

وضعیت بازیابی اطلاعات در دو پایگاه نمایه و نما و سنجش اثربخشی استفاده از واژگان کنترل شده در نمایه‌سازی این دو پایگاه

امیر غائبی

استادیار گروه کتابداری و اطلاع‌رسانی دانشگاه الزهرا
ghaebi@gmail.com

فریبرز خسروی

عضو هیئت علمی بازنشسته سازمان اسناد و کتابخانه ملی ج.ا.
kh.class@gmail.com

آزیتا مالمیر (نویسنده مسئول)

کارشناس ارشد کتابداری و اطلاع‌رسانی دانشگاه شهید بهشتی
az.malmir@gmail.com

تاریخ دریافت: ۱۳۸۹/۴/۱۲؛ تاریخ پذیرش: ۱۳۹۰/۲/۱۸

چکیده

هدف: هدف از این پژوهش تعیین میزان جامعیت، مانعیت و مدت‌زمان جست‌وجو در دو پایگاه نما و نمایه و همچنین مقایسه دو ابزار نمایه‌سازی اصطلاحنامه و رده‌بندی ده‌دهی دیویی در این دو پایگاه است.

روش: این پژوهش با استفاده از روش پیمایشی-تحلیلی صورت پذیرفته است. اطلاعات مورد نیاز آن با مشاهده مستقیم منابع پایگاه مقالات نما و نمایه گردآوری شده است. جامعه آماری، مقالات مشترک در دو پایگاه مقالات فارسی نما و نمایه در دو حوزه موضوعی کتابداری و اطلاع‌رسانی و تاریخ (۶۱۰ مقاله) است که ۲۴۰ مقاله به روش نمونه‌گیری نسبی و تصادفی انتخاب گردید.

یافته‌ها: یافته‌های پژوهش نشان می‌دهد که میزان جامعیت (در حوزه کتابداری و اطلاع‌رسانی و در پایگاه نما ۵۶ درصد و در پایگاه نمایه ۴۰ درصد و میزان مانعیت (در حوزه کتابداری و اطلاع‌رسانی و تاریخ) در پایگاه نما ۵۸ درصد و در پایگاه نمایه ۳۵ درصد، و نیز مدت‌زمان جست‌وجو در پایگاه نما ۱ دقیقه و ۱۳ ثانیه و در پایگاه نمایه ۲ دقیقه و ۱۳ ثانیه است. استفاده از آزمون آماری تی برای نمونه‌های مستقل در این پژوهش بیانگر این است که از نظر جامعیت، مانعیت و مدت‌زمان جست‌وجو اختلاف معنی‌داری بین پایگاه نما و پایگاه نمایه وجود دارد؛ اما همبستگی در این معیارها در دو پایگاه مستقیم و معنی‌دار است. نتایج به‌دست آمده از پژوهش نشان می‌دهد که پایگاه نما از نظر میزان جامعیت و مانعیت و مدت‌زمان جست‌وجو در وضعیت مطلوب‌تری نسبت به پایگاه نمایه قرار دارد و می‌توان گفت استفاده از اصطلاحنامه در نمایه‌سازی پایگاه نما بر نتایج بازیابی از آن پایگاه تأثیر مثبت و موثر گذاشته است.

اصالت/ارزش: براساس نتایج پژوهش جست‌وجوگران اطلاعات می‌توانند در انتخاب پایگاه متناسب با نیاز خود بهتر تصمیم‌گیری کنند و نیز طراحان پایگاه‌ها دریابند که برای تولید پایگاه از کدام زبان کنترل‌شده در سازمان‌دهی منابع استفاده کنند.

کلیدواژه‌ها: جامعیت، مانعیت، مدت‌زمان جست‌وجو، اصطلاحنامه، ارزیابی بازیابی اطلاعات، پایگاه مقالات نما، پایگاه مقالات نمایه

مقدمه

امروزه با وقوع پدیده انفجار اطلاعات و به تبع آن گسترش اطلاعات و ایجاد پایگاه‌های اطلاعاتی متعدد و روبه‌رو شدن پژوهشگران با حجم وسیعی از اطلاعات و اتلاف وقت و هزینه در یافتن اطلاعات مطلوب‌شان موجب شده صاحب‌نظران در پی راه چاره باشند تا این اطلاعات را سازماندهی و شیوه‌ای برای بازیابی آن‌ها یابند. نمایه‌سازی یکی از این راه‌حل‌ها است. نمایه‌سازی یکی از رهیافت‌های سازماندهی مواد است.

نقش عمده نظام بازیابی اطلاعات برقراری ارتباط بین تولیدکنندگان و مصرف‌کنندگان اطلاعات است و این ارتباط از طریق همپوشانی و تطابق اطلاعات موجود در مجموعه با نیازهای کاربران اطلاعات رخ می‌دهد. برای انجام این عمل باید به توصیف مناسب محتوای مدارک پرداخت. البته باید نیاز کاربر هم به نوعی به نظام اطلاعاتی عرضه شود که عمل انطباق بین مدارک موجود در پایگاه و پرسش مطرح شده کاربر تسریع شود. این کار به وسیله استفاده از نوعی زبان نمایه‌سازی انجام می‌گیرد. زبان نمایه‌سازی در نظام‌های اطلاع‌رسانی به زبان ساختگی و قراردادی اطلاق می‌شود که برای مقاصد نمایه‌سازی به‌ویژه قابلیت بازیابی اطلاعات و مدارک به کار گرفته می‌شود.

با گسترش روبه‌رو رشد سامانه‌های اطلاعاتی و کاربرد شیوه‌های گوناگون نمایه‌سازی، پژوهشگران آزمون‌هایی را انجام دادند که در یابند کدام یک از این زبان‌های نمایه‌سازی مؤثرترند. استفاده از واژگان کنترل‌شده که یکی از انواع این زبان‌ها است، علی‌رغم بحث‌های نظری و تأکیدی که بر نقش آن در بازیابی شده، در حاله‌ای از ابهام قرار دارد و در این پژوهش سعی شده به تاثیرات آن پرداخته شود.

اهداف پژوهش

هدف اصلی پژوهش حاضر مقایسه میزان جامعیت، مانعیت و مدت‌زمان جست‌وجو مدارک بازیابی شده در دو پایگاه نما و نمایه است. همچنین سنجش اثربخشی اصطلاحنامه و رده‌بندی دیویی در بازیابی مدارک از این پایگاه‌ها و تایید بر تغییر رویکرد استفاده از اصطلاحنامه در نمایه‌سازی مدارک پایگاه نمایه از دیگر اهداف این پژوهش محسوب می‌گردند.

پرسش‌های اساسی

۱. استفاده از اصطلاحنامه در نمایه‌سازی پایگاه نما و استفاده از رده‌بندی ده‌دهی دیویی در نمایه‌سازی پایگاه نمایه چه تأثیری بر جامعیت جست‌وجو داشته است؟
۲. استفاده از اصطلاحنامه در نمایه‌سازی پایگاه نما و استفاده از رده‌بندی ده‌دهی دیویی در نمایه‌سازی پایگاه نمایه چه تأثیری بر مانعیت جست‌وجو داشته است؟
۳. استفاده از اصطلاحنامه در نمایه‌سازی پایگاه نما و استفاده از رده‌بندی ده‌دهی دیویی در نمایه‌سازی پایگاه نمایه چه تأثیری بر مدت‌زمان جست‌وجو داشته است؟

پیشینه

چشمه‌سهرابی (۱۳۷۸) در پایان‌نامه کارشناسی ارشد خود با عنوان «تأثیر استفاده از اصطلاحنامه در بانک‌های اطلاعاتی کتابشناختی بر میزان جامعیت، مانعیت و مدت‌زمان جست‌وجوی اطلاعات بازیابی شده» به بررسی تأثیر اصطلاحنامه و زبان طبیعی بر ۷ بانک اطلاعاتی (WRA, Eric, Medline, IP, Cab Agris, Inspec) در ۶ گروه علوم انسانی، علوم پزشکی، علوم پایه، کشاورزی، فنی و مهندسی و هنر پرداخته‌اند. روش پژوهش تجربی و ابزار گردآوری اطلاعات مشاهده مستقیم و فرم محقق ساخته است. براساس یافته‌های پژوهش، استفاده از اصطلاحنامه به هنگام بازیابی اطلاعات نسبت به زبان طبیعی باعث افزایش چشم‌گیر میزان مانعیت و کاهش میزان جامعیت و مدت‌زمان جست‌وجوی اطلاعات می‌شود.

شوره‌زاری (۱۳۷۹) در پایان‌نامه کارشناسی ارشد خود به بررسی جست‌وجوی استفاده‌کنندگان نهایی پایگاه اطلاعاتی رایانه‌ای کتابخانه منطقه‌ای علوم و تکنولوژی شیراز و مقایسه آن با جست‌وجوی کارشناسان اطلاع‌رسانی از نظر ضرایب جامعیت و مانعیت اطلاعات بازیابی شده و زمان جست‌وجو می‌پردازد. جامعه آماری پنج نفر از کارشناسان اطلاع‌رسانی و روش پژوهش نیمه‌تجربی و شیوه گردآوری پرسشنامه و مشاهده با مشارکت است. نتایج حاصل از پژوهش نشان می‌دهد زمان جست‌وجوی کارشناسان کم‌تر از زمان جست‌وجوی استفاده‌کنندگان است. جامعیت مدارک بازیابی شده افزایش و مانعیت مدارک بازیابی شده کاهش یافته است. جوکار و انواری (۱۳۸۵) در مقاله‌شان تحت عنوان «بررسی رویکردهای موضوعی (زبان

طبیعی و کنترل شده) در بازیابی اطلاعات از پایگاه‌های پیوسته کتابشناختی» کارایی دو رویکرد جست‌وجوی موضوعی زبان طبیعی و واژگان کنترل شده را با استفاده از معیارهای ربط، دقت، بازیافت و ریزش کاذب در دو پایگاه کتابشناختی پیوسته کتابخانه کنگره و اریک مقایسه کردند. نتایج نشان می‌دهد جامعیت رویکرد جست‌وجوی زبان طبیعی در هر دو پایگاه بیش از رویکرد موضوعی کنترل شده است. دقت رویکرد زبان طبیعی در پایگاه اریک بیش از کنترل شده است، اما در فهرست پیوسته کتابخانه کنگره تفاوتی معنی‌دار مشاهده نشد. رویکرد زبان طبیعی در پایگاه اریک ریزش کاذب کم‌تری به همراه داشته است؛ درحالی‌که در فهرست کتابخانه کنگره تفاوت معنی‌دار نیست.

شاکری (۱۳۸۷) در مقاله‌اش تحت عنوان «میزان جامعیت و مانعیت ابزارهای کاوش فارسی اینترنت در بازیابی اطلاعات در حوزه کتابداری و اطلاع‌رسانی» به بررسی میزان جامعیت و مانعیت ابزارهای کاوش اینترنت با واسط جست‌وجوی فارسی پرداخته است. از نظر میزان مانعیت موتور کاوش وبلاگ در بین ابزارهای کاوش مورد بررسی با مانعیتی حدود ۳۳/۵ درصد دارای بالاترین میزان مانعیت است و سپس موتور کاوش دهیو با ۳۲ درصد میزان مانعیت در رتبه دوم و موتور کاوش گوگل فارسی با مانعیتی حدود ۳۰ درصد در رتبه سوم قرار دارد. همچنین راهنمای موضوعی آفتاب با ۹ درصد مانعیت از کم‌ترین میزان مانعیت در بین ابزارهای کاوش مورد مطالعه برخوردار است. علاوه بر این، ابزارهای کاوش با واسط جست‌وجوی فارسی از نظر میزان جامعیت بازیابی اطلاعات نیز مورد ارزیابی قرار گرفتند. این مطالعه نشان می‌دهد که موتور کاوش وبلاگ در بین ابزارهای کاوش مورد بررسی، دارای بالاترین میزان جامعیت (۴۳/۵ درصد) و سپس موتور کاوش دهیو با ۴۱/۵ درصد میزان جامعیت و موتور کاوش گوگل فارسی با جامعیتی حدود ۳۸/۹ درصد به ترتیب در رتبه‌های بعد قرار دارند. راهنمای موضوعی آفتاب نیز با ۱۱/۶ درصد از کم‌ترین میزان جامعیت در بین ابزارهای کاوش مورد مطالعه برخوردار است. همچنین بین مانعیت و جامعیت ابزارهای کاوش مورد مطالعه، اختلاف معنی‌داری وجود دارد.

محمدی (۱۳۸۸) در مقاله‌اش با عنوان «توسعه بهینه پرسش‌های جست‌وجو با استفاده از اصطلاحنامه پیوسته اریک و تأثیر آن در مانعیت نتایج و زمان جست‌وجوی کاربر» به بررسی تأثیر استفاده از اصطلاحات اصطلاحنامه پیوسته اریک در انتخاب اصطلاحات جست‌وجو و در نتایج

حاصل از جست‌وجو در پایگاه‌های اطلاعاتی مرتبط با این اصطلاحنامه‌ها پرداخت. روش این پژوهش تجربی و برای اجرا از پرسشنامه سنجش اولیه مهارت‌های جست‌وجو در اینترنت، نرم افزار Track4Win و اصطلاحنامه اریک (وابسته به پایگاه اطلاعاتی اریک) به‌عنوان ابزار گردآوری داده‌ها استفاده شده است. جامعه پژوهش آن ۶۵ نفر از دانشجویان کارشناسی ارشد رشته روان‌شناسی و علوم تربیتی دانشگاه تهران است. مقایسه مراحل پیش‌آزمون و پس‌آزمون جست‌وجوی دو گروه آزمایش و گواه نشان داد که استفاده از اصطلاحنامه پیوسته اریک در بهبود پرسش‌ها و زمان جست‌وجو نقش به‌سزایی دارد و استفاده از اصطلاحات اعم، اخص، و وابسته اصطلاحنامه، دقت بازیابی را تغییر می‌دهد. همچنین استفاده از اصطلاحات اعم و وابسته اصطلاحنامه، زمان جست‌وجو را افزایش می‌دهد.

علاوه بر پژوهش‌های ذکر شده، در کشورهای مختلف جهان نیز طرح‌های پژوهشی در این رابطه انجام شده که در اینجا به برخی از آن‌ها اشاره می‌شود.

ناپ^۱ و همکارانش (۱۹۹۸) در پژوهشی برای تعیین ارزش بالقوه اصطلاحنامه‌ای که با اصطلاحات متن آزاد یکپارچه شده و توسط متخصصان حوزه علوم انسانی و جامعه‌شناسی مورد استفاده قرار می‌گیرد؛ به این نتیجه رسیدند که ترکیب هر دو رویکرد یعنی استفاده هم‌زمان از اصطلاحنامه و اصطلاحات متن آزاد منجر به بازیابی موارد مرتبط‌تر و همین‌طور افزایش جامعیت می‌شود.

مودامیل^۲ (۱۹۹۸) در تحقیقی که با هدف بررسی کارآمدی اصطلاحنامه و زبان طبیعی در نظام بازیابی اطلاعات انجام داده است، با در نظر گرفتن نتایج جامعیت و مانعیت اظهار می‌دارد که جست‌وجو به زبان طبیعی جایگاه بهتری نسبت به جست‌وجوی اصطلاحنامه‌ای به‌دست آورده است. این مطالعه به‌طور مقایسه‌ای نشان داده است که جست‌وجو به زبان طبیعی تفاوت ۰/۰۵ درصدی در جامعیت و ۰/۰۲ درصدی در مانعیت نسبت به جست‌وجوی اصطلاحنامه‌ای دارد. این پژوهشگر در نهایت با توجه به عملکرد یکسان اصطلاحنامه و زبان طبیعی در بازیابی اطلاعات ترکیبی از این دو را برای انجام جست‌وجوها پیشنهاد می‌کند.

گرینبرگ^۳ (۱۹۹۸) در پایان‌نامه دکترای خود تأثیر روابط معناشناختی اصطلاحنامه بر

1. Knapp

2. Muddamale

3. Greenberg

کارآمدی بازیابی را بررسی و تأثیر این روابط بر فرآیندهای سوال‌گستری خودکار و سوال‌گستری تعاملی را مقایسه کرده است. محیط آزمایش این پژوهش شامل بانک اطلاعاتی ادواری‌های تجاری ABI/Inform و واژگان مهارشده پروکوئست^۱ است. نتایج این تحقیق به‌طور کلی نشان می‌دهد که سوال‌گستری خودکار با استفاده از روابط معناشناختی موجود در اصطلاحنامه باعث افزایش نسبی جامعیت می‌شود و برای افزایش مانعیت در سوال‌گستری تعاملی اصطلاحات مرتبط معرف‌های خوبی به‌شمار می‌آیند.

هریسون^۲ (۱۹۹۸) در پایان‌نامه دکترای خود که با اهداف ارزیابی اثر استفاده یا عدم استفاده از اصطلاحنامه بر نتایج کاوش، آگاهی از کارآمدی اصطلاحنامه‌های نمایه‌سازی بر کاوش پیوسته، و ارزیابی تأثیر تفاوت‌های فردی کاوشگران بر استفاده از اصطلاحنامه انجام گرفت، به همراه ۱۲ کاوشگر جست‌وجو‌هایی را در بانک اطلاعاتی «روسوپس»^۳ که از اصطلاحنامه امتری^۴ بهره می‌گیرد انجام داد. نتایج نشان می‌دهد که هیچ تفاوت معنی‌داری در نتایج جست‌وجو چه با استفاده از اصطلاحنامه و چه بدون استفاده از آن وجود ندارد. کاوشگران غیرمغرب در این پژوهش از اصطلاحنامه بیش‌تر استفاده کرده‌اند. در برخی موارد به‌نظر رسید کاوشگران مغرب بدون نیاز به رجوع به اصطلاحنامه از واژه‌های آن استفاده کرده‌اند. در رفتار کاوش‌هایی که با اصطلاحنامه و بدون آن صورت گرفت تفاوت معنی‌داری وجود داشت. یعنی عبارت‌های کاوشی بیش‌تری با کاوش‌هایی که با اصطلاحنامه انجام می‌گرفت همراه می‌شد و کاوش‌های انجام شده با اصطلاحنامه رکوردهای کم‌تری نسبت به کاوش‌های انجام شده بدون اصطلاحنامه بازیابی کرد. میلی^۵ و همکارانش (۲۰۰۳) به‌منظور کشف روش‌هایی برای توسعه نرم‌افزار قابل استفاده مجدد و ابزاری برای ساختن سیستم اطلاعاتی خبره با ترکیبات قابل استفاده مجدد تحقیقی انجام دادند. پژوهشگران برای این کار بر روی ترکیبات مختلف نمایه‌سازی و روش‌های بازیابی آزمایش کردند. نتایج این تحقیق نشان می‌دهد که نمایه‌سازی تمام‌متن از جامعیت بیش‌تری نسبت به نمایه‌سازی واژگان کنترل‌شده برخوردار است و مانعیت آن دو با هم تفاوت معنی‌داری ندارد. بنک^۶ و همکارانش (۲۰۰۶) در مقاله‌شان یک روش جدیدی پیشنهاد کردند و از روابط مفاهیم اساسی اصطلاحنامه در طبقه‌بندی k-NN (روش مقوله‌بندی اسناد ملی با عملکرد نسبتاً

1. ProQuest
4. Emtree

2. Harrison
5. Milli

3. Rosoopse
6. Bankg

خوب علی‌رغم سادگی آن) استفاده کردند. در هنگامی که یک سند بیش‌تر از یک مقوله داشت، k -NN دچار ابهام می‌شد و به‌طور معنی‌داری مانعیت را کاهش می‌داد. برای رفع این مشکل از مفاهیم اساسی و ساختار سلسله‌مراتبی اصطلاحنامه در مقوله‌بندی سند استفاده شد. با ارجاعات مختلف در اصطلاحنامه، k -NN توانست به‌طور آشکاری مانعیت را بهبود بخشد و ابهامات را از بین ببرد. نتایج تجربی این پژوهش نشان داد که استفاده از اصطلاحنامه در k -NN مانعیت را از ۱۳/۸۶ درصد بالاتر می‌برد و در جامعیت تغییری نمی‌دهد.

گهنو^۱ و همکارانش (۲۰۰۷) به ارزیابی بازیابی اطلاعات در دروازه اطلاعاتی کنترل کیفیت سلامت CISMef با استفاده از دو شیوه به‌کارگیری فرااصطلاح (اصطلاحات ویژه‌ای که از نظر معنایی با یک یا چند اصطلاح سرعنوان‌های پزشکی، عنوان‌های فرعی و منابع دیگر مرتبطند) و اصطلاحات سرعنوان‌های پزشکی Mesh پرداختند. برای این کار ۱۶ جفت سؤال در یک موضوع که یک سؤال مربوط به حداقل یک فرااصطلاح و سؤال دیگر مربوط به یک اصطلاح از سرعنوان موضوعی بود تهیه شد. نتایج این پژوهش نشان می‌دهد که به‌کارگیری فرااصطلاح در بازیابی اطلاعات در این دروازه اطلاعاتی میزان جامعیت را نسبت به استفاده از اصطلاحات سرعنوان‌های موضوعی پزشکی بالاتر می‌برد، اما میزان مانعیت مشابه است و تغییری پیدا نمی‌کند.

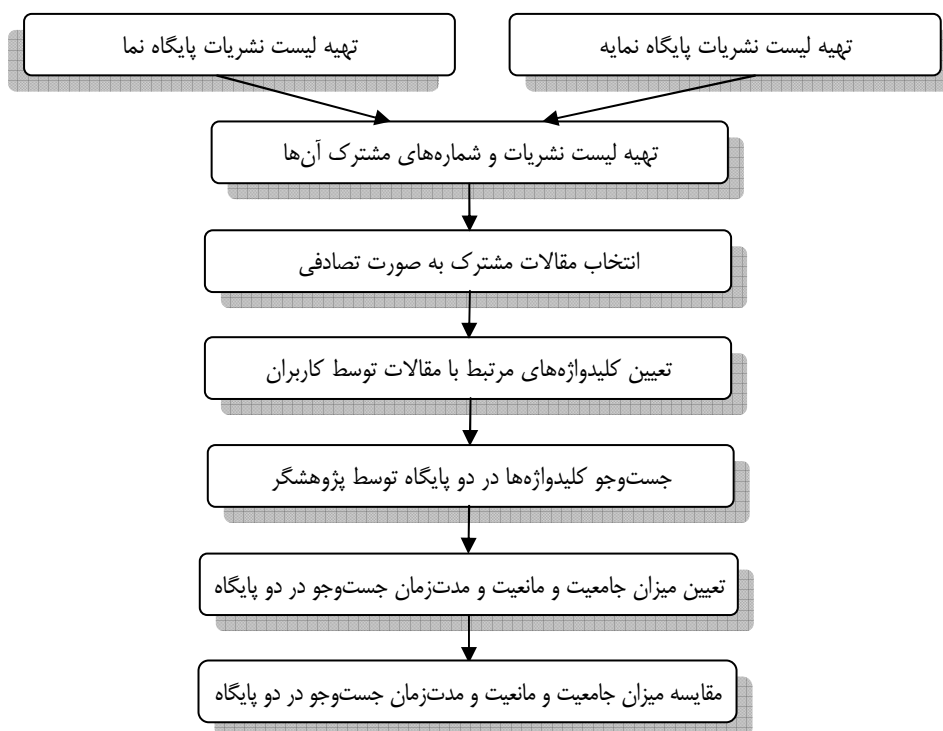
صحبت در مورد اصطلاحنامه‌ها و کاربردهای آن در بازیابی بحثی گسترده است و براساس پژوهش‌های موجود می‌توان ارزشمندی بالقوه اصطلاحنامه‌ها را در بازیابی مدارک مرتبط اثبات نمود. این ارزشمندی به‌خصوص در کمک به کاربران در ساخت و تدوین پرسش‌ها و نیز در اصلاح راهبردهای جست‌وجوهایشان مشخص است. پژوهش‌ها نشان می‌دهد که اتفاق نظری بر روی اثرات استفاده از اصطلاحنامه در پایگاه‌های اطلاعاتی وجود ندارد، تنها برخی از آن‌ها به این نکته اشاره دارند که اصطلاحنامه باعث افزایش مانعیت و کاهش زمان جست‌وجو خواهد شد و بیش‌تر آن‌ها پیشنهاد می‌کنند که پایگاه‌های اطلاعاتی برای نمایه‌سازی اطلاعات خود هم از زبان طبیعی و هم از اصطلاحنامه استفاده کنند.

روش پژوهش

این پژوهش با استفاده از روش پیمایشی-تحلیلی انجام گرفته است. جامعه آماری دو پایگاه

1. Gehanno

مقالات فارسی نما و نمایه است و پژوهش بر روی آن قسمتی از مقالات که در دو پایگاه در دو حوزه موضوعی کتابداری و اطلاع‌رسانی و تاریخ مشترک بوده (۶۱۰ مقاله) و براساس رده‌بندی دیویی نمایه‌سازی شده است، انجام شد. حجم نمونه با استفاده از جدول مورگان ۲۴۰ مقاله تعیین گردید که مقالات به روش نمونه‌گیری تصادفی طبقه‌ای انتخاب شد. اطلاعات مورد نیاز این پژوهش از طریق مشاهده مستقیم از منابع پایگاه داده‌های نما و نمایه تهیه لیست‌هایی از شماره نشریات هم‌سان در دو پایگاه گردآمده است. در تجزیه و تحلیل یافته‌های پژوهش از آزمون تی مستقل با درجه اطمینان ۰/۹۵ و نیز برای بیان شدت ارتباط متغیرهای وابسته در دو پایگاه از ضریب همبستگی پیرسون استفاده شده است. قابل ذکر است که از دو نرم افزار Spss و Excel نیز بهره گرفته شده است. نمودار ۱ نشان‌دهنده روش گردآوری داده‌ها و سیر عملیاتی پژوهش است:



نمودار ۱. روش گردآوری داده‌ها و سیر عملیاتی پژوهش

یافته‌ها

مشاهده نتایج بازیابی نشان می‌دهد که در حوزه کتابداری و اطلاع‌رسانی و تاریخ جامعیت در پایگاه نما ۵۶ درصد و در پایگاه نمایه ۴۰ درصد است. براین اساس، مشاهده می‌شود که جامعیت پایگاه نما که از اصطلاحنامه در نمایه‌سازی پایگاه خود استفاده می‌کند، بیش‌تر از نمایه است.

جدول ۱. میزان جامعیت جست‌وجو در دو حوزه کتابداری و اطلاع‌رسانی و تاریخ در دو پایگاه

پایگاه‌های اطلاعاتی	درصد جامعیت در دو حوزه کتابداری و اطلاع‌رسانی و تاریخ
نما	۵۶
نمایه	۴۰

برای آزمون معنی‌داری اختلاف میزان جامعیت دو پایگاه نما و نمایه از آزمون تی مستقل با درجه اطمینان ۰/۹۵ استفاده شده و نیز برای بیان شدت ارتباط متغیرهای وابسته در دو پایگاه، ضریب همبستگی پیرسون به کار رفته است. همان‌طوری که در جدول ۲ مشخص است، نتایج آزمون نشان می‌دهد که مقدار بحرانی این آزمون صفر است و به سبب این که این مقدار از ۰/۰۵ کوچک‌تر است، بین جامعیت در دو پایگاه اختلاف معنی‌داری وجود دارد. همچنین میزان ضریب همبستگی بین جامعیت در پایگاه‌های مقالات نما و نمایه بنا به جدول ۳ برابر ۰/۴۱ است و علامت مثبت در این همبستگی بیانگر ارتباط مستقیم و معنی‌دار بین جامعیت در دو پایگاه داده است.

جدول ۲. آزمون تی (برای نمونه‌های مستقل) جامعیت در دو پایگاه

	بر اساس واریانس	t	df	p	اختلاف میانگین	اختلاف انحراف استاندارد	اختلاف فاصله اطمینان ۹۵٪	
							کم‌ترین	بیش‌ترین
جامعیت دو پایگاه	با فرض برابری واریانس	۴۸۹	۷۵۰	۰.۰۰	۰.۱۵	۰.۰۳	۰.۰۹	۰.۲۲
	با فرض عدم برابری واریانس	۴۸۹	۷۴۸.۹۸	۰.۰۰	۰.۱۵	۰.۰۳	۰.۰۹	۰.۲۲

جدول ۳. ضریب همبستگی بین جامعیت در دو پایگاه

	آماره‌ها	جامعیت نما	جامعیت نمایه
جامعیت نما	ضریب همبستگی	۱	.۴۱۸ (**)
	مقدار P	.	.۰۰۰
	N	۳۷۶	۳۷۶
جامعیت نمایه	ضریب همبستگی	.۴۱۸ (**)	۱
	مقدار P	.۰۰۰	.
	N	۳۷۶	۳۷۶

$p < 0.05$: * و $p < 0.01$: **

در پاسخ به پرسش دوم پژوهش، مشاهده نتایج بازیابی نشان می‌دهد که در حوزه کتابداری و اطلاع‌رسانی و تاریخ میزان مانعیت در پایگاه نما ۵۸ درصد و در پایگاه نمایه ۳۵ درصد است که نشان می‌دهد مانعیت پایگاه نما بیش از پایگاه نمایه است.

جدول ۴. میزان مانعیت جست‌وجو در دو حوزه کتابداری و اطلاع‌رسانی و تاریخ در دو پایگاه

درصد مانعیت در دو حوزه کتابداری و اطلاع‌رسانی و تاریخ	پایگاه‌های اطلاعاتی
۵۸	نما
۳۵	نمایه

برای نشان دادن اختلاف معنی‌دار میان مانعیت در دو پایگاه از آزمون تی مستقل استفاده شد. براساس نتایج آزمون تی مقدار بحرانی این آزمون صفر است و به سبب این که این مقدار از ۰/۰۵ کوچک‌تر است، بین مانعیت در دو پایگاه اختلاف معنی‌داری وجود دارد. میزان ضریب همبستگی بین مانعیت در پایگاه‌های مقالات نما و نمایه برابر ۰/۴۲ است و علامت مثبت در این همبستگی بیانگر ارتباط مستقیم و معنی‌دار بین مانعیت در دو پایگاه داده است.

جدول ۵. آزمون تی (برای نمونه‌های مستقل) مانعیت در دو پایگاه

	بر اساس واریانس	t	df	p	اختلاف میانگین	اختلاف انحراف استاندارد	اختلاف فاصله اطمینان ۹۵٪	
							کم‌ترین	بیش‌ترین
مانعیت دو پایگاه	با فرض برابری واریانس	۷.۲۹	۷۵۰	۰.۰۰	۰.۲۳	۰.۰۳	۰.۱۷	۰.۳۰
	با فرض عدم برابری واریانس	۷.۲۹	۷۴۳.۶۹	۰.۰۰	۰.۲۳	۰.۰۳	۰.۱۷	۰.۳۰

جدول ۶. ضریب همبستگی بین مانعیت در دو پایگاه

	آماره‌ها	جامعیت نما	جامعیت نمایه
مانعیت نما	ضریب همبستگی	۱	.۴۲۱(**)
	مقدار P	.	.۰۰۰
	N	۳۷۶	۳۷۶
مانعیت نمایه	ضریب همبستگی	.۴۲۱(**)	۱
	مقدار P	.۰۰۰	.
	N	۳۷۶	۳۷۶

** : $p < 0.01$ و * : $p < 0.05$

در پاسخ به پرسش سوم پژوهش، مشاهده نتایج بازیابی نشان می‌دهد که در حوزه کتابداری و اطلاع‌رسانی و تاریخ مدت‌زمان جست‌وجو در پایگاه نما ۱ دقیقه و ۱۳ ثانیه و در پایگاه نمایه ۲ دقیقه و ۱۳ ثانیه است. بنابراین با توجه به یافته‌ها می‌توان استنباط کرد که مدت‌زمان جست‌وجو در

پایگاه نما (که از اصطلاحنامه در نمایه‌سازی پایگاه خود استفاده می‌کند) کم‌تر از پایگاه نمایه (که از رده‌بندی ده‌دهی دیوئی در نمایه‌سازی پایگاه خود استفاده می‌کند) می‌باشد.

جدول ۷. میزان مدت‌زمان جست‌وجو در دو حوزه کتابداری و اطلاع‌رسانی و تاریخ در دو پایگاه

مدت‌زمان جست‌وجو در دو حوزه کتابداری و اطلاع‌رسانی و تاریخ	پایگاه‌های اطلاعاتی
۰:۰۱:۱۳	نما
۰:۰۲:۱۳	نمایه

برای نشان دادن اختلاف معنی‌دار میان مدت‌زمان جست‌وجوی اطلاعات در دو پایگاه از آزمون تی مستقل استفاده شد. براساس نتایج آزمون تی مقدار بحرانی این آزمون بنا به جدول ۸ صفر است و به سبب این که این مقدار از ۰/۰۵ کوچک‌تر است، بین مدت‌زمان جست‌وجو در دو پایگاه اختلاف معنی‌داری وجود دارد، و میزان ضریب همبستگی بین مدت‌زمان جست‌وجو در پایگاه‌های مقالات نما و نمایه برابر ۰/۶۳ است و علامت مثبت در این همبستگی بیانگر ارتباط مستقیم و معنی‌دار بین مدت‌زمان جست‌وجو در دو پایگاه داده است.

جدول ۸. آزمون تی (برای نمونه‌های مستقل) مدت‌زمان جست‌وجو در دو پایگاه

مدت‌زمان جست‌وجو در دو پایگاه	بر اساس واریانس	t	df	p	اختلاف میانگین	اختلاف انحراف استاندارد	اختلاف فاصله اطمینان ۹۵٪	
							کم‌ترین	بیش‌ترین
مدت‌زمان جست‌وجو در دو پایگاه	با فرض برابری واریانس	-۷.۲۷	۷۵۰	۰.۰۰	-۰:۰۱:۰۰	۰:۰۰:۰۸	-۰:۰۱:۱۷	-۰:۰۰:۴۴
	با فرض عدم برابری واریانس	-۷.۲۷	۶۴۲.۲۷	۰.۰۰	-۰:۰۱:۰۰	۰:۰۰:۰۸	-۰:۰۱:۱۷	-۰:۰۰:۴۴

جدول ۹. ضریب همبستگی بین مدت زمان جست‌وجو در دو پایگاه

	آماره‌ها	جامعیت نما	جامعیت نمایه
مدت زمان جست‌وجو در نما	ضریب همبستگی	۱	.۶۳۲ (**)
	مقدار P	.	.۰۰۰
	N	۳۷۶	۳۷۶
مدت زمان جست‌وجو در نمایه	ضریب همبستگی	.۶۳۲ (**)	۱
	مقدار P	.۰۰۰	.
	N	۳۷۶	۳۷۶

$$p < 0.05: * \quad \text{و} \quad p < 0.01: **$$

نتیجه‌گیری

نتایج به‌دست آمده از این پژوهش نشان می‌دهد که پایگاه مقالات نما نسبت به پایگاه نمایه از نظر جامعیت و مانعیت در وضعیت مطلوب‌تری قرار دارد و می‌توان گفت استفاده از اصطلاحنامه در نمایه‌سازی مقالات پایگاه نما بر نتایج بازیابی از این پایگاه موثر بوده است. در پایگاه نما از دو تا هفت کلیدواژه و در پایگاه نمایه یک تا دو کلیدواژه در نمایه‌سازی مقالات به کار رفته است که این تعداد کلیدواژه‌ها موجب افزایش جامعیت در پایگاه نما و کاهش جامعیت در پایگاه نمایه می‌شود.

رده‌بندی ده‌دهی دیویی به‌صورت کلی به موضوع می‌پردازد و می‌بایست برای نمایه‌سازی کتاب‌ها مورد استفاده قرار گیرد نه مقالات که به‌طور جزئی و تخصصی موضوع را بررسی می‌کنند. این کل‌نگری توسط رده‌بندی دیویی سبب شده که مانعیت در پایگاه نمایه کاهش یابد. علاوه بر موارد ذکر شده نتایج مانعیت و جامعیت در هر پایگاه نشان می‌دهد که ارتباط معکوسی بین جامعیت و مانعیت وجود ندارد و استفاده از اصطلاحنامه بیش‌ترین تأثیر را در معیار مانعیت نسبت به جامعیت دارد و باعث افزایش آن به‌طور محسوس می‌شود.

با این‌که نتایج پژوهش استفاده از پایگاه نما را برای جست‌وجو پیشنهاد می‌کند؛ اما باید به این نکته توجه داشت که پایگاه نمایه، پایگاهی تمام‌متن است و می‌توان با پرداخت هزینه به متن

مقالات آن دسترسی پیدا کرد و در مقایسه با پایگاه نما که یک پایگاه کتابشناختی است، پژوهشگر راحت‌تر و سریع‌تر می‌تواند به متن کامل مقالات خود دست یابد. همچنین هم‌اکنون پایگاه نمایه از اصطلاحنامه و سرعنوان‌های موضوعی فارسی جهت نمایه‌سازی مقالات خود استفاده می‌کند و با گذر زمان این رویکرد در همه پایگاه اجرا خواهد شد و این موضوع نوبدبخش بالا رفتن کیفیت و بهبود وضعیت بازیابی مقالات در این پایگاه خواهد شد.

به‌طور کلی می‌توان از این پژوهش و نتایج پژوهش‌های مشابه آن دریافت که اصطلاحنامه نسبت به دیگر واژگان کنترل‌شده نتایج جست‌وجو را بهبود می‌بخشد. البته این به‌تنهایی کافی نیست؛ بلکه برای کسب نتایج جامع در جست‌وجوهای موضوعی می‌بایست ترکیبی از دو رویکرد زبان طبیعی و کنترل‌شده مورد استفاده قرار گیرد. برای انجام جست‌وجوهای که مانعیت نتایج حاصل از آن‌ها اهمیت چندانی ندارد و هدف، بالا بردن جامعیت و به‌دست آوردن تعدادی پیشینه مربوط است، رویکرد زبان طبیعی مزیت بیش‌تری دارد.

پیشنهادها

کل‌نگری که شناسایی کلیه موضوعات موجود در مدرک و جزءنگری که توصیف دقیق واژه‌ها و مفاهیم مدرک است دو عامل اساسی در نمایه‌سازی هستند که بر مقیاس‌های مانعیت و جامعیت اثر می‌گذارد.

فنونی از قبیل کاهش تعداد مفاهیم مختلف در یک کاوش، افزایش واژه‌های مترادف جهت بازنمایی مفاهیم کاوش، و کوتاه کردن عبارات موجب افزایش جامعیت و کاهش مانعیت می‌شود. اضافه کردن مفاهیم منحصر به فرد، کاهش تعداد واژه‌های مترادف، استفاده از عملگرهای مجاورت و منحصر کردن کاوش به میدان‌های خاص در پیشینه‌ها مانند میدان عنوان و میدان اصطلاحات نمایه، به‌طور معمول منجر به توسعه مانعیت و کاهش جامعیت می‌شود.

تولید کنندگان پایگاه‌های اطلاعاتی به‌منظور کمک به کاربران جهت دستیابی بهتر و سریع‌تر به مدارک مورد نیازشان می‌بایست بر به‌کارگیری اصطلاحنامه در نمایه‌سازی مقالات پایگاه‌ها و طراحی سیستم تأکید بیش‌تری داشته باشند. همچنین برای به‌دست آوردن بازیابی مطلوب‌تر بهتر است ترکیبی از دو رویکرد زبان طبیعی و کنترل‌شده برای نمایه‌سازی مورد استفاده قرار گیرد.

منابع

- جوکار، عبدالرسول؛ انواری، سعیده (۱۳۸۵). بررسی رویکردهای موضوعی (زبان طبیعی و کنترل شده) در بازیابی اطلاعات از پایگاه‌های پیوسته کتابشناختی. *کتابداری و اطلاع‌رسانی*، ۳۶ (۴)، ۱۵۲-۱۶۴.
- چشمه سهرابی، مظفر (۱۳۷۸). تأثیر استفاده از اصطلاحنامه در بانکهای اطلاعاتی کتابشناختی بر میزان جامعیت، مانعیت، و مدت‌زمان جست‌وجوی اطلاعات بازیابی شده. پایان‌نامه کارشناسی‌ارشد، دانشگاه تربیت مدرس، تهران.
- شاگری، صدیقه (۱۳۸۷). میزان جامعیت و مانعیت ابزارهای کاوش فارسی اینترنت در بازیابی اطلاعات در حوزه کتابداری و اطلاع‌رسانی. *فصلنامه کتاب*، ۷۳، ۱۷۷-۲۰۰.
- شوره زاری، علی (۱۳۷۹). بررسی جست‌وجوی استفاده‌کنندگان نهایی پایگاه‌های اطلاعاتی رایانه‌ای کتابخانه منطقه‌ای علوم و تکنولوژی شیراز و مقایسه آن با جست‌وجوی کارشناسان اطلاع‌رسانی از نظر ضرایب جامعیت و مانعیت اطلاعات بازیابی شده و زمان جست‌وجو. پایان‌نامه کارشناسی‌ارشد، دانشگاه شیراز، شیراز.
- محمدی، فخرالسادات (۱۳۸۸). توسعه بهینه پرسش‌های جست‌وجو با استفاده از اصطلاحنامه پیوسته اریک و تأثیر آن در مانعیت نتایج و زمان جست‌وجوی کاربر. *علوم و فناوری اطلاعات*، ۲۴ (۴)، ۲۹-۵۲. بازیابی ۲۳ فروردین ۱۳۸۹ از: <http://jlist.irandoc.ac.ir/browse.php>

References

- Bankg, S. L.; Yang, Jae Dong; Yang, Hyung Jeong (2006). Hierarchical document categorization with k -NN and concept-based thesauri. *Information Processing & Management*, 42 (2), 387.
- Gehanno, J. F.; Thiriona, B.; Darmoni, S. J. (2007). Evaluation of Meta-Concepts for Information Retrieval in a Quality- Controlled Health Gateway. *AMIA Annual Symposium Proceedings*, pp. 269-273. Retrieved May 23, 2010, from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655924/?report=abstract>
- Greenberg, J. (1998). *An Examination of the Impact of lexical-semantic relationships on Retrieval Effectiveness During the Query Expansion Process*. Ph.D Thesis, University of Pittsburg.
- Harrison, L. D. (1998). *The Impact of Thesauri on Informatio Retrieval*. Ph.D Thesis, University of New Jersey.
- Knapp, S. D.; Cohen, L. B; Juedes, D. R. (1998). A National Language Thesaurus for the Humanities: the Need for a Database search Aid. *The Library Quarterly*, 68 (4), 406-430.
- Mili, H. et al. (2003). An Experiment Software Component Retrieval. *Information and Software Technology*, 45 (10), 633.

Muddamale, M. R. (1998). Natural Language versus Controlled Vocabulary in Information Retrieval: A Case Study in Soil Mechanics. *Journal of American Society for Information Science*, 49 (10), 881-887.

به این مقاله این گونه استناد کنید:

غائبی، امیر؛ خسروی، فریبرز؛ المیر، آریتا (۱۳۹۰). وضعیت بازیابی اطلاعات در دو پایگاه نمایه و نما و سنجش اثربخشی استفاده از واژگان کنترل شده در نمایه‌سازی این دو پایگاه. *تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی*، ۱۷ (۲)، ۲۴۷-۲۶۲.