شناسایی خلاءهای نمایه‌سازی موضوعی پایان‌نامه‌ها و رساله‌های پایگاه گنج و راهکارهای بهبود آن

فریبرز درودی[1*]
استادیار، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، تهران، ایران

مدیریت اطلاعات
دوره ۱۱ شماره ۲
پاییز و زمستان ۱۴۰۴

چکیده: هدف این مطالعه، تحلیل خلاءهای موضوعی پایان‌نامه‌ها و رساله‌های موجود در پایگاه اطلاعات علمی ایران (گنج) و ارائه راهکارهایی برای سازماندهی مؤثر این منابع اطلاعاتی است. روش پژوهش کیفی است و با استفاده از یک شیوهٔ مقایسه‌ای، خلاءهای موضوعی در نمایه‌سازی پایان‌نامه‌ها و رساله‌ها (پارساها) را بر اساس معانی اصلی آن‌ها شناسایی می‌کند. جامعه آماری شامل ۱,۱۱۲,۵۳۱ پایان‌نامه و رساله (پارسا) در پایگاه اطلاعات علمی ایران (گنج) بود. از روش نمونه‌گیری تصادفی-طبقه‌ای با حجم نمونه ۱۵۰۰ سند برای تجزیه و تحلیل داده‌ها استفاده شد. داده‌ها با استفاده از یک چک لیست حاوی ۲۱ مؤلفه مربوط به ارزیابی و تحلیل حوزه موضوعی جمع‌آوری شد. برای تقویت داده‌های به‌دست‌آمده، مصاحبه‌هایی با ۴ متخصص سازماندهی اطلاعات در ایرانداک انجام شد. روایی ابزار جمع‌آوری اطلاعات با استفاده از روش اعتبار صوری و با نظر ۵ تن از متخصصان دانشگاهی در علم اطلاعات ارزیابی شد. پایایی سیاههٔ وارسی با استفاده از روش آماری آلفای کرونباخ تعیین شد و مقدار ۰.۸۷۲ به دست آمد. تجزیه و تحلیل یافته‌های پژوهش، چندین خلاء در سازماندهی موضوعی را نشان داد، از جمله مسائل مربوط به زیرشاخه‌های موضوعی، عمق نمایه‌سازی، برچسب‌های موضوعی، تطبیق مفهوم، ترجیح موضوعات بر رشته‌های دانشگاهی، توسعه طبقه‌بندی موضوعی، طبقه‌بندی روزآمد، ساختار طبقه‌بندی، مدیریت زمان و پردازش زبان طبیعی. این مطالعه کاستی‌هایی را در ساختار نمایه‌سازی و استفاده از کلمات کلیدی نشان داد. مشخص شد که رابطه بین سطوح ساختار طبقه‌بندی موضوعی به‌طور قابل توجهی بر عمق نمایه‌سازی تأثیر می‌گذارد.همچنین، چالش‌هایی مانند تطبیق مفهوم-موضوع و شناسایی دقیق محتوا مشخص شد. تشخیص روابط بین موضوعات و منابع بین رشته‌ای به سازماندهی منابع کمک می‌کند. توسعه و گسترش طبقه‌بندی‌های موضوعی برای گنجاندن کلمات کلیدی مرتبط، بسیار مهم است، زیرا موضوعات جدیدی با پیشرفت‌های علمی پدیدار می‌شوند. علاوه برآن، نتایج نشان داد که ساختار طبقه‌بندی موضوعی نقش کلیدی در تعیین کلیدواژه‌های مناسب دارد. نتایج نشان داد که توسعه موضوع‌های فرعی در زیر موضوع‌های کلان، سازماندهی اطلاعات را بهبود می‌بخشند و روابط بهتر کلمات کلیدی و هم‌ترازی مفاهیم را تسهیل می‌کنند. مشخص شد که عمق نمایه‌سازی در سطوح موضوعی حیاتی است و نظام‌های واژگان سلسله‌مراتبی از توسعه طبقه‌بندی موضوعی پشتیبانی می‌کنند. برچسب‌گذاری صحیح موضوع، بازیابی پایان‌نامه را بهبود می‌بخشد. از جمله چالش‌های سازماندهی موضوعی، شناسایی دقیق محتوای اطلاعات است. تطبیق حوزه‌های موضوعی با پیشرفت علمی نقش مهمی داشته و طبقه‌بندی علمی بر اساس روابط میان کلیدواژه‌ها می‌تواند اثربخشی بازیابی را افزایش دهد.

کلیدواژه‌ها: پایان‌نامه‌ها و رساله‌ها، پایگاه اطلاعات علمی ایران (گنج)، خلاء موضوعی، طبقه‌بندی موضوعی، نمایه‌سازی موضوعی

[1] نویسنده مسئول: doroudi@irandoc.ac.ir

مقدمه

پایان‌نامه یکی از فعالیت‌های ضروری و نیز تاثیرگذار اجزای یک دوره درسی است که اگر به درستی انجام شود، نه تنها می‌تواند کمک ارزشمندی به توسعه یادگیری خود دانشجویان کند، بلکه در کل رشته اثربخش باشد (Borden & Rüedi, 2006). در واقع، می‌توان این منبع را یک کار علمی همراه با علاقه دانست که نیازمند فعالیت، تلاش و همچنین مهارت‌های بنیادی و استفاده از منابع گسترده از سوی کسانی است که درگیر این فرآیند هستند (Grant & Osanloo, 2014). باید اظهار کرد که پایان‌نامه‌ها و رساله‌های (پارساهای) الکترونیکی بخش مهمی از سابقهٔ علمی در گسترهٔ جهانی هستند (Klair et al., 2025) و منطق گنجاندن این منبع به عنوان جزء اصلی دوره‌های کارشناسی‌ارشد و دکتری این است که توسعه مهارت‌های فکری، فراهم شود (Naoum, 2013). بنابراین، پایان‌نامه به‌عنوان یک نگارش پیچیده مطرح می‌شود که مطالبات زیادی را از دانشجویان و استادان آنها می‌طلبد (Paré, 2011) و خدمات اطلاع‌رسانی مهم نظیر خلاصه‌سازی اسناد طولانی، تحلیل معنایی، طبقه‌بندی و فهرست‌نویسی برای بازیابی از زمره توانایی‌هایی است که در ساختار منابع پایان‌نامه‌ای و رساله‌های الکترونیکی قابل استفاده است (Manzoor, 2023). یکی از مسائل مهم در زمینه بهره‌گیری از منابع اطلاعاتی پارساها، روش‌های طبقه‌بندی و نمایه‌سازی آن‌هاست که نقش مهمی در بازیابی و استفاده سودمند از این منابع تخصصی دارد. انواع رویکردهای سازماندهی در دسترس‌پذیری بهتر و مناسب پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی حائز اهمیت است که در ادامه به برخی از این روش‌های اساسی اشاره می‌شود.

روش‌های آرشیو و سازماندهی سنتی برای توصیف پایان‌نامه‌ها و رساله‌های (پارساهای) الکترونیکی به طرح‌های فراداده‌ای گسترده و سطح بالا متکی هستند (Banerjee et al., 2024) و طبقه‌بندی موضوعی دقیق پارساهای الکترونیکی در اثربخشی آنها نقش بنیادی دارد (Klair, 2025). در طبقه‌بندی موضوعی منابع اطلاعاتی چندین راهکار وجود دارد که توسط پژوهشگران توضیح داده و تحلیل شده است. نظریه ریزساختاری[1] یکی از آنهاست که به رابطه بین مفاهیم در پیکرهٔ نظری دانش می‌پردازد (Bastos & Fujita, 2005). دیگری، درخت‌های موضوعی (سلسله‌مراتب) است (Wheatley, 2024). بعدی، نظریه سلسله‌مراتب شناخت اجتماعی است که تشخیص موضوع بخش مهمی از پژوهش است (Wang, 2021). همچنین، بهره‌گیری از رده‌های موضوعی به منظور در اختیارداشتن ویژگی‌های جست‌وجو چندوجهی از دیگر موارد به شمار می‌آید (Kandimalla, 2020). علاوه بر آن، طبقه‌بندی موضوعی بر مبنای سطوح مختلف تخصصی (Datta, 2023) و بهره‌گیری از فراداده‌های منابع اطلاعاتی مانند: عنوان، چکیده و کلیدواژه از زمره روش‌های کاربردی در سازماندهی منابع پارساها به شمار می‌آید (George, 2022). بر اساس رویکردهای مطرح شده باید بیان کرد که در این پژوهش به روش‌های سازماندهی بر مبنای رده‌ها و نمایه‌سازی موضوعی و بهره‌گیری از کلیدواژه‌ها توجه شده است. این رویکرد سبب تاثیرگذاری در شناخت

[1] Microstructural theory

بهتر مشکلات نمایه‌سازی پارساها از منظر کاربردی می‌شود. علاه بر آن ابعاد دیگری در این میان اهمیت دارد که به آنها اشاره می‌شود.

از سویی، مدل‌های بازیابی اطلاعات، مبتنی بر مفهوم و کلیدواژه است که در بازیابی اطلاعات اهمیت دارد (Roostaee et al., 2020) و از سوی دیگر، این فرایند با طبقه‌بندی خودکار تقویت می‌شود. از این رو، نمایه‌سازی موضوعی خودکار نقش مهمی در سازماندهی دانش و بازیابی اطلاعات ایفا می‌کند (Xu et al., 2025) و نیز این نوع از نمایه‌سازی به طور فزاینده‌ای با استفاده از مدل‌های زبانی بزرگ[1] پیاده‌سازی می‌شود (Junzhi & Rundong, 2025). در کنار آن، شیوه‌هایی که برای نمایه‌سازی موضوعی خودکار مطرح می‌شود، شامل رویکردهای آماری، واژگانی و تلفیقی برای طبقه‌بندی [مفاهیم] است (Dermentzi et al., 2025). نمایه‌سازی خودکار نقش مهمی در پر کردن شکاف بین تولیدکنندگان اطلاعات، ناشران محتوای فکری و پژوهشگرانی که به دنبال دسترسی به آن هستند، ایفا می‌کند (Dawi, 2024). بر همین اساس، از آنجا که روش سنتی طبقه‌بندی [منابع] از طریق شیوهٔ دستی، زمان‌بر و پرزحمت است، برخی مطالعات، روش‌های یادگیری ماشینی را معرفی کرده‌اند (Ouyang et al. , 2022). بر این اساس توجه به ساختار طبقه‌بندی موضوعی خودکار و بهره‌گیری از ابزارهای مرتبط با آن، یکی از ابعاد مهم سازماندهی منابع اطلاعاتی پارساهاست که شایسته توجه بیشتر است. این مقوله با مشکلات سازماندهی ارتباط داشته و ضرورت دارد تا به ابعاد و جوانب آن بیشتر توجه شود.

پایگاه اطلاعات علمی ایران (گنج) که مخزن اصلی سازماندهی و اشاعهٔ پارساهای کشور محسوب می‌شود، یک پایگاه مهم برای انجام پژوهش‌های دانشگاهی است که حجم عظیمی از منابع پایان‌نامه‌ای و رساله‌های تحصیلات تکمیلی را پوشش می‌دهد. با توجه به استفاده گسترده‌ای که استادان، دانشجویان، پژوهشگران و متخصصان در بهره‌گیری از پایگاه یاد شده انجام می‌دهند، درستی نمایه‌سازی، سازماندهی و مدیریت اطلاعات علمی در این مرجع علمی نقش مهمی دارد. بر این اساس، شناسایی مشکلات سازماندهی اطلاعات در اینجا می‌تواند، به رفع موانع این حوزه یاری رساند و میزان دسترس‌پذیری و بازیابی اطلاعات مرتبط را افزایش دهد. با توجه به مطالعات انجام گرفته در زمینه مورد نظر، مطالعات محدودی به جنبه‌های فنی و تخصصی نمایه‌سازی موضوعی در پایگاه‌های اطلاعاتی مرتبط با پارساها پرداخته‌اند و از همین رو، در پژوهش حاضر خلاءهای موضوعی در سازماندهی اطلاعات پایان‌نامه‌ها و رساله‌های (پارساهای) تحصیلات تکمیلی که در پایگاه اطلاعات علمی ایران (گنج) سازماندهی می‌شوند، به منظور آماده‌ساختن این منابع برای ایجاد شرایط مناسب در راستای خودکارسازی سازماندهی منابع اطلاعاتی بررسی و تحلیل شده است. منظور از خلاءهای موضوعی در پژوهش حاضر، کاستی‌های کلیدواژه‌ها و اصطلاح‌های به کار رفته در سازماندهی مدارک به منظور انجام فرایند نمایه‌سازی پارساها بر اساس روش‌های استاندارد نمایه‌سازی موضوعی و تطبیق آن با اصطلاح‌نامه و نیز واژگان تخصصی متناسب با مفاهیم حرفه‌ای است. باید اظهار کرد که نوآوری این مطالعه در شناخت عملیاتی خلاءهای موجود در ساختار موضوعی یک پایگاه ملی و فراهم آوردن چارچوبی برای طراحی نظام خودکار سازماندهی پارساهاست.

[1] Large Language Model (LLM)

همچنین باید بر آن افزود که در مطالعه حاضر، بررسی خلاءهای موضوعی نمایه‌سازی از منظر ساختار و ویژگی‌های تخصیص کلیدواژه منطبق با مفهوم بنیادی پارساها تحلیل شده و بر اساس آن، حوزه‌های موضوعی مورد مطالعه عبارت است از موضوع‌های کلانی که در سامانه برای تفکیک طبقه‌بندی موضوعی پارساها در نظر گرفته شده است. از جمله مصادیق آن می‌توان به علوم اجتماعی، علوم و علوم کاربردی؛ و نیز هنر و علوم انسانی اشاره کرد که در حال حاضر در سامانه سازماندهی به کار گرفته شده است. البته امکان توسعه آن به حوزه‌هایی چون کشاورزی، علوم پزشکی، علوم انسانی، فنی و مهندسی، علوم پایه و نظایر آن نیز وجود دارد. از این‌رو، سازماندهی با سه سطح مطرح شده در کنار دیگر موضوع‌ها که به صورت کلیدواژه‌های مفهومی به منابع اطلاعاتی اختصاص می‌یابد، در فرایند پردازش و سازماندهی اطلاعات قرار می‌گیرد. هدف پژوهش حاضر، مشخص ساختن خلاء موجود در حوزۀ موضوعی استفاده شده در این پایگاه برای ایجاد شرایط مناسب در مسیر خودکارسازی سازماندهی منابع اطلاعاتی است. پرسش اصلی پژوهش نیز عبارت است از: خلاءهای موضوعی به منظور نمایه‌سازی پایان‌نامه‌ها و رساله‌های (پارساها) تحصیلات تکمیلی در پایگاه اطلاعات علمی ایران (گنج) با توجه به ساختار و کارکرد حوزه سازماندهی اطلاعات کدامند؟

پیشینه پژوهش

پایان‌نامه‌های کارشناسی ارشد و رساله‌های دکتری، به عنوان محصول علمی و در نقش فرآیند نگارش پژوهشی مطرح می‌شوند (Thompson, 2025) و برای سازماندهی آنها از طرح‌های طبقه‌بندی موضوعی بهره گرفته می‌شود (Lyu & Ke, 2025). علاوه بر آن، امکان اشتراک‌گذاری شفاف اطلاعات را فراهم می‌کنند (Klair, 2025). در حالی که نظام‌های تخصصی [استنادی] مانند اسکوپوس[1] طبقه‌بندی‌های پرکاربردی را ارائه می‌دهند، اغلب از جزئیات، ذهنیت‌گرایی و سازگاری محدود با حوزه‌های میان‌رشته‌ای نوظهور رنج می‌برند (Lyu & Ke, 2025) و از همین رو، طبقه‌بندی‌های ارائه شده توسط نویسندگان اغلب متناقض یا نادرست هستند و معمولا نیاز به بررسی متخصصان دارند و نیز جست‌وجو و گزارش‌دهی را پیچیده می‌کنند (Klair, 2025). به همین دلیل، بازیابی مرتبط‌ترین [منابع] بر اساس درخواست کاربران هنوز یک امر انتزاعی است. دلیل اصلی آن این است که [منابع] به طور مناسب بر اساس سلسله مراتب طبقه‌بندی موضوعی جزئی نمایه نمی‌شوند (Sajid et al., 2023). بر این اساس راهکارهای متفاوتی ارائه شده که می‌توان بیان کرد، بازیابی اطلاعات اینترنتی عمدتاً به حفظ موتورهای جست‌وجو و حتی درخت‌های موضوعی محبوب‌تر مربوط می‌شود (Wheatley, 2024) و علاوه بر آن، طبقه‌بندی بین‌موضوعی از ارزش عملی بالایی در پایش ذهنی برخوردار است (Sun & Li, 2024). تمامی فعالیت‌های تخصصی طبقه‌بندی موضوعی برای ایجاد شرایط بهتر در بازیابی اطلاعات و دستیابی به این منابع در نظر گرفته شده است. با این رویکرد، در ادامه پاره‌ای از پژوهش‌های مهم در زمینه طبقه‌بندی موضوعی، نمایه‌سازی و بازیابی اطلاعات معرفی می‌شود.

---

[1] Scopus

پژوهش «فتاحی» (۱۳۸۶) مشخص ساخت که چالش‌های مهم مورد بررسی در سازماندهی اطلاعات عبارتند از افزونگی منابع اطلاعاتی در شکل‌های گوناگون، دشواری در تعیین مبنای توصیف منابع، محدودیت سازماندهی در مبدأ تولید منابع، دشواری کنترل مستند نام‌ها و موضوع‌ها در سطح ملی و جهانی، تضاد میان شیوه‌های پیش‌همارایی و پس‌همارایی، کاستی‌های فهرست‌های رایانه‌ای در بازیابی و نمایش اطلاعات و تضاد میان فهرستنویسی و نمایه‌سازی در موتورهای کاوش. یافته‌های پژوهش «ناخدا» و «شرفی» (۱۳۹۸) نشان داد که مهم‌ترین عوامل ایجاد آشوب در سازماندهی اطلاعات شامل: ۱. عوامل انسانی ۲. عوامل سیستمی ۳. عوامل بنیادی ۴. عوامل اقتصادی و ۵. عامل زبان‌شناختی است. همچنین یافته‌ها نشان داد که اثر پروانه‌ای، خودمانایی، جاذبه‌های غریب و سازگاری پویا به ترتیب از مهم‌ترین عوامل بازدارندهٔ آشوب در سازماندهی اطلاعات هستند و سازماندهی‌کنندگان با داشتن تخصص، دقت و زمان کافی، عملکرد بهتری را در جلوگیری از آشوب خواهند داشت. یافته‌های پژوهش «نوروزی» و «جعفری‌فر» (۱۳۹۹) نشان داد که کتابخانه‌های دیجیتالی در بخش سازماندهی منابع خود در هفت حوزهٔ توصیف منابع، تعیین ساختار منابع، مدیریت منابع، ساماندهی محتوای منابع، تعیین قالب محلی در نظام مدیریتی پایگاه داده کتابخانه دیجیتالی، تعیین بستر معنایی و مبادله از استانداردهای فراداده‌ای و پروتکل‌ها استفاده می‌کنند. همچنین مشخص شد که میزان توجه به استانداردهای حوزه سازماندهی منابع چندرسانه‌ای و غیرمتنی در متون علمی پایین است و می‌تواند موجب مهجور ماندن سازماندهی در این حوزه باشد. نتایج پژوهش «زره‌ساز» (۱۴۰۰) نشان داد که در حوزه سازماندهی اطلاعات چالش‌های مضمون «موارد آموزشی»، بیشتر به ضعف آموزش‌های روزآمد به حرفه‌مندان و آموزش‌های دانشگاهی مربوط می‌شود. در رابطه با مضمون «موارد پژوهشی»، چالش‌هایی همچون کافی نبودن متخصصان و نبود پیوستگی لازم میان پژوهش‌های این موضوع مطرح شده است. در رابطه با مضمون «موارد حرفه‌ای»، عمده‌ترین چالش‌ها به مشکلات نرم‌افزارهای تخصصی، روزآمدنبودن حرفه‌مندان و بی‌توجهی به نیازهای کاربران مربوط می‌شوند. در رابطه با مضمون «موارد راهبردی»، بیشترین چالش‌ها به ضعف نهادهای تأثیرگذار و نبود همکاری‌های لازم ملی و بین‌المللی برمی‌گردد. نتایج پژوهش «شادانپور» و همکاران (۱۴۰۱) نشان داد که تعدد ابزارهای تحلیل موضوعی و انتساب کلیدواژه‌ها، میزان پایین کاربرد طرح‌های استاندارد توصیف، عدم همگونی در تحلیل موضوعی، عدم همخوانی عبارت‌های جستجوی کاربران با کلیدواژه‌های انتساب داده شده، و روزآمد نبودن توصیف منابع از نقاط ضعف رویکرد تحلیل موضوعی توسط انسان است. روش‌های خودکار هوش مصنوعی و پردازش زبان طبیعی توان افزایش سرعت و یکدستی در فرایندهای مختلف توصیف و سازماندهی اطلاعات و از جمله استخراج کلیدواژه های موضوعی را دارند، اما ارزیابی کارآمدی خروجی این روش‌ها، به تنهایی، یا در مقایسه با کلیدواژه‌های تولید انسان، و از دیدگاه‌های مختلف ضروری است. پژوهش «طاهری» (۱۴۰۴) نشان داد که چالش‌های مهم روزآمدسازی و پیاده‌سازی سازماندهی دانش در کشور موارد متعددی هستند. چالش‌هایی که از ضعف در به‌روزرسانی دانش و مهارت‌های تخصصی، مقاومت در برابر فناوری‌های نو، کاستی در اجرای استانداردهای پیشین، فقدان سیاست‌ها و اسناد بالادستی منسجم، نارسایی در بومی‌سازی معناشناختی، نگرش تقلیل‌گرایانه به نقش سازماندهی دانش، تا ضعف زیرساخت‌های فناورانه و نرم‌افزاری

را دربرمی‌گیرد. تداوم این وضعیت، مانع ایفای نقش راهبردی سازماندهی دانش در حکمرانی داده، تصمیم‌سازی کلان و توسعهٔ زیست‌بوم داده‌ای کشور خواهد شد.

پاره‌ای از پژوهش در بارهٔ ساختار رده‌بندی‌های علمی به انجام رسیده که حاوی نتایج سودمندی دربارهٔ کاربرد، فرایند و روش‌های این گونهٔ مفید طبقه‌بندی علمی است. از جمله مطالعات مربوط به طبقه‌بندی موضوع پایان‌نامه بر اساس ساختار تخصصی رده‌بندی که در این زمینه به ابعاد فنی و روش‌های موثر نمایه‌سازی و بهره‌گیری از اصطلاح‌نامه‌ها و کلیدواژه‌های تخصصی اشاره شده است. این گونه از پژوهش‌ها با بررسی جنبه‌های فرایندی و شیوه‌های اختصاص واژگان و حفظ رابطهٔ ارگانیک میان متن منبع و برچسب موضوعی اختصاص یافته، فرایند اجرایی آن را نیز تحلیل می‌کند (Howard & Knowlton, 2018; Hadi, 2022; Fujita et al., 2022)؛ از سوی دیگر برخی از مطالعات به طبقه‌بندی‌های کتابخانه‌ای و کاربردهای حرفه‌ای آن در ارتباط با تبیین موضوعات خاص اشاره دارند. این پژوهش‌ها جنبه‌هایی از سیستم‌های طبقه‌بندی موضوعی را در ارتباط با کارکردهای حرفه‌ای در میان منابع اطلاعاتی و نقشی که در بازیابی اطلاعات ایفا می‌کنند را مورد بررسی و توجه قرار داده‌اند (Haider & Sharma, 2017; Avila & Guimarã, 2013; Lund, 2019; Lund et al., 2019)؛ همچنین، مطالعات مرتبط با چالش‌های محاسباتی پردازش پارساهای الکترونیکی بخش دیگری از پژوهش‌های این عرصه به شمار می‌روند. این گونه از مطالعات با مطرح ساختن مشکلات اساسی و بنیادی در خصوص روش‌های سازماندهی پارساهای الکترونیکی در فضای مجازی و بهره‌گیری از ابزارهای هوشمند با تکیه بر انواع الگوریتم‌ها به سازه‌های مهم فرایندهای سازماندهی پرداخته‌اند و پیشنهادهای کاربردی برای بهبود نظام نمایه‌سازی خودکار و رفع موانع فنی مطرح کرده‌اند (Ahuja, 2023; Katagi & Kumbar, 2023;.Klair et al, 2025). از سوی دیگر برخی از مطالعات به مقولهٔ بهره‌گیری از واژگان کنترل سلسله‌مراتبی در قالب طبقه‌بندی‌های موضوعی توجه کرده‌اند. این دسته از مطالعات با بررسی دقیق و عمیق جایگاه واژگان تخصصی در نظام‌های درختی و بررسی روابط میان آنها در کنار کاربردهای حرفه‌ای، به زمینه‌های عملیاتی آنها اشاره دارند. در این میان تحلیل جایگاه مترادف‌ها و متضادها، گروه‌بندی‌ها، روابط معنایی، ارجاع‌ها و دیگر مباحث مرتبط به بررسی ویژگی‌های آنها پرداخته و راهکارهای اجرایی را معرفی می‌کنند (Mitrou et al., 2025; Holstrom, 2023). پاره‌ای از پژوهش‌ها نیز به حیطهٔ مطالعاتی برچسب‌گذاری خودکار موضوعی، سازماندهی مفهومی و پردازش متون پارساها مبتنی بر مدل‌های زبانی بزرگ پرداخته‌اند. در این گونه از تحقیقات، مسائل مهمی از جمله فنون موثر خلاصه‌سازی متن، انجام ترجمه ماشینی، معادل‌گذاری مفهومی، طبقه‌بندی متن و ربط واژگانی مورد تحلیل قرار گرفته و شیوه‌های مدیریت اطلاعات علمی برای انجام فرایندهای سازماندهی در قالب نمایه‌سازی، طبقه‌بندی موضوعی و دیگر روش‌های سودمند معرفی شده است (Jeangirard, 2021; O'Leary, 2023; Kluge & Kähler, 2025; Junzhi & Rundong, 2025).

از دیگر حوزه‌های مرتبط می‌توان به مطالعات در بارهٔ طبقه‌بندی موضوعی خودکار، بر اساس الگوهای از پیش تعیین‌شده اشاره کرد. این مطالعات به جنبه‌های مهمی از سازماندهی اطلاعات پرداخته‌اند که با روش‌های خودکار و ماشینی قرابت داشته و از اصطلاحات و واژگان تخصصی در راستای سازماندهی متون

و نمایه‌سازی حرفه‌ای بهره می‌گیرند (Artene et al., 2020; Kazi et al., 2021, Schumann & Mergel, 2025). علاوه بر آن، مطالعاتی نیز به موضوع رده‌بندی علمی برای سازماندهی و بازیابی منابع اطلاعاتی مبتنی بر طبقه‌بندی و سنجش روابط کلیدواژگان و نیز چالش‌های فرایندها و روش‌های پردازش موضوع پرداخته‌اند که با ابزارهای سازماندهی و بررسی جایگاه و نیز وضعیت واژگانی رابطه دارند (Zhang et al., 2017; Posigha et al., 2019; Katagi & Kumbar, 2023). همچنین پاره‌ای دیگر از تحقیقات به پژوهش‌های انجام شده در موضوع رده‌بندی منابع اطلاعاتی در محیط وب و بر مبنای ارزیابی نظام‌های طبقه‌بندی اختصاص یافته است. این گونه از مطالعات با رتبه‌بندی مفاهیم، تنظیم کلیدواژه‌های مفهومی و اختصاص اصطلاح‌های تخصصی برای منابع اطلاعاتی رابطه داشته و راهکارهای اجرایی فرایندهای آن را بررسی و تحلیل می‌کند. همچنین به جنبه‌های فنی موضوع نیز اشاره دارند (Aviv-Reuven & Rosenfeld, 2023; Suciati et al., 2023). بخشی دیگری از مطالعات نیز در بارهٔ ساختار رده‌بندی‌های خودکار و نیز کاربرد ابزارها، برنامه‌ها و فناوری‌های اطلاعاتی است که به مقایسه، تحلیل و ارزیابی منابع اطلاعاتی و کاربرد استانداردهای این حوزه می‌پردازند. در این رابطه جنبه‌های مهمی از استانداردهای فراداده‌ای تحلیل شده و کاربرد فناوری‌های هوشمند معرفی شده است (Arisukwu & Ogbonna. 2020; Yaman et al., 2021; O'Leary, 2023).

نتایج مطالعات انجام شده نشان می‌دهد که توجه به طبقه‌بندی موضوعی و نیز خودکارسازی روش‌های سازماندهی اطلاعات در محیط وب، پارساها و منابع اطلاعاتی قابل ملاحظه است. در این رابطه، فعالیت‌های پژوهشی متعددی در ابعاد مختلف به انجام رسیده است. از همین‌رو، حوزه‌های مطالعاتی شامل موضوع‌های مهم و اثربخش در این عرصه به زمینه‌های شکل‌گیری ساختار سازماندهی و خودکارسازی توجه جدی داشته‌اند. مطالعات مهمی که در پژوهش‌های یاد شده به انجام رسیده‌اند، بیشتر به ساختار تخصصی رده‌بندی، طبقه‌بندی‌های کتابخانه‌ای، چالش‌های پردازش اطلاعات، و نیز به ابعاد تخصصی آن پرداخته‌اند. همچنین برخی از جنبه‌های خاص نیز در این میان مورد توجه قرار گرفته است. موضوع‌های کاربردی چون واژگان کنترل سلسله‌مراتبی، برچسب‌گذاری خودکار موضوعی، طبقه‌بندی موضوعی خودکار بر اساس الگوهای از پیش تعیین شده، طبقه‌بندی و سنجش روابط میان کلیدواژگان، ارزیابی نظام‌های طبقه‌بندی، ساختار رده‌بندی‌های خودکار و ابزارها، و نیز برنامه‌ها و فناوری‌های اطلاعاتی در خودکارسازی سازماندهی اطلاعات در این دسته قرار می‌گیرند. یکی از زمینه‌های مهم و موثر در رابطه با مطالعه ابعاد نمایه‌سازی موضوعی، توجه به خلاءهای موجود در ساختار اختصاص موضوع به منابع ارزشمند پارساهاست که کمتر به آن توجه شده است. با این نگرش، پژوهش حاضر بر آن است تا با مطالعه و بررسی خلاءهای موضوعی به سنجش این زمینه سودمند بپردازد و از آن برای ایجاد بستر مناسب به منظور ایجاد ساختار سازماندهی خودکار اطلاعات بهره گیرد.

روش پژوهش

در پژوهش پیش‌رو، خلاءهای موجود در ارتباط با سازماندهی و پردازش پایان‌نامه‌ها و رساله‌های (پارساهای) تحصیلات تکمیلی مورد مطالعه قرار گرفته و مولفه‌های آن ارزیابی شده است. روش پژوهش آمیخته است، و از رویکرد کیفی و کمّی به شیوهٔ تطبیقی در ارتباط با مشکلات نمایه‌سازی و سازماندهی از ابعاد موضوعی بهره گرفته است. در این پژوهش خلاءهای موضوعی موجود در اختصاص کلیدواژه‌های توصیفی برای پارساها در پایگاه گنج با مفهوم اصلی آنها در رشته تخصصی خود مورد بررسی قرار گرفته و ابعاد آن تعیین شده است. جامعه آماری پژوهش، پایان‌نامه‌ها و رساله‌های پایگاه اطلاعات علمی ایران (گنج) است. این منابع شامل تعداد ۱/۱۱۲/۵۳۱ مدرک است. منابع اطلاعاتی در این پژوهش شامل پایان‌نامه‌ها و رساله‌های (پارساهای) تحصیلات تکمیلی است و پیشنهاده‌ها را در بر نمی‌گیرد. با عنایت به حجم مجموعه بر مبنای جدول مورگان نمونه آماری ۳۸۴ مدرک تعیین شد. ولی به دلیل بالا بردن ضریب اطمینان میزان درستی اطلاعات سازماندهی و با توجه به تنوع موضوعی پارساها، که رشته‌ها و موضوع‌های متعددی را در بر می‌گرفت، نمونه آماری از میان پایان‌نامه‌ها و رساله‌ها به تعداد ۱۵۰۰ مدرک و به صورت تصادفی طبقه‌ای انتخاب شد تا سنجش مناسب‌تری از وضعیت نمایه‌سازی موضوعی به انجام رسد. بررسی بر روی این نمونه آماری متناسب با حجم رشته و گرایش‌های تحصیلی به انجام رسید. ابزار گردآوری اطلاعات، سیاههٔ وارسی محقق‌ساخته بود. این سیاهه با تعیین ۲۱ مولفه در زمینه‌های مرتبط چون تعیین حوزهٔ موضوعی، کلیدواژه، چکیده، ساختار مفهومی، بسامد تکرار واژگان، بیان مساله و فصل‌بندی پارساها، برای تحلیل و ارزیابی ساختار موضوعی طراحی شد. این سنجش با طراحی سه گزینه «مرتبط»، «تا حدودی مرتبط» و نیز «غیرمرتبط» انجام پذیرفت. هر یک از مولفه‌ها بر اساس نقشی که سازماندهی اطلاعات و نمایه‌سازی موضوعی دارند تبیین شدند. سنجش روایی ابزار پژوهش به شیوهٔ 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. بر این مبنا میزان ربط میان کلیدواژه‌های موضوعی با مفهوم اساسی پارساها تحلیل شده تا در سنجش ساختار جدید به منابع اطلاعاتی موجود نیز توجه شده باشد. روش گردآوری اطلاعات، مراجعه به پایگاه اطلاعات علمی ایران (گنج) بود. برای تحلیل اطلاعات از روش ارزیابی بر مبنای مقایسه میان فیلدهای موجود بر اساس مولفه‌های سیاههٔ وارسی بهره گرفته شد. علاوه بر آن برای بالا بردن ضریب اعتبار و اعتبارسنجی نتایج در تحلیل خلاءهای موضوعی پارساها، مصاحبه با ۴ نفر از همکاران نمایه‌ساز در ایرانداک انجام گرفت. این مصاحبه به روش نیمه‌ساختاریافته به انجام رسید. همکاران بر اساس میزان سابقه کاری انتخاب شدند. مطالب بیان شده توسط آنان در قالب ۵ محور اساسی تنظیم و مشخص شد. نکات بیان شده شامل موارد مربوط به مشکلات سازماندهی و ابعاد آن بود. در خصوص سطوح حوزه‌های موضوعی سطح، 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 و نظر آنان در اصلاح تحلیل نهایی لحاظ شد.

یافته‌های پژوهش

در راستای پاسخگویی به پرسش اصلی پژوهش، یعنی خلاءهای موضوعی به منظور نمایه‌سازی پایان‌نامه‌ها و رساله‌های (پارساهای) تحصیلات تکمیلی در پایگاه اطلاعات علمی ایران (گنج) با توجه به ساختار و کارکرد حوزه سازماندهی اطلاعات کدامند؟ در بخش مربوط به نظر نمایه‌سازان نکات مهم در جدول ۱ مطرح شده است.

جدول ۱. خلاءهای نمایه‌سازی پارساها از دیدگاه نمایه‌سازی متخصص

محور اصلی	نوع خلاء موضوعی	شرح و توصیف
R1	کاستی در واژگان تخصصی	نبود واژگان مرتبط با موضوع‌های پارساها؛ مشخص نبودن اصطلاح اخص
R2	اشتباه در نگارش اصطلاح	خطای نگارشی؛ خطای مفهومی در بانک واژگان کمکی نمایه‌سازی
R3	تکرار واژگان تخصصی	تکرار اصطلاح مرتبط با مفاهیم متعدد و اختصاص نامناسب اصطلاح به مفهوم
R4	استفاده از اصطلاح عام و کلی	اصطلاح به کار رفته کلی و فراگیر بوده و فاقد تخصیص مرتبط با مفهوم است.
R5	در هم‌تنیدگی موضوع با رشته تحصیلی	توجه به نگرش کاربر و نوع کاوش اطلاعات توسط او برای بازیابی مرتبط
R6	زمان مناسب برای تخصیص کلیدواژه	رعایت زمان بر اساس نوع، ساختار، کاربرد و ویژگی‌های کلیدواژه‌ها

یافته‌های جدول ۱ نشان می‌دهد که ۶ مورد از نگاه نمایه‌سازان از اهمیت بالاتری برخوردار بوده است. کمبود واژگان تخصصی در انجام فرایند نمایه‌سازی یکی از ابعاد مهم به شمار می‌آید. در این رابطه باید اظهار کرد که برخی از اصطلاح‌های موجود در برنامه نمایه‌سازی ارتباط بالایی با محتوای مدرک نداشته و نمی‌تواند توصیفگر مناسبی برای موضوع به شمار آید. از این جنبه، توجه به اصطلاح مرتبط بر اساس نظام اصطلاح‌نامه حائز اهمیت سطح ۱ است. علاوه بر آن، در استفاده از بانک واژگان کمکی در کنار اصطلاح‌نامه‌های استاندارد به منظور پوشش کمبود واژگان تخصصی، برخی خطاهای مفهومی یا نگارشی وجود دارد که کار نمایه‌سازی را با چالش مواجه می‌سازد. این مورد در سطح ۲ قرار می‌گیرد. در سطح ۳، از نظر اهمیت خلاءهای موضوعی، مقوله تکرار واژگان تخصصی قرار دارد. این مورد با عنایت به تکرار اصطلاح مرتبط با مفاهیم متعدد در میان پارساها، نقش مهمی در پائین آوردن میزان خطا در نمایه‌سازی دارد که در کنار آن باید به اختصاص نامناسب اصطلاح به مفهوم نیز اشاره کرد که می‌تواند سبب بروز چالش در این زمینه شود. همچنین در سطح ۴، بهره‌گیری از اصطلاح عام و کلی مطرح می‌شود که نمی‌تواند پوشش مناسبی برای مفهوم اساسی مدرک باشد. در این زمینه ضروری است تا تخصیص کلیدواژه‌های موضوعی به

صورت خُرد و ویژه در نظر گرفته شود تا تطبیق مناسب با مفهوم مدرک داشته باشد. در سطح ۵، بحث در هم‌تنیدگی موضوع با رشته تحصیلی بیان شده که توجه به اولویت هر یک در سنجش موضوع پارسا حائز اهمیت است. زیرا بسیاری از کاربران، پارساها را از طریق رشتهٔ تحصیلی کاوش می‌کنند و در برخی موارد، میان موضوع و رشتهٔ تحصیلی همبستگی بالایی وجود دارد. در سطح ۶، نیز بحث فرصت و زمان مناسب برای تخصیص کلیدواژه مطرح شده که بیشتر در ابعاد مدیریتی قابل ملاحظه است.

همچنین، بر اساس نتایج حاصل از سیاههٔ وارسی با سنجش نمونه آماری پژوهش، پیرامون خلاءهای موضوعی در ساختار فعلی موضوعی سازماندهی در نرم‌افزار ویرایش مدیریت سازماندهی و تحلیل اطلاعات، برخی از مسائل اساسی و تاثیرگذاری که در این زمینه نقش دارند مطرح شده‌اند. نتایج بررسی و استخراج نکات اساسی حاصل از تحلیل اولیه، در جدول ۱ مطرح شده‌اند.

جدول ۲. خلاءهای موضوعی در ساختار سازماندهی اطلاعات در پایان‌نامه‌ها و رساله‌ها (پارساها)

عنوان چالش	توضیح آن
«زیرشاخه‌های موضوعی»	برای تعیین موضوع مناسب نیاز است تا زیرشاخه‌های مرتبط با موضوع به درستی مشخص شده و انطباق لازم برای تخصیص موضوع صورت گیرد.
«عمق نمایه‌سازی»	ارتباط میان زیربخش‌های موضوعی با حوزهٔ کلان با عنایت به توسعهٔ علمی و گسترش موضوع‌های متعدد انجام گیرد.
«برچسب موضوعی»	کافی نبودن دقت در تعیین کلیدواژهٔ مرتبط با مفهوم اساسی منبع اطلاعاتی و شناسایی مناسب حوزهٔ موضوعی با توجه به تعیین تعداد بیشتری از موضوع‌های تخصصی.
«تطبیق مفهوم با موضوع»	کافی نبودن شناخت نمایه‌ساز از مفهوم کلان منابع اطلاعاتی و شناسایی بهتر آن. تسلط نمایه‌سازی در اختصاص موضوع.
«ترجیح موضوع بر رشتهٔ تحصیلی»	بیش از آنکه حوزهٔ موضوعی با مفهوم موضوعی ارتباط داشته باشد، با رشتهٔ تحصیلی مرتبط است.
« توسعه و گسترش طبقه‌بندی موضوعی »	ضرورت استفاده از طبقه‌بندی موضوعی تا ابعاد کوچک و زیرشاخه‌های فرعی برای تخصیص موضوع مناسب با مفهوم اساسی پارسا.
« طبقه‌بندی موضوعی روزآمد »	توجه به توسعهٔ علمی و شناسایی موضوع‌های جدید. ضرورت همسو بودن طبقه‌بندی موضوعی با پیشرفت‌های علمی و تولید کلیدواژه‌های نوین.
«ساختار طبقه‌بندی موضوعی»	توجه بیشتر به ساختار طبقه‌بندی موضوعی متناسب با نیاز کاربران و بازیابی بهتر اطلاعات.
«مدیریت زمان در سازماندهی»	حجم ورودی منابع اطلاعاتی پارساها بیشتر از ظرفیت زمانی نمایه‌سازان است. توجه به فرایند نمایه‌سازی ماشینی.
«پردازش زبان طبیعی»	استخراج داده‌ها از متن است برای فرایند سازماندهی اطلاعات و طبقه‌بندی موضوعی دقیق.

براساس نتایج به دست آمده، باید بیان کرد که ساختار نمایه‌سازی و بهره‌گیری از کلیدواژه‌های موجود، دارای برخی کاستی‌های بنیادی است. این بدان معناست که طبقه‌بندی موضوعی امکان موضوع‌دهی گسترده برای منابع اطلاعاتی پارساها در نظر نگرفته است. طبقه‌بندی موضوعی معمولا دارای زیرشاخه‌های خُرد و توسعهٔ وسیع موضوعی است که برای تمامی موضوع‌های متعدد کلیدواژهٔ خاص در نظر می‌گیرد. این امر سبب می‌شود تا در هنگام سازماندهی منبع اطلاعاتی، موضوع مورد نظر به صورت دقیق و مشخص تعیین شده و در زمان بازیابی اطلاعات، انطباق لازم میان کلیدواژه و مفهوم برقرار باشد.

ارتباط میان ساختار طبقه‌بندی موضوعی در سطح اول و دوم نقش مهمی در عمق نمایه‌سازی دارد. سطوح موضوعی کنونی، ارتباط میان زیربخش‌های موضوعی با حوزهٔ کلان را در سطح حدود ۳۰۰ کلیدواژهٔ موضوعی اصلی مشخص ساخته است. در حالی که توجه به این مساله حائز اهمیت است که با توسعهٔ علمی، رشد رشته‌های تخصصی از شتاب بیشتری برخوردار شده و موضوع‌های فرعی بسیاری را شامل می‌شود. این موضوع‌ها همواره نیاز به اختصاص برچسب موضوعی مناسب دارند تا با دقت مفهوم اساسی منبع اطلاعاتی تبیین شده و برای کاربر شناسایی مناسبی را به همراه داشته باشد. لذا به تعیین تعداد بیشتری از موضوع‌های تخصصی نیاز است.

از زمره موارد و چالش‌های پیش‌روی سازماندهی اطلاعات و طبقه‌بندی موضوعی، تطبیق مفهوم با موضوع اختصاص داده شده است که نیاز به آگاهی نمایه‌ساز دارد. شناخت نمایه‌ساز از مفهوم کلان منابع اطلاعاتی و شناسایی بهتر آن، سبب توسعه و رشد بهتر فرایند نمایه‌سازی خواهد شد. از این منظر، به هر میزان که نمایه‌ساز تسلط بیشتری نسبت به تشخیص موضوع اصلی پارسا داشته باشد، می‌تواند در این مسیر کلیدواژه‌های مناسب‌تری را برای توصیف پارساها ارائه دهد. البته پاره‌ای از موضوع‌ها در نظام فعلی وجود ندارد، لذا نمایه‌ساز مجبور است تا نزدیک‌ترین موضوع را برای آن انتخاب کند. ضعف در این زمینه سبب خواهد شد تا شناخت مفهوم منابع به درستی انجام نپذیرد و در نظر گرفتن کلیدواژهٔ موضوعی نیز به درستی انتخاب نشود و بین نمایه‌سازی و مفهوم اصلی منبع اطلاعاتی فاصله ایجاد شود.

همچنین شناسایی دقیق مفهوم و محتوای اصلی منبع اطلاعاتی از دیگر چالش‌های سازماندهی و طبقه‌بندی موضوعی است. برای انجام سازماندهی مناسب باید توجه کرد که مفهوم اساسی منبع اطلاعاتی به درستی تشخیص داده شود. شناسایی مفهوم اصلی می‌تواند شرایط بهتری برای اختصاص کلیدواژه مناسب فراهم آورد. این به آشنایی نمایه‌ساز با آن حوزهٔ موضوعی باز می‌گردد، که در زمینهٔ مورد نظر و محتوای اصلی آشنایی مناسبی وجود داشته باشد. از سوی دیگر، شناخت نمایه‌ساز با دامنهٔ موضوعی می‌تواند به درک بهتری از معنای کلان متن یاری رساند. متون اطلاعاتی در زمینه‌های مختلفی قرار می‌گیرند. شناخت ارتباط میان موضوع‌ها و برخی منابع میان‌رشته‌ای به سازماندهی بهتر منابع یاری می‌رساند.

این نکته را باید بیان کرد که روش موجود بیشتر مبتنی بر رشته است تا موضوع. از همین‌رو، انتخاب کلیدواژه به رشتهٔ تحصیلی نزدیکتر است. البته تخصیص موضوع با عنایت به شاخص قرار گرفتن رشته تحصیلی، می‌تواند مفهوم را به صورت مناسب‌تری برساند؛ اما اختصاص موضوع خُرد را با دشواری مواجه می‌سازد. این امر سبب می‌شود که انتخاب کلیدواژه بیشتر مبتنی بر مفهوم رشته باشد که البته در زمان

بازیابی برای کاربر هم سهولت خواهد داشت. اختصاص کلیدواژه‌های موضوعی نیز باید بیشتر براساس مفهوم اساسی پارساها در نظر گرفته شود. اختصاص موضوع دقیق البته با توجه به مفهوم بنیادی منبع در نظر گرفته می‌شود.

از چالش‌های دیگر پیش‌رو، امکان توسعه و گسترش طبقه‌بندی موضوعی برای استفاده بیشتر از کلیدواژه‌هاست. این بدان معناست که ضرورت دارد تا طبقه‌بندی موضوعی تا ابعاد کوچک و زیرشاخه‌های فرعی ادامه پیدا کند. به هر میزان که تقسیم‌های فرعی کلیدواژه‌های موضوعی توسعه پیدا کند، امکان تطبیق موضوع با پارساها بیشتر خواهد بود. برای دستیابی به یک منبع اطلاعاتی، توجه به موضوع اصلی می‌تواند شرایط بهتری را برای تخصیص کلیدواژه مهیا سازد. گسترش موضوعی همواره با تخصیص موضوع مرتبط، شرایط مطلوب برای بازیابی بهتر منابع اطلاعاتی ایجاد می‌کند. این مقوله به تعیین بهتر موضوع در ارتباط با مفهوم کلان پارسا منجر خواهد شد. از همین‌رو، ضرورت دارد تا بتوان با تکیه بر تقسیم‌های فرعی موضوعی شرایط بهتری را برای تخصیص موضوع فراهم ساخت.

نکتهٔ دیگر در رابطه با طبقه‌بندی موضوعی روزآمد است. همواره با توسعهٔ علمی، موضوع‌های جدید نیز شناسایی می‌شوند. این امر بدان معناست که ضرورت دارد تا طبقه‌بندی موضوعی همواره با پیشرفت‌های علمی همسو باشد. انواع طبقه‌بندی‌های موضوعی مورد استفاده، معمولا با تکیه بر پیشرفت‌های علمی، توسعه یافته و ویرایش‌های جدیدی را ارائه می‌دهند. طبقه‌بندی موضوعی بسته به نوع منبع برای سازماندهی می‌تواند متفاوت باشد. میزان و نوع کلیدواژه‌ها در یک طبقه‌بندی موضوعی که برای سازماندهی مجلات استفاده می‌شود، متفاوت از طبقه‌بندی برای کتاب‌ها یا پارساهاست. لذا دقت در انتخاب طبقه‌بندی موضوعی شرایطی را ایجاب می‌کند که مطابق با وضعیت منابع اطلاعاتی مورد نظر باشد.

ساختار طبقه‌بندی موضوعی در تخصیص کلیدواژهٔ مناسب نقش به‌سزایی دارد. این ساختار از نظر شکلی و نیز محتوایی قابل اهمیت است. علاوه بر آن ضرورت دارد تا بتوان با تکیه بر ساختار مدون طبقه‌بندی موضوعی، منابع را متناسب با نیاز کاربران سازماندهی کرد. اساسا، نیاز کاربران به بازیابی موثر منابع اطلاعاتی، سبب می‌شود تا سازماندهی به سوی کاربرمدار شدن پیش رود. اکنون پاره‌ای وبگاه‌های اطلاعاتی و موتورهای کاوش فراتر از نظام طبقه‌بندی موضوعی استاندارد کتابخانه‌ای از فولکسونومی تبعیت می‌کنند. همچنین روش‌های دستیابی به اطلاعات مراحل بالاتر که می‌تواند فواید قابل توجهی به همراه داشته باشد. به عنوان مثال امکان کشف پژوهشگران[1] در پایگاه اسکوپوس[2] که از طریق موضوع پژوهش و کلیدواژه‌های کاوش می‌توان به سطوح بالاتری از داده‌ها و از جمله پژوهشگران مرتبط با موضوع‌ها دست یافت. کاربران با کاربرد آن می‌توانند شبکه‌های علمی خود را گسترش داده و با پژوهشگران در حوزه‌های موضوعی مشترک ارتباط علمی داشته باشند. بخشی از طراحی این فرایند با بهره‌گیری از الگوریتم‌های جست‌وجوی معنایی در قالب فرادادههای چندسطحی امکان‌پذیر می‌شود.

1 Researcher Discovery
2 Scopus

یکی از چالش‌های مهم نمایه‌سازی مساله مدیریت زمان برای انجام فرایند سازماندهی اطلاعات است. این بدان معناست که معمولا حجم ورودی منابع اطلاعاتی پارساها بیشتر از ظرفیت زمانی نمایه‌سازان است. برای بهینه‌سازی فرایند نمایه‌سازی معمولا از نمایه‌سازی ماشینی یا نیمه ماشینی استفاده می‌شود. در نظام نمایه‌سازی ماشینی به روش‌های گوناگون، با استخراج کلیدواژه‌های اصلی از دل عنوان، چکیده، کلیدواژهای موضوعی و نیز در مرحله بعدی با توجه به بسامد تکرار واژگان از درون متن می‌توان در زمان و وقت نمایه‌سازی صرفه‌جویی قابل ملاحظه‌ای انجام داد. البته نمایه‌سازی ماشینی دقت نمایه‌سازی دستی را ندارد و از این منظر سطح پایین‌تری را به خود اختصاص می‌دهد، ولی با روش‌های ترکیبی می‌توان دقت آن را افزایش داده و میزان درستی نمایه‌سازی را تضمین کرد. بهره‌گیری از روش‌های خودکار امروزه تا حد زیادی در پایگاه‌های اطلاعاتی افزایش یافته و فرایند سازماندهی در نظام اطلاع‌رسانی را تسریع کرده است.

بحث

نتایج پژوهش نشان داد که کاستی‌هایی وجود دارد که طبقه‌بندی موضوعی حاضر، امکان موضوع‌دهی گسترده برای منابع اطلاعاتی پارساها را ندارد. زیرشاخه‌های خُرد و توسعهٔ وسیع موضوعی باعث دقت بیشتر سازماندهی اطلاعات می‌شود و می‌توان ارتباط مناسب‌تری میان کلیدواژه، موضوع‌های استاندارد و مفهوم برقرار کرد. این یافته به این دلیل قابل توجه است که در هنگام بازیابی اطلاعات اثربخشی بالاتری دارد که می‌توان با بهره‌گیری از طبقه‌بندی موضوع پایان‌نامه بر مبنای ساختار تخصصی رده‌بندی و نیز استفاده از طبقه‌بندی‌های کتابخانه‌ای و کاربردهای تخصصی آن در تعیین موضوع‌های خاص نقش مهمی در بازیابی مناسب اطلاعات پارساها ایفا کرد. این یافته با نتایج پژوهش «شومان» و «مرگل» (۲۰۲۵) (Schumann & Mergel, 2025) و «لین» و «کاهاندا» (Lane & Kahanda, 2021) از منظر ساختار تخصصی رده‌بندی و نیز با پژوهش‌های «حیدر» و «شارما» (Haider & Sharma, 2017) و «لوند» (Lund, 2019) از ابعاد رده‌های کتابخانه‌ای دارای اشتراک است.

در تحلیل یافته‌های پژوهش باید به این نکته مهم اشاره کرد که نقش عمق نمایه‌سازی در سطوح موضوعی دارای اهمیت است. سطوح موضوعی کنونی، ارتباط میان زیربخش‌های موضوعی با حوزهٔ کلان را در سطحی کمتر در بهره‌گیری از کلیدواژه‌های موضوعی در نظر گرفته است. با توسعهٔ علمی، رشد پارساهای دانشگاهی دارای سرعت بیشتری شده و می‌تواند موضوع‌های فرعی بسیاری را در بر بگیرد. از این منظر توسعه موضوع‌های فرعی در ساختار طبقه‌بندی موضوعی فعلی نقش تعیین‌کننده‌ای دارد. از همین رو، بهره‌گیری از نظام واژگان کنترل‌شده و متناسب با ساختار پارساها حائز اهمیت است. نظام‌های سازماندهی با تکیه بر نظام واژگان سلسله‌مراتبی و نیز ساختارهای مفهومی می‌توانند شرایط مناسبی را برای توسعه طبقه‌بندی موضوعی ایجاد کنند. این یافته با نتایج پژوهش «هولستروم» (Holstrom, 2023) و نیز مطالعه انجام گرفته توسط «میترو» و همکاران (Mitrou et al., 2025) از منظر ارتباط ساختاری در حوزهٔ جایگاه طبقه‌بندی موضوعی هم‌سویی دارد.

علاوه بر آن، اختصاص برچسب موضوعی مناسب نیز تاثیرگذاری مطلوبی در بازیابی منابع پارسا دارد، تا با دقت مفهوم بنیادی مدارک اطلاعاتی مشخص شود. برچسب‌گذاری معنایی سبب رفع ابهام در تعیین کلیدواژه‌های موضوعی می‌شود که می‌تواند بازیابی را از سمت واژه‌محور به سوی محتوامحور هدایت کند. این یافته با نتایج برخی از پژوهش‌های پیشین، از جمله «جون‌ژی» و «رون‌دونگ» (Junzhi & Rundong, 2025)، «کلوگه» و «کِلِر» (Kluge & Kähler, 2025) و «اُولیری» (O'Leary, 2023) از جنبهٔ نقش برچسب‌گذاری موضوعی در بازیابی اطلاعات مرتبط دارای اشتراک است. همچنین شناسایی دقیق محتوای اصلی مدارک اطلاعاتی چالشی را برای سازماندهی حوزهٔ موضوعی ایجاد می‌کند. لذا همواره باید توجه کرد که در سازماندهی مناسب باید مفهوم بنیادی منبع اطلاعاتی به صورت درست مشخص شود؛ زیرا شناسایی مفهوم اصلی یک مدرک می‌تواند شرایط بهتری برای تخصیص کلیدواژه مناسب فراهم کند. این یافته با نتایج پژوهش «سوچیاتی» و همکاران (Suciati et al., 2023) و نیز مطالعه علمی «آویو-رووِن» و «روزِنفِلد» (Aviv-Reuven & Rosenfeld, 2023) از جنبه درک و شناخت مناسب موضوع و ربط معنایی در نمایه‌سازی هم‌سویی دارد. ولی با یافته‌های پژوهش «میترو» و همکاران (Mitrou et al., 2025) از ابعاد ساختار طبقه‌بندی اشتراکی ندارد.

از سویی، آشنایی نمایه‌ساز با دامنهٔ موضوعی منجر به درک بهتر از مفهوم اساسی مدرک می‌شود؛ و از سوی دیگر، شناخت ارتباط میان موضوع‌ها و منابع میان‌رشته‌ای نیز سبب سازماندهی بهتر خواهد شد. برای تقویت مهارت‌های متخصصان سازماندهی اطلاعات، بازیابی پارساها بر اساس طبقه‌بندی و سنجش روابط کلیدواژگان و نیز آگاهی از چالش‌های فرایندها و روش‌های پردازش موضوع نقش مهمی دارد. این نیازمند آموزش و توسعه مهارت‌های نمایه‌سازان است. نتایج پژوهش‌های «ژانگ» و همکاران (Zhang et al., 2017) و همچنین «کاتاگی» و «کومبار» (Katagi & Kumbar, 2023) در ارتباط با دامنهٔ موضوعی و شناخت مفهوم مدرک در تخصیص کلیدواژه‌های تخصصی با یافته‌های پژوهش حاضر هم‌سویی دارد. ولی از ابعاد مفهومی و ساختاری در تعیین دامنهٔ موضوعی با پژوهش «لین» و «کاهاندا» (Lane & Kahanda, 2021) هم‌سویی ندارد. اما از سوی دیگر، گسترش تقسیم‌های فرعی کلیدواژه‌های موضوعی، امکان تطبیق موضوع با پارساها را افزایش می‌دهد. مدل‌های زبانی بزرگ راهکاری مناسب برای این توسعه به شمار می‌آید. پژوهش «کلوگه» و «کِلِر» (Kluge & Kähler, 2025) و همچنین مطالعه عمیق «جون‌ژی» و «رون‌دونگ» (Junzhi & Rundong, 2025)، زمینه استفاده از مدل‌های زبانی در تخصیص کلیدواژه‌های تخصصی و بازیابی موثر اطلاعات با نتایج پژوهش حاضر دارای اشتراک است. ولی پژوهش «کاتاگی» و «کومبار» (Katagi & Kumbar, 2023) و نیز «ژانگ» و همکاران (Zhang et al., 2017) از ابعاد بازیابی اطلاعات ساختاری و به کارگیری سازه‌های مفهومی زبان با این یافته اشتراکی ندارد.

همچنین، بر اساس نتایج حاصل از پژوهش، ایجاد شرایط مناسب برای همسان‌سازی حوزه‌های موضوعی با پیشرفت‌های علمی از اهمیت بالایی برخوردار است. از این منظر، استفاده از طرح‌های طبقه‌بندی موضوعی خودکار، بر اساس الگوهای از پیش تعیین‌شده می‌تواند حاوی سودمندی مطلوبی باشد. نتایج پژوهش «لین» و «کاهاندا» (Lane & Kahanda, 2021) و نیز مطالعات «کازی» و همکاران (Kazi et

al., 2025) و «آریسوکو» و «اُگبونا» (Arisukwu & Ogbonna. 2020) از ابعاد طراحی نمایه‌سازی خودکار در بافت معنایی با یافته‌های پژوهش پیش‌رو، دارای هم‌سویی است. علاوه بر آن، کاربرد رده‌بندی علمی مبتنی بر طبقه‌بندی و سنجش روابط کلیدواژگان می‌تواند تا حد زیادی سبب بازیابی بهتر شود. روش‌های پردازش موضوع خودکار بر اساس چنین طبقه‌بندی معرفی شده نقش موثری در دستیابی به منابع پارساها خواهد شد. نتایج پژوهش «پوزیگا» و همکاران (Posigha et al., 2019) و «کاتاگی» و «کومبار» (Katagi & Kumbar, 2023) و نیز «یامان» و همکاران (Yaman et al., 2021) در زمینه ساختار پردازش خودکار نمایه‌سازی با یافته‌های پژوهش حاضر دارای اشتراک است.

در جمع‌بندی یافته‌های پژوهش باید اظهار کرد که برای تقویت ساختار نمایه‌سازی توجه به زیرشاخه‌های خُرد و توسعهٔ وسیع موضوعی از اهمیت بالایی برخوردار است. از این منظر توسعهٔ عمق نمایه‌سازی و بهره‌گیری از دامنهٔ موضوعی مرتبط با رشته و مفهوم پارسا جایگاه مهمی دارد. بر این اساس، ربط معنایی و برچسب‌گذاری موضوعی می‌تواند شرایط مناسبی را برای بهینه‌سازی فعالیت مرتبط با نمایه‌سازی ایجاد کند. لذا شایسته است تا شرایط سازماندهی به گونه‌ای برنامه‌ریزی شود تا نظام طبقه‌بندی موضوعی با تکیه بر عوامل یادشده تقویت شده و نمایه‌سازی پارساها به صورت مناسب به انجام رسد. از محدودیت‌های پژوهش آن بود که سنجش و تحلیل پارساها در رشته‌های مختلف با توجه به ویژگی‌ها و تفاوت‌های رشته‌های تحصیلی با دشواری همراه بود. همچنین بر اساس یافته‌های پژوهش پیشنهاد می‌شود تا این فعالیت‌ها در دستور کار قرار گیرد:

- بر اساس ویژگی توجه به عمق نمایه‌سازی، مناسب آن است تا شناسایی انواع نام‌ها و ابهام‌زدایی از معنای واژگان در رده‌بندی مفهومی موضوع به انجام رسد.
- بر مبنای یافته پژوهش مبنی بر توسعهٔ زیرشاخه‌های موضوعی، شایسته است تا آماده‌سازی بانک کلیدواژگان پارساها بر اساس نظام طبقه‌بندی موضوعی سلسله‌مراتبی در سطوح تعریف شده انجام گیرد.
- با توجه به نتیجه دیگر حاصل از پژوهش مبنی بر ترجیح موضوع بر رشتهٔ تحصیلی، پیشنهاد می‌شود تا استفاده مرتبط از رشته و گرایش‌های تحصیلی در کنار طبقه‌بندی موضوعی پارساها به منظور تقویت سیستم سازماندهی مدارک اجرایی شود.
- با توجه به یافته بعدی پژوهش یعنی پردازش زبان طبیعی، که ارتباط بالایی با سیستم‌های فناوری دارد، پیشنهاد می‌شود تا بهره‌گیری از سیستم‌های خبره برای استفاده از نظام موضوعی رده‌بندی مفهومی در دستور کار قرار گیرد. علاوه بر آن شایسته است تا خودکارسازی سیستم نمایه‌سازی موضوعی با استفاده از فرایندهای هوش مصنوعی نیز مورد توجه قرار گیرد.
- بر مبنای یافته‌های مهم پژوهش شامل توسعه و گسترش طبقه‌بندی موضوعی و نیز طبقه‌بندی موضوعی روزآمد، پیشنهاد می‌شود تا با بهره‌گیری از مدل‌های زبانی بزرگ در سازماندهی مفهومی پارساها، به تقویت این حوزه مبادرت ورزید.

- با عنایت به یافته دیگر پژوهش، یعنی بهره‌گیری از برچسب موضوعی، شایسته است تا طراحی پیکره‌های زبانی مرتبط با پارساها به منظور سازماندهی محتوای آنها در دستور کار قرار گیرد.
- همچنین بر اساس راهکارهای سیاستی پیشنهاد می‌شود تا از استانداردهای روزآمد حوزه سازماندهی برای نمایه‌سازی بهره گرفته شود و نیز به بومی‌سازی بخش‌های مرتبط استانداردها در ساختار سازماندهی اقدام شود. همچنین فراداده‌ها موجود در گنج بررسی و نسبت به استانداردسازی آنها فعالیت مناسب انجام پذیرد.
- علاوه بر آن، در ارتباط با پیشنهادهای نهادی باید بیان کرد که شایسته است تا همکاری میان پژوهشکده‌ها برای انجام پژوهش‌های مرتبط افزایش یافته و از تجارب دیگر مراکز اطلاع‌رسانی در این زمینه بهره گرفته شود.
- شایسته است تا با برگزاری دوره‌های آموزشی مرتبط در زمینه نمایه‌سازی و روش‌های موثر سازماندهی پارساها و نیز ارائه رهنمودهای سودمند برای دانشجویان و استادان در تنظیم نهایی پارساها، اقدامات مناسب انجام گیرد.

منابع

زره‌ساز، محمد . (۱۴۰۰). سازماندهی اطلاعات در ایران: فرصت‌ها و راهکارهای پیشنهاد شده برای رفع چالش‌ها. *مطالعات کتابداری و سازماندهی اطلاعات*، ۳۲(۳)، ۱۰۳-۱۲۳. https://doi.org/10.30484/nastinfo.2021.2941.2068

شادانپور، فرزانه؛ ریاحی نیا، نصرت؛ برنا، کیوان؛ و منتظر، غلامعلی. (۱۴۰۱). درباره «موضوع» و «تحلیل موضوعی»: نکته‌ها و چشم اندازها. *کتابداری و اطلاع‌رسانی*، ۲۵(۳)، ۳۶-۶۴. https://doi.org/10.30481/lis.2022.349952.1979

طاهری، سید مهدی . (۱۴۰۴). ایستایی یا پویایی؟: چالش‌های روزآمدسازی و پیاده‌سازی سازماندهی دانش در ایران. *مطالعات کتابداری و سازماندهی اطلاعات*، ۳۶(۴)، ۷-۱۶. https://doi.org/10.30484/nastinfo.2026.4022.2382

فتاحی، رحمت الله. (۱۳۸۶). از آرمانها تا واقعیت: تحلیلی از مهمترین چالشها.. و رویکردهای سازماندهی اطلاعات در عصر حاضر. *کتابداری و اطلاع‌رسانی*، ۱۰(۴)، ۵-۲۶.

ناخدا، مریم، و شرفی، علی. (۱۳۹۸). عوامل ایجاد آشوب در سازماندهی اطلاعات و راهکارهای جلوگیری از آن. تحقیقات کتابداری و اطلاع‌رسانی دانشگاهی، ۵۳(۳)، ۶۴-۸۶.

نوروزی، یعقوب، و جعفری‌فر، نیره. (۱۳۹۹). استانداردهای فراداده‌ای سازماندهی منابع اطلاعاتی در کتابخانه دیجیتالی مروری نظامند. *فصلنامه بازیابی دانش و نظام‌های معنایی*، ۷(۲۴)، ۱۱۱-۱۳۷. https://doi.org/10.22054/jks.2020.50159.1281

References

Ahuja, A. (2023). *Analyzing and navigating electronic theses and dissertations*. Virginia Polytechnic Institute and State University, [Doctoral Dissertations]

Arisukwu, F. E. A., & 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), 133-146.

Artene, C. G., Vecliuc, D. D., Tibeică, M. N., & Leon, F. (2022). An experimental study of Convolutional Neural Networks for functional and subject classification of web pages. *Vietnam Journal of Computer Science*, *9* (04), 435-453. https://doi.org/10.1142/S2196888822500245

Avila, D. M., & J. A. C. Guimarães. .2013. Library classifications criticisms: universality, poststructuralism and ethics. *Scire: representación y organización del conocimiento*, 21-26.

Aviv-Reuven, S., & Rosenfeld, A. (2023). A logical set theory approach to journal subject classification analysis: intra-system irregularities and inter-system discrepancies in Web of Science and Scopus. *Scientometrics*, *128* (1), 157-175. DOI: https://doi.org/10.1007/s11192-022-04576-3

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. 10.1109/BigData62323.2024.10825418

Bastos, F. M., & Fujita, M. S. L. (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.

Borden, I., & Rüedi, K. (2006). *The dissertation: an architecture student's handbook*. London: Routledge.

Datta, E. (2023). Attempts at automating journal subject classification. *Upstream*. https://doi.org/10.54900/n6dnt-xpq48

Dawi, M. (2024). PhD dissertation review: Automatic indexing of Arabic periodicals in Al Mandhoumah databases. *Cybrarians Journal*, (72), 120-128. https://doi.org/10.70000/cj.2024.72.606

Dermentzi, M., Bryant, M., Rovigo, F., & García-González, H. (2025). Multilingual Automated Subject Indexing: a comparative study of LLMs vs alternative approaches in the context of the EHRI project.

Fujita, M. S. L., Ribas, R., Rodrigues, M., & Silveira, T. (2022). Thesis and dissertation subject metadata in repositories: an exploratory study on vocabulary control.

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. https://doi.org/10.32628/CSEIT228516

Grant, C., & Osanloo, A. (2014). Understanding, selecting, and integrating a conceptual framework in dissertation research: Creating the blueprint for your House. *Administrative Issues Journal: Connecting Education, Practice, and Research, 4* (2), 11-26.

Hadi, R. (2022). A Implementation of the Support Vector Machine Algorithm in the Student Thesis Subject Classification System at the University of Lampung. *Journal of Applied Science, Engineering and Technology*, *2* (2), 47-54. https://doi.org/10.47355/aset

Haider, S., & 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.

Holstrom, C. (2023). *Controlled, in control, and out of control: The effects of different forms of vocabulary control on the subject indexing and subject tagging processes*. University of Washington. [Doctoral Dissertations]

Howard, S. A., & S. A. Knowlton (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.

Jeangirard, E. (2021). Content-based subject classification at article level in biomedical context. *arXiv preprint arXiv:2104.14800.*

Junzhi, J., and Rundong, H. (2025). Disciplinary Gaps in Subject Indexing: A Structural Analysis of Controlled Vocabularies' Breadth and Depth. In *International Conference on Asian Digital Libraries* (pp. 484-493). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-95-4861-3_41

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. https://doi.org/10.5281/zenodo.8267818

Kazi, N., Lane, N., & 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).

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). https://doi.org/10.1145/3711670.3764654

Klair, H., German, F., Banerjee, B., & Ingram, W. A. (2025). 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

Kluge, L., & Kähler, M. (2025). DNB-AI-Project at SemEval-2025 Task 5: An LLM-Ensemble approach for automated subject indexing. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)* (pp. 1118-1128).

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. https://doi.org/10.4148/2160-942X.1072

Lund, B., D. Agbaji, S. Tijani, & I. Omame (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. 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*, 1-21. https://doi.org/10.48550/arXiv.2512.22524

Manzoor, J. A. (2023). *Segmenting electronic theses and dissertations by chapters* (Doctoral dissertation, Virginia Tech).

Mitrou, N., Koutsileou, S., & Makris, N. (2025). Relating Study Courses with Open Educational Resources using hierarchical controlled vocabularies and advanced subject classification tools. *I HE2024*, 311.

Naoum, S. G. (2013). *Dissertation research and writing for construction students*. London: Routledge.

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. https://doi.org/10.1002/isaf.1547

Ouyang, T., Zhou, J., & Xu, T. (2022). Subject classification algorithm of Chinese materials in university libraries based on the BERT model. In *Fourth International*

Conference on Computer Science and Educational Informatization (CSEI 2022) (Vol. 2022, pp. 232-237). IET. https://doi.org/10.1049/icp.2022.1477

Paré, A. (2011). Speaking of writing: Supervisory feedback and the dissertation. In *Doctoral education: Research-based strategies for doctoral students, supervisors and administrators* (pp. 59-74). Dordrecht: Springer Netherlands.

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. https://doi.org/10.51983/ajist-2019.9.3.294

Roostaee, 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. https://doi.org/10.1016/j.ipm.2019.102150

Sajid, N. A., Ahmad, M., Rahman, A. U., Zaman, G., Ahmed, M. S., Ibrahim, N., & Ahmed, R. (2023). A Novel Metadata-Based Multi-Label Document Classification Technique. *Computer Systems Science and Engineering*, *46* (2), 2195-2214. 10.32604/csse.2023.033844

Schumann, A. L., & Mergel, I. (2025). AI-Supported Automated Subject Indexing and Metadata Management with Annif and Finto AI at the National Library of Finland. In *AI Innovations in Public Services: The Case of National Libraries* (pp. 43-57). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-01344-6_4

Suciati, N., Fabroyir, H., & Pardede, E. (2023). Educational data mining clustering approach: case study of undergraduate student thesis topic. *IEEE Access*, *11*, 130072-130088.

Sun, W., & Li, J. (2024). Deep subdomain adaptation network improves cross-subject mental workload classification. In *2024 International Conference on Cyberworlds (CW)* (pp. 138-143). IEEE. https://doi.org/10.1109/CW64301.2024.00049

Thompson, P. (2025). Thesis and dissertation writing. *The handbook of English for specific purposes*, 403-419. https://doi.org/10.1002/9781119985068.ch21

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).

Wheatley, A. (2024). Subject trees on the Internet: A new role for bibliographic classification? In *Internet Searching and Indexing* (pp. 115-141). CRC Press.

Xu, Y. S., Mohd Khalid, Y. I. A., & Kassim, M. S. S. (2025). Automatic Subject Indexing: How has it Evolved and Where is AI Taking it? In *International Conference on Asian Digital Libraries* (pp. 3-21). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-95-4861-3_1

Zhang, H., Xiao, Y., & Bu, Z. (2017). Personalized book recommender system based on chinese library classification. In *2017 14th Web Information Systems and Applications Conference (WISA)* (pp. 127-131). IEEE. https://doi.org/10.1109/WISA.2017.42

Identifying Gaps in Subject Indexing of Theses and Dissertations in the Ganj Database and Ways to Improve It

Fariborz Doroudi[*] [1]
Assistant Professor, Iranian Research Institute for Information Science & Technology (IranDoc), Tehran, Iran.

Abstract

This study aims to analyze the thematic gaps present in the theses and dissertations available in the Iranian Scientific Information Database (Ganj) and propose solutions for organizing these information resources effectively. This qualitative research utilized a comparative method to identify thematic gaps in the indexing of theses and dissertations based on their core meanings. The statistical population included 1,112,531 theses and dissertations in the Iranian Scientific Information Database (Ganj). A random-stratified sampling method was employed, with a sample size of 1,500 documents for analysis. Data were collected using a checklist containing 21 components related to subject area evaluation and analysis. Interviews with 4 information organization experts at IranDoc were conducted to strengthen the data obtained. The validity of the data collection tool was assessed using the face validity method with input from 5 academic experts in information science. The reliability of the checklist was determined using Cronbach's alpha statistical method, yielding a value of 0.872. The analysis revealed several thematic organization gaps, including issues related to subject sub-branches, depth of indexing, subject labels, concept matching, preference for subjects over academic fields, development of thematic classification, up-to-date classification, classification structure, time management, and natural language processing. The study highlighted shortcomings in the indexing structure and utilization of keywords. The relationship between subject classification structure levels significantly impacts indexing depth. Challenges such as concept-subject matching and accurate identification of content were identified. Recognizing relationships between topics and interdisciplinary sources aids in resource organization. Developing and expanding subject classifications to incorporate more keywords is crucial, as new topics emerge with scientific advancements. The structure of subject classification plays a key role in assigning appropriate keywords. The results showed that developing subtopics within the main subject enhances information organization and fosters stronger keyword relationships and conceptual alignment. Indexing depth at subject levels is vital, and hierarchical vocabulary systems support subject classification development. Proper subject labeling improves thesis retrieval. Identifying information content accurately poses challenges in subject organization. Keyword-based retrieval and awareness of the challenges of subject processing are crucial in this context. Matching subject areas to scientific progress is essential, and scientific classification based on the relationships between keywords can enhance the effectiveness of retrieval.

Keywords: Dssertations and Theses, Iran Scientific Information Database (Ganj), Subject Gap, Subject Classification, Subject Classification, Subject Indexing

[1] Corresponding Author: doroudi@irandoc.ac.ir