

عنوان: اشاعه گزینشی اطلاعات بافت آگاه برای پایان‌نامه‌ها و رساله‌های الکترونیکی

نام نویسندگان:

عباس صادقی پوریانی، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، تهران، ایران، دانشجوی دکتری،

sadeghi@students.irandoc.ac.ir

مرضیه زرین بال، دانشگاه صنعتی امیرکبیر؛ پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، استادیار،

zarinbal@irandoc.ac.ir

مقدمه: در سامانهٔ پایان‌نامه‌ها و رساله‌های الکترونیکی (پارسا)، اشاعه گزینشی اطلاعات، رساندن اسناد پارسا به کاربرانی است که آن اسناد احتمال سودمندی بیشتری برای کاربران داشته باشند. هدف پژوهش حاضر، ارائه مدلی برای ارسال پارساهای گزینش شده به کاربران بر اساس بافت‌های ذخیره شده از هر دو موجودیت کاربران و پارساها بوده است. سوال اصلی پژوهش این بود که چگونه می‌توان پارساهای گزینش شده مبتنی بر بافت را به سمت کاربرانی ارسال کرد که بیشترین سودمندی را برای ایشان داشته باشد. هسته این مدل، یک سیستم پیشنهاددهنده بافت آگاه محتوا محور است که در آن بافت پارسا با بافت کاربر مقایسه و کاربران هدف با احتمال سودمندی بیشتر برگزیده می‌شوند.

روش‌شناسی: نوع تحقیق از نظر جهت‌گیری اصلی، کاربردی- توسعه‌ای و ماهیت تحلیل، تحقیق به شیوه علم طراحی بوده که در آن یک مصنوع (سامانه نرم‌افزاری) دارای نوآوری ساخته شده است. نوآوری پژوهش: مدل بافت کاربران، مدل بافت پارساها و روش وزن‌دهی به بافت‌ها برای سامانه‌ی پارسا است. در اشاعه گزینشی اطلاعات، بافت‌های کاربر و پارسا براساس روش پیشنهادی مقاله حاضر، «سی‌اف-آی‌سی‌اف» وزن‌دهی و مقایسه شدند تا با مقایسه بافت‌های وزن‌دار نزدیکترین کاربران به پارساهای قابل انتشار گزینش شوند. هرگاه هر دو نوع بافت بر اساس جدول انطباق همگن بوده و دو مفهوم متناظر بافت یکسان باشند آنگاه یک بافت از کاربر با یک بافت از پارسا برابر است. برای مثال دو بافت (شهر محل مطالعه پارسا، رشت) و (شهر محل سکونت کاربر، رشت) دو بافت قابل انطباق فرض شده‌اند. برای ارزیابی مدل، از مجموع رساله‌های تولید شده سال ۱۴۰۱-۱۴۰۲ دانشگاه‌های ایران که در «سامانه ثبت» پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) ثبت شده‌اند، حدود ۱۰۸۶۰ رساله به صورت تصادفی انتخاب و برای اشاعه گزینشی به ۱۶۳ کاربر ارسال شد. این کاربران، از بین افرادی هستند که پژوهشگران رساله حاضر، امکان ارتباط با ایشان را داشته و برای آزمون مدل در دسترس بودند. ۵۳ نفر از بین آنها پارساهای ارسال شده را رتبه‌بندی کردند. مفید بودن n -پیشنهاد اول ($n=5,10,20,30$) در پیشنهاددهنده، با روش‌هایی چون تمرکز، فراخوانی، معیار-اف و سود تجمعی کاهنده نرمال شده ارزیابی شده است.

یافته‌های اصلی: مدل اشاعه گزینشی بافت آگاه براساس یک روش اکتشافی مبتنی بر تسخیر بافت‌های کاربر و بافت‌های پارسا، استنتاج و یادگیری اولیه بافت از پایگاه مفاهیم، وزن‌دهی و ذخیره بافت، گزینش اولیه

کاربران و پارساها، پالایش بافت محور برای تعیین مفیدترین پارسای مورد نیاز هر کاربر، انتشار و دریافت بازخورد از کاربران ارائه شده است. پس از ارسال پارساها برای کاربران، آنها میزان مفید بودن را در بازه ۰ تا ۱۰ امتیازدهی کردند. بررسی معیارهای تمرکز، فراخوانی و معیار اف، در نتایج آزمون نشان داد که برای ۵- پیشنهاد اول بهترین نتایج تمرکز به میزان ۰.۷۳ از ۱ به دست آمده است و بیش از ۵۶ درصد ۵-پیشنهاد اول از نظر کاربران خیلی مفید بوده‌اند که نمایانگر پیشنهاددهی مناسب است. در این آزمون، با افزایش تعداد پیشنهاد، سرعت رشد فراخوانی بیشتر از سرعت کاهش تمرکز است، از این رو، در تعداد پیشنهاد بالاتر (یعنی: ۳۰-پیشنهاد اول) معیار اف (کارآمدی همزمان دو متغیر تمرکز و فراخوانی) بهتری نسبت به پیشنهادها کمتر دارد. معیارهای سه گانه گفته شده، اهمیت میزان امتیاز داده شده کاربران را در نظر نمی‌گیرد از این رو، از معیار سود جمعی کاهنده نرمال شده (NDCG) نیز بهره گرفته شده است که بازهم تایید می‌کند در تعداد پیشنهادها بیشتر، امتیاز NDCG بیشتر است.

بحث و نتیجه گیری: در این رساله مدلی مبتنی بر انطباق بافت‌های پارسا با بافت‌های کاربر برای پیشنهاددهی ارائه شد. در این مدل نه تنها انطباق یک اصطلاح در محتوای پارسا با همان اصطلاح از پروفایل کاربر مد نظر است، بلکه علاوه بر آن انطباق بافتی از سوی هر دو موجودیت پارسا و کاربر ملاک تصمیم است. این پژوهش مدلی را برای اشاعه‌گزینی اطلاعات ارائه می‌دهد که می‌تواند برای توسعه‌دهندگان و مدیران سامانه پارسا مفید باشد. در شباهت‌یابی کاربران می‌توان از ترکیب بافت آگاه با روش‌های دیگری مانند خوشه‌بندی نیز استفاده و پارساها را پس از مقایسه با بافت‌های مشترک کاربران در یک خوشه، به سایر کاربران همان خوشه ارسال و نتایج را با این پژوهش مقایسه کرد. با لحاظ روابط معنایی اصطلاحات (شامل وابستگی، مفاهیم اعم و مفاهیم اخص) می‌توان توسعه جدیدی در سامانه پیشنهاددهنده داد که در فرآیند یادگیری به بافت‌ها افزوده شود. مدل کسب و کار در سامانه گنج می‌تواند چنان تغییر کند که بتوان پارساهای ذخیره شده را برای کاربران غیردانشگاهی (مانند: کسب و کارها، صنعت، دستگاه‌های اجرایی و نهادها) براساس بافت شرکت یا سازمان آنها (مانند: محل حضور، محتوای مورد علاقه سازمانی، وظایف و پست‌های سازمانی) ارسال کرد.

کلیدواژه‌ها: پیشنهاد دهنده بافت آگاه، پایگاه اطلاعاتی پایان‌نامه‌ها و رساله‌ها، اشاعه‌گزینی اطلاعات.

Title: A Context-aware system for selective dissemination of information in electronic theses and dissertations.

Names of Authors: Abbas Sadeghi Pouryani, PhD Candidate in Information Technology Management, Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran, sadeghi@students.irandoc.ac.ir, ORCID: 0009-0003-4924-134X.

Marzieh Zarinbal, *Corresponding author*, PhD in Industrial Engineering, Assistant Professor, Amirkabir University of Technology, Faculty of Industrial Engineering and Management Systems, Tehran, Iran; Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran, mzarinbal@aut.ac.ir, ORCID: 0000-0001-9284-7329

Introduction: In an electronic dissertation and thesis (ETDs) system, selective dissemination of information involves delivering ETD documents to users who are most likely to find them useful. This research aims to present a model for selective dissemination of information using both user and ETD contexts in an ETDs database. The primary research question is how to deliver selected ETDs to users in a manner that aligns with their contexts. 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 finding the useful are selected.

Methodology: The research focused on applied development and followed the design science approach, in which a software system is constructed. This research contributed through the model of user contexts, the model of ETD contexts, and the method of weighting contexts for the ETDs system. In selective dissemination of information, a user's contexts and an ETD's contexts were weighted based on the proposed method outlined in the paper "CF-ICF." Two contexts from User and ETD are considered the same if the context types are homogeneous according to the correspondence table and their context values are identical. For example, two contexts (Case study in ETD, Rasht city) and (user's city of residence, Rasht city) are considered equivalent. The vector of the user's contexts was compared to the vector of the ETD's contexts using the cosine similarity. To validate the model, approximately 10860 ETDs from the total number produced in academic year 1401-1402 by Iranian universities registered with the Iranian Institute of Information Science and Technology were randomly selected and sent to 163 users for dissemination of information. users were individuals with whom the researcher of the present thesis had the opportunity to communicate and were available to test the model. Fifty-three of them rated the submitted ETDs. The usefulness of the first n-recommendations (n=5, 10, 20, 30) in the

recommendation evaluated using methods such as precision, recall, F-measure, and normalized discounted cumulative gain.

Main findings: The selective dissemination of information model was utilized as a heuristic approach that captured user contexts and ETD contexts, extracted new context by comparing it to a database of concepts, selected ETDs for dissemination, and used context-based filtering to the most relevant users. The ETD was then sent to these users, who provided feedback by rating its usefulness on a scale of 0 to 10. The results indicated that the recommendations were highly beneficial to users in at least 56% of cases. For these recommendations, the precision was 0.73 out of 1, and the Normalized Discounted Cumulative Gain index was 0.69 out of 1.

Discussion and conclusions: In this study, a model based on matching ETD contexts with user contexts was presented for recommendation. In this model, not only is the matching of a term in ETD content with the same term from the user profile considered, but also the contextual matching from both ETD and the user serves as the decision criterion. 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. In user similarity detection, context-awareness can be combined with clustering methods. The ETD can be sent to other users within the same cluster after comparing them with the common contexts of users in that cluster, and the results can be compared with this research. By considering the semantic relationships of terms (including related, broader, and narrower concepts), a new development can be made in the recommender system that can be added to the contexts in the learning process. The business model of the Iranian ETDs system (GANJ) can be modified to deliver stored ETDs to non-academic users (such as businesses, industries, and institutional users) by aligning with the context of their company or organization (such as location, organizational interests, tasks, and positions).

Keywords: Context-aware recommender, dissertation and thesis database, selective dissemination of information.