

جایگاه موضوع به منظور سازماندهی منابع اطلاعاتی در پایگاه‌های داده

فریبرز درودی^۱*

استادیار، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، تهران، ایران

مدیریت
اطلاعات

دوره ۱۱، شماره ۱

بهار و تابستان ۱۴۰۴

چکیده: هدف پژوهش شناخت بررسی وضعیت فیلد موضوع در پایگاه‌های اطلاعاتی است. در این مطالعه، روش‌های بهره‌گیری از فیلد موضوع تحلیل و ارزیابی شده و دلایل اصلی برای استفاده از آن و نیز کاربرد و عملکرد موضوع در حوزه فعالیت ساماندهی اطلاعات بررسی شده است. روش پژوهش، مطالعه تطبیقی در کاربرد فیلدهای مرتبط با موضوع است. جامعه آماری شامل ۱۰ پایگاه اطلاعاتی فارسی و ۱۲ پایگاه اطلاعاتی خارجی است که از بالاترین اهمیت در میان کاربران و نهادهای علمی برخوردارند. یافته‌های پژوهش نشان می‌دهد که کاوش اطلاعات در وبگاه‌های اطلاعاتی مبتنی بر انواع فیلدهای مختلف بوده که از عنوان، نویسنده، نوع مدرک، شابک، شاپا، دی او آی، نام مجله، نوع مجله، نوع مقله، و برخی از فیلدهای دیگر بهره گرفته‌اند. هر پایگاه متناسب با نوع اطلاعات موجود خود و نیز نیاز کاربران و فعالیت تخصصی از فیلدهای متعددی استفاده کرده است. در کنار این فیلدها، استفاده از فیلد موضوع در ۹۰ درصد پایگاه‌های ایرانی به روش‌های مختلف بهره گرفته شده است. همچنین در میان پایگاه‌های خارجی ۱۰۰ درصد از آنها از فیلد موضوع و شیوه‌های گوناگون ارائه آن استفاده کرده‌اند. پایگاه‌های اطلاعاتی در بیشتر موارد در کنار فیلد موضوع و یا در برخی موارد به عنوان جانشین موضوع از فیلد کلیدواژه برای پشتیبانی موضوعی استفاده کرده‌اند. نتایج نشان می‌دهد که نیاز کاربران، اهداف سازمانی و نوع خدمات اطلاعاتی نیاز به فیلد موضوع را تأیید می‌کند. بهره‌گیری از ابزارهای استاندارد در اعتبار بخشی به موضوع نقش مهمی دارد. طبقه‌بندی‌های موضوعی در قالب سلسله‌مراتبی، استنادی، خوشه‌بندی، استانداردهای نمایه‌سازی و کاربرد هوش مصنوعی در تنظیم و سازماندهی موضوعی نقش مهمی برعهده دارد.

کلیدواژه‌ها: بازیابی اطلاعات، پایگاه اطلاعاتی، ساختار موضوعی، فیلد موضوع.

^۱ نویسنده مسئول: doroudi@irandoc.ac.ir

مقدمه

سازمان‌های بزرگ، حجم عظیمی از داده‌ها را تولید، جمع‌آوری و پردازش می‌کنند (Nambiar & Mundra, 2022) و ذخیره‌سازی، مدیریت و پردازش داده‌های گسترده برای همه حوزه‌های کاربردی از اهمیت بالایی برخوردار است (De Lange, 2023). از همین‌رو، پایگاه داده سازوکاری برای مدیریت عملیات داده‌ها در یک سیستم یکپارچه اطلاعاتی است (Elias & Nguyen, 2020) و انباری است که داده‌ها را بر اساس ساختار داده‌ای که بیش از چندین دهه پیش ایجاد شده است، سازماندهی، ذخیره و مدیریت می‌کند (Yan et al., 2024). باید اظهار کرد که سیستم مدیریت پایگاه داده یک برنامه نرم‌افزاری است که ذخیره‌سازی، بازیابی و به‌کارگیری سازمان‌یافته داده‌ها را تسهیل می‌کند (Malakar et al., 2025). همچنین، پس از دهه‌ها توسعه، سیستم‌های پایگاه داده امروزی همگی دارای ویژگی‌های بی‌شماری هستند که انتخاب این ویژگی‌ها را متناسب با نیاز برنامه‌های خاص استفاده‌کننده از آنها بسیار دشوار می‌کند (Yadav, 2023). می‌دانیم که رشد سریع انتشارات پژوهشی، تقاضاهای زیادی را برای فناوری‌های پیشرفته مدیریت اطلاعات بر کتابخانه‌های دیجیتال ایجاد کرده است (Jiang et al., 2022) و بر همین اساس، جست‌وجوی اطلاعات و داده‌های علمی، از یک سو، گسترده‌تر شده، در حالی که از سوی دیگر، به دلیل افزایش روزافزون منابع، دشوارتر می‌شود (Redkina, 2024). همچنین اهمیت مداوم جست‌وجوی اطلاعات و پژوهش‌های رفتاری در اطلاع‌رسانی، در عمل، بر سیاستگذاری و طراحی سیستم‌های اطلاعاتی در عصر دیجیتال تأکید دارد (Onyancha, 2025).

رشد انتشارات علمی، نیاز به ابزارهای قوی برای سازماندهی و بازیابی مؤثر پژوهش‌ها را افزایش داده است (Mamedov et al., 2025) و یکی از ابزارهای کاوش حوزه موضوعی، وب معنایی است، که جستجو را به جای انسان، برای ماشین مناسب می‌کند (Naik & Oza, 2023) و یا مدل‌سازی موضوعی یک کار مهم پردازش زبان طبیعی است و به ویژه در زمینه ترکیب و تحلیل متون، محبوبیت بیشتری پیدا کرده است (Ogunleye et al., 2025). علاوه بر آن، مسائل اصطلاح‌شناسی مربوط به ارزیابی کیفیت ابزارهای طبقه‌بندی سلسله‌مراتبی حوزه‌های علمی مورد توجه قرار گرفته است (Kalachikhin, 2025). اینها نشان می‌دهد که پیشرفت تحقیقات سیستم‌های بازیابی اطلاعات تحت تأثیر فناوری‌های موجود قرار گرفته است (Maña et al., 2024). بنابراین، پایگاه داده مجموعه‌ای منظم از داده‌ها است که به فرایند ذخیره‌سازی و سازماندهی الکترونیکی اطلاعات می‌پردازد. در واقع پایگاه‌های داده سبب تسهیل مدیریت اطلاعات می‌شود. از همین‌رو، توجه به ساختار موضوعی در پایگاه‌های اطلاعاتی و محیط وب، یکی از مسائل مهم و قابل اعتنا در دسترسی به اطلاعات ارزشمند و علمی است.

باید توجه کرد که توصیف موضوع، در بیشتر موارد، به طور پیش فرض در مفهوم سند تعریف می‌شود (Dutta & Dutta, 2013). اساساً، امروزه مدل‌های بازیابی اطلاعات مبتنی بر مفهوم و کلیدواژه است که نقشی مهم در دستیابی به اطلاعات ایفا می‌کند. این مدل‌های مفهومی در انتخاب موضوع از ترکیب کلمات کلیدی بهره می‌گیرند. بنابراین از منظر بازیابی اطلاعات تخصیص کلیدواژه مناسب و مبتنی بر موضوع مرتبط نقشی مؤثر در بازیابی اطلاعات دارد. در بسیاری از موارد موضوع به طور ضمنی در عنوان مدرک

بیان شده (Roostaee et al., 2020). همچنین، مشخصات جست‌وجوی مبتنی بر کلیدواژه برای استخراج داده‌ها از پایگاه داده‌های ساختاریافته مورد توجه پژوهشگران مختلف قرار گرفته است و بسیاری از پیشنهادهاى جالب در متون علمی یافت می‌شود (Ramada et al., 2020). این رویکردها شامل مواردی چون: نزدیکی بین کلیدواژه‌ها، تقسیم‌بندی پرس‌وجو، استفاده از توابع جمع منطقی، و نیز برخی موارد دیگر است. پاره‌ای از مسائل مربوط به کاوش کلیدواژه‌ها مربوط به ارزیابی تجربی در خصوص ارتباط نتایج است که به تحلیل مفهومی و کلیدواژه‌های مرتبط می‌پردازد.

از همین‌رو تاکید می‌شود که اصطلاح موضوع برای تبیین معنای مفهومی در نمایه‌سازی و طبقه‌بندی اسناد و بازیابی اطلاعات دارای اهمیت بالایی است. موضوع یکی از فیلدهای موثر در ساختار پایگاه اطلاعاتی و نرم‌افزار کتابخانه‌ای است که می‌تواند نقش مهمی در بازیابی موثر داشته باشد. با توجه به لزوم بهره‌گیری مناسب از موضوع‌های مرتبط با مدارک در پایگاه‌های اطلاعاتی بر اساس نیاز سازمان‌ها و کاربران در بهره‌گیری از این پایگاه‌ها، ضرورت وجود فیلد موضوع در پایگاه‌های اطلاعاتی یکی از مسائلی است که باید مورد مطالعه و توجه قرار گرفته و مشخص شود تا چه میزان به کارگیری و روش‌های ارائه آنها در پایگاه ضروری است. زیرا علاوه بر فیلد موضوع در نمایه‌سازی منابع اطلاعاتی، از کلیدواژه نیز برای توصیف موضوعی استفاده می‌شود. با عنایت به جایگاه مهم موضوع در پایگاه اطلاعاتی، پژوهش حاضر به بررسی کاربرد فیلد موضوع در بازیابی منابع اطلاعاتی در پایگاه‌های داده مختلف می‌پردازد تا ابعاد و جنبه‌های مختلف آن تحلیل و ارزیابی شود. هدف این پژوهش مشخص ساختن وضعیت کاربرد فیلد موضوع و جنبه‌های کاربرد آن در پایگاه‌های اطلاعات علمی است، تا بتواند موقعیت موضوع و جایگاه آن در این پایگاه‌ها را مشخص کرده و به ارائه پیشنهادهاى کاربردی برای تبیین بهره‌گیری از فیلدهای موضوع بپردازد. پرسش اساسی نیز عبارت است از: بهره‌گیری از فیلد موضوع تا چه اندازه و چگونه در پایگاه‌های اطلاعات علمی ایرانی و خارجی به کار رفته و به چه شیوه‌ای می‌توان به بهینه‌سازی آن کمک کرد؟

پیشینه پژوهش

سومینن و تووانن^۱ (2016) در پژوهشی در باره نقشه علم با الگوسازی موضوع: مقایسه یادگیری بدون نظارت و طبقه‌بندی موضوع اختصاص یافته به انسان به ترسیم مختصات برای نقشه برداری علوم اساسی پرداختند. نتایج نشان داد که طبقه‌بندی دقیق و معتبر موضوعی دانش علمی چالش مداوم در این زمینه است. مجموعه داده‌های جدید در مقیاس بزرگ برای استفاده در طرحهای طبقه‌بندی خودکار در فضاهایی چون پایگاه اطلاعاتی قابل استفاده است. چاوداری و شارما^۲ (2017) در بررسی تحلیلی و گردآوری اسناد زیست‌پزشکی با استفاده از روش‌های آماری و تحلیل واژگانی و موضوعی به این نتیجه رسید که تحلیل موضوعی مربوط به اسناد بسیار طولانی است و اگر این اسناد خلاصه شوند، می‌توان ارزیابی آسان‌تری را انجام داد. اعتبارسنجی خلاصه اسناد پزشکی تولید شده نشان می‌دهد که اختصاص موضوع به این اسناد با

¹ Suominen & Toivanen

² Chaudhary & Sharma

تکیه بر روش خوشه‌بندی کاربرد بهتری در بازیابی اطلاعات دارد. یوآن و زو^۱ (2018) در پژوهش خود به مقوله استخراج کلمات کلیدی به روش فراآموزشی پرداختند. نتایج پژوهش آنها نشان داد که با بهره‌گیری از فنون خوشه‌بندی و طبقه‌بندی متن در کنار ترجمه ماشینی، می‌توان نتایج مطلوبی ارائه داد. برای بهبود دقت استخراج کلیدواژگان مهم در متن که می‌تواند به دسترسی بهتر به منابع مرتبط منجر شود، چارچوبی مدون و الگویی کاربردی برای استخراج کلمات کلیدی را بر اساس روش فرایادگیری ارائه دادند. ناصر^۲ و همکاران (2019) با انجام پژوهشی در باره استخراج خودکار کلیدواژه و خلاصه‌سازی متن در نظام‌های برخط و شبکه‌های اجتماعی به این نتیجه دست یافتند که برای استخراج مفاهیم و موضوع‌ها از درون شبکه‌های اجتماعی، رویکردهای ماشینی و فنون یادگیری خودکار دارای کاربرد عمیقی است. رویکرد پیشنهادی آنها بهره‌گیری از طراحی معماری مناسب سیستمی در مسیر خلاصه‌سازی متن و استخراج کلیدواژه‌های بنیادی برای مفاهیم و موضوع‌های مطرح شده در متون اطلاعاتی است. ژنگ^۳ و همکاران (2020) در پژوهش خود در خصوص بررسی سیستم مبتنی بر پردازش زبان طبیعی^۴ در دسترسی به منابع اطلاعاتی و کاوش فیلدهای مبتنی بر کلیدواژه‌های موضوع و عنوان به این نتیجه رسیدند که استخراج کلمات کلیدی یک مسئله مهم در بسیاری از برنامه‌های پردازش زبان طبیعی است و می‌تواند عملکرد بسیاری از سیستم‌های بازیابی اطلاعات را بهبود بخشد. آنها دریافتند که روش‌های سنتی استخراج کلمات کلیدی مبتنی بر یادگیری ماشین یا مدل نمودار به کیفیت مطلوب بازیابی یاری می‌رساند. فیلیپ^۵ و همکاران (2020) در پژوهش خود بر نقش عملکرد مهم فراداده برای پایان‌نامه‌های الکترونیکی پرداختند. نتایج پژوهش آنها نشان داد که مرور سوابق ابرداده در خصوص پایان‌نامه‌های الکترونیکی سبب بازیابی بهتر موضوع‌های مورد نظر می‌شود. در این مطالعه روش تحلیل شبکه‌ای به کار گرفته شد و ویژگی‌های داده برای ذخیره‌سازی اطلاعات فراداده مبتنی بر کلیدواژه و کنترل شده بر اساس واژگان گزارش می‌دهد و می‌توان از این رویکرد به تجزیه و تحلیل کیفیت ابرداده‌ها دست یافت. یولیانتی و راهادیانتی^۶ (2021) به پژوهش در باره نقش سرعنوان‌های موضوعی اسناد با استفاده از مدل‌های بازیابی اطلاعات مبادرت ورزیدند. نتایج نشان داد که اثربخشی مدل‌های بازیابی اطلاعات، یعنی مدل زبانی و مدل فضای برداری، برای تولید خودکار فهرستی رتبه‌بندی‌شده از سرعنوان‌های موضوعی مرتبط، با هدف ارائه توصیه‌ای برای کتابداران جهت تعیین مؤثر و کارآمد سرعنوان‌های موضوعی در سطح بالایی قرار دارد. مشخص شد که تعداد زیادی از پرس‌و‌جوهای (تا ۶۱٪) دارای سرعنوان‌های موضوعی مرتبط در ده توصیه برتر هستند؛ و به طور متوسط، اولین سرعنوان موضوعی مرتبط در موقعیت اولیه (رتبه سوم) یافت می‌شود. این نشان می‌دهد که روش‌های بازیابی سند می‌توانند به فرآیند اختصاص سرعنوان موضوعی کمک کنند. گوسنباور^۷ (2022) در پژوهش

¹ Yuan & Zou

² Nasar

³ Zhang

⁴ Natural language processing (NLP)

⁵ Phillips

⁶ Yulianti & Rahadianti

⁷ Gusenbauer

خود به ارائه روش جدید علم‌سنجی به منظور تخمین پوشش موضوعی بسیاری از پایگاه‌های داده کتابشناختی محبوب انگلیسی‌زبان پرداخت. این روش از نتایج پرس‌وجو به عنوان یک مخرج مشترک برای مقایسه طیف گسترده‌ای از موتورهای جست‌وجو، مخازن، کتابخانه‌های دیجیتال و سایر پایگاه‌های داده کتابشناختی استفاده می‌کند. یافته‌ها نه تنها تفاوت‌ها در پوشش موضوعی گوگل اسکالر، اسکوپوس یا وب آو ساینس، بلکه همچنین تفاوت‌ها در پایگاه‌های داده‌ای که کمتر مورد تجزیه و تحلیل قرار می‌گیرند را نشان می‌دهد. شائو و یو^۱ (2023) فنون بومی‌سازی خطا مبتنی بر بازیابی اطلاعات به عنوان راه‌حلی برای شناسایی فایل‌هایی که احتمالاً حاوی خطاهایی هستند که علل ریشه‌ای خرابی‌های گزارش شده توسط کاربران هستند، پیشنهاد شده‌اند. در این پژوهش فرایند فرمول‌بندی مجدد پرس‌وجوهای گزارش که حاوی مشکلات هستند، برای شناسایی دقیق‌تر فایل‌های منبع پیشنهاد شده است. ایده اصلی این راه‌کار مبتنی بر یک نمودار دانش جدید است که نشان‌دهنده موجودیت‌های دامنه استخراج شده از گزارش‌های اشکال همزمانی و روابط معنایی آنها است. پس از آن نمودار دانش به پرس‌وجوی مرتبط با بازیابی اطلاعات منابع تبدیل می‌شود تا بومی‌سازی خطا را انجام دهد. در این زمینه از تکنیک‌های پردازش زبان طبیعی و مدل‌سازی مفهومی برای ساخت نمودار دانش استفاده می‌کند. پیکوس و پاسی^۲ (2024) در پژوهشی به بررسی عامل ربط موضوعی در بازیابی اطلاعات پرداختند. آنان با تحلیل مدل‌های ربط چندبعدی، ویژگی دامنه و جنبه‌های ربط متمایز به کار رفته برای تخمین ربط چندبعدی را دسته‌بندی کردند. علاوه بر آن، رویکردهای مورد استفاده برای تجمیع امتیاز مربوط به این عوامل و رتبه‌بندی اقلام اطلاعاتی را مشخص ساختند. نتایج مشخص ساخت که روش‌های یکپارچه برای تخمین عوامل ربط در داخل و بین دامنه‌ها نقش موثری دارد و مجموعه‌های معیار برای ارزیابی‌ها بر اساس جنبه‌های ربط چندگانه در بازیابی اثرگذار است. حسن-مونته‌رو و هررو-سولانا^۳ (2024) در پژوهشی به بررسی بازیابی اطلاعات دیداری مبتنی بر روش برچسب‌گذاری در فضای ابری پرداختند. نتایج این تحقیق نشان داد که سیستم‌های مبتنی بر برچسب‌گذاری، کاربران را قادر می‌سازند تا منابع وب را با استفاده از برچسب‌ها (کلمات کلیدی که آزادانه انتخاب می‌شوند) دسته‌بندی کنند تا بعداً بتوانند این منابع را دوباره بازیابی کنند. برچسب‌گذاری به طور ضمنی یک فرآیند نمایه‌سازی اجتماعی نیز هست، زیرا کاربران برچسب‌ها و منابع خود را به اشتراک می‌گذارند و یک نمایه برچسب اجتماعی، به اصطلاح فولکسونومی، ایجاد می‌کنند. چوپیکووا^۴ (2025) در پژوهش خود به ارزیابی توسعه مدلی از زبان بازیابی اطلاعات در قالب یک طبقه‌بندی‌کننده ترکیبی بین‌رشته‌ای پرداخت. نتایج پژوهش او نشان داد که این ابزار موثر می‌تواند به عنوان مجموعه‌ای از نمادهای به هم پیوسته و پیوندهای چندسلسله‌ای شاخه‌های طبقه‌بندی در حوزه موضوعی «نوآوری‌ها و مدیریت نوآوری» عمل کند. همچنین نتیجه این بررسی حاکی از آن بود که مجموعه کلی کلمات کلیدی و عبارات، و همچنین سرعنوان‌های موضوعی طبقه‌بندی‌کننده، که به آرایه اطلاعات مورد مطالعه مرتبط بودند، موثر

¹ Shao & Yu

² Peikos & Pasi

³ Hassan-Montero & Herrero-Solana

⁴ Chuikova

عمل کردند. پژوهش دورکین و سیرتس^۱ (۲۰۲۵) در باره کاربرد برچسب‌گذاری موضوعی در بازبایی دو مرحله‌ای به انجام رسید. نتایج نشان داد که این شیوه برچسب‌گذاری سبب بازبایی بهتر منابع اطلاعاتی در پایگاه داده می‌شود که در آن از محتوای سند برای بازبایی برچسب‌های موضوعی از یک طبقه‌بندی موضوعی بزرگ استفاده شده است. همچنین مشخص شد که با بهره‌گیری از دو نوع مدل رمزگذار برای ساخت یک سیستم بازبایی اطلاعات دو مرحله‌ای می‌توانیم شرایط مناسبی برای بازبایی اطلاعات به انجام رساند که مبتنی بر فرایند رمزگذاری دوگانه برای استخراج کلیدواژه‌های کلان در مرحله اول و یک رمزگذار متقاطع برای رتبه‌بندی مجدد کلیدواژه‌های خرد در مرحله دوم است. کان^۲ و همکاران (۲۰۲۵) در باره روش‌های نوین و بهره‌گیری از مدل‌های هوش مصنوعی برای انجام وظایف اصلی پردازش هوشمند متون اسناد علمی و فنی به مطالعه پرداختند. نتایج نشان داد که منطق گذار تاریخی از روش‌های زبانی پردازش اسناد به روش‌های مدرن مبتنی بر فناوری‌های هوش مصنوعی، از جمله استفاده از کتابخانه‌های پایتون، نقش مهمی برعهده دارد. در این مطالعه، قدرتمندترین ابزارهایی که به‌طور مؤثر وظایف پردازش موضوعی هوشمند متن را انجام می‌دهند، برجسته شده‌اند. خورشیدی^۳ و همکاران (۲۰۲۶) در بررسی تحلیلی خود که در باره ساختار کتابسنجی مجله کلان‌داده انجام دادند، نشان دادند که شاخص‌های عملکرد را با تکنیک‌های نگاشت علم، از جمله استناد مشترک، اتصال کتاب‌شناختی و هم‌آوایی کلمات کلیدی که در مرورگر وب آو ساینس^۴ و برنامه کاربردی کتابسنجی^۵ پیاده‌سازی شده‌اند، ترکیب می‌کند. همچنین ساختارهای استنادی و موضوعی، سه هسته موضوعی اصلی را نشان می‌دهند که بر زیرساخت‌های کلان‌داده مربوط هستند و از مجموعه‌ای از موضوعات متنوع و حوزه‌های تحقیقاتی نوظهور پشتیبانی می‌کند. دان^۶ و همکاران (۲۰۲۶) در پژوهش خود در باره سیستم مبتنی بر هوش مصنوعی برای طبقه‌بندی مقالات پژوهشی در حوزه دانشگاهی مرتبط و توصیه مقالات مشابه به مطالعه پرداختند. نتایج نشان داد که با استفاده از یک رابط برنامه‌نویسی کاربردی برای استفاده پژوهشگران، این سیستم به طور خودکار طبقه‌بندی موضوعی-رشته‌ای را بهینه‌سازی می‌کند. این سیستم، فراداده‌هایی مانند نقل قول‌ها، اطلاعات نویسنده و کاوش‌های فشرده چون نشانی اینترنتی مستقیم مقاله را تسهیل می‌کند. یک رابط وب کاربرپسند، تعاملات خودجوش را امکان‌پذیر می‌سازد و با تسریع کشف متون و بهبود دقت جست‌وجو، بهره‌وری آموزشی را افزایش می‌دهد. نتایج پیشنهادی نشان می‌دهد که پژوهش‌های انجام یافته در حوزه‌های مختلفی به انجام رسیده است. حوزه‌های موضوعی چون ارتباط با ساماندهی کلیدواژه‌های پایگاه‌های اطلاعاتی، تأثیر شیوه‌های نمایه‌سازی در بازبایی اطلاعات، جایگاه کاربردی کلیدواژه‌ها در نرم‌افزارهای کتابخانه‌ای، همخوانی کلیدواژه‌ها و توصیفگرهای نمایه‌سازی، عبارت‌های جست‌وجوی کاربران، دامنه ربط و نیز تحلیل موضوع باز، بهره‌گیری

¹ Dorkin & Sirts

² Kan

³ Khorshidi

⁴ VOSviewer

⁵ Bibliometrix

⁶ Done

استانداها در کلیدواژه‌های موضوعی، طبقه‌بندی موضوعی در کاربرد جست‌وجوی اطلاعات، مشخص ساختن و اصلاح کلیدواژه‌ها بر اساس نتایج و بازخورد کاربران، بهره‌گیری از سیستم اصطلاح‌نامه و کاربرد کلیدواژه‌ها، رتبه‌بندی موضوعی، تحلیل واژگانی و موضوعی، خوشه‌بندی در ساختار موضوع، استخراج کلمات کلیدی، روش‌های بهینه‌سازی نمایه‌سازی، خلاصه‌سازی متن، روش‌های یادگیری ماشینی، پردازش زبان طبیعی و فیلدهای مبتنی بر کلیدواژه‌های موضوع از زمره تحقیقاتی است که مورد عنایت پژوهشگران بوده است.

روش پژوهش

باتوجه به بررسی دلایل توجیهی استفاده از فیلدهای مرتبط با موضوع در این پژوهش، از روش مطالعه تطبیقی استفاده شده است. بر این اساس با بررسی وبگاه‌ها و پایگاه‌های اطلاعاتی معتبر که به ارائه منابع اطلاعاتی می‌پردازند، و نوع استفاده از فیلدهای مختلفی که در ساختار پایگاه به‌کار برده‌اند، انواع فیلدها مشخص و در یک بررسی تطبیقی استفاده از فیلدهای مرتبط با موضوع مورد ارزیابی قرار گرفته است. همانطور که می‌دانیم یکی از شاخص‌های مهم در ارتباط با مدیریت پایگاه داده بر اساس تحلیل حجم کار از همه درخواست‌های یک سیستم است (Smith & Pant, 1995). بَی بی‌ایر^۱ و همکاران (2003) توجه کاربران به پایگاه اطلاعاتی را با اهمیت ارزیابی می‌کنند. همچنین براساس نظر اوینگ و کنتس^۲ (1997) امکان جست‌وجوی هدایت شده توسط وبگاه و پایگاه اطلاعاتی یکی از معیارهای مهم در انتخاب آن است. علاوه بر آن براساس نظر برازل^۳ و همکاران (2012) عملکردهای مختلف امتیازدهی در پایگاه اطلاعاتی و توجه به نقاط نقاط قوت و ضعف آن دارای اهمیت است. بر مبنای بیان کالویان و هوسیتالو^۴ (2007) داشتن پروتکل‌های مهم برای ارزیابی وبگاه مهم است.

با توجه به معیارهای بیان شده، جامعه آماری پژوهش، حیطه پایگاه‌های اطلاعاتی ایرانی و خارجی علمی است که به ارائه اطلاعات تخصصی می‌پردازند. در این گروه پایگاه‌های مهم و اصلی در خصوص ارائه منابع اطلاعات تخصصی و علمی شناسایی و مورد بررسی قرار گرفت. معیار انتخاب اهمیت پایگاه‌ها، جایگاه آنها در میان جامعه علمی، داشتن مرجعیت برای پژوهشگران، شهرت، اهمیت و به ویژه بالا بودن استفاده، و تاثیرگذاری آنها در فعالیتهای علمی و تخصصی بود. پایه اصلی این کار در سال ۱۴۰۰ انجام گرفت. سپس در سال ۱۴۰۵ یک بار دیگر بازنگری شده و مولفه‌ها بررسی و تحلیل دقیق شد. در بخش پایگاه‌های فارسی ۱۰ پایگاه اصلی شامل: مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری، سازمان اسناد و کتابخانه ملی ج.ا. ایران، علم نت یا جویشگر علمی فارسی، کنسرسیوم محتوای علمی، سامانه پایان‌نامه‌های دانشگاه تهران، سامانه پایان‌نامه‌ها و رساله‌های دانشگاه تربیت مدرس، سامانه پایان‌نامه‌های دانشگاه اصفهان، کتابخانه مرکزی دانشگاه فردوسی مشهد، سیستم یکپارچه کتابخانه‌های دانشگاه آزاد اسلامی (سیکا)، بانک اطلاعات پایان‌نامه‌ها، دانشگاه آزاد اسلامی، تهران مرکز بررسی شد. در بخش پایگاه‌های خارجی ۱۲ پایگاه مورد

¹ Bailly-Baillière

² Ewing & Kuntz

³ Brozell

⁴ Kälviäinen & Uusitalo

ارزیابی قرار گرفت. این پایگاه‌ها عبارتند از: پایان‌نامه‌ها و رساله‌های الکترونیکی دانشگاه فلوریدا^۱، رساله‌های دانشگاه ام آی تی^۲، مخزن فناوری هوشمند دانشگاه جورجیا^۳، مرکز رساله‌ها و پایان‌نامه‌های الکترونیکی دانشگاه ایالتی اوهایو^۴، پایگاه دسترسی آزاد به رساله‌ها و پایان‌نامه‌ها^۵، پایگاه خدمات برخط رساله‌های الکترونیکی کتابخانه بریتانیا^۶، پایگاه رساله‌های کانادا^۷، پایگاه رساله‌های دیجیتال استرالیا^۸، پایگاه رساله‌های دانشگاه هنگ کنگ^۹، پایگاه رساله‌های فرانسه^{۱۰}، پروکوئست^{۱۱}، و نیز وب آو ساینس^{۱۲}. در مجموع ۲۲ پایگاه اطلاعاتی فارسی و خارجی مورد بررسی و تحلیل قرار گرفت. نکته مهم در انتخاب جامعه آماری این است که برای سنجش وضعیت فیلد موضوع صرفاً به پایگاه‌هایی که به ارائه رساله و یا پایان‌نامه می‌پردازند بسنده نشد و تعدادی از پایگاه‌ها که منابع اطلاعاتی پژوهشی چون مقاله‌ها، طرح‌های پژوهشی، کنفرانس‌ها، فصول کتاب و نظایر آن را ارائه می‌دهند نیز مورد بررسی قرار گرفت. این امر از آن رو دارای اهمیت است که جامعیت بررسی را افزایش داده و رویکردهای متعدد پایگاه‌های اطلاعاتی مهم و پراستفاده شناخته می‌شود. همچنین مقایسه‌ای هم میان آنها انجام شده و نقاط افتراق و اشتراک آنها مشخص شده است. در این میان علاوه بر فیلد موضوع، موارد دیگر از جمله: کلیدواژه، رشته تحصیلی، چکیده، الگوی کاوش گوگل و نظایر آن نیز مورد بررسی قرار گرفت تا مفهوم موضوع در پژوهش تقویت شود. همانطور که می‌دانیم الگوی کاوش گوگل، در بخش جست‌وجوی تخصصی شامل موارد چهارگانه: با تمام کلمات؛ با عین عبارت؛ با حداقل یکی از کلمات؛ و بدون کلمات، مشخص می‌شود که از طریق به‌کارگیری منطق بولی، و بهره‌گیری از پالایش عبارت‌های کاوش به بازایی دقیق و منطبق با نیاز کاربر اطلاعاتی منجر می‌شود.

روش گردآوری اطلاعات، مراجعه مستقیم به پایگاه‌های اطلاعاتی و مطالعه وضعیت آنها از نظر دارا بودن فیلدهای اطلاعاتی است. در این روش با توجه به معیارهای ذکر شده در خصوص پایگاه‌های اطلاعاتی ابتدا مواردی از جمله حجم اطلاعات پایگاه داده، وضعیت کاربران در بهره‌گیری از پایگاه، جایگاه پایگاه و بهره‌مندی از اطلاعات تخصصی مورد توجه قرار گرفت. با عنایت به معیارهای مطرح شده پس از شناسایی پایگاه‌های مورد نظر، فیلدهای اطلاعاتی هر پایگاه شناسایی شد. سپس این فیلدها به صورت مدون ثبت و وضعیت هر پایگاه از نظر دارا بودن فیلدهای اطلاعاتی تعیین گردید. از میان فیلدهای مورد نظر فیلد موضوع که به صورت مشخص مربوط به انجام این پژوهش بود، تعیین شده و وضعیت و کاربرد آن مورد بررسی قرار

¹ Electronic Thesis or Dissertation, University of Florida

² MIT University, Theses

³ SMARTech Repository (Georgia University)

⁴ Electronic Thesis or Dissertation Center, Ohio State University

⁵ Open Access Theses and Dissertations

⁶ Electronic Theses Online Service, a British Library

⁷ Theses Canada

⁸ Australasian Digital Theses

⁹ Theses of The University of Hong Kong

¹⁰ Theses.fr

¹¹ ProQuest

¹² Web Of Science

گرفت. در واقع باید بیان کرد که داده‌های حاصل از پژوهش به شیوه میدانی و از طریق مراجعه به وبگاه‌ها و پایگاه‌های اطلاعاتی به دست آمده است.

یافته‌های پژوهش

در پاسخ به پرسش پژوهش یعنی بهره‌گیری از فیلد موضوع تا چه اندازه و چگونه در پایگاه‌های اطلاعاتی توجیه‌پذیر است؟ (فیلد موضوع تا چه اندازه و چگونه در سایر پایگاه‌های داخلی و خارجی استفاده شده‌اند؟) به بررسی پایگاه‌های اطلاعاتی، در دو بخش پایگاه‌های فارسی و خارجی که دارای پایان‌نامه و رساله هستند پرداخته شده است. با توجه به اطلاعات به دست آمده از وضعیت پایگاه‌های اطلاعاتی فارسی و نیز خارجی مشخص می‌شود که جایگاه موضوع چگونه است. همچنین توجه به برخی فیلدهای دیگر که می‌تواند از نظر مفهومی دارای ارتباط با این فیلدها باشد مورد بررسی قرار می‌گیرد.

وضعیت پایگاه‌های فارسی

با توجه به اطلاعات به دست آمده از وضعیت پایگاه‌های اطلاعاتی فارسی مشخص می‌شود که جایگاه موضوع چگونه است. همچنین توجه به برخی فیلدهای دیگر که می‌تواند از نظر مفهومی دارای ارتباط با این فیلدها باشد مورد بررسی قرار می‌گیرد.

جدول ۱. بررسی فیلدهای کاوش پایگاه‌های اطلاعاتی فارسی حاوی پایان‌نامه و رساله

نام پایگاه	موضوع	رشته تحصیلی	کلیدواژه	طبقه‌بندی موضوعی	پدیدآور	عنوان	الگوی کاوش گوگل	نوع مدرک	تاریخ نشر	اعتبار مدرک	چکیده	مقطع تحصیلی
مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗
سازمان اسناد و کتابخانه ملی	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗
علم نت	✓	✗	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗
کنسرسیوم محتوای ملی	✓	✗	✓	✗	✓	✓	✗	✓	✓	✗	✗	✗
سامانه پایان‌نامه‌های دانشگاه تهران	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓
سامانه پارساهای دانشگاه تربیت مدرس	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓

نم پایگاه	موضوع	رشته تحصیلی	کلیدواژه	طبقه‌بندی موضوعی	پدیدآور	عنوان	الگوی کاوش گوگل	نوع مدرک	تاریخ نشر	اعتبار مدرک	چکیده	مقطع تحصیلی
سامانه پایان‌نامه‌های دانشگاه اصفهان	✓	✓	✓	×	✓	✓	×	✓	✓	×	✓	✓
کتابخانه مرکزی دانشگاه فردوسی مشهد	✓	×	✓	×	✓	✓	✓	✓	✓	×	✓	×
سیکا ^۱	✓	×	✓	×	✓	✓	×	✓	✓	×	✓	×
بانک پارساهای دانشگاه آزاد تهران مرکز	×	✓	✓	×	✓	✓	×	×	✓	×	✓	✓
مراکزی که از این فیلد استفاده می‌کنند	۹	۶	۱۰	۰	۱۰	۱۰	۳	۸	۹	۱	۸	۴
درصد	۹۰	۶۰	۱۰۰	۰	۱۰۰	۱۰۰	۳۰	۸۰	۹۰	۱۰	۸۰	۴۰

بررسی جدول ۱ در ارتباط با تحلیل وضعیت بهره‌گیری از فیلدهای مختلف ۱۰ پایگاه فارسی که دارای پایان‌نامه و رساله هستند نشان می‌دهد که در کل ۹۰ درصد از این پایگاه‌ها از فیلد موضوع استفاده می‌کنند، و همچنین ۶۰ درصد از آنها از فیلد رشته تحصیلی بهره می‌گیرند. در خصوص سایر فیلدهای مورد استفاده، به ترتیب میزان بهره‌گیری از کلیدواژه ۱۰۰ درصد، پدیدآور ۱۰۰ درصد، عنوان ۱۰۰ درصد، تاریخ نشر ۹۰ درصد، نوع مدرک ۸۰ درصد، چکیده ۸۰ درصد، مقطع تحصیلی ۴۰ درصد، الگوی کاوش گوگل ۳۰ درصد، و اعتبار مدرک ۱۰ درصد است. علاوه بر آن، هیچ پایگاهی از طبقه‌بندی موضوعی استفاده نمی‌کند.

وضعیت پایگاه‌های خارجی

با توجه به اطلاعات به دست آمده از وضعیت پایگاه‌های اطلاعاتی خارجی نیز مشخص می‌شود که وضعیت به‌کارگیری فیلدهای موضوع و چگونه است. همچنین توجه به برخی فیلدهای دیگر که می‌تواند از نظر مفهومی دارای ارتباط با این فیلدها باشد، مورد بررسی و مطالعه قرار می‌گیرد. فیلدهای دیگر همواره به عنوان مکمل موضوع در نظر گرفته می‌شود که می‌تواند در بازایی منابع اطلاعاتی سودمند باشد. به عنوان مثال: فیلدهای کلیدواژه، طبقه‌بندی موضوعی، الگوی کاوش گوگل، و چکیده می‌توانند در این راستا

^۱ سیستم یکپارچه کتابخانه‌های دانشگاه آزاد اسلامی

محسوب شوند. همچنین فیلدهای دیگر برای پوشش کامل اطلاعات و دسترسی بهتر به منابع طراحی می‌شوند.

جدول ۲. بررسی فیلدهای کاوش پایگاه‌های اطلاعاتی خارجی حاوی پایان‌نامه و رساله

نام پایگاه	موضوع	رشته تحصیلی	کتابخانه	طبقه‌بندی موضوعی	پدیدآور	عنوان	الگوی کاوش کوئل	نوع مدرک	تاریخ نشر	اعتبار مدرک	چکیده	مبلغ تحصیلی
پارساهای الکترونیکی دانشگاه فلوریدا	✓	✗	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓
رساله‌های دانشگاه ام آی تی	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓
مخزن فناوری هوشمند دانشگاه جورجیا	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓
مرکز پارساهای، دانشگاه ایالتی اوهایو	✓	✗	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗
پایگاه دسترسی آزاد به پارساها	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓
پایگاه پارساها یکتابخانه بریتانیا	✓	✗	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗
پایگاه رساله‌های کانادا	✓	✗	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗
پایگاه رساله‌های دیجیتال استرالیا	✓	✗	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗
پایگاه رساله‌های دانشگاه هنگ کنگ	✓	✗	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓
پایگاه رساله‌های فرانسه	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓
پایگاه پروکوئست	✓	✗	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓
پایگاه وب آو ساینس	✓	✗	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗

۷	۱۲	۰	۱۲	۱۲	۲	۱۲	۱۲	۸	۱۲	۴	۱۲	مراکزی که از این فیلد استفاده می‌کنند
۵۸/۳۱	۱۰۰	۰	۱۰۰	۱۰۰	۱۶/۶۶	۱۰۰	۱۰۰	۳۲	۱۰۰	۳۳/۳۳	۱۰۰	درصد

داده‌های جدول ۲ نشان می‌دهد که در میان ۱۲ پایگاه خارجی که دارای رساله و پایان‌نامه هستند استفاده از فیلد موضوع به میزان ۱۰۰ درصد انجام می‌پذیرد و در ارتباط با بهره‌گیری از فیلد رشته تحصیلی ۳۳/۳۳ درصد از این پایگاه‌ها از این امکان استفاده می‌کنند. همچنین میزان بهره‌مندی از فیلدهای کلیدواژه، پدیدآور، عنوان، نوع مدرک، تاریخ نشر و چکیده به میزان ۱۰۰ درصد است. علاوه بر آن، استفاده از فیلد مقطع تحصیلی ۵۸/۳۱ درصد و طبقه‌بندی موضوعی ۳۲ درصد است.

مقایسه اجمالی پایگاه‌های فارسی و خارجی

در مقایسه اجمالی پایگاه‌های فارسی و خارجی نتایج به دست آمده به شرح زیر است:

جدول ۳. مقایسه کلی وضعیت پایگاه‌های اطلاعاتی فارسی و خارجی

پایگاه‌ها	موضوع	رشته تحصیلی	کلیدواژه	طبقه‌بندی موضوعی	پدیدآور	عنوان	مؤلف	الگوی کاوش	نوع مدرک	تاریخ نشر	اعتبار مدرک	چکیده	مقطع تحصیلی
پایگاه‌های اطلاعاتی فارسی (درصد استفاده از فیلدها)	۹۰	۶۰	۱۰۰	۰	۱۰۰	۱۰۰	۳۰	۸۰	۹۰	۱۰	۸۰	۴۰	
پایگاه‌های اطلاعاتی خارجی (درصد استفاده از فیلدها)	۱۰۰	۳۳/۳۳	۱۰۰	۳۲	۱۰۰	۱۰۰	۱۶/۶۶	۱۰۰	۱۰۰	۱۰۰	۰	۱۰۰	۵۸/۳۱

بر اساس مقایسه میان پایگاه‌ها بر اساس اطلاعات جدول ۳ ملاحظه می‌شود که در سطح تمامی پایگاه‌ها بهره‌گیری از موضوع در بین پایگاه‌های خارجی بیشتر از پایگاه‌های فارسی است. در ارتباط با کاربرد کلیدواژه شرایط تمامی پایگاه‌ها یکسان است و همگی به میزان ۱۰۰ درصد از این فیلد استفاده کرده‌اند. در ارتباط

با طبقه‌بندی موضوعی پایگاه‌های فارسی از این امکان استفاده نمی‌کنند و نتایج بهره‌گیری از این فیلد کاملاً منتفی است و هیچ پایگاهی در این گروه از طبقه‌بندی موضوعی استفاده نمی‌کند. اما در میان پایگاه‌های اطلاعاتی خارجی ۳۲ درصد از طبقه‌بندی موضوعی بهره می‌گیرند. نتایج نشان می‌دهد که پایگاه‌های اطلاعاتی خارجی نسبت به این فیلد توجه بیشتری دارند. همچنین در خصوص بهره‌گیری از الگوی کاوش گوگل، پایگاه‌های اطلاعاتی فارسی به میزان ۳۰ درصد از این امکان استفاده می‌کنند. در میان پایگاه‌های اطلاعاتی خارجی میزان استفاده از این فیلد ۱۶/۶۶ درصد است. همانطور که ملاحظه می‌شود پایگاه‌های اطلاعاتی فارسی به میزان بیشتری از الگوی کاوش گوگل بهره می‌برند. در خصوص بهره‌گیری از فیلد چکیده در میان پایگاه‌های اطلاعاتی فارسی به میزان ۸۰ درصد از آن بهره می‌گیرند. ولی در میان پایگاه‌های اطلاعاتی خارجی به میزان ۱۰۰ درصد از فیلد چکیده بهره می‌گیرند.

در ارتباط با دیگر فیلدها مشاهده می‌شود که امکان استفاده از فیلد پدیدآور و عنوان در تمامی پایگاه‌ها به میزان ۱۰۰ درصد است. همچنین در زمینه تاریخ نشر، نوع مدرک و مقطع تحصیلی پایگاه‌های اطلاعاتی خارجی از پایگاه‌های اطلاعاتی فارسی بیشتر استفاده می‌کنند. ولی در زمینه اعتبار مدرک پایگاه‌های اطلاعاتی فارسی به میزان بیشتری نسبت به پایگاه‌های اطلاعاتی خارجی بهره می‌برند.

بحث و تحلیل

همانطور که یافته‌ها نشان داد در پایگاه‌های اطلاعاتی خارجی در کل وضعیت بهره‌گیری از موضوع در سطح بالاتر است که نشانگر توجه ویژه به فیلد موضوع است که نتایج دیگر فیلدهای کمکی حاکی از آن است که اساساً روش الگوی کاوش گوگل که در پایگاه‌های فارسی مورد توجه بیشتر بوده است، چندان در این پایگاه‌ها استفاده نمی‌شود. البته توجه به طبقه‌بندی موضوعی بیش از الگوی ذکر شده است و دلیل آن می‌تواند پیشنهادهایی بر اساس نیاز از قبل تعیین شده کاربر باشد. در واقع بهره‌گیری از نتایج رتبه‌بندی موضوعی مورد عنایت قرار گرفته است. این یافته با نتایج پژوهش یوان و زو (۲۰۱۸) همگونی دارد. همچنین در خصوص بهره‌گیری از الگوی کاوش گول پژوهش‌های متعددی به انجام رسیده که اهمیت و سומندی آن را از جنبه‌های مختلف بررسی کرده‌اند. چنانچه علاء^۱ (۲۰۱۶) نقش این الگو را از جنبه جذب بیشتر کاربر در تبلیغات و نیازهای اطلاعاتی موثر می‌داند. همچنین بیتیریم^۲ (۲۰۲۲) در خصوص نقش الگوی کاوش گوگل در بازیابی تصاویر به نتایج مثبتی در این زمینه دست یافت. علاوه بر آن پژوهش ليو^۳ (۲۰۱۵) نیز نشان داد که در زمینه دسترسی به اخبار نیز کاوش گوگل تا حد قابل توجهی موفقیت دارد.

نتیجه مهمی که می‌شود از طراحی وضعیت موضوعی پایگاه‌های خارجی گرفت آن است که آنها به تخصیص فیلد موضوع به صورت مشخص تمرکز کرده‌اند و معتقدند که این روش نتایج بهتری به همراه دارد. یکی از نکات مهم استفاده از فیلد کلیدواژه است که با بررسی انجام گرفته در تمامی پایگاه‌های خارجی، اطلاعات موجود در فیلد موضوع از طریق کلیدواژه قابل بازیابی است و این امر کمکی به بازیابی بهتر مفاهیم

^۱ Ala'a

^۲ Bitirim

^۳ Liu

مورد نظر کاربران می‌کند که با رویکرد نمایه‌سازی ماشینی به انجام می‌رسد. این نتایج با یافته‌های پژوهش ناصر و همکاران (2019) و گوسنباوئر (2022) همسویی دارد. در این پایگاه‌ها همچنین به دلیل نقش موثر چکیده در بازیابی اطلاعات، به میزان ۱۰۰ درصد از آن بهره گرفته‌اند. می‌دانیم که همواره در کاوش کلیدواژه‌ای که پوشش دهنده موضوع نیز هست، اطلاعات چکیده می‌تواند حاوی داده‌های ارزشمندی برای کاربر باشد.

نکته‌ای که در خصوص سنجش وضعیت اعتبار مدارک وجود دارد آن است که برخلاف پایگاه‌های ایرانی، در پایگاه‌های خارجی اساساً سنجش اعتبار مدارک در سطح پایینی قرار گرفته و شاید به این دلیل باشد که پیش فرض پایگاه‌های اطلاعاتی خارجی دارا بودن اعتبار مناسب برای مدارکی است که وارد این پایگاه‌ها شده است. همچنین این پایگاه‌ها به تقسیم‌بندی‌های اعتباری که در پایگاه‌های فارسی وجود دارد توجه ندارند. همچنین در ارتباط با طبقه‌بندی موضوعی به عنوان فیلدی کمکی برای بازیابی موضوع می‌توان بیان کرد، از آنجا که حیطه موضوعی برخی از پایگاه‌های خارجی گسترده است، لذا طبقه‌بندی موضوعی از همان ابتدا می‌تواند مسیر کاوش را تعیین کند. این بدان معناست که قبل از شروع کاوش در فیلد موضوع می‌توان به شیوه طبقه‌بندی اولیه مسیر کلان را به سمت موضوع‌های فرعی پیمود. تقریباً شبیه همان کاری که معمولاً در فولکسونومی انجام می‌پذیرد. لذا وجود طبقه‌بندی موضوعی از این حیث یاری رسان فیلد موضوع است که تا حد زیادی نیز با فنون پردازش زبان طبیعی تقویت می‌شود این یافته با نتایج پژوهش ژنگ و همکاران (2020) دارای همسویی است. از سوی دیگر استفاده از امکان ابرداده در پیوند موضوعی نقش مهمی برای دستیابی به اطلاعات کلیدواژه‌های موضوعی دارد. این مورد با نتایج پژوهش فیلیپ و همکاران (2020) دارای اشتراک است. علاوه بر آن بهره‌گیری از ساختار استنادی در کاربردهای موضوعی، می‌تواند نقش مهمی در بهبود کاوش اطلاعات داشته باشد. این مقوله با نتایج پژوهش خورشیدی و همکاران (2026) دارای همسویی است.

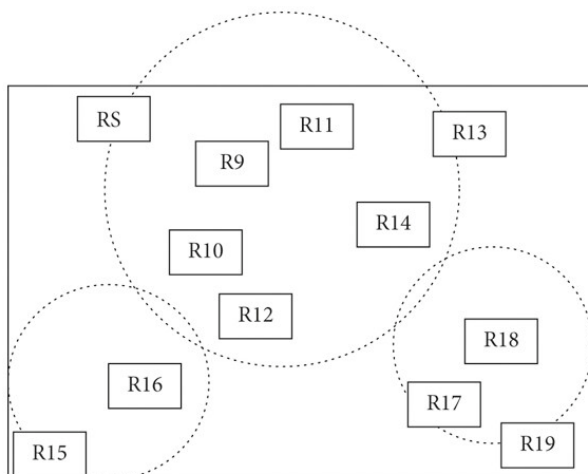
پایگاه‌های اطلاعاتی فارسی نسبت به پایگاه‌های اطلاعاتی خارجی از شیوه الگوی کاوش گوگل بیشتر بهره می‌گیرند. این امکان با توجه به تکیه بر منطق بولی مورد عنایت این پایگاه‌ها بوده است. در پایگاه‌های خارجی بیشتر تفکیک بر اساس منطق بولی در فیلدهای مختلف محور توجه بوده است. هر دو روش می‌توانند زمینه مناسبی برای بازیابی بهتر اطلاعات مربوط به موضوع رساله و پایان‌نامه باشند. همچنین روش‌های خوشه‌بندی موضوعی نیز در سیستم اطلاعاتی پایگاه داده پشتیبان الگوهای کاوش است. این یافته با نتایج پژوهش چاوداری و شارما (2017) دارای اشتراک است. علاوه بر آن، موضوع در طبقه‌بندی موضوعی نیز از اهمیت بالایی برخوردار است. بهره‌گیری از مدل‌های زبانی هوشمند نیز می‌تواند در این میان موثر باشد. این یافته با نتایج پژوهش چویکوا (2025) دارای همسویی است. می‌توان بیان کرد که امکان طبقه‌بندی موضوعی انتخاب کاربر را تسریع می‌کند و نیز به نوعی پیشنهاد استفاده از موضوع نیز هست. بر این اساس هر گونه دستیابی بهتر به موضوع با استفاده از فیلدهای کمکی شرایط بهتری برای استفاده از مفهوم اصلی منابع رساله و پایان‌نامه است. این مورد با نتایج پژوهش سومین و تواوانن (2016) دارای اشتراک است.

با مطالعه وضعیت کاربرد موضوع در پایگاه‌های اطلاعاتی فارسی و خارجی می‌توان بیان کرد که استفاده از فیلد موضوع کاملاً ضروری است و تردید در آن وجود ندارد. الگوهایی مانند طرح کاوش گوگل که در بسیاری از پایگاه‌ها مورد بهره‌برداری قرار می‌گیرد، می‌تواند موجب غنی‌سازی فرایند کاوش در دستیابی به اطلاعات موضوعی رساله‌ها و پایان‌نامه‌ها شود. علاوه بر آن بهره‌گیری از امکانات هوش مصنوعی در سازماندهی و تحلیل اطلاعات موضوع، نقش مهمی در ساختار اطلاعاتی پایگاه‌های معتبر پیدا کرده است. نتایج پژوهش کان و همکاران (2025) و نیز دان و همکاران (2026) در این زمینه با نتایج پژوهش حاضر دارای هم‌سویی است. علاوه بر آن، بررسی‌های انجام شده بر اساس پایگاه‌ها و وبگاه‌های اطلاعاتی که به ارائه منابع اطلاعات علمی می‌پردازند نشان می‌دهد که تعداد زیادی از این پایگاه‌های در زبان فارسی و انگلیسی از روش‌های مشخص پیروی کرده و بر اساس مولفه‌های نزدیک به هم به کاوش و بازیابی اطلاعات می‌پردازند. انواع فیلدهای مختلف شامل: نویسنده، عنوان، ناشر، تاریخ نشر، شابک، شابا، نوع ماده اطلاعاتی، دی او آی، بهره‌گیری از منطق بولی، زبان اثر، نام موسسه تولیدکننده مدرک، نام مجله، مدارک استناد داده و دیگر فقره‌های اطلاعاتی در این پایگاه‌ها وجود دارد.

بر اساس تحلیل انجام گرفته، راهکار ارائه شده در بازیابی بهتر اطلاعات در پایگاه‌های اطلاعاتی بر اساس ساختار موضوعی و سازماندهی اطلاعات مرتبط، استفاده از مدل خوشه‌بندی پویای درخت- R و الگوریتم پیشنهادی آن بر اساس پژوهش دی و لیو^۱ (2022) است. در این مدل، گره‌های موضوعی در ساختار پردازش هدف نهایی، یعنی سازماندهی بر اساس روش درختی و سلسله‌مراتب موضوعی است. علاوه بر آن، از الگوریتم خوشه‌بندی برای سازماندهی موضوعی استفاده می‌شود تا بهینه‌سازی پویای مدل خوشه‌بندی درخت- R در مقیاس بزرگ محقق شود. برای تولید مدل خوشه‌بندی درخت- R برای هر مجموعه داده مکانی، فرآیند اصلی به شرح زیر است: ابتدا، مستطیل مرزی حداقل برای همه اشیاء مکانی (برچسب موضوعی منابع اطلاعاتی) در قالب موضوع ایجاد می‌شود. سپس، مستطیل پایه طبق الگوریتم خوشه‌بندی پویای سلول^۲ گروه‌بندی می‌شود. به عنوان مثال، در نمودار ۱، ابتدا R12 را که در واقع به نقطه میانگین نزدیک است به عنوان مرکز خوشه‌بندی اولیه انتخاب می‌کنیم. سپس، R19 را که دورترین فاصله را از R12 دارد و R8 را که دورترین فاصله را از R19 دارد، به عنوان مراکز خوشه‌بندی انتخاب می‌کنیم و فرایند خوشه‌بندی را شروع می‌کنیم (نمودار ۱).

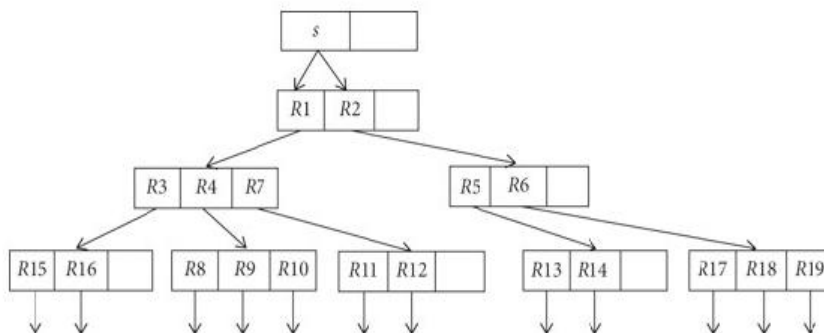
¹ Dai & Liu

² Dynamic Cell Clustering Algorithm



نمودار ۱. آماده‌سازی فرایند خوشه‌بندی براساس روش دی و لیو

در مرحله دوم که قصد خوشه‌بندی موضوعی نهایی را داریم، R13، R14، R17 و R18 به R19 و R9، R10، R11، R12، R15 و R16 به R8 تقسیم می‌شوند. در این زمان، دو خوشه تشکیل می‌شوند و مرکز خوشه به همراه تابع اندازه‌گیری خوشه محاسبه می‌شوند. متعاقباً، خوشه‌ای را که بزرگترین شعاع و مرکز خوشه آن R12 را از بین دو خوشه دارد انتخاب می‌کنیم. سپس، R15 را که از R11، R12 و R18 که از R15 دورترین هستند، به عنوان مرکز خوشه برای خوشه‌بندی مجدد موضوع انتخاب می‌کنیم. پس از آن، مرکز خوشه و تابع اندازه‌گیری خوشه آن را محاسبه می‌کنیم. در یک چرخه، همگرایی تابع خوشه‌بندی افزایش می‌یابد. در نهایت، یک مستطیل شامل اشیاء کاملاً فضایی در کل فضای مجموعه منابع اطلاعاتی در خصوص سازماندهی اطلاعات برای به‌دست آوردن مدل خوشه‌بندی درخت R در ساختار موضوعی شکل می‌گیرد (نمودار ۲).



نمودار ۲. انجام خوشه‌بندی براساس روش دی و لیو

در تحلیل نهایی این مدل خوشه‌بندی برای سازماندهی موضوع در پایگاه اطلاعاتی باید بیان کرد که ساختار شاخص مدل خوشه‌بندی درخت -R، لایه به لایه، از بالا به پایین از طریق تعیین پویای مرکز خوشه‌بندی با استفاده از الگوریتم پیشنهادی ساخته می‌شود. این نظام خوشه‌بندی پیشنهادی می‌تواند با استقرار مدل خوشه‌بندی درخت -R کارایی بازیابی اطلاعات را به طور مؤثر بهبود بخشد. بهتر است که با بهره‌گیری از این مدل، پایگاه داده از خدمات‌دهنده ابری^۱ نیز استفاده کند.

نتیجه‌گیری

با عنایت به جایگاه خاص مفاهیم منابع اطلاعاتی و بازیابی آنها در پایگاه‌داده، فیلدهای اطلاعاتی پایگاه‌های داده بر اساس نقش موضوع و زمینه‌های وابسته به آن از جمله کلیدواژه، چکیده، الگوی کاوش گوگل، رشته تحصیلی و ... مورد بررسی قرار گرفت. نتایج نشان داد که کاربرد مناسب و دقیق فیلد موضوع در پایگاه‌های اطلاعاتی نقش مهمی در ساختار بازیابی اطلاعات دارد. در این راستا هرگونه بهره‌گیری از الگوهای شناخته‌شده و مناسب در سازماندهی اطلاعاتی مبتنی بر موضوع و ساختار مدون آن، می‌تواند در نظام‌های کاوش دقیق اطلاعات سودمند باشد. همچنین مناسب آن است که طبقه‌بندی‌های موضوعی در قالب سلسله‌مراتبی، استنادی، خوشه‌بندی به انجام رسد و ساختار نمایه‌های ایجاد شده، بر اساس استانداردهای نمایه‌سازی نوین شکل بگیرد. بهره‌گیری از امکانات هوش مصنوعی در تنظیم و سازماندهی موضوعی نقش مهمی برعهده دارد. هوش مصنوعی قابلیت موثری برای نمایه‌سازی و اختصاص واژگان کلیدی از خود نشان داده است. همچنین، توجه به بسامد اصطلاح‌ها قابلیت نمایه‌سازی را افزایش می‌دهد. بهره‌گیری از الگوریتم خوشه‌بندی درخت -R راهکار پیشنهادی برای پردازش نظام موضوعی در پایگاه داده است که می‌تواند به بهینه‌سازی ساختار موضوعی در آن منجر شود. علاوه بر آن، بهره‌گیری از پاره‌ای از امکانات مانند الگوی کاوش گوگل و نیز نمایه‌سازی معنایی پنهان می‌تواند تا حد زیادی در این حوزه مؤثر باشد. در تمامی موارد یاد شده، بهره‌گیری از توانایی نیروی انسانی متخصص و کارآمد برای بهینه‌سازی ساختار موضوعی در پایگاه‌های اطلاعاتی سبب توسعه این فرایند شده و می‌تواند برخی از مشکلات را مرتفع سازد. همواره استخراج ماشینی موضوع‌ها و تنظیم الفبایی از میان اسناد و مدارک اطلاعاتی و سپس انطباق با سیاهه الفبایی بانک واژگان موضوعی در پایگاه‌های اطلاعاتی مبتنی بر اصطلاح‌نامه‌ها و واژگان اختصاصی نیز یکی از راهکارهای مهم در بهبود نظم موضوع به شمار می‌رود. با توجه به یافته‌های پژوهش پیشنهاد می‌شود:

واژگانی که به صورت غیر اصطلاحنامه‌ای در پایگاه‌های اطلاعاتی استفاده می‌شوند مورد بازنگری قرار گیرند. معادل مناسب در اصطلاحنامه و یا سرعنوان موضوعی داشته باشند.

روابط معنایی و ارتباط میان واژگان از نظر ارتباط سلسله‌مراتبی، اعم و اخص، ارتباط ترجیحی، رابطه همبستگی، تعاریف دقیق و جز آن به دقت در پایگاه‌های اطلاعاتی تعیین شوند.

¹ Cloud Server

از روش طبقه‌بندی موضوعی در پایگاه برای مشخص ساختن حیطه موضوعی و دسترسی به اطلاعات فیلد موضوع استفاده شود. در این روش مناسب آن است که موضوع‌های استاندارد مبتنی بر سرعنوان و تزاروس شناسایی و نسبت به تعیین موضوع اعم و اخص اقدام مناسب انجام گیرد. سپس طبقه‌بندی موضوعی با توجه به موضوع‌های اصلی و فرعی مشخص و از موضوع‌های اصلی در ساختار کلان طبقه‌بندی موضوعی استفاده شود.

در بخش کاوش اطلاعات پایگاه‌های اطلاعاتی از الگوی کاوش گوگل بهره گرفته شود (گرچه در پایگاه‌های اطلاعاتی خارجی به صورت مستقیم از الگوی یاد شده استفاده نشده است، ولی در ساختار پایگاه داده، با بهره‌گیری از منطق بولی، به نوعی از این الگو بهره گرفته شده است). راهکار مناسب برای انتخاب موضوع دقیق‌تر، می‌تواند بسامد کاربرد اصطلاح‌های مهم در کل متن، در کنار کلیدواژه‌های استاندارد مبتنی بر اصطلاح‌نامه نیز باشد.

در عرصه استخراج کلیدواژه‌های موضوعی اصلی و نیز موضوع‌های فرعی در متن، بر اساس الگوریتم‌های داده‌کاوی می‌تواند فواید قابل توجهی داشته باشد.

منابع

- Ala'A, A. (2016). The impact of using Google search engine and Yahoo Bing network on attracting new customers online and on the effectiveness of online advertisements. *International Journal of Marketing Studies*, 8(3), 163-172. <https://doi.org/10.5539/ijms.v8n3p163>
- Bailly-Baillié, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Mariétoz, J., Matas, J., Messer, K., Popovici, V., Porée, F., Ruiz, B., & Thiran, J. P. (2003). The BANCA database and evaluation protocol. In *International conference on Audio-and video-based biometric person authentication* (pp. 625-638). Springer, Berlin, Heidelberg.
- Bitirim, Y. (2022). Retrieval effectiveness of Google on reverse image search. *Journal of Imaging Science and Technology*, 66(1), 1-6. <https://doi.org/10.2352/J.ImagingSci.Technol.2022.66.1.010505>
- Brozell, S. R., Mukherjee, S., Balius, T. E., Roe, D. R., Case, D. A., & Rizzo, R. C. (2012). Evaluation of DOCK 6 as a pose generation and database enrichment tool. *Journal of computer-aided molecular design*, 26(6), 749-773. <https://doi.org/10.1007/s10822-012-9565-y>
- Chaudhary, H., & Sharma, A. (2017). Bio-medical mining and summarization using K-means and lexical analysis. *International Journal for Technological Research in Engineering*, 4(9), 1753-1757.
- Chuikova, N. A. (2025). Interdisciplinary Synthetic Classifier in the Subject Area Innovations and Innovation Management. *Scientific and Technical Information Processing*, 52(2), 110-118. <https://doi.org/10.3103/S0147688225700145>
- Dai, Y., & Liu, H. (2022). Application of the R-Tree Clustering Model in Medical Information Retrieval. *Mobile Information Systems*, 2022(1), 2887395. <https://doi.org/10.1155/2022/2887395>
- De Lange, N. (2023). Data Organisation and Database Systems. In *Geoinformatics in Theory and Practice: An Integrated Approach to Geoinformation Systems, Remote Sensing and Digital Image Processing* (pp. 329-373). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-65758-4_8
- Done, S. L., Schandra, J., Rishitha, G., & Rani, M. J. (2026). Automated Research Paper Classification and Recommendation System for Efficient Literature Discovery. In: Gunjan, V. K., & Zurada, J. M. (eds), *Proceedings of 6th International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications. ICMISC 2025. Lecture Notes in Networks and Systems*, vol 1736. Springer, Singapore. https://doi.org/10.1007/978-981-95-5079-1_21

- Dorkin, A., & Sirts, K. (2025). Tartunlp at semeval-2025 task 5: Subject tagging as two-stage information retrieval. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)* (pp. 2449-2454).
- 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*, 60(2), 78-87.
- Elias, F., & Nguyen, F. (2020). *U.S. Patent No. 10,650,021*. Washington, DC: U.S. Patent and Trademark Office.
- Ewing, T. J., & Kuntz, I. D. (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. *Journal of computational chemistry*, 18(9), 1175-1189.
- Gusenbauer, M. (2022). Search where you will find most: Comparing the disciplinary coverage of 56 bibliographic databases. *Scientometrics*, 127(5), 2683-2745. <https://doi.org/10.1007/s11192-022-04289-7>
- Hassan-Montero, Y., & Herrero-Solana, V. (2024). Improving tag-clouds as visual information retrieval interfaces. *arXiv preprint arXiv*, 2401.04947. <https://doi.org/10.48550/arXiv.2401.04947>
- Jiang, M., D'Souza, J., Auer, S., & Downie, J. S. (2022). Evaluating BERT-based scientific relation classifiers for scholarly knowledge graph construction on digital library collections. *International Journal on Digital Libraries*, 23(2), 197-215. <https://doi.org/10.1007/s00799-021-00313-y>
- Kalachikhin, P. A. (2025). Qualimetry of Hierarchical Classifiers of Fields of Science. *Scientific and Technical Information Processing*, 52(1), 53-62. <https://doi.org/10.3103/S0147688225700042>
- Kälviäinen, R. V. J. P. H., & Uusitalo, H. (2007). DIARETDB1 diabetic retinopathy database and evaluation protocol. In *Medical image understanding and analysis* (Vol. 2007, p. 61). Citeseer.
- Kan, A. V., Khoroshilov, A. A., Rykalin, A. V., Chechulin, I. A., & Khoroshilov, A. A. (2025). Analytical Review of the Possibilities of Modern Methods and Models of Artificial Intelligence in Relation to Intelligent Processing of Scientific and Technical Documents. *Automatic Documentation and Mathematical Linguistics*, 59(6), 391-399. <https://doi.org/10.3103/S0005105525701134>

- Khorshidi, M. S., Merigó, J. M., Beydoun, G., Furht, B., & Khoshgoftaar, T. M. (2026). A bibliometric analysis of the first decade of the Journal of Big Data. *Journal of Big Data*, 13(80), 1-47. <https://doi.org/10.1186/s40537-026-01412-w>
- Liu, M. (2015). Evaluating the Effectiveness of Health Awareness Events Based on Google Search Frequency. *Electronic Theses and Dissertations*. [South Dakota State University], 1807.
- Malakar, K. D., Roy, S., & Kumar, M. (2025). Database Management System: Foundations and Practices. In *Geospatial Technologies in Coastal Ecologies Monitoring and Management* (pp. 191-255). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-92017-2_7
- Mamedov, V. Y. E., Kovalevsky, D. A. E., Morozov, D. A., Stolyarov, S. S., & Ospichev, S. S. (2025). Hierarchical classification of scientific articles using deep learning (using the UDC hierarchy as an example). *Automatic Control and Computer Sciences*, 59(7), 1181-1192. <https://doi.org/10.3103/S0146411625700440>
- Maña, N. J., Babiera, J., Bayloces, K. L., Palmer, X. L., Potter, L., Lavilles, R., & Velasco, L. C. (2024). Information Retrieval Systems: A Methodological Review. In Arai, K. (eds), *Proceedings of the Future Technologies Conference (FTC) 2024*, Volume 3. FTC 2024. Lecture Notes in Networks and Systems, vol 1156. Springer, Cham. https://doi.org/10.1007/978-3-031-73125-9_36
- Naik, P. G., & Oza, K. S. (2023). Design and Development of Multithreaded Web Crawler for Efficient Extraction of Research Data. In Kumar, A., Mozar, S., & Haase, J. (eds), *Advances in Cognitive Science and Communications. ICCCE 2023*. Cognitive Science and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-19-8086-2_56
- Nambiar, A., & Mundra, D. (2022). An overview of data warehouse and data lake in modern enterprise data management. *Big data and cognitive computing*, 6(4), 132. <https://doi.org/10.3390/bdcc6040132>
- Nasar, Z., Jaffry, S. W., & Malik, M. K. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management*, 56(6), 102088.
- Ogunleye, B., Lancho Barrantes, B. S., & Zakariyyah, K. I. (2025). Topic modelling through the bibliometrics lens and its technique. *Artificial Intelligence Review*, 58(3), 74. <https://doi.org/10.1007/s10462-024-11011-x>
- Onyancha, O. B. (2025). Which Way Information Seeking Behaviour as a Research Area in the Library and Information Science Field? An Informetric Study. In Ocholla, D., Onyancha, O. B., & Adesina, A. O. (eds), *Information, Knowledge,*

- and Technology for Teaching and Research in Africa. Synthesis Lectures on Information Concepts, Retrieval, and Services. Springer, Cham. https://doi.org/10.1007/978-3-031-65745-0_1
- Peikos, G., & Pasi, G. (2024). A systematic review of multidimensional relevance estimation in information retrieval. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 14(5), e1541. <https://doi.org/10.1002/widm.1541>
- Phillips, M. E., Tarver, H., & Zavalina, O. L. (2020) Using metadata record graphs to understand controlled vocabulary and keyword usage for subject representation in the UNT theses and dissertations collection. *Cadernos BAD*, (1), 61-76.
- Ramada, M. S., da Silva, J. C., & de Sá Leitão-Júnior, P. (2020). From keywords to relational database content: A semantic mapping method. *Information Systems*, 88, 101460.
- Redkina, N. S. (2024). Impact of Open Science infrastructure on the development of the world information resources market. *Scientific and Technical Information Processing*, 51(2), 161-172. <https://doi.org/10.3103/S0147688224700096>
- 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.
- Shao, S., & Yu, T. (2023). Information retrieval-based fault localization for concurrent programs. In *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (pp. 1467-1479). IEEE. <https://doi.org/10.1109/ASE56229.2023.00122>
- Smith, G. S., & Pant, S. (1995). *U.S. Patent No. 5,404,510*. Washington, DC: U.S. Patent and Trademark Office.
- Suominen, A. & Toivanen, H. (2016). Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *Journal of the Association for Information Science and Technology (JASIST)*, 67(10): 2464-2476.
- Yadav, S. K. (2023). Database and Research Metrics. In *Research and Publication Ethics*. Springer, Cham. https://doi.org/10.1007/978-3-031-26971-4_12
- Yan, Z., Yang, H., Xi, Y., & Ramesh, G.P. (2024). A Study on the Application of Cloud-Based Database Construction in Language Subject. In Jansen, B. J., Zhou, Q., & Ye, J. (eds), *Proceedings of the 3rd International Conference on Cognitive Based Information Processing and Applications–Volume 1*. CIPA

2023. Lecture Notes on Data Engineering and Communications Technologies, vol 196. Springer, Singapore. https://doi.org/10.1007/978-981-97-1975-4_1
- Yuan, M., & Zou, C. (2018). Text keyword extraction based on meta-learning strategy. In *2018 International Conference on Big Data and Artificial Intelligence (BDAI)* (pp. 78-81). IEEE.
- Yulianti, E., & Rahadiani, L. (2021). Determining subject headings of documents using information retrieval models. *Indonesian Journal of Electrical Engineering and Computer Science*, 23(2), 1049-1058. <https://doi.org/10.11591/ijeecs.v23.i2.pp1049-1058>
- Zhang, Y., Tuo, M., Yin, Q., Qi, L., Wang, X., & Liu, T. (2020). Keywords extraction with deep neural network model. *Neurocomputing*, 383, 113-121. <https://doi.org/10.1016/j.neucom.2019.11.083>

Subject Position for Organizing Information Resources in Databases

Fariborz Doroudi^{*1}

Assistant Prof., Iranian Research Institute for Information Science & Technology (IranDoc), Tehran, Iran

Abstract

The aim of the research is to understand the status of the subject field in databases. This study analyzes and evaluates the methods of using the subject field, examines the main reasons for its use, and explores the application and performance of the subject in the field of information organization activities. The research method is a comparative study on the use of subject-related fields. The statistical population includes 10 Persian databases and 12 foreign databases that are highly important among users and scientific institutions. The research findings indicate that information exploration on information websites is based on various types of fields such as title, author, document type, ISBN, ISSN, DOI, journal name, journal type, article type, and other fields. Each database has utilized multiple fields based on the available information and user needs. Additionally, the use of the subject field has been observed in 90% of Iranian databases in various ways. Among foreign databases, 100% have incorporated the subject field in different formats. In most cases, databases have utilized the keyword field to support the subject field or as a substitute for it. The results highlight that user needs, organizational goals, and the type of information services emphasize the importance of the subject field. The use of standard tools is crucial in validating the subject. Subject classifications in the form of hierarchy, citation, clustering, systematic classifications, indexing standards, and the use of artificial intelligence all play significant roles in organizing and structuring the subject.

Keywords: Information Retrieval, Database, Subject Structure, Subject Field.

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