

Challenges of Subject Classification of Theses and Dissertations in the Iran Scientific Information Database (GANJ)

Fariborz Doroudi

*Corresponding author, Assistant Professor; Iranian Research Institute for Information Science & Technology (IranDoc), Tehran, Iran. Email: doroudi@irandoc.ac.ir. Orcid No.: 0000-0003-0386-5301

Tahereh Sadeghian khoshoei

MSc. in Knowledge and Information Science, Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran. Email: sadeghian@irandoc.ac.ir. Orcid No.: 0009-0008-7620-5041

Abstract

Objective: The purpose of this study was to analyze the topics of theses and dissertations in the Iran Scientific Information Database (GANJ) and determine criteria for selecting topics based on information organization standards. Additionally, the study aimed to identify the main challenges in selecting relevant topics.

Methodology: The research method was qualitative and its approach was comparative. Data collection involved conducting interviews and extracting relevant data from postgraduate theses and dissertations available in the Iran Scientific Information Database (GANJ). The statistical population of the database consisted of 1,112,531 documents. Sampling was done according to the Morgan table, and 1,500 documents were selected for review. The validity of the research tool was assessed using the face validity method with the input of 5 academic experts in information science. Additionally, the reliability of the assessment tool, a checklist to determine the thematic connection between the structure of theses and dissertations and the allocation of related keywords, was evaluated using the Cronbach's alpha statistical method, with a value of 0.872. The study was conducted on this sample in proportion to the size of the field and academic orientations. The analysis focused on the relevance between thematic keywords and the core concept of theses and dissertations, taking into account existing information sources to evaluate the new structure.

Findings: The findings revealed significant challenges in organizing topics, including mismatches between "topic areas" and "topic levels", absence of some "main topics", non-standardization of topics, inability to "add a topic", lack of clarity in the concept of the topic area, and insufficient "supplementary keywords". The findings show that subject keywords in a subject classification system play a crucial role in efficiently searching for information and retrieving information sources. Users typically try to choose the most relevant keyword when searching for information. If they do not find the desired results, they refine their keywords based on feedback and results to further explore by adjusting their search queries. Subject ranking is also important, as determining high-level and low-level relationships in the indexing structure can help identify relevant keywords for accurate retrieval. Using specialized dictionaries can enhance the subject area, However, this is currently not feasible.

Conclusion: The study concluded that using pre-determined keywords outside the thesaurus structure complicates subject assignment for indexers. Utilizing citation networks, techniques such as clustering in subject structure, understanding the relationship between general and specific terms, dependent terms, and the reference system can improve keyword assignment accuracy. Implementing automated information organization methods can enhance information analysis and processing, improve keyword assignment accuracy, and optimize information retrieval. The results suggest that the similarity between keywords and visualization descriptors is vital, often demonstrated through the use of

specialized tools such as thesauruses. Additionally, utilizing the Web of Science classification system can significantly improve retrieval results. However, this may not be adequate for exploring smaller trends that require detailed examination. Considering other classification systems such as Scopus can be beneficial. Incorporating synonym keywords can improve the selection of subject keywords. Folksonomy systems can also enhance the subject structure effectively.

Keywords: Information Evaluation, Iran Scientific Information Database (GANJ), Dissertations and Theses, Subject Classification

چالش‌های طبقه‌بندی موضوعی پایان‌نامه‌ها و رساله‌ها در پایگاه اطلاعات علمی ایران (گنج)

فریبرز درودی

* نویسنده مسئول، استادیار، پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)، تهران، ایران. رایانامه: doroudi@irandoc.ac.ir شماره ارجید: 0000-0003-0386-5301

طاهره صادقیان خشویی

کارشناس ارشد علم اطلاعات و دانش‌شناسی، پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)، تهران، ایران. رایانامه: sadeghian@irandoc.ac.ir شماره ارجید: 0009-0008-7620-5041

تاریخ ارسال: ۱۴۰۵/۰۲/۲۶

تاریخ ویرایش: ۱۴۰۵/۰۴/۰۷

تاریخ پذیرش: ۱۴۰۵/۰۶/۰۱

چکیده

مقدمه: هدف این پژوهش بررسی موضوع‌های موجود پایان‌نامه‌ها و رساله‌ها (پارساه‌ها) و تطبیق آنها با معیارهای مناسب انتخاب موضوع بر اساس استانداردهای سازماندهی اطلاعات و مشخص ساختن چالش‌های اصلی در انتخاب موضوع مرتبط در پایگاه اطلاع علمی ایران (گنج) بود.

روش پژوهش: روش پژوهش کیفی و شیوه آن تطبیقی بود. جامعه آماری پارساهای پایگاه اطلاعات علمی ایران (گنج) شامل تعداد ۱/۱۱۲/۵۳۱ مدرک بود. نمونه‌گیری براساس جدول مورگان به تعداد ۱۵۰۰ مدرک تعیین شد. ابزار گردآوری اطلاعات سیاهه واری و مصاحبه بود. سنجش روایی ابزار پژوهش به شیوه روایی صوری و پایایی با روش آماری آلفای کرونباخ به انجام رسید و میزان آن ۰/۸۷۲ تعیین شد.

یافته‌ها: یافته‌ها نشان داد که چالش‌های مهم سازماندهی موضوعی عبارتند از: نداشتن تناسب برخی «حوزه‌های موضوعی» با «سطوح موضوعی»، فقدان پاره‌ای از «موضوع‌های اصلی»، «استاندارد نبودن» برخی موضوع‌ها، عدم امکان «افزودن موضوع»، «نارسانا بودن مفهوم» در حوزه موضوعی، نبود «کلیدواژه‌های مکمل».

نتیجه‌گیری: نتایج نشان داد که بهره‌گیری از کلیدواژه‌های از قبل تعیین شده و خارج از ساختار اصطلاح‌نامه، تا حدودی شرایط کار تخصیص موضوع را برای نمایه‌ساز از منظر تطبیق موضوعی دشوار می‌کند. مشخص شد که استفاده از شبکه‌های استنادی در نمایه‌سازی، امکان بهره‌گیری از فنون مختلف چون خوشه‌بندی در ساختار موضوع، ارتباط میان اصطلاح اعم و اخص، اصطلاح وابسته و نیز نظام ارجاعی سبب تقویت میزان درستی تخصیص کلیدواژه به حوزه موضوعی می‌شود. علاوه بر آن، انجام روش‌های سازماندهی اطلاعات خودکار، تحلیل و پردازش اطلاعات را تسریع کرده و میزان دقت تخصیص کلیدواژگان مرتبط را نیز افزایش می‌دهد و شرایط بهتری برای بازیابی اطلاعات ایجاد می‌کند.

کلیدواژه‌ها: ارزیابی اطلاعات، پایگاه اطلاعات علمی ایران (گنج)، پایان‌نامه‌ها و رساله‌ها (پارساه‌ها)، طبقه‌بندی موضوعی

مقدمه

پایان نامه ها و رساله های (پارساهای) الکترونیکی برای مدیریت و بازیابی مؤثر آرشیو، به طبقه بندی موضوعی دقیق نیاز دارند (کلر و دیگران^۱، ۲۰۲۵ A) و مؤسسات آموزش عالی معمولاً از مخازن سازمانی برای گردآوری و در دسترس قرار دادن پارساهای الکترونیکی استفاده می کنند (فیری^۲، ۲۰۲۰). باید تاکید کرد که این منابع، حاوی دانش ارزشمندی هستند که می توانند برای طیف وسیعی از اهداف مفید باشند (آهوچا، دورا و فاکس^۳، ۲۰۲۲) و رساله ها قدرتمندترین و عینی ترین معیار موفقیت برنامه دکترای یک گروه آموزشی را ارائه می دهند (لویتس^۴، ۲۰۲۳). همانطور که می دانیم، طرح های طبقه بندی موضوعی، اساس سازماندهی، ارزیابی و پیمایش دانش علمی هستند (لیو و که^۵، ۲۰۲۵) و طبقه بندی موضوعی دقیق پارساهای الکترونیکی در اثربخشی آنها نقش اساسی دارد (کلر^۶، ۲۰۲۵). همچنین، طبقه بندی موضوعی بخش ضروری همه موتورهای کاوش دانشگاهی برای تسهیل جستجو و بازیابی سریع تر مقالات علمی بر اساس پرس و جوهای کاوش است (جورج^۷، ۲۰۲۲). بر همین اساس، موتورهای کاوش می توانند از رده های موضوعی برای ایجاد ویژگی های جستجو چندوجهی استفاده کنند (کاندیمالا^۸، ۲۰۲۰). اما باید توجه کرد که برچسب های موضوعی تعیین شده توسط نویسندگان در پارساهای الکترونیکی اغلب متناقض، بیش از حد گسترده یا ناهماهنگ با تمرکز برای پژوهش هستند (کلر و دیگران^۹، ۲۰۲۵ B). به همین دلیل، شیوه های آرشیو و سازماندهی سنتی برای توصیف پارساهای الکترونیکی به طرح های فراداده ای گسترده و سطح بالا متکی هستند که نمی توانند عمق، پیچیدگی و ماهیت بین رشته ای این آثار علمی وسیع را به تصویر بکشند (بنرجی، اینگرام و فاکس^۹، ۲۰۲۴).

توصیف موضوع، در بیشتر موارد، به طور پیش فرض در مفهوم سند تعریف می شود (دوتا و دوتا^{۱۰}، ۲۰۱۳). به عنوان نمونه، در پایگاه های داده دانشگاهی مختلف، عمده تاً دو نوع رویکرد طبقه بندی موضوعی وجود دارد: مبتنی بر مجله بر اساس سطح منبع و مبتنی بر مقاله بر اساس سطح انتشار (سینق و دیگران^{۱۱}، ۲۰۲۰) و از همین رو، طبقه بندی موضوعی برای کمک به گردآوری منابع بر اساس موضوع انجام می شود و کاربر را قادر می سازد تا نشریات را بر اساس سطوح مختلف تخصصی موضوع شناسایی کند (داتا^{۱۲}، ۲۰۲۳). با همین رویکرد، توانایی تعیین ساختار موضوعی منابع بنیادی پارساهای الکترونیکی می تواند توسعه بسیاری از خدمات را تسهیل کند (منظور^{۱۳}، ۲۰۲۳).

¹ Klair et al.

² Phiri

³ Ahuja, Devera & Fox

⁴ Lovitts

⁵ Lyu & Ke

⁶ Klair

⁷ George

⁸ Kandimalla

⁹ Banerjee, Ingram & Fox

¹⁰ Dutta & Dutta

¹¹ Singh et al

¹² Datta

¹³ Manzoor

اما چندین روش سودمند برای سازماندهی منابع اطلاعاتی وجود دارد؛ یکی از آنها نظریه ریزساختاری^۱ است این نظریه به مطالعه در باره رابطه میان مفاهیم در پیکره نظری دانش می‌پردازد و ابعاد آن را مشخص می‌سازد. در کنار آن، روندهای مشاهده شده در این نظریه در نظریه کلان ساختاری نیز در نقشه مفهومی، حوزه موضوعی، حوزه‌های معنایی و نظریه‌های اصطلاحی مورد بررسی قرار می‌گیرد (باستوز و فوجیتا^۲، ۲۰۰۵). دیگری، نظریه سلسله‌مراتب شناخت اجتماعی^۳ است که تشخیص موضوع بخش مهمی از پژوهش است و می‌تواند اطلاعات بیشتری بیشتری را در اختیار خوانندگان قرار دهد و آنها را قادر می‌سازد تا تجربه شناختی بهتری به دست آورند (ونگ^۴، ۲۰۲۱). مورد بعدی درخت‌های موضوعی (سلسله‌مراتب) است. در آینده‌ای قابل پیش‌بینی، درخت‌های موضوعی تنها کاربرد عملی روش‌های طبقه‌بندی در اینترنت برای بازیابی اطلاعات باقی خواهند ماند (ویتلی^۵، ۲۰۲۴). همچنین در مجموع باید بیان کرد که امروزه مدل‌های بازیابی اطلاعات مبتنی بر مفهوم و کلیدواژه است که نقشی مهم در دستیابی به اطلاعات ایفا می‌کند (روستایی، صدرالدینی و فخر احمد^۶، ۲۰۲۰).

برای انجام پژوهش حاضر، خلاءهای موجود در حوزه موضوعی مورد شناسایی قرار گرفته و کاستی‌های ساختار موضوعی در ارتباط با موضوع‌های در نظر گرفته شده شناسایی و متناسب با هر موضوع تقسیم‌بندی‌های فرعی آن مشخص می‌شود و یک ساختار روزآمد و متناسب با طبقه‌بندی علمی نوین ارائه خواهد شد. این کار بر مبنای ساختار اصطلاح‌نامه‌های استاندارد به انجام می‌رسد تا ارتباط میان ساختار نوین و وضعیت کنونی نیز مشخص شود. با این رویکرد، مشکلات شناخته شده و راهکارهای اجرایی برای اصلاح ساختار موضوع و روش متناسب‌سازی موضوع‌های اختصاص یافته با مفهوم اصلی پارساها نیز بیان می‌شود. منظور از متناسب‌سازی موضوع‌ها، تطبیق مناسب کلیدواژه‌های موضوعی با محتوای اصلی پارساهاست که ربط دقیق میان آن دو برقرار شود. از همین‌رو، هدف اصلی این پژوهش شناخت چالش‌ها و مشکلات ساختار طبقه‌بندی موضوعی^۷ پارساها در پایگاه اطلاع علمی ایران (گنج) است. پرسش اصلی نیز عبارت است از: چالش‌های وضعیت کنونی تعیین حوزه موضوعی در فرایند نمایه‌سازی و ویراستاری پارساها در پایگاه اطلاع علمی ایران (گنج) چیست؟

پیشینه پژوهش

ابتدا باید بیان کرد که در بررسی پیشینه پژوهش به ابعاد و جوانب مطالعات مختلف در ارتباط با حوزه رده‌بندی^۸ و نیز طبقه‌بندی موضوعی پرداخته شده است. اصطلاح رده‌بندی، تقسیم‌بندی منظم دانش به موضوع‌های مرتب شده در رشته‌های علمی است، که هر کدام بخش خاص خود را از یک سیستم سازماندهی دانش دارند (هاوارد و نولتون^۹،

¹ Microstructural theory

² Bastos & Fujita

³ Social cognition

⁴ Wang

⁵ Wheatley

⁶ Roostaei, Sadreddini & Fakhrahmad

⁷ Subject Categories

⁸ Classification

⁹ Howard & Knowlton

۲۰۱۸) و منابع اطلاعاتی بر اساس دسته‌بندی‌های ثابت موضوعی سازماندهی می‌شوند (درابینسکی^۱، ۲۰۱۹). مانند طرح‌های رده‌بندی کتابخانه کنگره آمریکا و یا طرح رده‌بندی دهدهی دیویی. ولی طبقه‌بندی موضوعی شامل تقسیم‌بندی‌هایی است که در ساختار کلان موضوعی برای دسته‌بندی‌های موثر انجام می‌شود. معمولاً تعداد آنها، محدودتر است و کلیت موضوعی را در ارتباط برخی از مولفه‌های دیگر، مانند رشته‌های تحصیلی، حوزه‌های مطالعاتی و نظایر آن مشخص می‌سازد. مانند طبقه‌بندی موضوعی وبگاه وب آو ساینس^۲. در ادامه پاره‌ای مطالعات انجام گرفته گرفته در خصوص روش‌های رده‌بندی، طبقه‌بندی و ساختار آنها معرفی می‌شود.

پژوهش‌های مرتبط با نظام‌های رده‌بندی پر استفاده

نتایج پژوهش گلوب، لیک و تودهوپ^۳ (۲۰۱۴) نشان‌دهنده اهمیت پیشنهادها و واژگان کنترل‌شده برای نمایه‌سازی و بازیابی بود. استفاده از برجسب‌های موضوعی و اطمینان از ثبات و افزایش تعداد نقاط دسترسی در بازیابی، از جمله یافته‌های مورد توجه در اختصاص کلیدواژه‌ها بود. همچنین، حیدر و شارما^۴ (۲۰۱۷) به جنبه‌های مختلف رده‌بندی کنگره و مناسب بودن آن به عنوان یک نظام رده‌بندی کتابخانه‌ای برای طبقه‌بندی منابع کتابخانه‌ای در هند پرداختند. نتایج نشان می‌دهد که نمونه‌های شماره‌سازی برای موضوعات خاص در زمینه هند، مانند تاریخ ادبیات هندی با رمزنویسی مارک ۲۱ برای اعداد رده‌بندی کنگره ارائه شد، کارآمد هستند. پژوهش لوند^۵ (۲۰۱۹) نشان دهنده آن بود که مقایسه‌های استنباطی بین پاسخ‌های ادراکی کتابداران فهرست‌نویس و غیرفهرست‌نویس نسبت به رده‌بندی دهدهی دیویی و رده‌بندی کتابخانه کنگره دارای تفاوت معناداری نیست و برداشت‌های آنان از این نظر نزدیک به هم است. نتایج مطالعه لوند و دیگران^۶ (۲۰۱۹) درباره استفاده و ترجیح دو طرح اصلی رده‌بندی کتابخانه دانشگاهی یعنی دیویی و کنگره حاکی از آن بود که تقریباً ۸۱ درصد از کتابخانه‌های دانشگاهی در ایالات متحده و ۹۳ درصد در نیجریه از رده‌بندی کتابخانه کنگره برای سازماندهی مجموعه‌های خود استفاده می‌کنند و آن را سودمند ارزیابی کرده‌اند. کلر و دیگران^۷ (۲۰۲۵) نیز درباره چالش‌های محاسباتی پردازش پارساهای الکترونیکی به مطالعه مبادرت ورزیدند. نتایج نشان داد که پیاده‌سازی یک رده‌بندی موضوعی مؤثر در سازماندهی آنها، یکی از مسائل اصلی در مخازن دیجیتال نوین است.

پژوهش‌های مرتبط با ساختار رده‌بندی‌های منطقه‌ای و چالش‌ها

ژنگ، ژیاو و بوو^۸ (۲۰۱۷) نوعی رده‌بندی علمی برای سازماندهی و بازیابی منابع اطلاعاتی را طراحی کردند که به روش رده‌بندی کتابخانه‌های چین^۹ مشهور است. نتایج نشان داد که این رده‌بندی از مدل روابط سطح بالا و پایین

¹ Drabinski

² Web of Science

³ Golub, Lykke & Tudhope

⁴ Haider & Sharma

⁵ Lund

⁶ Lund et al.

⁷ Klair et al.

⁸ Zhang, Xiao & Bu

⁹ Chinese Library Classification Method (CLCM)

برای توصیف ساختار رده‌بندی به خوبی استفاده می‌کند. پژوهش پوسیکا، اوژوو و اوبرهیری-اروما^۱ (۲۰۱۹) حاکی از آن بود که افزایش حمایت از اتوماسیون کتابخانه‌ها، فرصت‌های آموزشی موجود، معرفی سیستم مدیریت اطلاعات، تسریع حرکت از روش دستی به روش رایانه‌ای، به‌عنوان نشانه‌ای از آینده روشن‌تر و چشم‌انداز مناسب برای بخش‌های فنی کتابخانه‌های مورد مطالعه است. نتایج مطالعه کاتاگی و کومبار^۲ (۲۰۲۳) نشان می‌دهد که وضعیت فعلی پروژه‌های سازماندهی پارساها در سطح جهانی رضایت‌بخش نیست، زیرا اکثر این پروژه‌ها از تخصیص بودجه مناسب، ابزارهای لازم و امکانات زیرساخت فناوری اطلاعات مناسب، به ویژه در کشورهای توسعه نیافته و در حال توسعه، برخوردار نیستند.

پژوهش‌های مرتبط با ساختار رده‌بندی مقالات و منابع اینترنتی

در پژوهش آوویرون و روزنفیلد^۳ (۲۰۲۱) که درباره طبقه‌بندی مجله‌های علمی در وب‌آوساینس و اسکوپوس^۴ به انجام رسیده است، یافته‌ها نشان می‌دهد که رده‌های موضوعی جنبه مهمی در ارزیابی پژوهش‌های علمی و نیز تحلیل کتاب‌سنجی دارد. نتایج حاکی از آن است که اختلاف‌های قابل توجهی در وب‌آوساینس و اسکوپوس وجود دارد که تعداد بیشتری از طبقه‌بندی‌های انجام شده با افزایش دامنه و تنوع رتبه‌بندی در آنها ارتباط دارد. نتایج مطالعه جینگیرارد^۵ (۲۰۲۱) درباره طبقه‌بندی موضوعی مقاله‌ها نشان داد که با توجه به آنکه در این فرایند از دو نوع نوع رویکرد استفاده می‌شود که شامل روش طبقه‌بندی در سطح مجله و نیز طبقه‌بندی در سطح مقاله است، طبقه‌بندی موضوعی نقش مهمی در تحلیل محتوای نشریات علمی دارد. در پژوهش کازی، لین و کاهاندا^۶ (۲۰۲۱) که به منظور سازماندهی خودکار مقالات عملی انجام پذیرفت، نتایج نشان داد که مؤسسات علمی و فرهنگی تلاش می‌کنند تا مقالات خود را با استفاده از موضوع‌های مرتبط و مناسب فهرست‌نویسی کنند تا کاربران بتوانند به راحتی مقالات مرتبط را از مخازن آنها بازیابی کنند. همچنین استفاده از وب اسکریپینگ^۷ برای استخراج کلمات کلیدی نتایج مطلوبی به همراه دارد. سوچیاتی، فابروییر و پردده^۸ (۲۰۲۳) در پژوهش خود به مقوله داده‌کاوی پارساها پرداختند. نتایج نشان داد که تحلیل خوشه‌بندی برای ایجاد خوشه‌هایی از موضوعات پارساها نقش موثری در سازماندهی آنها دارد. لیو و که^۹ (۲۰۲۵) در پژوهشی به بررسی طرح طبقه‌بندی علمی پرداختند. نتایج نشان داد که جایگزین‌های داده‌محور مبتنی بر شبکه‌های استناد، نویدبخش بوده اما فاقد اعتبارسنجی خارجی دقیق در برابر محتوای معنایی ادبیات علمی هستند.

¹ Posigha, Ojohwoh & Oberhiri-Oruma

² Katagi & Kumbar

³ Aviv-Reuven & Rosenfeld

⁴ Scopus

⁵ Jeangirard

⁶ Kazi, Lane & Kahanda

⁷ Web scraping (عمل استخراج داده‌ها از وبگاه‌ها برای ذخیره‌سازی محلی آنها)

⁸ Suciati, Fabroyir & Pardede

⁹ Lyu & Ke

پژوهش‌های مرتبط با ساختار رده‌بندی خودکار و کاربرد فناوری اطلاعات

نتایج پژوهش گلوب و همکاران^۱ (۲۰۱۶) نشان داد که ابزارهای تخصیص موضوعی خودکار در مقابله با مقیاس پایداری در ایجاد و غنی‌سازی فراداده، ایجاد ارتباط بیشتر بین منابع و افزایش ثبات سازماندهی کمک می‌کند. پژوهش آریسوکو و اوگبونا^۲ (۲۰۲۰) نیز در باره کاربرد فناوری اطلاعات در رده‌بندی منابع کتابخانه‌ای حاکی از آن بود که چالش‌های کاربرد فناوری اطلاعات در فهرست‌نویسی و رده‌بندی مواد کتابخانه‌ای در کتابخانه‌های دانشگاهی شامل نبود امکانات کافی فناوری، کمبود فرصت‌های آموزشی و نیز عدم نگهداری ابزارها، نوسان‌های شبکه و خرابی خدمات اینترنتی است. در پژوهش یامان، تونسر و ارتام^۳ (۲۰۲۱) مشخص شد که با توجه به تعداد بالای منابع اطلاعاتی، مجموعه داده‌های شاخص قدرت سیگنال دریافتی (آر.اس.اس.آی.)^۴ و یک مدل طبقه‌بندی مکانی خودکار جدید موثر است. این مدل با استفاده از ابزار رده‌بندی درخت تصمیم^۵ با اعتبارسنجی متقاطع ده برابری موفق ارزیابی شد. اولیری^۶ (۲۰۲۳) در پژوهش خود به بررسی میزان «استفاده» پژوهشگران دانشگاهی از مدل‌های زبانی بزرگ، مانند: چت جی پی تی^۷ و گوگل بارد^۸ در استفاده از پایان‌نامه‌ها، پروپوزال‌ها و رساله‌ها پرداخت. نتایج نشان داد که توجه محدودی به میزان استفاده دانشجویان از این ابزارها برای نگارش فعالیت‌های پژوهشی شده است. یافته‌ها حاکی از آن بود که استدلال‌ها در زمینه بهره‌گیری از ابزارهای یادشده، سبب شده تا مقالات مجلات بهتر سازماندهی شده و تشکیل جلسات فنی برای بررسی پایان‌نامه‌ها و رساله‌ها سودمند بوده است.

نتایج پژوهش‌های انجام شده نشان می‌دهد که بررسی ساختارها و اثربخشی طبقه‌بندی علمی در ابعاد مختلفی به توجه به کارایی و کاربرد آنها مورد توجه قرار گرفته است. این طبقه‌بندی‌ها سودمندی خود را در ارتباط با سازماندهی منابع اطلاعاتی دارند و هر یک متناسب با وضعیت منابع، نقش موثری در این زمینه ایفا می‌کنند. با عنایت به آنکه در حال حاضر، برخی از موضوع‌های اختصاص یافته به پارساها متناسب با مفهوم اساسی آن انجام نشده است؛ همچنین ارتباط میان بخش حوزه موضوعی در نرم‌افزار سازماندهی با موضوع‌های اختصاصی در برخی موارد منطبق نیست، در پاره‌ای از موارد موضوع مرتبط با پارسای مورد نظر در بخش حوزه موضوعی وجود ندارد؛ علاوه بر آن، موضوع‌های استفاده شده، مبتنی بر اصطلاح‌نامه و یا سرعنوان موضوعی نبوده و از این نظر نوعی چالش در سازماندهی پارساها وجود دارد. لذا با توجه به وجود این خلاء در فرایند سازماندهی پارساها، توجه به وضعیت موضوع و بررسی ابعاد و جوانب به کارگیری آن می‌تواند حائز اهمیت بالایی در پژوهش حاضر باشد.

روش پژوهش

¹ Golub et al

² Arisukwu & Ogbonna

³ Yaman, Tuncer & Ertam

⁴ Received Signal Strength Indicator (RSSI)

⁵ Decision Trees

⁶ O'Leary

⁷ ChatGPT

⁸ Google Bard

در این پژوهش به مطالعه چالش‌ها و خلاءهای موجود در سامانه ویرایش و سازماندهی اطلاعات علمی پرداخته و این موارد مشخص و تعیین شد. لذا از مطالعه کیفی، به شیوه تطبیقی در ابعاد حوزه موضوعی و ساختار آن بهره گرفته شد. مطالعه تطبیقی، فرایندی عملی و مطالعاتی به منظور مقایسه مفاهیم و یا متغیرهایی است که با هدف کشف و بازنمود ساختار، ویژگی‌ها و یا دیگر خصوصیات است که در کنار هم بررسی می‌شوند. در مطالعات تطبیقی، علاوه بر مقایسه، مقولات مورد نظر، توصیف، ارزیابی و یا تفسیر نیز می‌شوند. در این میان می‌توان وجوه مشترک و یا تفاوت‌های موجود را بررسی کرد. همچنین می‌توان به دلایل برتری، نقاط ضعف و قوت، ساختار شکل‌دهنده متغیرها و دیگر ابعاد و جوانب مطالعاتی پرداخت. بنابراین به دلیل سودمندی و مناسب بودن این روش برای چنین سنجشی از آن بهره گرفته شد. جامعه آماری پژوهش، پارساهای پایگاه اطلاعات علمی ایران (گنج) است. این منابع شامل تعداد ۱/۱۱۲/۵۳۱ مدرک است. این مدارک شامل پارساهای تحصیلات تکمیلی است و پیشنهادها را در بر نمی‌گیرد. بر این اساس با توجه به حجم مجموعه بر مبنای جدول مورگان نمونه آماری در ابتدا به میزان ۳۸۴ مدرک انتخاب شد. با عنایت به حجم بالایی مجموعه پارساها که بیش از ۱ میلیون رکورد را پوشش می‌دهد، و نیز به دلیل تقویت فرایند روش‌شناسی پژوهش، در کنار عامل بالا بردن ضریب اطمینان میزان درستی در ارتباط با اطلاعات سازماندهی‌شده و با توجه به تنوع موضوعی پارساها در حوزه‌های مفهومی متعدد و گوناگون، یک دسته نمونه آماری از میان پارساها به تعداد ۱۵۰۰ مدرک و به صورت تصادفی طبقه‌ای انتخاب شد. بررسی بر روی این نمونه آماری متناسب با حجم رشته و گرایش‌های تحصیلی به انجام رسید. برای انجام این فرایند از یک سیاهه واریسی محقق‌ساخته استفاده شد. این سیاهه با تعیین ۲۱ مولفه در زمینه‌های مرتبط چون: تعیین حوزه موضوعی، کلیدواژه، چکیده، ساختار مفهومی، بسامد تکرار واژگان، بیان مساله و فصل‌بندی پارساها، برای تحلیل و ارزیابی ساختار موضوعی طراحی شد.

سنجش روایی ابزار پژوهش به شیوه روایی صوری توسط ۵ تن از متخصصان موضوعی علم اطلاعات دانشگاهی انجام شد. پایایی ابزار گردآوری اطلاعات نیز با روش آماری آلفای کرونباخ به انجام رسید و میزان آن ۰/۸۷۲ تعیین شد. برای تکمیل فرایند پژوهش، پس از سنجش ساختار موضوعی توسط سیاهه واریسی، نظر کارشناسی ۴ تن از متخصصان نمایه‌سازی در ایراندک، با انجام مصاحبه دریافت شده تا در تقویت فرایند پژوهش از آن استفاده شده و شناخت بهتری نسبت به چالش‌های نمایه‌سازی در حوزه موضوعی حاصل شود. این مصاحبه شامل موارد مربوط به مشکلات سازماندهی و ابعاد آن بود. در خصوص سطوح حوزه‌های موضوعی، جنبه‌های مفهومی در اختصاص موضوع، تطبیق موضوع با مفاهیم کلان پارساها، استفاده از کلیدواژه‌های کمکی و نیز روش تخصیص موضوع نهایی و همچنین ویراستاری پارساها، ابعاد و جوانب کاربرد موضوع مورد بحث و بررسی قرار گرفت. علاوه بر آن، برخی از اشکالات موجود و خلاءهایی که معمولاً نمایه‌سازان در استفاده از برنامه ویرایش با آن مواجه هستند مورد کنکاش بیشتری واقع شد. در نهایت ارزیابی وضعیت کلیدواژه‌های موضوعی در سامانه انجام شد تا میزان درستی و انطباق آنها با ساختار مفهومی و موضوعی پارساها تعیین شود. بر این مبنای میزان ربط میان کلیدواژه‌های موضوعی با مفهوم اساسی پارساها تحلیل شده تا در سنجش ساختار جدید به منابع اطلاعاتی موجود نیز توجه شده باشد.

یافته‌های پژوهش

در پاسخ به پرسش اصلی پژوهشی یعنی: چالش‌های وضعیت کنونی تعیین حوزه موضوعی در فرایند نمایه‌سازی و ویراستاری پارساها چیست؟ این موارد شامل بررسی موضوع‌های وارد شده در سیستم سازماندهی اطلاعات از طریق بررسی داده‌های حاصل از سیاهه واری با محتوای اطلاعاتی شامل ساختار موضوعی در پارساهای نمونه و تطبیق آن با پیشینه‌های سازماندهی شده در پایگاه اطلاعات علمی ایران (گنج) به انجام رسیده است. علاوه بر آن با کسب نظر نمایه‌سازان در مدیریت سازماندهی و تحلیل اطلاعات، بررسی میدانی شرایط پایگاه‌های اطلاعاتی، و مطالعه نظری درباره موضوع نمایه‌سازی موارد مشخص شد. بر اساس سنجش انجام گرفته و بررسی پارساهای نمونه، چالش‌های سازماندهی به شرح جدول ۱ تعیین شدند.

جدول ۱. چالش‌های پارساها در تعیین طبقه‌بندی موضوعی

عنوان چالش	توضیح آن
تناسب «حوزه‌های موضوعی» با «سطوح موضوعی»	در این زمینه ارتباط میان حوزه‌های موضوعی با سطوح موضوعی در ۲ سطح آن مطرح می‌شود. پاره‌ای از موضوع‌ها با سطوح مطرح شده متناسب نیستند.
فقدان پاره‌ای از «موضوع‌های اصلی»	برخی از موضوع‌های اصلی و مهم که با کلیدواژه‌های مرتبط خود باید مطرح شوند، در این سطوح وجود ندارد.
توجه به «اهمیت موضوع»	پاره‌ای از موضوع‌های با اهمیت کمتر مطرح شده‌اند و گرایش‌های موضوعی محدودی دارند. همچنین برخی بیشتر به موضوع‌های خردتر تقسیم شده‌اند
«استاندارد» بودن موضوع	تطابق موضوع مورد استفاده با ابزارهای استاندارد موضوعی از جمله تزاروس‌ها و سرعنوان‌های موضوعی که مبتنی بر خط‌مشی دقیق واژه‌گزینی و اصطلاح‌های استاندارد باشد.
«فریز» بودن عبارت موضوعی	بهره‌گیری از عبارت‌های موضوعی محدود بوده و در این محدودیت تمامی ابعاد موضوعی از نظر مفهومی و اطلاق به مفهوم اصلی آن نیز لحاظ نشده است.
عدم امکان «افزودن موضوع»	امکان افزودن موضوع جدید در صورت تشخیص نیاز وجود ندارد.
«نارسا بودن مفهوم» در حوزه موضوعی	در برخی موارد مفهوم اساسی و کلان یک پارسا متناسب با کلیدواژه اختصاص یافته آن تعیین و تعریف نشده است. همچنین دارای نوعی نارسایی مفهومی از نظر انطباق میان مفهوم و کلیدواژه موضوعی است.
نبود «کلیدواژه‌های مکمل»	برخی موارد کلیدواژه‌های مکمل می‌تواند به انتخاب بهتر موضوع منجر شود. لذا ایجاد یک پایگاه از این موضوع‌ها کمک می‌کند تا استفاده از حوزه‌های موضوعی بهتر انجام گیرد.
لحاظ کردن عنصر «ربط»	در طراحی حوزه‌های موضوعی استفاده از عامل «ربط» نقش مهمی در درستی آن دارد.
«نیاز کاربر اطلاعاتی»	توجه به کاربر و نقش او در انتخاب کلیدواژه‌های موضوعی و «فرمول‌بندی جست‌وجو» در دستیابی به منابع اطلاعاتی پارساها

باید بیان کرد که در طراحی پایگاه اطلاعاتی سازماندهی اطلاعات توجه به دو سطح موضوعی شده است. با توجه به تعریفی که پیشتر در ارتباط طبقه‌بندی موضوعی به انجام رسید، سطح کلان آن در ساختار سازماندهی پارساها در سامانه گنج، با عنوان حوزه موضوعی مشخص شده است که در واقع، تقسیم‌بندی کلان سطوح خرد موضوعی به شمار می‌آید. در این زمینه، دو بخش اصلی حوزه موضوعی ۱ و حوزه موضوعی ۲ معرفی شده‌اند. در

حوزه اول ۳ سطح علوم اجتماعی، علوم و علوم کاربردی، هنر و علوم انسانی قرار دارد. در سطح دوم نیز هر یک از سطوح اول، شامل تعداد زیادی موضوع می‌شود. در واقع این شیوه دسته‌بندی برای تنظیم بهتر، و ارتباط میان رشته‌های تحصیلی و نیز زمینه‌های کلان موضوعی در ساختار سازمانی آموزش عالی و با توجه به نیازهای آتی برای گزارش‌گیری و تحلیل گسترده مورد بهره‌برداری قرار گرفته است. آمار موضوع‌های سطح ۲، ۲۸۳ فقره است. آنچه که در این بخش مورد توجه قرار می‌گیرد، آن است که هر یک از سطوح تعدادی از موضوع‌ها را در بر می‌گیرد که مرتبط با رشته‌های پارساها است. از این منظر ارتباط نزدیکی میان رشته و موضوع وجود دارد که تا حد زیادی می‌تواند مفهوم اصلی را بیان کند. این طرح در سازماندهی پارساها از وبگاه اطلاعاتی وب آو ساینس^۱ اخذ شده است.

علاوه بر آن، میزان ارزش یک متن علمی در پایگاهی که مبتنی بر سازماندهی منابع اطلاعات پژوهشی است، بیشتر آشکار می‌شود. نکته مهم در تعیین میزان اهمیت یک منبع علمی آن است که تا چه میزان به آن استناد داده می‌شود. به تعبیر دیگر، متون علمی انتشار یافته، تا چه حد از این منابع در تنظیم متن و برای افزایش اعتبار علمی خویش استفاده می‌کنند. برای نمایه‌سازی مناسب نیاز به تحلیل موضوع است. معمولاً نمایه‌ساز با توجه به بررسی متن با عنایت به چکیده، مقدمه و بخش‌هایی از متن، به تشخیص مفهوم اصلی آن مبادرت می‌ورزد. سپس با عنایت به مفهوم اصلی پارسا، به تخصیص موضوع مناسب اقدام می‌کند. موضوع نیز بر اساس کلیدواژه‌های استاندارد مشخص شده و به مدرک اختصاص داده می‌شود. این موضوع البته بر اساس رشته و گرایش موضوعی که پارسا به آن پرداخته است تعیین می‌شود. نمایه‌سازان و ویراستاران علمی برای انجام فعالیت سازماندهی پارساها از سطوح موضوعی ۱ و ۲ بهره می‌گیرند. این سطوح با توجه به ترجمه‌ای که از وبگاه اطلاعاتی وب آو ساینس به انجام رسیده، دارای پوشش موضوعی نسبت به حوزه‌های مختلف موضوعی در خصوص پارساهاست. از این منظر نمایه‌سازان با بهره‌گیری از کلیدواژه‌های موجود و با عنایت به مفهوم اساسی هر پارسا، نسبت به انتخاب موضوع مرتبط اقدام می‌کنند.

یک نکته و چالش در این باره، نداشتن امکان بهره‌گیری از ابزارهایی است که می‌تواند مکمل این حوزه‌های موضوعی باشد. بهره‌گیری از اصطلاح‌نامه و تزاروس در این محیط امکان‌پذیر نیست. گرچه بخش قابل توجهی از این حوزه‌های موضوعی با اصطلاح‌نامه‌ها دارای همگونی است، ولی فراتر از سطوح موضوعی مشخص شده، انتخابی برای نمایه‌سازان وجود ندارد. در واقع انتخاب اولیه عبارت موضوعی به صورت فریز در این سطوح درآمده و نمایه‌ساز الزام به استفاده از همین تعداد مشخص کلیدواژه موضوعی دارد. البته تعیین موضوع‌های انتخابی می‌تواند کار نمایه‌ساز را آسان سازد. زیرا انتخاب از قبل در یک بازه مشخص موضوعی انجام شده، زمینه‌های موضوعی پرستفاده و با بسامد بالا شناسایی شده، و در زمان نمایه‌سازی می‌توان به سرعت موضوع متناسب را انتخاب و از آن در سازماندهی مدرک مورد نظر استفاده کرد. همچنین با تکیه بر این کلیدواژه‌ها شناسایی محتوا نیز قابل بررسی بیشتر است. علاوه بر آن در زمان و وقت نمایه‌ساز صرفه‌جویی انجام می‌گیرد، زیرا او دسترسی سریع‌تر به کلیدواژه‌های منتخب دارد. اما از سوی دیگر باید بیان کرد که در این زمینه سازماندهی کلیدواژه‌های موجود در پایگاه اطلاعاتی مرتبط با مفهوم بنیادی منابع است. با توجه به آنکه این انتخاب از طریق پایگاه وب آو ساینس انجام گرفته است، حوزه‌های موضوعی

¹ Web of Science

اصلی در این پایگاه شناسایی شده و لذا مفهوم کلان یک منبع اطلاعاتی قابل تشخیص و اختصاص کلیدواژه مرتبط است. تاثیر این روش نمایه‌سازی در بازیابی اطلاعات تا حد زیادی به دسترسی به اطلاعات مرتبط منجر می‌شود. ولی باید به یک نکته مهم در اینجا توجه کرد و آن اینکه جایگاه کاربردی کلیدواژه‌ها همواره در متن و متناسب با مفهومی که منبع اطلاعاتی ارائه می‌دهد مورد سنجش قرار می‌گیرد. در بسیاری مواقع کاربران به دنبال موضوع‌های فرعی و خرد هستند که معمولاً در طبقه‌بندی موضوعی پایگاه وب آو ساینس این امکان چندان مورد حمایت قرار نمی‌گیرد. لذا همخوانی میان کلیدواژه‌ها و توصیفگرهای نمایه‌سازی یکی از مباحث مهمی است که معمولاً در استفاده از ابزارهای حرفه‌ای مانند اصطلاح‌نامه بیشتر قابل مشاهده است.

چالش دیگر آن است که چنانچه در بررسی بیشتر کلیدواژه‌ها مشخص شد که برخی موضوع‌های خرد در این فضا وجود ندارد، لذا نمایه‌ساز مجبور می‌شود تا از یک حیطه کلی‌تر بهره گیرد تا بتواند موضوع مورد نظر را شناسایی و به منبع اطلاعاتی اختصاص دهد. یکی از نکات مهم در زمینه اختصاص کلیدواژه به منبع اطلاعاتی، توجه کردن به پدیده ربط است. میزان ارتباط میان مفهوم و کلیدواژه موضوعی که به مدرک اختصاص داده می‌شود باید بالاترین میزان ارتباط را داشته باشد؛ زیرا عبارت‌های جست‌وجوی کاربران متخصص بر اساس میزان ربط میان مفهوم مدرک و موضوع مورد نظر است. البته برخی مواقع متخصصان نیز ممکن است از کلیدواژه‌هایی استفاده کنند که کاملاً منطبق با موضوع استاندارد نیست، ولی به احتمال زیاد همگونی و همراهی آن با مفهوم اساسی مدرک بیشتر است. در این موارد از نظام ارجاعی بهره گرفته می‌شود تا از کلیدواژه مرتبط غیر مرجح به کلیدواژه مرتبط مرجح ارجاع داده شود. نمایه‌سازان علاوه بر استفاده از میان اصطلاح‌نامه و سرعنوان‌های موضوعی، در برخی موارد از واژه‌نامه‌های تخصصی نیز بهره می‌گیرند. دلیل آن نیز نبود یک اصطلاح مرتبط و تخصصی در منابع یاد شده است. این اصطلاح‌ها در صورت فقدان کلیدواژه مورد نظر می‌تواند در یک پایگاه فرعی نگهداری شده تا در موارد بعدی نیز مورد استفاده قرار گیرد. در حال حاضر بهره‌گیری از این کلیدواژه‌ها برای نمایه‌سازان میسر نیست و آنها نمی‌توانند از چنین واژگانی برای تخصیص موضوعی استفاده کنند.

روش مهیاسازی این ابزار سودمند آن است که این اصطلاح‌ها معمولاً در چند واژه‌نامه تخصصی تطبیق داده شده و در صورت تایید توسط بیشینه واژه‌نامه‌ها که در واقع مورد تایید بیشتر متخصصان آن حوزه موضوعی است انتخاب شوند. کارکرد اساسی کلیدواژه‌های موضوعی در یک نظام طبقه‌بندی موضوعی، نقش مهمی در جست‌وجوی بهینه اطلاعات برای بازیابی بهتر منابع اطلاعاتی دارد. همانطور که می‌دانیم، معمولاً کاربران در زمان کاوش اطلاعات تلاش می‌کنند تا نزدیک‌ترین کلیدواژه مرتبط را انتخاب کنند. حتی زمانی که به نتیجه مورد نظر نمی‌رسند، با اصلاح کلیدواژه‌ها بر اساس نتایج و بازخورد آن تلاش می‌کنند تا با فرمول‌بندی دوباره پرسش به کاوش بیشتر بپردازند. در این میان رتبه‌بندی موضوعی از اهمیت بالایی برخوردار است. کاربرد اصطلاح‌نامه به این نکته مهم اشاره دارد که در انتخاب اصطلاح‌های مرجح، از تحلیل واژگانی و موضوعی استفاده شده است. این بدان معناست که از یک سو نیاز کاربر در نظر گرفته شده، و از سوی دیگر میزان ربط میان مفهوم و کلیدواژه تخصصی لحاظ شده است. علاوه بر آن از روش‌های مختلف چون خوشه‌بندی در ساختار موضوع نیز بهره گرفته شده است. ارتباط میان اصطلاح اعم و اخص، اصطلاح وابسته و نیز نظام ارجاعی همگی در این رابطه سودمند هستند.

بحث و نتیجه‌گیری

نتایج نشان داد که استفاده مستقیم از ابزارهای کمکی برای سازماندهی حوزه‌های موضوعی میسر نیست. این بدان دلیل است که امکان استفاده از اصطلاح‌نامه در این محیط محدود است و بیشتر از سطوح موضوعی تعیین‌شده شده، انتخاب دیگری برای اختصاص موضوع وجود ندارد. عبارت موضوعی به شکل تعریف شده از قبل در سیستم بارگذاری شده و استفاده از آن برای نمایه‌ساز قطعی است. در واقع تعداد معینی کلیدواژه موضوعی در اختیار نمایه‌سازی قرار دارد که گرچه کار او را تسریع می‌کند، ولی محدودیت نیز به همراه دارد. به رغم بهره‌گیری از اصطلاح‌نامه و تزاروس در طراحی ساختار نظام موضوعی فعلی، و داشتن همگونی با اصطلاح‌نامه‌ها، استفاده بیشتر از سطوح مشخص شده فعلی وجود ندارد. دلیل آن استفاده از ساختار طبقه‌بندی موضوعی وب آو ساینس است که در این طرح، با توجه به تعریف موضوع‌های کلان و نزدیک به حوزه‌های مفهومی مرتبط با رشته‌های علمی، نمی‌توان به سطوح خرد موضوعی وارد شد و این خود سبب تعیین بهره‌گیری محدود از نظام طبقه‌بندی موضوعی وسیع و گسترده است. لذا وجود همگونی بین کلیدواژه‌ها و توصیف‌گرهای نمایه‌سازی دارای اهمیت بالایی است که معمولاً در استفاده از ابزارهای تخصصی چون اصطلاح‌نامه قابل مشاهده است. گرچه دسترس‌پذیری در این فرایند میسر است، ولی بازیابی با پاره‌ای اختلال‌ها از منظر استاندارد کاوش اطلاعات مواجه می‌شود. این یافته با نتایج پژوهش پوسیکا، اوژوو و اوبرهیری-اروما (۲۰۱۹) دارای هم‌سویی است، ولی از منظر ساختار خاص سازماندهی پایگاه گنج، با نتایج پژوهش ژنگ، ژیاو و بوو (۲۰۱۷) اشتراکی ندارد. همچنین بهره‌گیری از طبقه‌بندی وب‌آوساینس تا حد زیادی می‌تواند به نتایج مفید در بازیابی منجر شود. ولی در ارتباط با گرایش‌های خردتر که نیاز به کاوش دقیق و محدود شده است، کافی نیست. انتخاب طبقه‌بندی وب آو ساینس به خاطر پوشش موضوعی مناسب و دربرگیری زمینه‌های مطالعاتی منطبق با رشته‌های علمی و گرایش‌های شناخته شده آن است. این طرح در عین محدود بودن، دامنه گسترده‌ای از موضوع‌های کلان را پوشش می‌دهد که سبب بالا رفتن سرعت نمایه‌سازی و نیز بازیابی مرتبط بیشتر برای متخصصانی است که می‌خواهند در حیطه مطالعاتی رشته دانشگاهی خود به دنبال مطلب بگردند و زمان دسترسی آنها را کوتاه می‌کند که راحت‌تر بتوانند منابع خود را بازیابی کنند.

در این میان توجه به برخی از طبقه‌بندی‌های نظام‌های مشابه از جمله اسکوپوس نیز می‌تواند حاوی فواید باشد. نتیجه پژوهش آوویورون و روزنفیلد (۲۰۲۱) در این زمینه با یافته‌های پژوهش حاضر دارای اشتراک است.

بهره‌گیری از کلیدواژه‌های از قبل تعیین شده تا حدودی شرایط کار تخصیص موضوع را برای نمایه‌ساز از منظر تطبیق موضوعی دشوار می‌کند. این بدان سبب است که بهره‌گیری از اصطلاحات تخصصی و واژگان مرتبط با آن، نیاز به تبیین دقیق دارد و کاربرد گسترده آنها توانایی مطلوبی برای نمایه‌ساز به وجود می‌آورد که سبب استفاده از کلیدواژه مناسب برای تعیین موضوع در ارتباط با مفهوم مدرک خواهد بود. علاوه بر آن این رویکرد سبب گسترش انتخاب نمایه‌ساز شده و نیز استفاده از امکاناتی چون کلیدواژه‌های مترادف باعث انتخاب دقیق‌تری از نظر کلیدواژه‌های موضوعی می‌شود. بنابراین محدودیت موضوعی یکی از چالش‌های فعلی سیستم نمایه‌سازی است. ابزارهای کاربردی و زیرساخت مناسب از زمره عوامل تسریع‌کننده این فعالیت است. نتایج پژوهش کاتاگی و کومبار

(۲۰۲۳) در این عرصه با این یافته دارای هم‌سویی است. در این میان راهکارهای استفاده از نظام‌های تکمیلی فولکسونومی یا رده‌بندی‌های مردم نهاد در محیط وب می‌تواند نقش مطلوبی در تقویت ساختار موضوعی داشته باشد. نتایج پژوهش اینیوکو و آگونوبی (۲۰۱۶) موید این نکته در پژوهش حاضر است. علاوه بر آن، بهره‌گیری از نظام‌های رده‌بندی شناخته شده نیز می‌تواند در پیوند با ساختار موضوعی اصلی پایگاه، نقش اصلاحی و تقویتی داشته باشد. پژوهش لوند (۲۰۱۹) نیز تاییدکننده این مطلب است.

از زمره موارد دیگر در خصوص کاستی‌های کلیدواژه‌های موضوعی آن است که در زمان انتخاب کلیدواژه مرتبط، امکان استفاده از برخی موضوع‌های خرد در این فضا وجود ندارد. لذا همین مورد نمایه‌ساز را در شرایطی قرار می‌دهد که او ناگزیر است تا از یک حیطه کلی‌تر بهره گیرد تا امکان شناسایی موضوع وجود داشته باشد. همچنین بحث رعایت عامل ربط نیز در این میان حائز اهمیت است. ربط سبب ایجاد پیوند مناسب میان نیاز کاربر و نظام سازماندهی می‌شود. بهره‌گیری از این عامل در گرو استفاده عمیق‌تر از ابزارهای استاندارد مانند اصطلاح‌نامه در حین فرایند نمایه‌سازی است. در این رابطه نوع بهره‌گیری از ساختار موضوعی و زیرمجموعه‌های آن نقش مهمی در بازیابی اطلاعات دارد. پژوهش آوویورون و روزنفیلد (۲۰۲۱) و نیز جینگیرارد (۲۰۲۱) در این رابطه با نتایج مطالعه حاضر دارای اشتراک است. علاوه بر آن استفاده از رویکردهای خوشه‌بندی در سازماندهی اطلاعات یکی از راهکارهای مطلوب در رفع مشکلات مرتبط با پایین بودن میزان ربط کلیدواژگان با محتوای پارساهاست. در این رابطه، پژوهش سوچیاتی، فابرویر و پرده (۲۰۲۳) با یافته‌های پژوهش حاضر هم‌سویی دارد. در کنار آن استفاده از شبکه‌های استنادی در نمایه‌سازی هم می‌تواند فوایدی در بر داشته باشد. این شبکه‌های نظام ساختاری موضوعی را به صورت کلیدواژگان جدید و علمی پوشش می‌دهند. نتایج پژوهش لیو و که (۲۰۲۵) موید این نکته است.

یکی از نکات مهم استفاده از فناوری‌های موثر وب در باره سازماندهی منابع اطلاعاتی است. به این منظور انجام روش‌های سازماندهی اطلاعات خودکار، توانایی انجام فرایند تحلیل و پردازش اطلاعات را تسریع می‌کند. همچنین میزان دقت تخصیص کلیدواژگان مرتبط را نیز افزایش داده و شرایط بهتری برای بازیابی اطلاعات ایجاد می‌کند. در این زمینه برنامه‌ریزی برای ارتقای نظام سازماندهی در چارچوب خودکارسازی نمایه‌سازی و چکیده‌نویسی یکی از اقدام‌های موثر به شمار می‌آید. نتایج پژوهش گلوب، لیک و تودهوپ (۲۰۱۴)، یامان، تونسرو و ارتام (۲۰۲۱)، گلوب و همکاران (۲۰۱۶) و نیز کلیور و دیگران (۲۰۲۵) در زمینه استفاده از فرایندهای خودکارسازی سازماندهی با یافته‌های پژوهش حاضر دارای اشتراک است. همچنین استفاده از برخی فناوری‌هایی که کار طبقه‌بندی موضوعی را تسریع می‌کنند تا میزان توانایی پردازش دقیق و نیز هماهنگی با ساختار موضوعی روشمند را تقویت کنند، در این زمینه سودمند خواهد بود. یافته‌های پژوهش حیدر و شارما (۲۰۱۷) و آریسوکو و اوگبونا (۲۰۲۰) در باره بهره‌گیری از فناوری‌های طبقه‌بندی با نتایج حاصل شده در این پژوهش هم‌سو هستند.

استفاده همزمان از اصطلاح‌نامه‌ها و سرعنوان‌های موضوعی نقش عمیق‌تری در تخصیص کلیدواژه و حوزه موضوعی دارد که اکنون به صورت مستقیم این امکان وجود ندارد. به تعبیر دیگر استفاده همزمان از نظام پیش‌همارا و پس‌همارا ترکیب مناسبی را فراهم می‌سازد. در این میان تعیین روابط سطح بالا و پایین در ساختار نمایه‌سازی می‌تواند به تعیین کلیدواژه‌های مرتبط و بازیابی دقیق منجر شود. نتایج پژوهش ژنگ، ژیانو و بوو (۲۰۱۷) در این

زمینه با یافته‌های پژوهش حاضر دارای هم‌سویی است. همچنین در پاره‌ای از موارد بهره‌گیری از واژه‌نامه‌های تخصصی مکمل برای غنی‌سازی حوزه موضوعی حاوی فوایدی است که اکنون امکان‌پذیر نیست. این بدان دلیل است که در ساختار کنونی امکان افزودن این کلیدواژه‌ها و مفاهیم در سیستم در نظر گرفته نشده است. علاوه بر آن، انتخاب اصطلاح‌های مرجع در نظام اهمیت دارد. این امر در تحلیل واژگانی و موضوعی نقش مهمی دارد. به این معنا که با در نظر گرفتن نیاز کاربر و میزان ربط میان مفهوم و کلیدواژه تخصیص معمولاً اثربخشی بیشتری را شاهد هستیم. امکان بهره‌گیری از فنون مختلف چون خوشه‌بندی در ساختار موضوع، ارتباط میان اصطلاح اعم و اخص، اصطلاح وابسته و نیز نظام ارجاعی سبب تقویت میزان درستی تخصیص کلیدواژه به حوزه موضوعی می‌شود. پژوهش سوچیتی، فابروی و پرده (۲۰۲۳) موید این مطلب است.

بر اساس یافته‌های پژوهش پیشنهاد می‌شود:

با توجه به شرایط موجود، از نظر سطح اول مبتنی بر روش اجرایی وب آو ساینس می‌تواند ظرفیت بالایی برای سازماندهی اطلاعات داشته باشد. این امر با عنایت به پیشینه این فعالیت در ارتباط با منابع اطلاعاتی وب آو ساینس، کاربردپذیری مناسب را نشان می‌دهد. ولی در خصوص سطح دوم، باید اظهار کرد که این سطح محدودیت داشته و نیاز به توسعه بیشتر دارد.

پیشنهاد می‌شود موضوع‌های فرعی متناسب با هر یک از این موضوع‌های سطح دوم، از طریق اصطلاح‌نامه‌های تخصصی به نظام سازماندهی فعلی افزوده شود.

پیشنهاد می‌شود که سطح سومی را برای نرم‌افزار تعریف کرد و تمامی موضوع‌های مرتبط استاندارد از طریق اصطلاح‌نامه‌ها تعریف شده و مورد استفاده قرار گیرد.

نظام سازماندهی خوشه‌بندی موضوع‌های مرتبط با پارسا برای سیستم تعریف شود.

پیشنهاد می‌شود بانک اطلاعاتی فرعی از عبارت‌های جست‌وجوی کاربران تهیه شود. این بانک می‌تواند برای تحلیل نیاز اطلاعاتی کاربران و نیز تقویت نظام ارجاعی با توجه به ارائه به برنامه مورد بهره‌برداری قرار گیرد. طراحی یک نظام ارجاع خودکار در زمان کاوش اطلاعات توسط واژگانی که کاربران استفاده می‌کنند، برای ارجاع به کلیدواژه‌های تخصیص یافته به پارساها می‌تواند بازایی بهتر منابع را در پی داشته باشد.

همچنین پیشنهاد می‌شود با تشکیل یک پایگاه فرعی در نرم‌افزار ویرایش، در مواردی که کلیدواژه‌ای در میان کلیدواژه‌های اصطلاح‌نامه‌ای توسعه یافته وجود نداشت، از طریق واژگان موجود در واژه‌نامه‌های تخصصی و با توجه به کاربرد پاره‌ای از واژگان در متون اطلاعاتی فنی، حرفه‌ای و تخصصی، و یا نظر کارشناسان و متخصصان موضوعی، واژگان برابر و مرتبط تعریف و در این پایگاه وارد شود.

در زمان تخصیص کلیدواژه موضوعی توجه شود که از توصیفگرهای تخصصی استفاده شود.

در ختم‌شی سازماندهی اطلاعات و نمایه‌سازی میانگینی برای تعداد کلیدواژه‌ها در نظر گرفته شود تا توجه به موضوع‌های فرعی افزایش یافته و فرایند تخصیص کلیدواژه در نمایه‌سازی تقویت شود.

پیشنهاد می‌شود در مراحل بعدی برنامه‌ریزی برای نمایه‌سازی ماشینی و نیمه‌ماشینی انجام گیرد.

سپاسگزاری

پدیدآوران بر خود لازم می‌دانند از همکاران مدیریت سازماندهی و تحلیل اطلاعات در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) که در بررسی چالش‌های سازماندهی اطلاعات پارساها با ارائه نظرهای ارزشمندشان به بهبود یافته‌های پژوهش یاری رساندند، سپاسگزاری کنند.

References

- Ahuja, A., Devera, A., & Fox, E. A. (2022, November). Parsing electronic theses and dissertations using object detection. In *Proceedings of the first Workshop on Information Extraction from Scientific Publications* (pp. 121-130). DOI: [10.18653/v1/2022.wiesp-1.14](https://doi.org/10.18653/v1/2022.wiesp-1.14)
- Arisukwu, F. E. A., and I. M. Ogbonna. (2020). Information and Communication Technology Application in Cataloguing and Classification of Library Materials in Public University Libraries in Abia State, Nigeria: Information and Communication Technology Application in Cataloguing and Classification of Library Materials in Public University Libraries in Abia State, Nigeria. *International Journal of Knowledge Dissemination*, 1 (1). <http://ijkd.uniabuja.edu.ng/index.php/ijkd/article/view/11> (accessed January, 12, 2022)
- Aviv-Reuven, S., and, A. Rosenfeld (2021). Journal subject classification: intra-and inter-system discrepancies in Web Of Science and Scopus. *arXiv preprint arXiv:2107.12222*. DOI: [10.48550/arXiv.2107.12222](https://doi.org/10.48550/arXiv.2107.12222)
- Banerjee, B., Ingram, W. A., & Fox, E. A. (2024). Automating chapter-level classification for electronic theses and dissertations. In 2024 IEEE International Conference on Big Data (*BigData*) (pp. 2400-2409). IEEE. DOI: [10.1109/BigData62323.2024.10825418](https://doi.org/10.1109/BigData62323.2024.10825418)
- Bastos, F. M., and M. S. L. Fujita (2005). Subject-field classification of theses and dissertations in digital libraries. In *La dimensió humana de l'organització del coneixement* (pp. 77-90). Facultat de Biblioteconomia i Documentació.
- Datta, E. (2023). Attempts at automating journal subject classification. *Upstream*. DOI: <https://doi.org/10.54900/n6dnt-xpq48>
- Drabinski, E. (2019). What is critical about critical librarianship?. *Art Libraries Journal*, 44(2), 49-57. DOI: <https://doi.org/10.1017/alj.2019.3>
- Dutta, B., & Dutta, C. (2013). Concept of 'subject' in the context of library and information science from a new angle. *Annals of Library and Information Studies (ALIS)*, 60 (2), 78-87.
- George, S. (2022). Comparison of Full-Text Indexing with Metadata Indexing Based Subject Classification Using Graph-Based Index. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 8 (5), 86-91. DOI: <https://doi.org/10.32628/CSEIT228516>
- Golub, K., D. Soergel, G. Buchanan, D. Tudhope, M. Lykke, and D. Hiom. 2016. A framework for evaluating automatic indexing or classification in the context of retrieval. *Journal of the Association for Information Science and Technology*, 67 (1), 3-16. DOI: <https://doi.org/10.1002/asi.23600>
- Golub, K., M. Lykke, and D. Tudhope. 2014. Enhancing social tagging with automated keywords from the Dewey Decimal Classification. *Journal of Documentation*. 70 (5), 801-828. DOI: <https://doi.org/10.1108/JD-05-2013-0056>
- Haider, S., and R. K. Sharma (2017). Library of Congress Classification (LCC): past, present and its future in the digital era. *Annals of Library and Information Studies (ALIS)*, 64 (3), 190-201.
- Howard, S. A., & Knowlton, S. A. (2018). Browsing through bias: The Library of Congress classification and subject headings for African American studies and LGBTQIA studies. *Library Trends*, 67(1), 74-88. DOI: <https://doi.org/10.1353/lib.2018.0026>

- Inyokwe, F. E., and J. Agwunobi (2016). Internet and library classification as determinants of students utilisation of information resources in University of Calabar Library. *Information Technologist (The)*, 13(1), 171-181
- Jeangirard, E. .2021. Content-based subject classification at article level in biomedical context. *arXiv preprint arXiv:2104.14800*. 1, 1-13. DOI: <https://doi.org/10.48550/arXiv.2104.14800>
- Kandimalla, B. (2020). *Subject Category Classification of Scholarly Papers Using Deep Attentive Neural Networks*. (Master of Science dissertation, The Pennsylvania State University).
- Katagi, S., & Kumbar, B. D. (2023). Access to electronic theses and dissertations: literature review with global perspective. *Journal of Advances in Library and Information Science*, 12 (3), 132-142. DOI: [10.5281/zenodo.8267818](https://doi.org/10.5281/zenodo.8267818)
- Kazi, N., Lane, N., and I. Kahanda. (2021). Automatically Cataloging Scholarly Articles using Library of Congress Subject Headings. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop* (pp. 43-49). DOI: <https://doi.org/10.18653/v1/2021.eacl-srw.7>
- Klair, H. (2025). Enhancing Digital Libraries as Communication Tools: LLMs for Automated Subject Classification of Electronic Theses and Dissertations. In *Proceedings of the 43rd ACM International Conference on Design of Communication* (pp. 243-245). DOI: <https://doi.org/10.1145/3711670.3764654>
- Klair, H., German, F., Aboelnaga, A. A., Banerjee, B., Eldardiry, H., & Ingram, W. A. (2025A). Retrieval-Augmented LLMs for ETD Subject Classification. In *2025 IEEE International Conference on Big Data (BigData)* (pp. 5995-6004). IEEE. DOI: [10.1109/BigData66926.2025.11402222](https://doi.org/10.1109/BigData66926.2025.11402222)
- Klair, H., German, F., Banerjee, B., & Ingram, W. A. (2025B). Evaluating Human-LLM Alignment in ETD Subject Classification. In *International Conference on Theory and Practice of Digital Libraries* (pp. 57-69). Cham: Springer Nature Switzerland. DOI: https://doi.org/10.1007/978-3-032-06136-2_6
- Lovitts, B. E. (2023). *Making the implicit explicit: Creating performance expectations for the dissertation*. Routledge.
- Lund, B. D. (2019). A Comparison of Cataloger and Non-Cataloger Responses to a Survey of Perceptions and Preference toward Dewey Decimal and Library of Congress Classification. *Kansas Library Association College and University Libraries Section Proceedings*, 9 (1), 1-8. DOI: <https://doi.org/10.4148/2160-942X.1072>
- Lund, B., D. Agbaji, S. Tijani, and I. Oname (2019). Evaluating Knowledge Organization in Developed and Developing Countries: A Comparative Analysis of Dewey Decimal and Library of Congress Classification Scheme Preference and use in the United States and Nigeria. *Technical Services Quarterly*, 36 (3), 249-268. DOI: <https://doi.org/10.1080/07317131.2019.1621563>
- Lyu, Z., & Ke, Q. (2025). Periodical embeddings uncover hidden interdisciplinary patterns in the subject classification scheme of science. *arXiv preprint arXiv:2512.22524*. DOI: <https://doi.org/10.48550/arXiv.2512.22524>
- Manzoor, J. A. (2023). *Segmenting electronic theses and dissertations by chapters* (Doctoral dissertation, Virginia Tech).
- O'Leary, D. E. (2023). Using large language models to write theses and dissertations. *Intelligent Systems in Accounting, Finance and Management*, 30 (4), 228-234. DOI: <https://doi.org/10.1002/isaf.1547>
- Phiri, L. (2020). Automatic classification of digital objects for improved metadata quality of electronic theses and dissertations in institutional repositories. *International Journal of Metadata, Semantics and Ontologies*, 14 (3), 234-248. DOI: <https://doi.org/10.1504/IJMSO.2020.112804>
- Posigha, B. E., Ojohwoh, R., & Oberhiri-Oruma, G. (2019). The Challenges and Prospects of Cataloguing and Classification in Academic Libraries in Bayelsa State, Nigeria. *Asian Journal of Information Science & Technology (AJIST)*, 9 (3), 14-18. DOI: [10.51983/ajist-2019.9.3.294](https://doi.org/10.51983/ajist-2019.9.3.294)
- Roostae, M., Sadreddini, M. H., & Fakhrahmad, S. M. (2020). An effective approach to candidate retrieval for cross-language plagiarism detection: A fusion of conceptual and keyword-based

- schemes. *Information Processing & Management*, 57 (2), 102150. DOI: <https://doi.org/10.1016/j.ipm.2019.102150>
- Singh, P., Piryani, R., Singh, V. K., & Pinto, D. (2020). Revisiting subject classification in academic databases: A comparison of the classification accuracy of Web of Science, Scopus & Dimensions. *Journal of Intelligent & Fuzzy Systems*, 39 (2), 2471-2476. DOI: <https://doi.org/10.3233/JIFS-179906>
- Suciati, N., Fabroyir, H., & Pardede, E. (2023). Educational data mining clustering approach: case study of undergraduate student thesis topic. *IEEE Access*, 11, 130072-130088. DOI: [10.1109/ACCESS.2023.3332818](https://doi.org/10.1109/ACCESS.2023.3332818)
- Wang, H. (2021). Subject Exploration in the Field of Library and Information from the Perspective of the Theory of Social Cognition. *International Journal of Frontiers in Engineering Technology*, 3 (10), 78-83. DOI: [10.25236/IJFET.2021.031010](https://doi.org/10.25236/IJFET.2021.031010)
- Wheatley, A. (2024). Subject trees on the Internet: A new role for bibliographic classification?. In *Internet Searching and Indexing* (pp. 115-141). CRC Press.
- Yaman, O., Tuncer, T., and Ertam, F. .2021. Automated book location and classification method using RFID tags for smart libraries. *Microprocessors and Microsystems*, 87, 104388. DOI: <https://doi.org/10.1016/j.micpro.2021.104388>
- Zhang, H., Xiao, Y., & Bu, Z. (2017, November). Personalized book recommender system based on chinese library classification. In *2017 14th Web Information Systems and Applications Conference (WISA)* (pp. 127-131). IEEE. DOI: [10.1109/WISA.2017.42](https://doi.org/10.1109/WISA.2017.42)