنکه برای سرعت و دقت بیشتر از ابتدا کاربران را بر اساس بافت‌های آن‌ها محدود کرد. شکل ۴، صفحه گزینش کاربران را نشان می‌دهد

شکل ٤. گزینش کاربران بر اساس بافت آن‌ها

٤-٦. مؤلفه پالایش محتوامحور: پس از گزینش «پارسا»ها و گزینش کاربران، لازم است که برای هر «پارسا»، کاربر نزدیک به آن شناخته شود و از این‌رو، مؤلفه دیگری به نام مؤلفه پالایش لازم است که اصلی‌ترین بخش مدل است. این مؤلفه از سیستم‌های پیشنهاددهنده بافت‌آگاه محتوامحور مبتنی بر بافت‌های «پارسا» و کاربر بهره می‌برد.

فرض کنیم که مجموعه گزینه‌های («پارسا»های) گزینش‌شده در بخش گزینش «پارسا» بر اساس اطلاعات بافت آن‌ها $I_{new} = \{d_1,d_2,\ldots,d_n\}$ باشد (ممکن است برخی «پارسا»های حاضر در این زیرمجموعه، پیش از این نیز برای یک پیشنهاد دیگر استفاده شده باشند. برای مثال: «پارسا»یی در مجموعه «مطالعات انجام‌شده در سال گذشته با مکان شهر سنندج» ممکن است در پیشنهاد جدید «مطالعات انجام‌شده برای فرهنگ کردی» دوباره امکان پیشنهاد داشته باشد). مجموعه کاربران می‌تواند همه کاربران سامانه یا بخشی از کاربران متناسب با بافت آن‌ها انتخاب شده باشد. فرض می‌کنیم این مجموعه کاربرانِ هدف برای ارسال پیشنهادها $U = \{u_1,u_2,\ldots,u_n\}$ است. تابع مطلوبیت $g: U \times I_{new} \to R$ تعریف می‌شود که در آن R مجموعه اعداد حقیقی است. هرچه عدد بزرگ‌تر باشد، به معنای مطلوبیت بیشتر برای ارسال یک «پارسا» به کاربر است.

$$g: U \times I_{new} \to R$$

$$g(u_i, d_{new}) = \begin{cases} 0 & if\ d_{new}\ recommended\ for\ user\ u_i \\ sim(u_i, d_{new}) & o.w \end{cases} \qquad (۱۰)$$

در رابطه (۱۰)، اگر «پارسا» پیش از این به کاربر پیشنهاد داده شده است، مقدار تابع مطلوبیت برابر صفر می‌شود و برای بار دوم پیشنهاد نمی‌شود. در حالت دیگر، تابع مطلوبیت، شباهت بردار بافت وزن‌دار کاربر با بردار بافت وزن‌دار «پارسا»ی به‌دست‌آمده از رابطه (۷) است. پس از رتبه‌بندی «پارسا»های گزینش‌شده برای هر کاربر، «پارسا»های آمادهٔ ارسال با دو شاخص میزان تابع مطلوبیت (مانند بیش از ۰/۳ از عدد ۱) و یا فراوانی ارسال (مانند ۳۰ «پارسا»ی دارای بیشترین تابع مطلوبیت برای هر کاربر) دسته‌بندی شده و پیوندی به‌صورت پیامک برای ایشان ارسال شد و آن‌ها به گزینه‌های پیشنهادی امتیاز ۰ تا ۱۰ را برای میزان مفید بودن «پارسا» دادند

۴-۷. مخزن پیشنهاد و بازخورد: برای به‌روزرسانی پروفایل کاربر فعال (کاربری که پیشنهاد برایش ارسال شده است) از مخزن بازخورد کمک گرفته شد. این بازخوردها می‌تواند صریح (پسند/ ناپسند، رتبه‌دهی، متن‌نویسی و ...) یا ضمنی (تعداد بار دانلودها، چاپ، نشانه‌گذاری، افزودن در قفسه و ...) باشد (Lops, de Gemmis, and Semeraro 2011b). در این مخزن اطلاعات پیشنهادهای قبلی به کاربر و بازخورد بر روی اسناد ذخیره شد. پس از پالایش کاربران دارای بیشترین تابع مطلوبیت، سند برای کاربر ارسال و یک رکورد در مخزن پیشنهاد و بازخورد ثبت شد. این بخش به‌صورت داده‌های یک جدول بانک اطلاعاتی مانند جدول ۶، شامل (سند، کاربر، تاریخ پیشنهاد، وضعیت بازخورد، رتبه بازخورد) ذخیره شد. امتیاز بازخورد کاربر، نمره پسند کاربر است که در یک طیف بازه‌ای [۱۰ ۰] انجام شد.

جدول ٦. نمونه جدول ذخیره بازخورد کاربران

کد پارسا	عنوان پارسا	کد کاربر	تاریخ پیشنهاد	وضعیت بازخورد	رتبه‌بندی سامانه	امتیاز بازخورد کاربر	تاریخ رتبه‌دهی کاربر
۱۰۰۰۰۱۳	شناسایی و رتبه‌بندی عوامل مؤثر بر موفقیت مدیریت دانش در شرکت فراسان	۱۷۷۱۴۵	Feb. 4, 2023	دریافت پاسخ	۱۵	۷	July 21, 2023

مخزن بازخورد دارای ارتباط با مؤلفه استنتاج بافت کاربر و بافت «پارسا» (در شکل ۲) است. پس از دریافت بازخورد از کاربران، برای «پارسا»های دارای حداقل امتیاز آستانه (مانند بیشتر از عدد ۸ از ۱۰) کلمات کلیدی، بافت‌های حاصل از عنوان یا چکیده «پارسا» به بافت علاقه‌مندی کاربر افزوده شد تا در کارهای بعدی مورد استفاده قرار گیرد.

همچنین کد کاربر و امتیاز پسندیدن او با «پارسا»ی پسندیده‌شده ذخیره می‌شود تا برای اشاعه گزینشی مشارکتی با انتخاب همسایگی از بافت‌های کاربران به‌جای ماتریس مشابهت به‌کار رود

۴-۸. ارتباط با کاربر: گرفتن بازخورد از کاربران بخش مهمی از سامانه پیشنهاددهنده است. برای این منظور برای هر کاربر (جامعه آماری منطبق با بخش ۳) پیامک زیر به همراه یک پیوند به صفحه طراحی‌شده (متفاوت برای هر کاربر) ارسال شد: «سلام، خواهشمندم تقبل زحمت فرموده و پایان‌نامه‌های دفاع‌شده اخیر که سامانه برای شما مفید تشخیص داده را ملاحظه و به مفید بودن آن امتیاز دهید. سپاس فراوان http://2.185.148.182/irandoc/2064453932»

کاربران نتایج مدل اشاعه گزینشی را در صفحه پیوند داده‌شده مشاهده کرده و در صورت تمایل به هر کدام از «پارسا»ها متناسب با مفید و مرتبط بودن در بازه ۰ تا ۱۰ امتیاز می‌دادند (مانند شکل ۵). پاسخ آن‌ها رتبه‌بندی مدارک ارسالی بود و در صورت تمایل می‌توانستند ایمیل و علاقه‌مندی‌های خود را برای ارسال پیشنهادهای دیگر در اختیار سامانه قرار دهند

لطفا میزان مفید بودن هر کدام از پیشنهادات برای شما را در بازه ۰ تا ۱۰ تعیین کنید

رساله ها یا پایان نامه های ثبت شده ۱۴۰۲ – ایرانداک که برای پیشنهاد به جنابعالی انتخاب شده است.

ردیف	شماره	عنوان	امتیاز (ثبت امتیازات)
۱	۷۰۰۸۱۰	تحلیل ساختاری پیشران‌های توسعه گردشگری رویداد و ارزیابی سناریوهای مرتبط در شهر یزد جزییات	۱۰.۰
۲	۷۰۰۸۱۱	شناسایی عوامل موثر برکاربست موفق بازاریابی مبتنی بر شبکه‌های اجتماعی در صنعت بوم‌گردی جزییات	۱۰.۰
۳	۷۰۰۸۱۲	بازطراحی آینده نگارانه ی فرایند نظام جامع آموزش گردشگری ایران با تاکید بر رویکرد کارآفرینی جزییات	۹.۰
۴	۷۰۰۸۱۳	طراحی سیستم مانای مزرعه کشاورزی جزییات	۴.۰
۵	۷۰۰۸۱۴	نقش ترویج ورزش های الکترونیک درتوسعه ورزش مدارس ایران جزییات	۵.۰

شکل ۵. صفحه پیشنهادها به کاربر

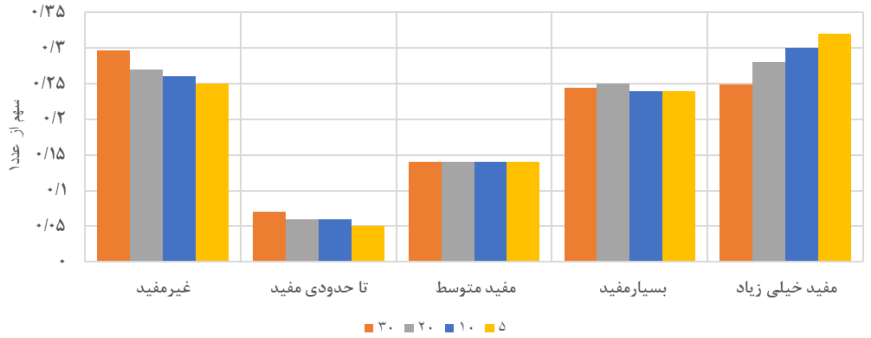
۴-۹. ارزیابی مدل پیشنهاددهنده محتوامحور: ارزیابی بازخوردهای کاربران بر اساس طبقه‌بندی، روش تمرکز، فراخوانی و معیار اف (معادلات ۸) و افزون بر آن‌ها با روش سود

تجمعی نرمال‌شده (معادله ۹) انجام شده است. با بازخورد گرفتن از کاربران پس از انتشار اولین نتایج ارزیابی مدل پیشنهاددهنده محتوامحور بافت‌آگاه، نتایج ارزیابی کاربران به‌طور میانگین برابر ۵/۰۸ از ۱۰ امتیاز برای ۵ پیشنهاد اول، میانگین برابر عدد ۴/۸۶ از ۱۰ امتیاز برای ۱۰ پیشنهاد اول، میانگین برابر ۴/۷۵ از ۱۰ برای ۲۰ پیشنهاد اول و میانگین ۴/۴۷ از ۱۰ برای ۳۰ پیشنهاد اول به‌شرح جدول ۶، به‌دست آمد که نشان می‌دهد هرچه تعداد پیشنهادها بیشتر از ۵ می‌شود، مفید بودن هم کاهش می‌یابد. در حالت کلی، پیشنهاددهنده‌ای با امکان پیشنهاد بیش از ۵۰ درصد مرتبط را باید در زمره پیشنهاددهنده‌های خوب رده‌بندی کرد. برای درک بهتر پاسخ‌های کاربران، پاسخ‌ها به طیف پنج-گزینه‌ای با طول بازه مساوی طبقه‌بندی شده است. چنانچه یک پیشنهاد دارای امتیاز در بازه [۲ ۰] از سوی کاربر بود، آن‌را غیرمفید دانستیم؛ اگر در بازه [۴ ۲] بود، تا حدودی مفید، اگر در بازه [۶ ۴] بود، مفید متوسط، اگر در بازه [۸ ۶] بود، بسیار مفید و برای پاسخ‌های [۱۰ ۸] مفید خیلی زیاد برچسب زده شد.

جدول۷. ارزیابی پیشنهاددهنده بافت‌آگاه محتوامحور

تعداد پیشنهاد	میانگین کل	سهم از عدد یک در ارزیابی کاربران				
		غیرمفید	تا حدودی مفید	مفید متوسط	بسیار مفید	مفید خیلی زیاد
۳۰	۴/۴۷	۰/۳۰	۰/۰۷	۰/۱۴	۰/۲۴	۰/۲۵
۲۰	۴/۷۵	۰/۲۷	۰/۰۶	۰/۱۴	۰/۲۵	۰/۲۸
۱۰	۴/۸۶	۰/۲۶	۰/۰۶	۰/۱۴	۰/۲۴	۰/۳۰
۵	۵/۰۸	۰/۲۵	۰/۰۵	۰/۱۴	۰/۲۴	۰/۳۲

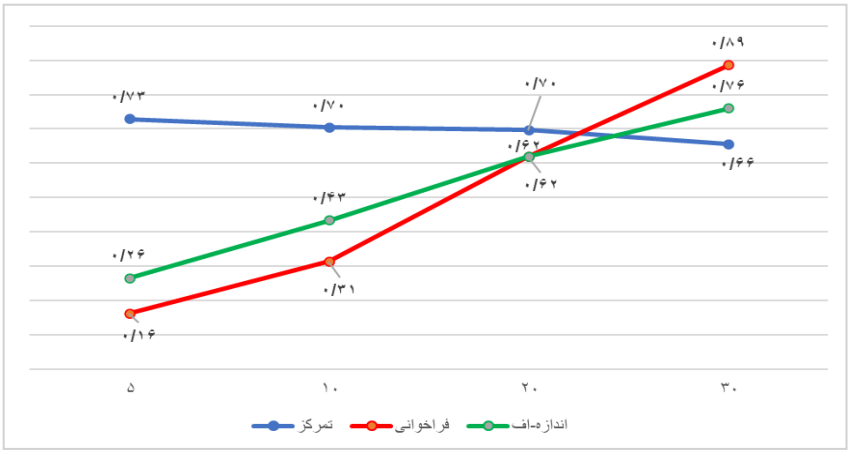
نتیجه نهایی در نمودار ۱، به‌تفکیک برای ۵ پیشنهاد اول، ۱۰ پیشنهاد اول، ۲۰ پیشنهاد اول و ۳۰ پیشنهاد اول آمده است. چنانکه در جدول ۷، و نمودار ۱، آمده، پیشنهادهای بسیار مفید و مفید خیلی زیاد در مجموع (با ۵۶ درصد سهم) به‌طور معناداری بیشتر از پیشنهادهای غیرمفید و تا حدودی مفید (با ۲۶ درصد سهم) است



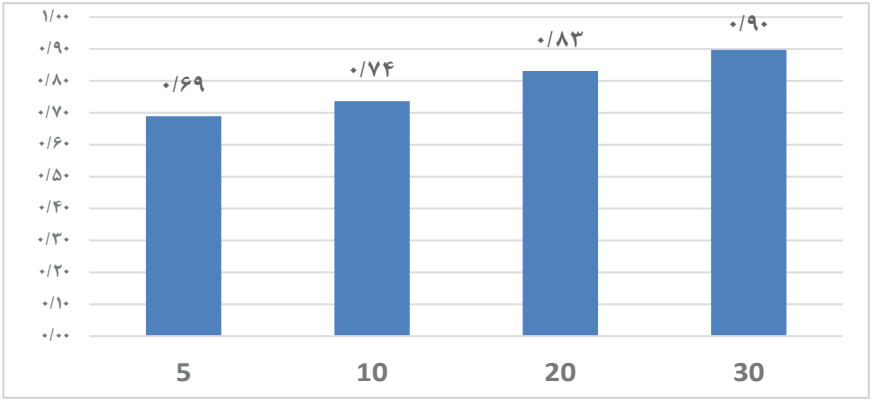
نمودار ۱. سهم نسبی میزان مفید بودن پیشنهادها برای کاربران بر حسب تعداد پیشنهاد

در نمودار ۲ (الف)، معیارهای تمرکز، فراخوانی و معیار اف، در نتایج آزمون نشان داده شده است. در این نمودار و برای پنج پیشنهاد اول بهترین نتایج تمرکز[1] با دقت ۰/۷۳ از ۱ به‌دست آمده است و نشان می‌دهد که بیش از ۷۳ درصد پنج پیشنهاد اول از نظر کاربران مفید بوده‌اند که نمایانگر پیشنهاددهی مناسب است. در این آزمون با افزایش تعداد پیشنهاد، سرعت رشد فراخوانی بسیار بیشتر از سرعت کاهش تمرکز است. از این‌رو، تعداد پیشنهاد بالاتر (۳۰ پیشنهاد اول)، معیار اف (کارآمدی همزمان دو متغیر تمرکز و فراخوانی) بهتری نسبت به پیشنهادهای کمتر دارد. معیارهای سه‌گانه گفته شده، اهمیت میزان امتیاز داده‌شدهٔ کاربران را در نظر نمی‌گیرد. از این‌رو، از معیار NDCG نیز بهره گرفته شده و نتایج در نمودار ۲ (ب) آمده است که باز هم تأیید می‌کند در تعداد پیشنهاد بیشتر، امتیاز NDCG بیشتر است

1. precision



نمودار ۲ (الف). شاخص‌های تمرکز، فراخوانی و معیار اف



نمودار ۲ (ب). ارزیابی بر اساس روش سود تجمعی کاهنده نرمال‌شده

۵. نتیجه‌گیری

با توسعهٔ سامانه‌های مرتبط با پایان‌نامه‌ها و رساله‌های الکترونیک و افزایش «پارسا»ها و کاربران این سامانه‌ها، اهمیت دخالت دادن اطلاعات بافتی کاربران و «پارسا»ها برای ارائهٔ خدمات، بیشتر از قبل احساس می‌شود. ترکیب اطلاعات بافتی کاربر (مانند مکان فعلی و پرس‌وجوهای قبلی کاربر) با اطلاعات بافتی «پارسا» (مانند مکان مورد مطالعه و کلیدواژه‌ها) می‌تواند به ارائه پیشنهادهای بافت‌آگاه کمک کند. یافته‌ها می‌تواند در طراحی سامانه‌های جدید دانشگاهی یا بازطراحی و بهبود عملکرد سامانه‌های موجود (مانند سامانهٔ

«گنج») مفید باشد. هدف این مطالعه یافتن روشی برای اشاعه گزینشی اطلاعات و هسته اصلی آن پیشنهاد «پارسا» به کاربران مبتنی بر داده‌های قبلی کسب‌شده از ایشان بود که در سه مرحلهٔ مدل بافت، پیشنهاددهی بر اساس مدل بافت، و ساخت یک نرم‌افزار مبتنی بر دانش‌افزایی برای دو مرحله قبل انجام شد. نوآوری این پژوهش در ارائه مدل بافت برای سامانهٔ «پارسا»، وزن‌دهی «سی‌اف-آی‌دی‌اف» برای شباهت‌یابی بافت کاربر و بافت «پارسا» و اعمال آن بر روی سامانه «گنج» است. استفاده همزمان از بافت کاربر و بافت «پارسا» در یک پیشنهاددهنده محتوامحور (بخش ۴) با هدف اشاعه گزینشی اطلاعات، کیفیت پیشنهادها را بهبود بخشید و نشان داد که این نوع کار می‌تواند سودمند باشد

روش‌های متفاوتی در مدل بافت و نحوه استفاده از آن برای پیشنهاددهی وجود دارد. یکی از کارهای نزدیک به مطالعه حاضر، مطالعه Liang et al. (2010) است که تفاوت مطالعه حاضر با آن مطالعه در توجه به سامانه‌های «پارسا»، و نمایه‌سازی بافت کاربران در مطالعه حاضر مشابه روش به‌کاررفته در نمایه‌سازی اسناد بود. استفاده از انواع بافت و وزن‌دهی به بافت به‌جای برچسب، تفاوت دیگر با آن است. از پیشنهادهای مطالعه «سعیدنیا و حسن‌زاده» (۱۴۰۱) در طراحی بهره برده شد. در آن مطالعه طراحی مدلِ تعاملی مد نظر بود. در مطالعه‌ای دیگر، «فدایی و فهیم‌نیا» (۱۳۹۰) به طرح ایجاد نظام اشاعه گزینشی اطلاعات برای اعضای هیئت علمی پرداختند که تفاوت آن با مطالعه حاضر، در روش اشاعه داده‌های پایگاه اطلاعات موجود «پارسا»ها، بافت‌آگاهی، و مدل عملی برای کاربست یافته‌هاست

پیشنهادهایی برای پژوهش و کاربرد:

در شباهت‌یابی کاربران می‌توان از ترکیب بافت‌آگاه با روش‌های دیگری مانند خوشه‌بندی، یافتن K- میانگین و سایر روش‌های داده‌کاوی استفاده کرد و «پارسا»ها را پس از مقایسه با بافت‌های مشترک کاربران در یک خوشه، به سایر کاربران همان خوشه ارسال و نتایج را با این پژوهش مقایسه نمود. این کار به احتمال، به کارآمدی منابع سامانه کمک می‌کند. هوش مصنوعی می‌تواند با استفاده از فناوری‌ها و الگوریتم‌ها، انتشار گزینشی اطلاعات را تا حد زیادی بهبود دهد (Soleymani et al. 2023)

مدل‌های عادی بازیابی اطلاعات بر اساس تعداد انطباق کلمات موجود در پرس‌وجو با سند، نتایج را بازمی‌گردانند و شواهد مربوط به «بودن» کلمات مرتبط را در سند نادیده می‌گیرند. بنابراین، هنگام رتبه‌بندی برای کلمه «ایران»، کلمات دیگری مانند «تهران» یا

«آبادان» ممکن است بسیار آموزنده باشد (Mitra and Craswell 2018). با لحاظ کردن روابط معنایی اصطلاحات (شامل وابستگی، مفاهیم اعم و مفاهیم اخص) می‌توان توسعه جدیدی در سامانه پیشنهاددهنده داد که در فرایند یادگیری به بافت‌ها افزوده شود. افزودن واژگان اصطلاحنامه‌ای یا نمونه‌های هستان‌نگار مرتبط با هر بافت در مرحله کسب و یادگیری بافت با افزودن روابط اعم، اخص، مرتبط در اصطلاحنامه یا روابط معنایی به غنای داده‌های بافت‌ها می‌افزاید. هستان‌نگار ستون فقرات وب معنایی است (میرعرب و خرم‌آبادی ۱۴۰۱). هستان‌نگارهای ایرانی مانند شاهنامه، هستان‌نگار قرآن‌جوی، حملات کامپیوتری، برهان، پیکره قرآنی فرقان و علم‌سنجی وجود دارند (همان)، اما حدود این مقاله اجازه نمی‌داد که از هستان‌نگارهای موجود و پیاده‌سازی آن‌ها در مفاهیمی که بافت کاربر و «پارسا» را تشکیل می‌دادند، استفاده شود. افزودن روابط اصطلاحنامه‌ای می‌تواند نرخ فراخوانی را افزایش دهد، ولی ممکن است نرخ تمرکز را نیز کاهش دهد.

توانایی بازیابی اطلاعات چون فهم هم‌معنایی، عملکرد در ورودی‌های ناکافی یا بدون تنوع، حساسیت به تنوع در پیکره اسناد، بهبود ورودی‌های نادرست و کارایی مدل‌ها (Mitra and Craswell 2018) همچنان موضوعات جدی در بازیابی بافت‌آگاه است و به‌ویژه کارایی سامانه‌ای مدل‌ها در این مقاله مطالعه نشده است و کارایی مصرف منابع برای پیشنهاددهی می‌تواند مبنای مطالعه دیگری باشد.

مدل کسب‌وکار در سامانه «گنج» می‌تواند چنان تغییر کند که بتوان «پارسا»های ذخیره‌شده را برای کاربران غیردانشگاهی (مانند نمایندگان کسب‌وکارها و صنعت، دستگاه‌های اجرایی و نهادها) با یک روش خاص درآمدی و برای کاربران دانشگاهی با روش درآمدی دیگری و همگی مبتنی بر بافت‌های مرتبط با هر یک از کاربران (مانند محل حضور، محتوای مورد علاقه سازمانی، وظایف و پست‌های سازمانی) ارسال کرد

محدودیت‌های پژوهش

پژوهش‌های علمی اغلب با محدودیت‌هایی مواجه هستند که آگاهی از آن‌ها در تفسیر و بهره‌برداری از نتایج اثرگذار است. این پژوهش نیز از این قاعده جدا نبوده است. به‌دلیل رعایت حریم خصوصی، ایمیل یا تماس کاربران از سوی «ایرانداک» در اختیار پژوهشگران نبود تا بتوانند نتایج را با کاربران به اشتراک بگذارند و بازخورد دریافت کنند. به این دلیل از کاربران در دسترس برای ارزیابی استفاده شد

قدردانی

از رئیس مرکز فناوری اطلاعات و امنیت فضای مجازی «ایرانداک» و کارکنان آن واحد برای همکاری در ارائه اطلاعات، قدردانی می‌شود

فهرست منابع

دهقانی، زهره. ۱۳۸۹. طراحی یک مدل برای سیستم‌های پیشنهاددهنده آگاه از زمینه در کتابخانه‌های دیجیتال. پایان‌نامه کارشناسی ارشد روان‌شناسی. دانشگاه اصفهان

حسان، رضا، رحمان شریف‌زاده، و امیرحسین صدیقی. ۱۳۹۹. روش‌شناسی پژوهشی علم طراحی به مثابهٔ یک روش‌شناسی راه‌حل‌محور. *روش‌شناسی علوم انسانی* ۲۶ (۱۰۵): ۳۵-۵۰.

رجبی، عباس، متینه‌السادات معین آزاد طهرانی، و ملیحه درخوش. ۱۳۹۳. نظام‌های بافت آگاه، مفهوم، کارکردها و کاربردهای آن در کتابخانه‌های دیجیتال. *تعامل انسان و اطلاعات* ۱ (۳): ۲۳۵-۲۴۵.

روحانی رانکوهی، سید محمدتقی. ۱۳۹۶. *مفاهیم بنیادی پایگاه داده*. ویراست چهارم. تهران: انتشارات جلوه.

سعیدنیا، حمیدرضا و محمد حسن‌زاده. ۱۴۰۱. طراحی مدلی تعاملی عملیاتی برای اشاعه گزینشی اطلاعات در کتابخانه‌های دانشگاهی (مطالعه موردی: دانشگاه تربیت مدرس). *فصلنامه علمی بازیابی دانش و نظام‌های معنایی* ۹ (۳۲): ۱-۳۴.

صادقی پوریانی، عباس و مرضیه زرین‌بال. ۱۳۹۹. شناسایی بافت‌ها برای کاربست در سامانه بافت‌آگاه پایان‌نامه‌ها و رساله‌های الکترونیکی. *پژوهشنامه پردازش و مدیریت اطلاعات* ۲ (۳۸): ۴۸۳-۵۱۴.

علیدوستی، سیروس، و مریم صابری. ۱۳۸۶. پایان‌نامه‌ها و رساله‌های الکترونیکی: نسل جدید مدارک علمی. *فصلنامه مطالعات ملی کتابداری و سازماندهی اطلاعات* ۲ (۱۸): ۸۵-۱۰۰.

فدایی، غلامرضا و فاطمه فهیم‌نیا. ۱۳۹۰. طرح ایجاد نظام اشاعه گزینشی اطلاعات برای اعضای هیئت علمی دانشکده روان‌شناسی و علوم تربیتی دانشگاه تهران. *فصلنامه پژوهشگاه علوم و فناوری اطلاعات ایران* ۱۰۹- ۱۲۶.

میرعرب، علی و محمدرضا خرم‌آبادی آرانی. ۱۴۰۱. مروری بر هستان‌نگارهای ایرانی به‌منظور بازیابی معنایی اطلاعات فارسی از منظر تحلیلی. *علوم و فنون مدیریت* ۸ (۳): ۲۰۱-۲۳۱.

References

Anjaian, M. 2013. Electronic Information services in digital environment: Need of the hour. *International Journal of Digital Library Services* 3 (2): 46-54.

Baldauf, Matthias, Schahram Dustdar, and Florian Rosenberg. 2007. A survey on context-aware systems. *International Journal of ad Hoc and ubiquitous Computing* 2 (4): 263277-.

Bauman, Konstantin, Alexey Vasilev, and Alexander Tuzhilin. 2022. The Long Tail of Context: Does it Exist and Matter? *arXiv preprint arXiv:2210.01023*.

Bobadilla, Jesús, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. 2013. Recommender systems survey. *Knowledge-based systems* 46: 109-132.

Burges, C., T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. 2005. Learning to rank using gradient descent. In Proceedings of the 22nd international conference on Machine learning (89-96).

Chaari, Tarak, Mohamed Zouari, and Frédérique Laforest. 2009. Ontology based context-aware adaptation approach. In Context-Aware Mobile and Ubiquitous Computing for Enhanced Usability: Adaptive Technologies and Applications (26-58). IGI Global.

Chen, Guanliang, and Li Chen. 2014. Recommendation based on contextual opinions. In User Modeling, Adaptation, and Personalization: 22nd International Conference, UMAP 2014, Aalborg, Denmark, July 7-11, 2014. Proceedings 22 (61-73). Springer International Publishing

2015 ._____. Augmenting service recommender systems by incorporating contextual opinions from user reviews. *User Modeling and User-Adapted Interaction* 25: 295-329.

Cyganiak, Richard, David Wood, and Markus Lanthaler. 2014. *RDF 1.1 Concepts and Abstract Syntax*. W3C.

De Giusti, Marisa R., Gonzalo L. Villarreal, Agustín Vosou, and Juan P. Martínez. 2010. An Ontology-based Context Aware System for Selective Dissemination of Information in a Digital Library. *Journal of Computing* 2 (5): 6-13. http://arxiv.org/abs/1005.4008%5Cnhttp://www.arxiv.org/pdf/1005.4008.pdf. (accessed ?)

Dey, Anind K. 2001. Understanding and using context. *Personal and ubiquitous computing* 5 (1): 4-7.

Gollagi, S. G., M. M. Math, and U. P. Kulkarni. 2019. Retracted Article: Research outlook and state-of-the-art methods in context awareness data modeling and retrieval. *Evolutionary Intelligence* 15 (2): 1025-1036. https://doi.org/10.1007/s12065-019-00274-x.

Hansen, Casper, Christian Hansen, Lucas Maystre, Rishabh Mehrotra, Brian Brost, Federico Tomasi, and Mounia Lalmas. 2020. Contextual and sequential user embeddings for large-scale music recommendation. In Proceedings of the 14th ACM Conference on Recommender Systems (53-62).

Hasan, Emrul, Mizanur Rahman, Chen Ding, Jimmy Xiangji Huang, and Shaina Raza. 2024. Review-based Recommender Systems: A Survey of Approaches, Challenges and Future Perspectives. *arXiv preprint arXiv:2405.05562*.

Liang, Huizhi, Yue Xu, Yuefeng Li, Richi Nayak, and Xiaohui Tao. 2010. Connecting users and items with weighted tags for personalized item recommendations In Proceedings of the 21st ACM conference on Hypertext and hypermedia (51-60).

Lops, Pasquale, Marco de Gemmis, and Giovanni Semeraro. 2011a. *Content-based Recommender Systems: State of the Art and Trends.* 73-105. Boston, MA: Springer US.

---. 2011b. Content-based Recommender Systems: State of the Art and Trends." In *Recommender Systems Handbook*, 73-105.

Luhn, H. P. 1958. A business intelligence system. *IBM J. Res. Dev.* 2 %@ 0018-8646 %U https://doi.org/10.1147/rd.24.0314 (4): 314–319. https://doi.org/10.1147/rd.24.0314.

Mezzi, Melyara, and Nadjia Benblidia. 2017. Study of Context Modelling Criteria in Information Retrieval. *International Journal of Information Technology and Computer Science* 9 (3): 28-39. https://doi.org/10.5815/ijitcs.2017.03.04.

Mitra, Bhaskar, and Nick Craswell. 2018. An introduction to neural information retrieval. *Foundations and Trends® in Information Retrieval* 13 (1): 1-126.

Mobasher, Bamshad. 2013. Context-Aware User Modeling for Recommendation. In Tutorial at the 21st Conference on User Modeling, Adaptation and Personalization (UMAP 2013).

Peffers, Ken, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. 2007. A design science research methodology for information systems research. *Journal of management information systems* 24 (3): 45-77.

Porcel, Carlos, Alberto Ching-lpez, Juan Bernabé-Moreno, Alvaro Tejeda-Lorente, and Enrique Herrera-Viedma. 2017. Fuzzy linguistic recommender systems for the selective diffusion of information in digital libraries. *JIPS (Journal of Information Processing Systems)* 13 (4): 653-667.

Raza, Shaina, and Chen Ding. 2019. Progress in context-aware recommender systems — An overview. *Computer Science Review* 31: 84-97. https://doi.org/10.1016/j.cosrev.2019.01.001. https://linkinghub.elsevier.com/retrieve/pii/S1574013718302120.

Recker, Jan. 2013. *Scientific Research in Information Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Saeidnia, Hamid Reza. 2023. Ethical artificial intelligence (AI): confronting bias and discrimination in the library and information industry. *Library Hi Tech News*. *Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/LHTN-10-2023-0182*.

Selvi, C., and E. Sivasankar. 2019. An efficient context-aware music recommendation based on emotion and time context. In Data Science and Big Data Analytics: ACM-WIR 2018 (215-228). Singapore: Springer Singapore.

Soleimani, Hooman, Bakthavachalam Elango, and Mohammad Hassanzadeh. 2024. A Systematic Review of a web-based System for Selective Dissemination of Information in Developing Nations: Benefits, Challenges, and Key Features. *InfoScience Trends* 1 (3): 43-55.

Soleymani, Hooman, Hamid Reza Saeidnia, Marcel Ausloos, and Mohammad Hassanzadeh. 2023. Selective dissemination of information (SDI) in the age of artificial intelligence (AI). *Library Hi Tech News*. Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/LHTN-08-2023-0156.

Strang, Thomas, and Claudia Linnhoff Popien. 2004. A context modeling survey. In Workshop Proceedings.

Uzohue, C., and J. Yaya. 2016. Provision of current awareness services and selective dissemination of information by medical librarians in technological era. *American Journal of Information Science and Computer Engineering* 2 (2): 8-14.

Knijnenburg, Bart P., and Martijn C. Willemsen. 2015. Evaluating recommender systems with user experiments. In Recommender systems handbook, 309352-. Boston, MA: Springer US.

Wu, Hao, Yijian Pei, Bo Li, Zongzhan Kang, Xiaoxin Liu, and Hao Li. 2015. Item recommendation in collaborative tagging systems via heuristic data fusion. *Knowledge-Based Systems* 75: 124-140.

Wu, Wenmin, Jianli Zhao, Chunsheng Zhang, Fang Meng, Zeli Zhang, Yang Zhang, and Qiuxia Sun. 2017. Improving performance of tensor-based context-aware recommenders using bias tensor factorization with context feature auto-encoding. *Knowledge-Based Systems* 128: 71-77.

عباس صادقی پوریانی

دانشجوی دکتری در رشتهٔ مدیریت فناوری اطلاعات- کسب‌وکار هوشمند در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) و دانش‌آموخته ریاضی از دانشگاه صنعتی امیرکبیر است.

برنامه‌ریزی سازمانی، مدل‌سازی ریاضی، گردشگری الکترونیک، سیستم‌های بافت‌آگاه، توسعهٔ نرم‌افزارهای رایانه‌ای از جمله علایق پژوهشی وی است.

مرضیه زرین‌بال

دارای مدرک تحصیلی دکتری در رشتهٔ مهندسی صنایع از دانشگاه صنعتی امیرکبیر است. ایشان هم‌اکنون استادیار دانشگاه صنعتی امیرکبیر و پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) است.

طراحی سیستم‌های اطلاعاتی، بازی‌وارسازی، پردازش تصویر و منطق فازی از جمله علایق پژوهشی وی است.

پژوهش‌نامه

پردازش و مدیریت اطلاعات