



Challenges of Storing and Retrieving Researchers' Names in Information Repositories: A Case Study (Scientific Information System of Ferdowsi University of Mashhad)

Hedayat Behroozfar 

*Corresponding author, Assistant Professor, Department of Knowledge and Information Science, Faculty of Literature & Humanities, University of Zabol, Zabol, Iran. E-mail: hedayatbehroozfar@uoz.ac.ir.

Abstract

Objective: One of the most important identifiers that users of any information system use in their searches is the author name identifier. Therefore, the designers and managers of any information system are always looking to improve the quality of content retrieval based on the search for the author's name. The content of information repositories and the quality of service delivery in them is affected by the quality and manner of storing and retrieving personal names. The purpose of this article is to investigate the problems and challenges of searching for the names of individuals in academic organizational repositories from its practical aspect.

Methodology: In order to investigate the practical challenges of searching for researchers in the organizational repository of Ferdowsi University of Mashhad, the evaluation and observation methods were used. In this method, performance measurement, which is a more specific type of evaluative research, is used. Accordingly, in this study, an attempt has been made to investigate the performance of the search engine of the reservoir system of Ferdowsi University of Mashhad in order to search for the names of ambiguous persons.

Findings: Despite the relatively small amount of information in this repository compared to very large organizational repositories, a search based on the name of the author in this organizational repository showed that there are issues and challenges related to name retrieval in this repository. A high percentage of similar surnames indicated that users had to be aware of people's first and last names in addition to surnames in many searches. About 40 percent of the names were multi-part or compound, which for various reasons lead to irrelevant or less relevant retrieving.

Conclusion: Given that all the ambiguous names extracted from the researchers' Name List in the Ferdowsi University of Mashhad repository were searched in order to examine the performance of the search, the result indicated that there were challenges in searching for the names of individuals and the designed system could not display related results. In general, it can be said that the organizational repository of Ferdowsi University of Mashhad, when searching

for the ambiguous names, has problems and confusion due to the similarity of the individual names, or due to language problems and its spelling. Authority control of the names of the researchers nationally and internationally is one of the most important and effective ways out of the challenge of the name searching. It is necessary to ask researchers to enter their information carefully in comprehensive systems and librarians to manually control the authority of their names. In this way, a national system of authority control of Iranian researchers in Persian and English languages emerges, which can have many uses and solve the problems of searching by name to a large extent.

Keywords: Authority control, Information retrieval, Organizational repositories, Names, Scientific information systems.

Article type: Research



ARTICLE INFO

Article history:

Received: 1/11/2021

Received in revised form: 30/11/2021

Accepted: 05/12/2021

Available online: 18/04/2022

Publisher: Central Library of Astan Quds Razavi
Library and Information Sciences, 2022, Vol. 25, No.1, pp. 221-243.

© The author



چالش‌های بازیابی نام‌های پژوهشگران در مخازن اطلاعاتی: مطالعه موردی (سامانه اطلاعات علمی دانشگاه فردوسی مشهد)

هدایت بهروزفر

*نویسنده مسئول، استادیار گروه علم اطلاعات و دانش‌شناسی، دانشکده ادبیات و علوم انسانی، دانشگاه زابل، زابل، ایران. رایانامه: hedayatbehroozfar@uoz.ac.ir

چکیده

هدف: یکی از مهمترین شناسه‌هایی که کاربران هر نظام اطلاعاتی در جستجوهای خود از آن بهره می‌برند، شناسه نام پدیدآور است. بنابراین طراحان و مدیران هر نظام اطلاعاتی همواره در پی آن هستند تا کیفیت بازیابی محتوا را بر اساس جستجوی نام پدیدآورنده بهبود بخشند. محتوای مخازن اطلاعاتی و کیفیت ارائه خدمات در آن‌ها متأثر از کیفیت و چگونگی ذخیره و بازیابی نام‌های اشخاص است. هدف مقاله حاضر، بررسی مشکلات و چالش‌های جستجوی نام اشخاص در مخازن سازمانی دانشگاهی از جنبه کاربردی آن است.

روش: به منظور بررسی چالش‌های عملی جستجوی نام پژوهشگران در مخزن سازمانی دانشگاه فردوسی مشهد، از روش ارزیابانه و مشاهده استفاده گردید. در این روش، از اندازه‌گیری عملکرد که نوع اختصاصی‌تری از پژوهش ارزیابانه است، استفاده شده است. بر این اساس، در این پژوهش تلاش شده است تا عملکرد موتور جستجوی سامانه مخزن دانشگاه فردوسی مشهد از جهت جستجوی نام اشخاص دارای ابهام مورد بررسی قرار گیرد.

یافته‌ها: با وجود حجم نسبتاً کم اطلاعات در این مخزن در مقایسه با مخازن سازمانی خیلی بزرگ، بررسی جستجو بر اساس نام پدیدآور در این مخزن سازمانی نشان داد مسائل و چالش‌های مربوط به جستجوی نام در این مخزن نیز وجود دارد. درصد بالایی از نام‌های خانوادگی مشابه نشان داد که کاربران مجبورند در بسیاری از جستجوها علاوه بر نام خانوادگی از نام کوچک افراد نیز آگاه باشند. حدود ۴۰ درصد از اسامی چندبخشی یا مرکب بودند که به دلایل مختلف باعث ایجاد بازیافت نامربوط یا کم مربوط می‌گردند.

نتیجه‌گیری: مستندسازی اسامی پژوهشگران در سطح ملی و بین‌المللی از جمله مهمترین و اثرگذارترین راهکارهای برون‌رفت از چالش جستجوی نام‌ها است. ضروری است تا از پژوهشگران خواسته شود اطلاعات خود را با دقت در سامانه‌ای جامع وارد نمایند و کتابداران نیز به صورت دستی نام آن‌ها را کنترل و مستند نمایند. به این ترتیب نظام ملی کنترل مستند نام‌های پژوهشگران ایرانی به زبان فارسی و انگلیسی پدید می‌آید که می‌تواند استفاده‌های فراوانی داشته و مشکلات جستجو بر اساس نام را تا حدود زیادی مرتفع سازد.

کلیدواژه‌ها: بازیابی اطلاعات، سامانه‌های اطلاعات علمی، کنترل مستند، مخازن سازمانی، نام‌ها.

نوع مقاله: پژوهشی

تاریخچه مقاله:

تاریخ دریافت: ۱۴۰۰/۸/۱۰

تاریخ ویرایش: ۱۴۰۰/۹/۹

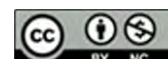
تاریخ پذیرش: ۱۴۰۰/۹/۱۴

تاریخ انتشار: ۱۴۰۱/۱/۲۹

ناشر: کتابخانه مرکزی آستان قدس رضوی

کتب‌داری و اطلاع‌رسانی، ۱۴۰۱، دوره ۲۵، شماره ۱، شماره پیاپی ۹۷، صص. ۲۲۱-۲۴۳

© نویسنده



مقدمه

با توسعه و گسترش روزافزون استفاده از فناوری اطلاعات در گردآوری، ذخیره و اشاعه اطلاعات، نظام‌های اطلاعاتی متعددی در اقصی نقاط دنیا به وجود آمده است که به کاربران در جستجو و دسترسی به اطلاعات مرتبط و مطلوب یاری می‌رساند. ناشران نیز در همراهی با این تحولات، فعالیت‌های گذشته خویش را متحول ساخته و با روزآمدسازی فرآیندهای سنتی نشر، پا به عرصه نشر الکترونیکی گذارده‌اند و بسیاری از انتشارات خویش را به صورت الکترونیکی در اختیار کاربران قرار می‌دهند. کارگزاران اطلاعات نیز اقدام به ایجاد پایگاه‌های اطلاعاتی متعددی نموده‌اند و با همکاری ناشران مختلف، مقاله‌ها و آثار علمی پژوهشگران و دانشمندان را به صورت الکترونیکی منتشر می‌سازند و به شیوه‌ها و روش‌های متنوع در دسترس علاقه‌مندان قرار می‌دهند. در این میان، سازمان‌ها و نهادهای دولتی و خصوصی در کشورهای مختلف دنیا نیز همان‌گونه که در گذشته انتشارات مخصوص به خود داشتند و آن را به شکل‌های مختلف توزیع می‌نمودند، به انتشارات الکترونیکی آثار خویش پرداخته و آن را به روش‌های مرسوم الکترونیکی در اختیار کاربران و مشتریان خویش قرار می‌دهند. این سازمان‌ها و نهادها در حقیقت آرشیو و مخزنی ایجاد می‌کنند و آثار علمی و مهم خویش را درون آن نگهداری و ذخیره می‌نمایند و تسهیلات لازم را برای جستجو و بازیابی اطلاعات فراهم می‌آورند. مخزن سازمانی، مجموعه‌ای از خدمات را ارائه می‌دهد که شامل برنامه، سازمان، دسترسی، توزیع و حفظ محتوای دیجیتال می‌شود (راتانیا^۱، ۲۰۱۷). مخزن سازمانی حکمران انتشارات علمی یک مؤسسه در قالب دیجیتالی است که این انتشارات حاصل فعالیت‌های علمی آن مؤسسه هستند (بنسود^۲، ۲۰۱۱؛ روزومند^۳، ۲۰۱۶). کلیفورد لینچ معتقد است مخزن سازمانی، مجموعه‌ای از خدمات ارائه شده به وسیله دانشگاه است که امکان مدیریت و اشاعه مواد دیجیتالی ارائه شده توسط سازمان و اعضای آن را فراهم می‌کند (لینچ^۴، ۲۰۰۳). مخزن سازمانی یک «نظام آرشیوی دیجیتال شامل مجموعه‌ای از پایگاه‌های مرتبط و یکپارچه است که در یک بخش از یک سازمان و یا در سازمان به طور کلی ایجاد و نگهداری می‌شود. این مخزن بستری برای گردآوری، ذخیره‌سازی، حفاظت و نگهداری و اشاعه یا به بیانی کلی مدیریت اطلاعات تولید شده به وسیله سازمان و نیز دسترس‌پذیر ساختن آن به شمار می‌آید» (زاهدی نوقابی، ۱۳۸۹). استفاده از مخزن سازمانی در دانشگاه‌ها و پژوهشگاه‌ها برای مدیریت اطلاعات علمی و افزایش و ارتقاء کیفیت دسترسی به اطلاعات و نیز فراهم‌آوری شرایط مناسب‌تر برای ارتباطات علمی است. البته باید توجه نمود که امروزه از عبارت سامانه‌های

1. Ratanya
2. Bansode
3. Rozumand
4. Lynch

اطلاعاتی برای مخازن سازمانی دانشگاه استفاده می‌شود. «مهمترین کارکرد مخازن سازمانی افزایش دسترسی به محتوای علمی تولید شده توسط اعضاء هیئت علمی است که می‌تواند به منظور تدریس، یادگیری یا همکاری علمی مورد استفاده قرار گیرد» (زاهدی نوقابی، ۱۳۹۴). مطابق تعریف‌های بالا و تعریف‌های فراوان دیگری که در متون برای این اصطلاح وجود دارد می‌توان گفت مخازن سازمانی یا سامانه‌های اطلاعات علمی در دانشگاه‌ها بستری نرم‌افزاری برای حفظ و نگهداری و بهره‌برداری از نتیجه تلاش‌های علمی اعضای هیئت علمی یک دانشگاه و یا اطلاعات موجود در یک سازمان هستند.

در دسترس بودن سیستم‌های بازیابی اطلاعات منبع باز، باعث گسترش مخازن سازمانی در سراسر جهان، به ویژه در میان نهادهای دانشگاهی و پژوهشی شده است. متناسب با میزان ایجاد مخازن سازمانی در چند سال گذشته به ویژه دهه اخیر، شیوه‌ها و خدمات بازیابی اطلاعات نیز برای سازمان‌های مختلف بسیار جذاب و گریزنایپذیر گردیده است. سودمندی مخازن سازمانی در محیط‌های دانشگاهی، به واسطه ایجاد دسترسی به دانش و توسعه علمی است. به عبارت بهتر، آن‌ها با عرضه یک پیشینه اطلاعاتی از خروجی‌های پژوهشی سازمان و دسترس‌پذیر نمودن جهانی آن، میزان رؤیت‌پذیری و استفاده از آن را چند برابر می‌نمایند که هم سبب ارتقاء وضعیت سازمان مربوطه در رتبه‌بندی‌های جهانی می‌شوند و هم به طور کلی باعث پیشرفت و توسعه علمی می‌گردند. از این روست که مخازن، تبدیل به بستری نرم‌افزاری برای ارائه یافته‌های علمی همه پژوهشگران و دانشگاهیان در سراسر دنیا شده است.

از آنجا که آموزش و پژوهش دو هدف بنیادین آموزش عالی در هر جامعه‌ای است و دانشگاه‌ها به عنوان مکان تحقق این اهداف هستند، از این رو نتیجه پژوهش‌های دانشگاهی به عنوان خروجی پژوهشی دانشگاه‌ها هم برای خود پژوهشگران و هم برای اعتبار دانشگاه‌ها (به ویژه رتبه‌بندی‌های علم‌سنجی) بسیار مهم و حیاتی است. بنابراین دانشگاه‌ها در هر کشوری تلاش می‌کنند تا با انجام تحقیقات بیشتر و نیز ارائه نتایج آن در اشکال گوناگون، تأثیرگذاری و اهمیت خود را به اثبات برسانند. با این وصف، مخازن سازمانی ابزاری مناسب برای پیشرفت پژوهشی دانشگاه‌ها خواهند بود. گرایش به مخازن سازمانی، به دلیل بهبود دسترسی به اطلاعات پژوهشی و گسترش تأثیر کتابخانه در داخل دانشگاه، تقویت شده است (کارور^۱، ۲۰۰۳).

ایجاد مخازن سازمانی در دانشگاه‌های ایران امری نوپاست و سابقه چندان زیادی ندارد. مخزن سازمانی یا همان سامانه مدیریت اطلاعاتی علمی دانشگاه فردوسی مشهد به عنوان یکی از دانشگاه‌های مهم و پر قدمت ایران از سال ۱۳۸۵ برای جمع‌آوری، نگهداری و اشاعه اطلاعات علمی تولیدی اعضای هیئت علمی این دانشگاه به وسیله مرکز اطلاعات، آمار و امور رایانه طراحی و اجرا گردیده است. تاکنون در این سامانه بیش از

۷۳۷۱۲ عنوان مقاله علمی از ۹۰۰ عضو هیئت علمی دانشگاه قرار داده شده است که دسترسی به متن کامل بیش از ۴۱۰۰۰ عنوان آن برای همگان فراهم شده است.^۱ به طور معمول، داده‌های مرتبط با مقاله‌ها و کتاب‌های اعضای هیئت علمی دانشگاه (از جمله شکل نام‌ها) توسط خود آن اعضاء وارد فرم‌های مربوط در مخزن سازمانی می‌شود و کنترلی روی درستی آن نیست. پریخ و زاهدی نوقابی در یکی از معدود پژوهش‌های انجام شده درباره مخازن سازمانی در ایران به بررسی وضعیت مخزن سازمانی دانشگاه فردوسی مشهد و میزان همخوانی آن با ویژگی‌های یک مخزن سازمانی مطلوب پرداخته است. مطابق با یافته‌های این پژوهش، این سامانه از نظر مدیریت، در وضعیت متوسط و در رابطه با چشم‌اندازها و سیاست‌ها و ویژگی‌های فنی در وضعیت مناسبی قرار دارد. در مقایسه با معیارهای یک مخزن سازمانی مطلوب، به طور تقریبی، ویژگی‌های مخزن سازمانی دانشگاه فردوسی مشهد در حد متوسط قرار دارد (پریخ و زاهدی، ۱۳۹۰).

همان‌گونه که بیان شد، مخازن سازمانی دانشگاهی آرشیوی دیجیتال از انواع و اقسام اسناد علمی است که پژوهشگران و استادان مختلف آن دانشگاه آن‌ها را پدید آورده‌اند. بدیهی است لازمه استفاده از این مخزن سازمانی، امکان انجام جستجو به روش‌های مختلف و بر اساس شناسه‌های (نقاط بازیابی) متعدد است. به عبارت بهتر، همچون سایر نظام‌های اطلاعاتی، باید بتوان با در دست داشتن یک کلید یا شناسه، اطلاعات موجود در مخزن سازمانی دانشگاه را با دقت بالا بازیابی نمود و به آن دسترسی پیدا کرد. کاربران عموماً سه نقطه دستیابی یا جستجوی عمده شامل نام پدیدآور، عنوان و موضوع را برای بازیابی اطلاعات در سامانه‌های الکترونیکی مورد استفاده قرار می‌دهند. به همین علت بود که یکی از اهداف تهیه فهرست‌های کتابخانه‌ای، بازیافتن منابع اطلاعاتی بر اساس یکی از مشخصات نام نویسنده، عنوان و موضوع آن بود چراکه افراد به طور معمول یا از پدیدآور اطلاعاتی دارند و یا این که از عنوان مطلبی که به دنبال آن هستند آگاه هستند و یا به طور کلی موضوعی در ذهن دارند و به دنبال مطالبی درباره آن موضوع هستند. اگرچه هر یک از این سه نقطه دستیابی بسیار ارزشمند و مهم است؛ اما، در این مقاله قصد داریم تا جستجوی نام افراد را در سامانه‌های اطلاعات علمی مورد بررسی قرار دهیم.

جستجو بر اساس نام پدیدآور از جمله شناسه‌ها یا کلیدهایی است که همواره از گذشته تا به امروز مورد توجه و استفاده بوده است. در گذشته، کاربران کتابخانه، کتاب‌های موجود در کتابخانه را بر اساس نام پدیدآور در برگه‌دان‌ها جستجو می‌کردند و امروز در مخزن سازمانی دانشگاه نیز، انتظار دارند تا بتوانند آثار منتشر شده مربوط به هر یک از پژوهشگران دانشگاه را بر اساس نام وی جستجو کنند و نتایج واقعی و صحیحی را بدون کمترین خطا و ریزش کاذبی دریافت کنند. از اواخر دهه ۹۰ میلادی مشخص شده بود که اهمیت نام‌ها در

کتابخانه دیجیتال درست همانند اهمیت شایک در نظام سنتی کتابخانه است (کلوند^۱، ۱۹۹۸). به عبارت بهتر، همان‌گونه که شایک عددی منحصر به فرد برای شناسایی یک کتاب است، نام افراد نیز می‌تواند به عنوان عنصری منحصر به فرد جهت شناسایی در محیط دیجیتالی مورد استفاده قرار گیرد. در چنین فضایی، پژوهشگران و کاربران عادی هر دو مایلند با داشتن نام یک فرد خاص در ذهن که می‌تواند نام خانوادگی و نام او و یا تنها نام خانوادگی وی باشد، جستجوی خود را انجام دهند و نتایج به دست آمده نیز همه مرتبط با آن چیزی باشد که آن‌ها در پی آن بوده‌اند. کاربران به دلیل آن که تنها نام پدیدآور یک مطلب را می‌دانند و یا این که تمایل دارند درباره یک شخص اطلاعاتی به دست آورند و مشاهده کنند درباره او چه اطلاعاتی وجود دارد، از جستجوی نام شخص استفاده می‌کنند چه در یک فهرست پیوسته کتابخانه‌ای و چه در یک مخزن سازمانی دانشگاهی. آنچه مشهود است آن که جستجوگران متخصص یا پژوهشگران و نیز کاربران معمولی هر دو هنگام بازیابی اطلاعات، استفاده از جستجوی نام پدیدآورندگان را مفید و مطلوب می‌دانند (مک‌کی، سنچز و پارکر^۲، ۲۰۱۰). این امر به ویژه در سامانه‌های اطلاعاتی دانشگاه‌ها و به منظور بررسی‌های علم‌سنجی اعضای هیئت علمی مهم است.

مخازن سازمانی دانشگاهی که حاوی منابع متنی مختلف و متنوع از پژوهشگران است و با نام‌های فراوانی در ارتباط است، بیشتر نگران کاهش میزان دقت نتایج بازیابی شده خود حاصل از جستجوی نام افراد است، اما مسئله اصلی در جستجوی نام اشخاص چیست؟ آیا در این راه چالش‌هایی وجود دارد؟ مسئله اصلی در انجام چنین جستجوهای آن است که چنانچه دو فرد دارای نام یکسان باشند، یا به عبارتی یک نام واحد مربوط به چند شخص باشد، چگونه افراد مختلف از هم تمیز داده شده و مشخص می‌شوند و چگونه کاربران می‌توانند به راحتی به فرد مورد نظر برسند و در کمترین زمان به مطالب درباره او و یا مجموعه آثار وی دسترسی یابند؟ اگر یک شخص دارای چندین نام مختلف باشد و یا چنانچه فردی دارای نام مرکب (ترکیب یافته از دو یا چند قسمت) باشد، چگونه می‌توان آثار او را که با نام‌های مختلف ثبت شده است در کنار هم پیدا نمود و مطمئن باشیم که توانسته‌ایم به همه آثار وی دسترسی داشته باشیم؟ به بیانی دیگر، در دنیای واقع، نام‌ها همواره منحصر به فرد نیستند و در نتیجه نمی‌توانند به شناسایی کمک مؤثری نمایند و خود باعث انحراف و ابهام می‌گردند. بدیهی است این واقعیت بدون شک در کیفیت بازیابی اطلاعات، تحلیل‌های استنادی و نیز ارتباط میان منابع اثر منفی می‌گذارد. آشکار است که تنوع، دوگانگی یا چندگانگی در نام افراد میزان دقت و جامعیت را در جستجوهای منابع دیجیتالی کاهش می‌دهد. مسئله آشکار آن است که در انجام تحقیق، به دلیل وجود

1. Cleveland

2. McKay, Sanchez & Parker

ابهام در نام پدیدآور یا پدیدآورندگان، پژوهشگر نمی‌تواند به منابع مورد نظر دسترسی یابد در حالی که این منابع در نزدیکی او قرار دارند.

پرواضح است که مشکلات مربوط به نام افراد در یک مخزن سازمانی با افزایش تعداد نویسندگان و پدیدآورندگان، باعث کاهش میزان دقت نتایج حاصل از جستجوی نام افراد و به همین میزان، افزایش ریزش کاذب خواهد شد که این خود سبب سردرگمی کاربران و افزایش نارضایتی آن‌ها می‌شود. مخازن سازمانی معمولاً بر روی میزبان یک دانشگاه یا مؤسسه پژوهشی به طور مستقل اجرا می‌شوند و آثار مربوط به همان دانشگاه یا مؤسسه را جمع‌آوری و نگهداری می‌کنند (لینچ^۱، ۲۰۰۳). بنابراین به دلیل تعداد محدود پژوهشگران در هر مؤسسه و یا دانشگاه، مسائل پیش‌گفته مربوط به جستجوی نام در اولویت مدیران مخازن سازمانی و کاربران آن‌ها قرار نمی‌گیرد و چندان توجهی به آن نخواهند داشت. هنگامی که یک مخزن سازمانی کوچک است و حاوی اطلاعات اندکی است، مشکلات مربوط به کاهش دقت و افزایش بازیافت چندان قابل ملاحظه نبوده و با کمک مدیر مخزن قابل حل است؛ اما هنگامی که مخزن سازمانی بزرگ است یا به تدریج رو به گسترش می‌نهد، همین مسئله باعث نگرانی و نارضایتی شدید کاربران خواهد شد (هیل، ۲۰۰۸). بنابراین، ابهام در شکل نام نویسنده، اثر منفی بر بازیابی اطلاعات می‌گذارد و چالشی در عرصه نظام بازیابی اطلاعات مخزن سازمانی ایجاد می‌کند. از این رو، هدف مقاله حاضر، بررسی مشکلات و چالش‌های جستجوی نام اشخاص در مخازن سازمانی دانشگاهی از جنبه عملی آن است. به این منظور، به عنوان نمونه وضعیت مخزن سازمانی دانشگاه فردوسی مشهد از نظر وجود چنین مشکلات و چالش‌هایی مورد بررسی قرار می‌گیرد.

پیشینه پژوهش

بررسی پژوهش‌های موجود در این حوزه نشان داد، به طور کلی تعداد پژوهش‌های انجام گرفته در ایران اندک است و بیش از همه محققان حوزه کامپیوتر و هوش مصنوعی به این موضوع به لحاظ فنی و عملیاتی توجه نموده‌اند و روش‌های مختلفی را به جهت ابهام‌زدایی از نام اشخاص ارائه نموده‌اند^۲ (مظفری، ۱۳۹۹؛ مرتضوی، ندیمی شهرکی و موسی خانی، ۱۳۹۶) که البته هدف مقاله حاضر پرداختن به این موضوع از این منظر نیست و بیشتر به آثاری اشاره می‌شود که از سوی متخصصان حوزه علم اطلاعات به آن‌ها اشاره شده است.

1. Lynch

2. Sanyal, bhowmick & Das (2019); Gu, Xu, Zhu & Ji (2016); Sun, Chang, Li & Chen (2017); Zhang, Zhang, Yao (2018)

جستجوی نام اشخاص

به طور کلی اشکالات مختلفی می‌تواند باعث اختلال در انجام جستجوهای موفق نام افراد گردد. برخی از این اشکالات عمومی و مشترک است یعنی مشکلی است که در همه نظام‌های بازیابی اطلاعات و در مورد همه انواع جستجوها مثل جستجوی موضوعی، کلیدواژه‌ای، عنوان و ... نیز ممکن است به وجود آید. اشکالات تایپی مثل افتادگی برخی حروف، اضافه شدن ناخواسته برخی حروف، امکان نگارش متفاوت یک کلمه مشابه در برخی زبان‌ها (نظیر زبان فارسی که در آن روش استاندارد جهت نوشتن نام‌های مرکب به صورت پیوسته یا جدا از هم وجود ندارد (آخشیک و فتاحی، ۱۳۹۱)، از جمله این موارد است که البته برخی از آن‌ها توسط رایانه قابل پیش‌بینی و اصلاح می‌باشد؛ اما نتایج نادرست در هنگام جستجوی نام افراد، اغلب به خاطر یکی از موارد ذیل است^۱.

نوشتن نام کامل یا نوشتن آن به صورت مخفف: برخی از نویسندگان در نوشته‌های خود گاه از نام کامل خود شامل نام و نام خانوادگی و گاه از نام مخفف خود شامل مخفف نام و نام خانوادگی کامل استفاده می‌کنند. برخی نیز در هنگام نوشتن منابع و مآخذ این کار را انجام می‌دهند و گاهی نام را به طور کامل و گاهی نیز به طور مخفف وارد می‌نمایند و همین امر باعث عدم یکدستی در ورود داده و اشکال در امر جستجو می‌گردد. البته به نظر می‌رسد در زبان فارسی مسئله مخفف نوشتن نام و نام خانوادگی افراد کمتر رایج باشد و به همین علت در جستجوی متون فارسی کمتر با چنین مشکلی مواجهیم. البته افرادی که نام آن‌ها دو قسمتی هست، نیز برخی اوقات دچار همین مشکل هستند. مثلاً فردی که نام او محمدرضا یا محمدحسین یا محمدعلی هست، برخی اوقات از نام محمد و برخی اوقات از نام رضا، حسین یا علی استفاده می‌کند و برخی اوقات نیز نام کامل خود یعنی محمدرضا، محمدحسین یا محمدعلی را مورد استفاده قرار می‌دهد.

پدیدآوردگانی با چند نام: برخی افراد دارای نام مستعار هستند مخصوصاً افرادی که در حوزه ادبیات فعالیت دارند و عده دیگری نیز هستند که ممکن است بنا به دلایلی نام خانوادگی خود را تغییر داده باشند. برخی از بانوان نیز -البته خارج از ایران- به دلیل ازدواج نام خانوادگی آن‌ها تغییر نموده و نام خانوادگی همسر آن‌ها نیز به نام خانوادگی قبلی آن‌ها افزوده می‌شود. این‌گونه اشخاص دارای چند نام خانوادگی بوده و ممکن است در آثار خود از هر یک از این اسامی استفاده کرده باشند؛ بنابراین، با نام‌های مختلفی مواجه خواهیم بود که همه آن‌ها به یک فرد مربوط می‌شوند و این مسئله باعث کاهش میزان دقت جستجوی ما خواهد شد. مثلاً

1. Xia (2006); Qiu (2008); McKay, Sanchez & Parker (2010)

فردی با نام خانوادگی کلدوی داریم که سال‌ها با این نام خانوادگی مقاله منتشر نموده است و سپس به فرنام تغییر نام خانوادگی داده و مقاله‌های بعدی خویش را به این نام منتشر می‌نماید.

پدیدآورندگانی با نام مشابه: افراد زیادی در هر جامعه وجود دارند که نام و نام خانوادگی مشابه و یکسانی دارند. چنانچه هر یک از این افراد دارای اثری نیز باشند و در مخزن داده‌های ما قرار گیرند باعث ایجاد مشکل می‌گردند. آن‌ها دو فرد متفاوت با یک نام یکسان و مشابه‌اند که آثار مختلف و متعددی دارند و حال باید برای تمیز میان آن دو و دسترسی مناسب کاربران به آثارشان، در جستجوی راه‌حلی بود. در برخی از نقاط ایران افراد زیادی با نام خانوادگی مشابه و حتی نام مشابه زندگی می‌کنند که تشخیص آن‌ها فقط از طریق نام پدر و یا شماره ملی امکان‌پذیر است. مثلاً دو فرد با نام علی حسینی که یکی در رشته الهیات و معارف اسلامی و دیگری در رشته شیمی آلی هستند.

پیوسته‌نویسی یا جدانویسی نام‌ها: برخی نام‌ها را می‌توان به دو شکل پیوسته یا جدا به کار برد. همین امر یک مسئله کلی در نظام‌های بازیابی به وجود می‌آورد زیرا فرد جستجوگر ممکن است نام مورد نظر خود را به صورت سرهم جستجو کند در حالی که نام مربوطه به شکل جدا در سامانه ذخیره شده باشد و نرم‌افزار از هوشمندی لازم برای بازیابی هر دو شکل نام برخوردار نباشد، مانند محمد پیش‌قدم یا محمد پیشقدم.

رویکردهای کنترل مستند نام‌ها

یکی از راه‌های پیشگیری از مشکل عدم یکدستی نام‌ها که از سال‌های دور در کتابخانه‌ها انجام می‌شده است، تدوین فهرست مستند اسامی مشاهیر و مؤلفان است. در نظام سنتی کتابخانه، کاربران برای آگاهی از تمام آثار موجود از یک پدیدآور در کتابخانه، به برگه‌دان پدیدآور مراجعه نموده و در نظم الفبایی موجود به جستجوی نام خانوادگی وی می‌پرداختند. در چنین نظامی، افراد با اسامی مشابه از طریق سال تولد و یا وفات آن‌ها قابل تشخیص بودند. مثلاً حسینی، علی ۱۳۲۹- و حسینی، علی ۱۳۴۵-. برای افرادی با اسامی مختلف و ناهماهنگ نیز یک نام به عنوان نام مرجح انتخاب می‌شد و از دیگر نام‌های وی به نام مرجح انتخابی ارجاع داده می‌شد. از این رو تمام نام‌های متفاوت پدیدآور در نظم الفبایی برگه‌دان پدیدآور وجود داشتند ولی از همه آن‌ها به نام مرجح انتخابی ارجاع داده شده بود. مثلاً برای فردی که دو نام خانوادگی کلدوی و فرنام دارد، نام خانوادگی کلدوی به عنوان نام مرجح پذیرفته می‌شود و این نام خانوادگی در نظم الفبایی خود یعنی «ک» قرار می‌گیرد و از نام خانوادگی «فرنام» که در نظم الفبایی ذیل حرف «ف» قرار می‌گیرد به نام خانوادگی «کلدوی» ارجاع داده می‌شود. به این منظور کتاب یا پایگاه فهرست مستند مشاهیر و مؤلفان در کتابخانه‌های بسیاری از کشورها وجود دارد که در آن نام تمام افرادی که کتابی منتشر ساخته‌اند و فرد مشهوری هستند، به

صورت مستند و استاندارد وجود دارد و مورد استفاده قرار می‌گیرد. با ورود رایانه به کتابخانه‌ها از اوایل دهه ۶۰ میلادی، به تدریج فایل دیجیتالی فهرست مستند مشاهیر و مؤلفان آماده شد و در فهرست‌های پیوسته کتابخانه‌ای مورد استفاده قرار گرفت؛ بنابراین، در پیشینه‌های دیجیتالی نیز از همان نظام کنترل مستند بهره گرفته شد و کار به همان روال سابق ادامه دارد. چنین است که کاربران فهرست‌های پیوسته در جستجوی اسامی کمتر دچار مشکلات پیش گفته در مورد مخازن سازمانی می‌گردند. به عنوان نمونه، بایگانی کنترل مستند کتابخانه کنگره (LCAF)^۱ که به صورت غیررسمی LCNAF یا NAF نامیده می‌شود، حاوی بیش از ۱۰ میلیون پیشینه‌های مستند برای توصیف و نامیدن اسامی افراد، خانواده‌ها، شرکت‌ها و حوزه‌های قضایی است که با استفاده از قالب استاندارد مارک (MARC) بیان شده است (کنان، فرانک و هاوکینز، ۲۰۱۹).

استفاده از فهرست مستند اسامی پژوهشگران و پدیدآورندگان، راه‌حلی است که بسیاری از کارشناسان برای حل مشکلات جستجوی اسامی در نظام‌های بازبایی پیوسته از جمله مخازن سازمانی ارائه می‌دهند؛ اما، بسیاری نیز با وجود تأیید آن که این روش بهترین است، آن را در شرایط حاضر ناممکن می‌دانند. تعداد پژوهشگران و پدیدآورندگان آثار غیرکتابی بسیار بیشتر از پدیدآورندگان کتاب در نظام‌های سنتی کتابخانه است؛ ضمن آن که تدوین فهرست مستند مشاهیر و مؤلفان با همکاری کتابخانه‌های دیگر و در یک فرایند تعاونی در طول چندین سال به وجود آمده است ولی در مورد پژوهشگران و مخازن سازمانی چنین فعالیت مشارکت‌جویانه‌ای وجود ندارد. آنچه مسلم است آن که همه کارشناسان به نوعی کنترل بر اسامی را قبول داشته و بر آن تأکید می‌کنند اما روش و شیوه‌های برخورد آن‌ها با این مسئله متفاوت است.

به طور کلی آنچه که در رابطه با ابهام‌زدایی از اسامی در سراسر جهان و در قالب تحقیقات مختلف انجام گرفته است، تأکید بر دو راهبرد مشخص است؛ راهبرد نخست آن است که نظام رایانه‌ای خود باید به صورت خودکار کار مستندسازی اسامی را انجام دهد. گروهی از پژوهشگران معتقدند با وجود اطلاعات فراوان تولید شده و امکانات موجود، نظام‌های رایانه‌ای خودکار باید کار مستندسازی را انجام دهند و در نتیجه بر روش‌های ابهام‌زدایی خودکار و هوشمند متمرکز شده‌اند. کتابخانه‌های دیجیتالی فراوانی که در سطح دنیا وجود دارد، هرکدام به شیوه‌ای کار مستندسازی خود را انجام می‌دهد و اغلب منابع خود را از سایر کتابخانه‌ها جدا می‌سازند بنابراین کار هماهنگی و یکدست‌سازی دستی این کتابخانه‌ها عملی نمی‌باشد، ضمن این که کاربران اینترنتی فقط از منابع کتابخانه‌های دیجیتالی استفاده نمی‌کنند؛ بنابراین باید به دنبال راهی بود که بتوان به شیوه‌ای خودکار در تمام اینترنت کار ابهام‌زدایی از اسامی را انجام داد (اسمالهیزر و تورویک^۲، ۲۰۰۹).

1. Library of Congress Authority File

2. Smalheiser & Torvik

در مقابل اسمال هایزر و تورویک، افراد دیگری از جمله وو^۱ (۲۰۰۹) معتقدند تطابق خودکار اسامی همیشه با شکست رو به رو خواهد بود و همواره به دخالت انسانی محتاجیم. او بیان می‌کند، تلاش‌های اندکی که برای خودکارسازی کنترل مستند اسامی صورت گرفته است، نظیر کنترل مستند اسامی در XML، تنها در استخراج اسامی از پیشینه‌های XML موفق بوده‌اند ولی نتوانسته‌اند آن‌ها را به نقاط دسترسی مطمئنی تبدیل سازند. همواره باید انسانی باشد تا پیشینه‌های مستند اسامی تولید شده به وسیله رایانه را مورد بررسی قرار دهد.

آنچه مسلم است آن که هر دوی این نظرات طرفدارانی برای خود دارد و پژوهش‌های مبتنی بر هر یک از این دو راهبرد در حال انجام است. نکته قابل تأمل آن است که در پاسخ به این که کدام راهبرد مناسب و بهتر است باید گفت این موضوع به شرایط و ویژگی‌های پروژه در دسترس بستگی دارد، به طوری که در برخی مواقع می‌بایستی از روش ترکیبی استفاده نمود (الیوت^۲، ۲۰۱۰). از این روست که تاکنون روش و رویه‌ای یکدست و هماهنگ و در حقیقت استاندارد مشخص برای این مسئله به وجود نیامده است و در طول سال‌های گذشته به خصوص دهه اخیر پروژه‌های متعددی در کشورهای مختلف تعریف شده و هر کدام با استفاده از تجربیات قبلی و مبتنی بر مشخصات خاص محلی، راه‌حل‌ها و ابتکاراتی را ارائه نموده‌اند که در ادامه به تعدادی از آن‌ها که در متون مختلف به آن‌ها پرداخته شده، اشاره می‌شود.

پروژه نام‌های کتابخانه موزه بریتانیا در اثر رشد مخازن سازمانی ضرورت یافت. در سال ۲۰۰۸ بیش از ۸۷ مخزن سازمانی با بیش از ۳۰۰۰۰۰ پیشینه در انگلستان وجود داشت که مستند نمودن نام‌های پدیدآورندگان آن‌ها اهمیت یافت (هیل، ۲۰۰۸)؛ اما مطابق با آمار سال ۲۰۲۱ در حال حاضر ۳۱۵ مخزن سازمانی در انگلستان وجود دارد.^۳ مطابق با توضیحاتی که راجع به این پروژه ارائه شده است، به هر نویسنده یک شماره شناسایی منحصر به فرد اختصاص می‌یابد که به سایر شکل‌های نام فرد ارتباط‌هایی را برقرار می‌سازد، بنابراین به آن «کنترل دسترسی»^۴ می‌گویند و نه کنترل مستند چرا که برخلاف روش کنترل مستند نام‌ها، در اینجا دیگر خبری از نام مرجح قراردادی و پذیرفته شده نیست.

فهرست مشترک نام‌های هنرمندان (ULAN)^۵ حاوی ۱۲۰۰۰۰ نام هنرمند است که به صورت دستی و براساس روش کنترل مستند کتابخانه‌ای ایجاد شده است تا در دسترس هنرمندان، کتابخانه‌های حوزه هنر، آرشيوها، فهرست‌نویسان منابع دیداری و شنیداری و جستجوگران حوزه هنر قرار گیرد. در این فهرست نیز با

1. Veve

2. Elliott

3. https://v2.sherpa.ac.uk/view/repository_visualisations/1.html

4. Access Control

5. Union List of Artists Names

اسامی پذیرفته شده و استاندارد مواجه‌ایم که از شکل‌های مختلف یک نام به شکل استاندارد آن ارجاع داده شده است. فهرست (ULAN) اسامی مشترکی با فهرست مستند کتابخانه کنگره (LCAF) دارد که از طریق فیلدی با عنوان (LC Flag) شناسایی می‌گردند. نکته قابل تأمل آن که با وجود این نقاط اشتراک، موردهایی نیز وجود دارند که یک فرد در دو نظام موجود دارای دو نام استاندارد متفاوت باشد که می‌تواند سبب بروز مشکلاتی گردد (تراست^۱، ۲۰۱۰).

در نظام FOAF^۲ برای شناسایی فرد از ثبت داده‌های مختلف مربوط به او مثل نام و نام خانوادگی، نشانی پست الکترونیکی که یک نشانه منحصر به فرد می‌باشد، تصویر و ... استفاده می‌شود. در این نظام برای توصیف ارتباط میان مدارک و نیز افراد از چارچوب توصیف منابع RDF^۳ استفاده می‌شود. برای شناسایی مدارک مربوط به یک فرد هر دو متن با هم مطابقت داده می‌شوند و چنانچه نشانی پست الکترونیکی هر یک از آن‌ها دقیقاً مشابه هم باشد، آن دو متن به یک فرد تعلق خواهد داشت. در حال حاضر از این نظام در شبکه‌های اجتماعی و موتورهای جستجو نظیر گوگل و یاهو استفاده می‌شود (الیوت، ۲۰۱۰). هر چند این روش در موتورهای جستجو و شبکه‌های اجتماعی به کار می‌رود و راه‌حلی نیز برای ابهام‌زدایی از نام‌ها ارائه می‌دهد، اما به دلیل این که افراد همواره خود را ملزم به درج نشانی پست الکترونیکی نمی‌بینند، چندان نمی‌تواند راه‌حل مطمئن و پرکاربردی باشد.

روش دیگری که اسمال‌هایزر و تروویک ارائه می‌دهند، استفاده از نظام سبک‌شناسی^۴ پدیدآورندگان است. مطابق با این شیوه، رایانه سبک نوشته‌های یک فرد را درک می‌کند و سپس از این منظر متن‌های مختلف را مورد بررسی و تحلیل قرار می‌دهد. اسمال‌هایزر و تروویک معتقدند هر چند این روش برای تعداد محدودی نویسنده قابل انجام است، اما امکان به کارگیری آن جهت ابهام‌زدایی از نام در پایگاه‌های اطلاعاتی بزرگ نیز وجود دارد و می‌توان بررسی‌هایی در این زمینه انجام داد (اسمال‌هایزر و تروویک^۵، ۲۰۰۹). سالو^۶ (۲۰۰۹) پیشنهاد می‌دهد که نرم‌افزارهای مخزن سازمانی به صورت استاندارد به وجود آید و پدیدآورندگان، خود نسبت به ورود اطلاعات شخصی اقدام نمایند و اصلاحات لازم را به عمل آورند. خوداظهاری پدیدآورندگان، اطمینان از صحت داده‌ها را افزایش می‌دهد اما عملیاتی شدن آن به میزان تمایل

1. Trust

2. Friend Of Friend

3. Resource Description Framework

4. Stylometry

5. Smalheiser & Torvik

6. Salo

پدیدآورندگان برای صرف وقت در این خصوص و نیز توانایی و مهارت آن‌ها در ورود درست، یکدست و استاندارد داده‌ها بستگی دارد.

هان^۱ و دیگران (۲۰۰۴) نیز مدلی را برای یادگیری رایانه جهت ابهام‌زدایی از نام‌ها پیشنهاد می‌دهند. در این روش، رایانه داده‌های مربوط به نام مجله، عنوان مقاله و نیز نام پدیدآورندگان مقاله را ذخیره می‌نماید. آنگاه با تحلیل نویسندگان همکار در مقاله‌ها، جهت ابهام‌زدایی از برخی نام‌ها، احتمال همکار شدن نویسنده مورد نظر را با نویسندگان دیگر در نظر می‌گیرد و بر این اساس میان دو نویسنده مشابه تمیز قائل می‌شود. توجه به موردهایی که در بالا به آن‌ها اشاره شد، نشان می‌دهد طراحان نظام‌های بازیابی اطلاعات به ویژه مخازن سازمانی به وجود مشکل در جستجوی نام‌ها پی برده‌اند و هر یک به طریقی سعی می‌کنند با ابتکاری این مشکل را حل نمایند. هر یک از این پروژه‌ها سعی دارند تا از یک عامل منحصر به فرد مربوط به اشخاص استفاده نماید و با عملیاتی نمودن آن در هنگام ذخیره و بازیابی از ایجاد مشکلات پیش‌گفته جلوگیری نماید، اما این عوامل منحصر به فرد مختلف در هر مخزن سازمانی قابل بهره‌برداری نمی‌باشد و لازم است تا با توجه به ویژگی‌های خاص داده‌های هر مخزن، نسبت به استفاده از هر یک از آن‌ها اقدام شود. پژوهش حاضر در پی یافتن پاسخ پرسش‌های زیر انجام شده است:

۱. نام‌های پژوهشگران موجود در مخزن سازمانی (سامانه اطلاعات علمی) دانشگاه فردوسی مشهد از نظر ابهام‌زایی چگونه هستند؟
۲. مخزن سازمانی دانشگاه فردوسی مشهد در رابطه با جستجوی نام‌های ابهام‌زا چگونه عمل می‌کند؟

روش پژوهش

به منظور بررسی مسائل و چالش‌های جستجوی نام اشخاص در مخزن سازمانی دانشگاه فردوسی مشهد به لحاظ مسائل پیش‌گفته، از روش ارزیابانه و مشاهده استفاده گردید. پژوهش ارزیابانه یا پژوهش ارزشیابی نوع دیگری از پژوهش کاربردی است که هدف اصلی آن به جای کشف دانش، آزمایش کاربرد دانش در یک طرح یا برنامه ویژه است (پاول، ۱۳۷۹). در این روش، از اندازه‌گیری عملکرد، که نوع اختصاصی‌تری از پژوهش ارزیابانه است، استفاده شده است. هدف از اندازه‌گیری عملکرد یا برون‌داد، تعیین دستاوردهای حاصل از برنامه‌ها، خدمات یا امکانات خاصی است که ارائه شده است. بر این اساس، در این پژوهش تلاش شده است تا عملکرد موتور جستجوی سامانه مخزن دانشگاه فردوسی مشهد از جهت جستجوی نام اشخاص مورد بررسی

قرار گیرد. برای انجام این کار، به وبگاه این سامانه به نشانی <http://profdoc.um.ac.ir> مراجعه گردید. در صفحه ورودی، دو قسمت مجزا برای جستجو در قالب مقاله و کتاب وجود دارد. قالب مقاله، دارای ۱۰ فیلد است که به ترتیب عبارتند از: عنوان مقاله، نویسنده، کلمات کلیدی، نام مجله، نام همایش، سال انتشار مجله، سال برگزاری همایش، موضوع مقاله، اعتبار علمی و زبان مقاله. در انتها هم کادری قرار دارد برای ورود عبارت امنیتی تا جستجو انجام گیرد، امری که در صفحات جستجوی وبگاه‌های مختلف دیده نمی‌شود و قدری غریب می‌نماید. درباره این مورد با مسئول مربوطه در سایت به صورت تلفنی گفتگو شد و ایشان دلیل این امر را رعایت نکات امنیتی و جلوگیری از نفوذ ربات‌های مختلف عنوان کردند. در فیلدهای سال انتشار، فقط می‌توان سال میلادی را وارد نمود، ضمن آن که فیلد موضوع مقاله، اعتبار علمی و زبان مقاله به صورت کشویی بوده و می‌بایستی از گزینه‌های موجود، در صورت نیاز یکی را انتخاب نمود. بخش کتاب نیز فیلدهایی شامل عنوان کتاب، نویسنده یا مترجم، کلمات کلیدی، موضوع کتاب و سال انتشار است که فیلد موضوعی نیز گزینه‌ای بوده و فقط می‌توان از میان عنوان‌های موضوعی یکی را انتخاب نمود. در این قسمت هم کادر مربوط به ورود عبارت امنیتی وجود دارد و قبل از انجام جستجو، حتما باید آن را وارد کرد. جستجوی پیشرفته‌ای وجود ندارد و برای ترکیب فیلدها، تنها می‌توان از عملگرهای (+) برای عمل (AND) و (|) برای عمل (OR) استفاده نمود و جستجوی ترکیبی انجام داد که با توجه به رابط کاربری آن برای کاربران بسیار مبهم و گمراه کننده است.

برای بررسی چالش‌های جستجوی نام در این مخزن سازمانی، ابتدا فهرست تمام ۹۰۰ نام و نام خانوادگی موجود در این سامانه که به صورت الفبایی نظم یافته است مرور گردید و فهرست اسامی که چالش برانگیز بودند (مانند کامل یا مخفف بودن نام، داشتن چند نام، داشتن نام مشابه و یکسان، پیوسته‌نویسی یا جدانویسی نام‌ها) به دست آمد و پس از آن مبتنی بر این نام‌ها، از طریق موتور جستجوی سامانه، جستجو برای بازیابی اشخاص مورد نظر انجام گرفت تا مشخص شود آیا جستجوی اسامی چالش برانگیز از این طریق به سهولت و بدون خطا انجام می‌گیرد یا خیر.

ارائه یافته‌ها

جهت تحلیل و بررسی اسامی موجود، تمام آن‌ها مرور شد و مطابق با طبقه‌بندی جدول ۱، دسته‌بندی شدند. همان‌گونه که پیش‌تر نیز بیان شد، یکی از دلایل کاهش دقت در جستجوی نام‌ها، وجود اسامی مشابه در یک مخزن سازمانی است که باعث می‌شود آثار آن‌ها با یکدیگر ادغام شده و با هم بازیابی شوند و ریزش کاذب اتفاق افتد. ۱۰۹ نام خانوادگی مشابه با فراوانی‌های مختلف وجود داشت که این موارد متعلق به ۲۶۹ نام خانوادگی یعنی ۲۹,۸۸ درصد از مجموع اسامی است. نام خانوادگی «حسینی» با فراوانی ۱۶ پرتکرارترین

است. وجود درصد بالایی از نام‌های خانوادگی مشابه نشان می‌دهد که کاربران مجبورند در بسیاری از جستجوها علاوه بر نام خانوادگی از نام کوچک افراد نیز آگاه بوده و آن را نیز اضافه نمایند؛ در غیر اینصورت با نتایج مناسبی روبه‌رو نخواهند بود. در این مخزن، تنها دو فرد وجود داشتند که نام و نام خانوادگی کاملاً مشابهی داشتند و هر دو «سید محمود حسینی» هستند؛ این در حالی است که یکی از آن‌ها عضو هیئت علمی گروه مهندسی صنایع و دیگری عضو هیئت علمی گروه مهندسی عمران است. برای مشاهده آثار این دو نفر از طریق جستجو در مخزن به دلیل چنین مشابهتی، هیچ راهی وجود ندارد؛ هر چند این مورد تنها ۰,۲ درصد از مجموع نام‌ها است، اما کاربرمداری اقتضا می‌کند که مدیران مخزن سازمانی درباره آن اندیشه کرده و راه‌حلی بیابند. لازم به ذکر است با افزایش تعداد اسامی و بزرگ شدن مخزن و یا ادغام مخازن سازمانی دانشگاه‌های مختلف کشور، مسلماً تعداد این مشابهت‌ها، فزونی خواهد یافت.

جدول ۱. تحلیل مجموع اسامی موجود در مخزن سازمانی دانشگاه فردوسی مشهد براساس مشابهت‌ها و نیز ساختار آن‌ها

نام‌های خانوادگی مشابه	نام و نام خانوادگی‌های مشابه	نام‌های خانوادگی چندبخشی	نام‌های خانوادگی مرکب	نام‌های کوچک چندبخشی	واژه «سید»
مورد ۱۰۹	۱ مورد	۳۴۱	۴۴	۲۲۹	۷۶
%۲۹,۸۸	%۰,۲	%۳۷,۸۸	%۴,۸۸	%۲۵,۴۴	%۸,۴۴

منظور از اصطلاح چندبخشی در طبقه‌بندی اسامی جدول ۱، نام‌های خانوادگی و یا نام‌های کوچکی است که از دو یا چند بخش مستقل و معنی‌دار تشکیل شده‌اند؛ به عنوان مثال نام کوچک «محمدعلی»، «محمدرضا»، «روح‌الله» و ... یا نام خانوادگی «جباری نوقابی»، «طباطبایی یزدی»، «کامیابی گل» و ... چندبخشی هستند. همچنین، نام‌های خانوادگی چون «کیوان‌فر»، «مشاوری‌نیا» نیز در دسته نام‌های خانوادگی مرکب قرار گرفته‌اند. دلیل توجه به این خصوصیات در این دسته‌بندی‌ها آن است که وجود چنین اسامی در زبان فارسی به دلیل مشکلاتی که در هنگام تایپ (ورود داده‌ها به فیلدهای مخزن سازمانی توسط خود اعضای هیئت علمی) به وجود می‌آورند مثل رعایت فاصله و یا عدم رعایت آن، باعث ایجاد بازیافت نامربوط یا کم مربوط می‌گردند. وجود حدود ۴۰ درصد نام خانوادگی چندبخشی و مرکب نشانگر آن است که مدیران این مخزن سازمانی و البته سایر مخازن سازمانی در ایران باید به مسئله نام‌ها و شیوه سازماندهی آن‌ها دقت ویژه‌ای کنند تا از بازیابی جامع و دقیق نظام بازیابی خود اطمینان حاصل کنند. نکته حائز اهمیت دیگر آن که حدود ۸ درصد از افراد، پیشوند «سید» داشتند که ورود و یا عدم ورود این واژه در هنگام جستجو، در

تعداد نتایج بازیابی شده مؤثر بود؛ بنابراین مدیران مخزن سازمانی لازم است تا از این امر نیز آگاه بوده و دقیق‌تر با این مسئله برخورد نمایند.

در مخزن مورد مطالعه، حالتی که یک فرد شکل‌های مختلف نام خانوادگی داشته باشد، یافت نگردید. درباره اشتباه تایپی و یا مسئله جدانویسی و پیوسته‌نویسی در زبان فارسی که یکی از دشواری‌های این زبان است و در تحقیقات متعددی به آن‌ها اشاره شده است (آخسیک و فتاحی، ۱۳۹۱) نیز باید گفت این مشکل در این مخزن سازمانی نیز وجود دارد. به عنوان نمونه در این مخزن دو فرد با نام خانوادگی مشابه «کهنسال» وجود دارد؛ یکی از آن‌ها «محمد رضا کهن‌سال» و دیگری «علیرضا کهنسال» دقیقاً با چنین املائی وجود دارد. نکته جالب آن که این هر دو دارای نام کوچک چندبخشی و هم نام خانوادگی چندبخشی هستند که کار جستجوی آن‌ها را دشوارتر می‌نماید، ضمن آن که به دلیل املائی متفاوت به کار رفته برای این دو نام خانوادگی، چنانچه کاربرد عبارت جستجو را برای فرد اول به صورت پیوسته «محمد رضا کهنسال» وارد کند، هیچ نتیجه‌ای یافت نمی‌گردد. قابل ذکر آن که واژه «الله» در این مخزن سازمانی به صورت «اله» ثبت شده است، بنابراین تمام اسامی مشتمل بر این واژه مثل «رحمت الله»، «روح‌الله» و ... حتماً باید به صورت «رحمت اله»، «روح‌اله» مورد جستجو قرار گیرند، در غیر این صورت نتیجه‌ای بازیابی نخواهد شد. به بیان دیگر، نرم‌افزار این مخزن سازمانی هوشمند نیست که بتواند اشتباه‌های تایپی را نادیده بگیرد.

در پاسخ به پرسش دوم پژوهش، با توجه به این که تمام نام‌های مستخرج از سیاهه نام‌های پژوهشگران موجود در مخزن که ابهام‌زا بودند مورد جستجو قرار گرفتند تا عملکرد جستجو مورد بررسی قرار گیرد، نتیجه حاصل بیانگر عدم استفاده از هر یک از روش‌های پیش‌گفته در به حداقل رساندن چالش‌های جستجوی نام اشخاص بود و نظام طراحی شده نتوانست نتایج مرتبطی را نمایش دهد.

در مجموع می‌توان گفت مخزن سازمانی دانشگاه فردوسی مشهد، در هنگام جستجوی نام‌های خانوادگی دچار مشکل و آشفتگی می‌گردد و این مسئله نیز به خاطر مشابهت‌های نام‌ها و تکرار آن‌ها و یا به دلیل مشکلات زبانی و املائی است. بنابراین با توجه به نوپا بودن این مخزن لازم است جهت هرچه بهتر شدن نواقص آن اصلاح شوند.

نتیجه‌گیری

نام اشخاص عنصری کلیدی در جستجوی منابع اطلاعاتی از قبیل کتاب و انواع مقاله‌های علمی است. مشابهت نام اشخاص با یکدیگر و یا وجود چند نام مختلف برای یک فرد و نیز دلایل دیگری سبب ابهام‌زایی این نام‌ها می‌شود و جستجو بر اساس نام اشخاص را با چالش‌هایی مواجه می‌سازد. پژوهشگران به خصوص

پژوهشگران حوزه رایانه و هوش مصنوعی تلاش دارند تا با ارائه روش‌های مختلف این چالش‌ها را مرتفع یا به حداقل برسانند و نتایج جستجوها را بهبود بخشند.

این پژوهش با هدف بررسی چالش‌های موجود در جستجوی نام اشخاص در مخزن سازمانی دانشگاه فردوسی مشهد به عنوان نمونه انجام گرفت. مخزن سازمانی دانشگاه فردوسی مشهد از جمله مخازن سازمانی در ایران است که بسیار مورد جستجو قرار می‌گیرد. با وجود حجم نسبتاً کم اطلاعات در این مخزن در مقایسه با مخازن سازمانی خیلی بزرگ، پژوهش نشان داد که این مخزن سازمانی حاوی نام‌های متعدد ابهام‌زا از انواع مختلف آن است و در مواجهه با جستجوی نام‌های ابهام‌زا دچار چالش می‌گردد و نتایج جستجو مورد نظر جستجوگر نمی‌باشد. بنابراین ضروری است تا راهکارها و راه‌حلهایی درباره آن اندیشه شود تا هرچه زودتر این مشکلات رفع شوند.

مطابق با آنچه بیان شد، جهت اصلاح مشکلات جستجوی نام‌ها در مخزن سازمانی دانشگاه فردوسی مشهد لازم است تا فعالیت‌هایی صورت پذیرد. بهتر است تا آموزش‌های اعضای هیئت علمی پیرامون اهمیت مخزن سازمانی دانشگاه افزایش یابد و چگونگی ورود درست داده‌ها و اهمیت هر فیلد و ضرورت آن آموزش داده شود و سعی شود از روش خوداظهاری برای ورود داده‌ها استفاده شود.

با توجه به تجربه کتابداران در رابطه با مستند نمودن نام‌ها، لازم است تا کتابداران شاغل در دانشگاه بخشی از وقت خود را صرف کنترل مستند نام‌های سامانه اطلاعات علمی دانشگاه نمایند و فهرست مستند نام‌های پژوهشگران دانشگاه را به زبان فارسی و لاتین آماده نموده و در اختیار مرکز امور رایانه دانشگاه و نیز خود پژوهشگران قرار دهند تا آن‌ها نیز آگاه باشند که همیشه به شکل استاندارد نام خود را بنویسند و از دوباره‌کاری و تکرار اشتباهات پرهیز نمایند. همچنین ضروری است تا مرکز امور رایانه دانشگاه جهت اصلاح ساختاری این سامانه با بهره‌گیری از تجربیات موجود در ساختار مارک برای داده‌های مستند، در نرم‌افزار مخزن سازمانی از ساختاری استاندارد برای ساماندهی و مستندسازی نام‌ها استفاده کند. بدیهی است چنانچه برخی مشکلات فنی جستجو به لحاظ برنامه‌نویسی برطرف شود و پژوهشگران نیز با آگاهی حاصل شده بار دیگر داده‌های خویش را کنترل نموده و مطابق با نام استاندارد خویش عمل نمایند، با چالش‌های اشاره شده مواجه نخواهیم بود.

با توجه به این که در ایران تجربه اندکی درباره مدیریت اطلاعات در مخزن سازمانی در دانشگاه‌ها وجود دارد و بسیاری از دانشگاه‌ها فاقد هرگونه مخزن سازمانی برای یافته‌های علمی و پژوهشی می‌باشند و نیز توجه به این مطلب که همه دانشگاه‌ها به تدریج سعی دارند تا مخزنی سازمانی برای خود ایجاد نمایند، پیشنهادی زیر ارائه می‌شود؛

کتابداران که سابقه بسیار طولانی در ذخیره و بازیابی اطلاعات هم به صورت سنتی و هم دیجیتال دارند، در این امر مهم پیشقدم شده و از طرف کتابخانه‌های دانشگاهی پیشنهاد ایجاد مخازن سازمانی را به مسئولان دانشگاه‌ها ارائه دهند و توانمندی‌های خود را در این زمینه به اطلاع آن‌ها برسانند.

کتابداران دانشگاهی به پژوهشگران دانشگاهی در دانشگاه‌های سراسر ایران کمک نمایند تا نام و نام خانوادگی خود را در زبان فارسی و نیز انگلیسی به صورت مستند و هماهنگ در نرم‌افزار مربوط وارد نموده و اهمیت انجام چنین کاری را برای آن‌ها تشریح نمایند. همچنین از آن‌ها خواسته شود تا همیشه در همه تولیدهای علمی خود از همین شکل نام خود استفاده نمایند.

صاحبان نرم‌افزارهای معتبر کتابخانه‌ای که تجربه زیادی در خصوص ذخیره و بازیابی اطلاعات کسب نموده‌اند، می‌توانند نرم‌افزارهای مناسبی برای مخازن دانشگاهی طراحی نمایند تا در تمام دانشگاه‌های کشور به صورت هماهنگ مورد استفاده قرار گیرد.

با توجه به این که بسیاری از پژوهشگران ایرانی در دانشگاه‌ها و پژوهشگاه‌های وابسته به وزارتخانه‌های «علوم، تحقیقات و فناوری» و «بهداشت، درمان و آموزش پزشکی» مشغول به فعالیت می‌باشند، معاونت‌های پژوهشی این دو وزارتخانه می‌توانند در یک طرح مشترک از تمام پژوهشگران خود به طور الزامی بخواهند تا شکل درست و یکدست نام و نام خانوادگی خود را مطابق شیوه‌نامه کتابخانه ملی، به زبان فارسی و نیز انگلیسی به همراه دیگر مشخصات کامل خود وارد نرم‌افزاری که برای این مهم تهیه شده است نمایند.

اهمیت کنترل مستند نام‌ها وقتی بیشتر درک می‌شود که بدانیم همه افراد و دانشگاه‌ها باید داده‌های خود را به گونه‌ای در مخازن سازمانی و سامانه‌های اطلاعات علمی وارد کنند که پایگاه‌های استنادی مشکلی در تشخیص درست نام‌ها نداشته باشند. در غیر این صورت، ضریب تأثیر هر عضو هیئت علمی به درستی قابل محاسبه نخواهد بود.

با توجه به حضور کتابداران در تمام کتابخانه‌های دانشگاهی و تعداد محدود محققان و پژوهشگران در هر دانشگاه، امکان مستند نمودن نام آن‌ها به صورت دستی در هر دانشگاهی وجود دارد. کتابخانه‌های مرکزی دانشگاه‌ها می‌توانند بر پایه رویه و استاندارد کتابخانه ملی به مستندسازی نام‌ها و ارسال داده‌های آن برای افزودن به پایگاه فهرست مستند اسامی اقدام کنند تا هماهنگی و انسجام بیشتری به وجود آید، مانند انجام فیپا در کتابخانه‌های مرکزی و فرستادن آن به کتابخانه ملی جهت بررسی و تأیید. با توجه به حجم بالای کار در کتابخانه ملی، چنانچه انجام این مهم در کتابخانه ملی امکان‌پذیر نباشد، پژوهشگاه علوم و فناوری اطلاعات ایران می‌تواند مرکزی برای انجام این هماهنگی باشد.

مطابق با آنچه به عنوان پیشنهاد آمد، لازم است تا نرم‌افزاری مناسب برای مخازن سازمانی دانشگاه‌ها پدید آید که این کار را شرکت‌های طراح نرم‌افزارهای کتابخانه‌ای که فایل مستند نام‌ها را نیز ایجاد کرده‌اند، می‌توانند انجام دهند. سپس از پژوهشگران خواسته شود اطلاعات خود را با دقت در این نرم‌افزار وارد نمایند و کتابداران نیز به صورت دستی نام آن‌ها را کنترل و مستند نمایند. به این ترتیب نظام ملی کنترل مستند نام‌های پژوهشگران ایرانی به زبان فارسی و انگلیسی پدید می‌آید که می‌تواند استفاده‌های فراوانی داشته و مشکلات جستجو بر اساس نام را تا حدود زیادی مرتفع سازد. با تحقق چنین امری، با جستجوی نام یک محقق در موتور جستجو، احتمالاً اولین نتیجه، خواسته ما خواهد بود و با کلیک بر نام او به صفحه شخصی او که با مخزن سازمانی متعلق به دانشگاهش مرتبط است راه خواهیم یافت و تمام آثار او را مشاهده خواهیم نمود. نمونه چنین کاری را می‌توانید با جستجویی به شکل «محمدحسین دینانی» در موتور جستجوی گوگل مشاهده نمایید.

سیاسگزاری

از استاد ارجمند و فرهیخته، جناب آقای دکتر فتاحی به خاطر بازبینی متن مقاله و ارائه نظرهای ساختاری تشکر و قدردانی می‌شود. از داوران محترم نیز به خاطر ارائه نظرهای ساختاری و علمی سپاسگزاری می‌شود.

منابع

- آخیشیک، سمیه؛ فتاحی، رحمت الله (۱۳۹۱). تحلیل چالش‌های پیوسته‌نویسی و جدانویسی واژگان فارسی در ذخیره و بازیابی اطلاعات در پایگاه‌های اطلاعاتی. *کتابداری و اطلاع‌رسانی*، ۱۵(۳)، ۹-۳۰.
- پاول، ار. رونالد (۱۳۷۹). *روش‌های اساسی پژوهش برای کتابداران (نحلا حریری، مترجم)*. تهران: مرکز انتشارات علمی دانشگاه آزاد اسلامی (نشر اثر اصلی ۱۹۹۷).
- پربرخ، مه‌ری؛ زاهدی نوقابی، مهدی (۱۳۹۰). بررسی مخزن سازمانی دانشگاه فردوسی مشهد و سنجش میزان همخوانی آن با معیارهای تخصصی. *علوم و فناوری اطلاعات*، ۲۶(۳)، ۵۱۹-۵۴۵.
- زاهدی نوقابی، مهدی (۱۳۸۹). *بررسی مخازن سازمانی دانشگاه فردوسی مشهد و سازگاری آن با معیارهای علمی (پایان‌نامه کارشناسی ارشد)*. دانشگاه فردوسی مشهد، دانشکده علوم تربیتی و روان‌شناسی.
- زاهدی نوقابی، مهدی (۱۳۹۴). تحلیل عامل‌های اثرگذار بر پیاده‌سازی مخازن سازمانی در دانشگاه‌ها و مراکز علمی بر پایه نظریه‌های سیستم‌های اطلاعاتی. *پژوهشنامه کتابداری و اطلاع‌رسانی*، ۱۵(۱)، ۲۸۰-۳۰۰.
- مرتضوی، سید محمد؛ ندیمی شهرکی، محمدحسین؛ موسی خانی، مصطفی (۱۳۹۶). بهبود صحت ابهام‌زدایی نام نویسنده با استفاده از خوشه‌بندی تجمعی. *پردازش علائم و داده‌ها*، ۳۴(۴)، ۱۱۷-۱۲۷.
- مظفری، نیلوفر (۱۳۹۹). ارائه روشی مبتنی بر ژنتیک برای رفع ابهام نام نویسندگان مقالات. *پردازش و مدیریت اطلاعات*، ۳۶(۳)،

References

- Akhshik, S., Fattahi, R. (2012). Analyzing the Challenges of writing Persian Vocabulary in Continuous and Separate format in Storing and Retrieving Information in Databases. *Library & Information Sciences*, 15(3), 9-30. (in Persian)
- Bansode, S. Y. (2011). Developing institutional repository in university library: A case study of University of Pune. *International Journal of Information Dissemination and Technology*, 4, 188-192. Available at: <https://www.ijdt.com/index.php/ijdt/article/view/32/32>.
- Cannan, J. P., Frank, P., & Hawkins, L. (2019) LC/NACO Authority File in the Library of Congress BIBFRAME Pilots, *Journal of Library Metadata*, 19, 1-2, 39-51. DOI: 10.1080/19386389.2019.1589693
- Carver, B. (2003). Creating an institutional repository: a role for librarians. *ExLibris*, 181, Retrieved May 18, 2014 from <http://marylaine.com/exlibris/xlib181.html>
- Cleveland, G. (1998). *Digital libraries: definitions, issues and challenges*. IFLA Universal Dataflow and Telecommunications Core Program, Occasional Paper 8, Retrieved March 18, 2014 from www.ifla.org/VI/5/op/udtop8/udtop8.htm
- Elliott, S. (2010). Survey of Author Name Disambiguation: 2004 to 2010. *Library Philosophy and Practice (e-journal)*, paper 473, Retrieved May 18, 2014 from <http://digitalcommons.unl.edu/libphilprac/473/>
- Fattahi, R. (1997). *AACR2 and Catalogue Production Technology: Relevance of cataloguing principles to the online environment in The Principles and Future Development of AACR: Proceedings of the International Conference on the principles and Future Development of AACR*, Toronto, 23-25 October 1997. Edited by Jean Weihs.-- Chicago: ALA, Canadian Library Association, Library Association, 17-43.
- Gu, S., Xu, X., Zhu, J., & Ji, I. (2016). Name Disambiguation Method based on Multi-step Clustering. *The 7th International Conference on Ambient Systems, Networks and Technologies*, 488-495. doi: 10.1016/j.procs.2016.04.237
- Han, H., et al. (2004). *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, Tuscon, AZ, 296-305. Retrieved April 19, 2018 from <http://doi.acm.org/10.1145/996350.996419>
- Hill, A. (2008). "What's in a name?" Prototyping a name authority service for UK repositories. *Proceedings of the 10th International Conference of the International Society of Knowledge Organization*. Montreal, Canada. Retrieved March 22, 2014 from http://names.mimas.ac.uk/documents/Names_ISKO2008_paper.pdf
- Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL Bimonthly Report*, 226, Retrieved March 18, 2014 from <http://www.arl.org/newsltr/226/ir.html>
- McKay, D., Sanchez, S., & Parker, R. (2010). *What's My Name Again? Sociotechnical Considerations for Author Name Management in Research Databases*. In *OZCHI 2010 Proceedings*. OZCHI 2010, November 22-26, Brisbane: ACM digital library.

- Mortazavi, S., Nadimi Shahraki, M., & Mosakhani, M. (2018). Improving the accuracy of the author name disambiguation by using clustering ensemble. *Signal & Data Processing*, 14(4), 117-127. (in Persian)
- Mozafari, N. (2021). A Genetic-based Approach for Author Name Disambiguation Problem. *Iranian Journal of Information Processing Management*, 36(3), 791-816. (in Persian)
- Qiu, J. (2008). Scientific publishing: Identity crisis. *Nature*, 451, 766-767.
- Parirokh, M., & Zahedi Nooghabi, M. (2011). A Study of Institutional Repository of Ferdowsi University of Mashhad (FUM), and its Compatibility with Scientific Criteria. *Iranian Journal of Information Processing Management*, 26(3), 519-545. (in Persian)
- Powell, R. R. (1997). Basic research methods for librarians. Tehran: Scientific Publishing Center of Azad Islamic University. (in Persian)
- Ratanya, F. C. (2017). Institutional repository: Access and use by academic staff at Egerton University, Kenya. *Library Management*, Vol.38, no.4/5, Available at: <http://dx.doi.org/10.1108/LM-02-2017-0018>.
- Rozum, B., & Thomas, B. (2016). *Populating your institutional repository and promoting your students: IRs and undergraduate research*. In making institutional repositories work (pp.311-318) WestLafayette, IN: Purdue University Press. Available at: https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1248&context=lib_pubs.
- Saló, D. (2009). Name Authority Control in Institutional Repositories. *Cataloging & Classification Quarterly*, 47(3), 249-261. Retrieved April 9, 2018 from <http://dx.doi.org/10.1080/01639370902737232>
- Sanyal, D. K., bhowmick, P. K. & Das, P. P. (2019). A review of author name disambiguation techniques for the PubMed bibliographic database. *Journal of Information Science*, 1-28. DOI: 10.1177/0165551519888605
- Smalheiser, N. & Torvik, V. (2009). Author name disambiguation. in B. Cronin, *Annual Review of Information Science and Technology*, 43, pp. 287-313.
- Sun, S., Chang, H., Li, N. & Chen, Y. (2017). Name Disambiguation for Chinese Scientific Authors with Multi-level Clustering. In: *EEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. DOI 10.1109/CSE-EUC.2017.39
- Veve, M. (2009). Supporting Name Authority Control in XML Metadata: A Practical Approach at the University of Tennessee. *Library Resources & Technical Services*, 53(1), 41-52. Retrieved May 15, 2018 from <http://search.ebscohost.com/login.aspx?direct=true&dblive>
- Wikipedia (2014). *Institutional repositories*. Retrieved May 18, 2018 from http://en.wikipedia.org/wiki/Institutional_repository
- Xia, J. (2006). Personal name identification in the practice of digital repositories. *Program: Electronic Library & Information Systems*, 40(3), pp. 256-267.
- Zahedi Nooghabi, M. (2010). *Investigation of organizational repositories of Ferdowsi University of Mashhad and its compatibility with scientific criteria* (master's thesis). Ferdowsi University of Mashhad, Faculty of Educational Sciences and Psychology. (in Persian)

- Zahedi Nooghabi, M. (2015). An Analysis of the Factors Affecting the Institutional Repositories Development in Universities based on the Information Systems Theories. *Library & Information Science Research*, 5(1), 280-300. (in Persian)
- Zhang, Y., Zhang, F., & Yao, P. (2018). Name disambiguation in AMiner: clustering, maintenance, and human in the loop. In: *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, London, 19–23 August 2018, pp. 1002–1011. New York: ACM.

