

تحلیل متن مقالات فارسی کتابداری و اطلاع‌رسانی و امکان نمایه‌سازی ماشینی آنها بر اساس قانون زیف

● انسیه ملکیان فرد

دانشجوی کارشناسی ارشد دانشگاه تربیت مدرس

واژگانی متن، کلیدواژه‌های عنوانی و توصیف‌گرهای موضوعی، می‌تواند از اهداف فرعی این پژوهش محسوب شود.

فرضیه‌های پژوهش

واژه‌های با بسامد بالای مقالات مورد بررسی از فرمول پیش‌بینی‌پذیر زیف تبعیت می‌کند. نسبت واژه‌های با بسامد پایین در مقالات مورد بررسی بر مبنای فرمول‌های زیف و بوث نسبتی پیش‌بینی‌پذیر است. واژه‌هایی که بین دو قطب واژه‌های با بسامد بالا و واژه‌های با بسامد پایین در محتوای مقالات می‌آید، احتمالاً، بیشترین توصیف را از محتوای آن مقاله به‌دست می‌دهد. نتایج حاصل از بررسی فرضیه سوم با کلیدواژه‌های عنوانی و توصیف‌گرهای موضوعی مقالات همخوانی دارد.

سؤال‌های پژوهش

چه تفاوتی بین معادله‌های زیف و بوث در پیش‌بینی واژه‌های با بسامد پایین وجود دارد؟ چه رابطه‌ای بین طول مقاله (تعداد واژه‌ها) و الگوی حاکم بر فراوانی واژه‌ها وجود دارد؟ چه تفاوتی بین الگوی حاکم بر فراوانی واژه‌ها در متن مقالات و چکیده‌ها وجود دارد؟ و همخوانی (انطباق) میان شیوه نمایه‌سازی ذهنی (توصیف‌گرهای موضوعی) و نمایه‌سازی عینی (بسامد واژگانی) چگونه است؟

پیشینه پژوهش در خارج از کشور

- مایک تلوال (۲۰۰۴) در پژوهشی با عنوان «ویژگی‌های متنی وب‌سایت‌های دانشگاهی انگلیسی‌زبان»

بسامد واژگانی در وب‌سایت‌های سه کشور انگلیسی‌زبان (استرالیا، نیوزلند و انگلستان) را بررسی و تحلیل کرد؛

- مارتینز و سیلوا (۲۰۰۴) در وب‌سایت‌های پرتغالی‌زبان ذخیره شده در موتور جستجوی tumba به تحلیل محتوای متن این وب‌سایت‌ها بر مبنای قانون زیف پرداخته و بسامد واژگانی و الگوی حاکم بر توزیع واژگان، حجم صفحات، زبان صفحات و واژگان صفحات را تجزیه و تحلیل کرد؛

- کانگولی و فانگ (۲۰۰۰) در تحقیق خود با عنوان «رده‌بندی ماشینی مفاهیم حسابداری:

● بلندیان کاشی، صدیقه. «تحلیل متن مقالات فارسی کتابداری و اطلاع‌رسانی و امکان نمایه‌سازی ماشینی آنها بر اساس قانون زیف». پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی دانشکده علوم تربیتی و روان‌شناسی دانشگاه فردوسی مشهد، ۱۳۸۵.

بیان مسئله

موضوع پژوهش حاضر بر این پرسش‌ها استوار است که آیا توزیع واژگان در متون زبان فارسی از الگوی زیف تبعیت می‌کند؟ تا چه اندازه میان کلیدواژه‌های عنوانی، توصیف‌گرهای موضوعی و بسامد واژگانی مقالات فارسی انطباق وجود دارد؟ و در نمایه‌سازی مقالات فارسی بسامد واژگانی تا چه اندازه می‌تواند در ارائه محتوای این مقالات نقش داشته باشد؟

اهمیت و ضرورت پژوهش

در نمایه‌سازی خودکار، روش‌های آماری مبتنی بر فراوانی نسبی کلمات مورد توجه است. همچنین سرعت بالا و ایجاد یکدستی در نمایه از ویژگی‌های این نوع نمایه‌سازی است. اما در پایگاه‌های اطلاعاتی فارسی تاکنون از شیوه‌های استخراجی خودکار بر مبنای بسامد واژگانی استفاده نشده است. از دلایل آن می‌توان به مشکلات مربوط به ویژگی‌های خط و زبان فارسی و همچنین عدم پژوهش در این زمینه اشاره کرد. بنابراین انجام یک پژوهش برای شناخت الگوی توزیع واژگانی متون فارسی و بررسی بسامد واژگانی ضروری به نظر می‌رسد. همچنین با انجام این پژوهش می‌توان تعیین کرد

که نمایه‌سازی خودکار بر مبنای بسامد واژگانی در زبان فارسی تا چه میزان کارایی دارد.

هدف‌های پژوهش

هدف این پژوهش شناسایی الگوی حاکم بر فراوانی واژه‌ها در مقالات فارسی حوزه کتابداری و اطلاع‌رسانی بر مبنای قانون زیف است. بررسی رابطه بین طول مقاله و الگوی حاکم بر فراوانی واژه‌ها و بررسی میزان همخوانی میان بسامد



پایه ادراکی دانش‌آموزان کلاس‌های سوم، چهارم و پنجم ابتدایی شهر تهران است؛ نادیا فکری (۱۳۸۲) در پایان‌نامه کارشناسی ارشد خود، «شناسایی و تحلیل واژگان نوشتاری کودکان دبستانی ایران»، نامه‌های ارسالی کودکان دبستانی سراسر کشور را به دفتر مجلات رشد نوآموز و رشد دانش آموز در طول سال‌های ۱۳۸۰-۱۳۸۱ بررسی کرده است.

*** در زمینه امکان‌سنجی نمایه‌سازی**

عبدالحسین آذرنگ (۱۳۵۴) در پایان‌نامه خود، «شیوه‌های نمایه‌سازی و انطباق آنها با ساختمان زبان فارسی»، به بررسی شیوه‌های نمایه‌سازی و انطباق آنها با ساختمان زبان فارسی می‌پردازد. این پایان‌نامه به بررسی مشکلات زبان فارسی در حوزه‌ای که نمایه‌ساز با آنها مواجه است، اختصاص دارد؛

معرف‌زاده (۱۳۶۶) در پژوهشی با عنوان «بررسی وضعیت ابزار بازیابی اطلاعات پزشکی ایران» به بررسی این پرسش می‌پردازد که کدام شیوه نمایه‌سازی گردان برای ذخیره و بازیابی اطلاعات پزشکی می‌تواند مفید باشد، تا از آن برای تهیه نمایه پزشکی ایران بتوان بهره گرفت؛

زهرا قریب (۱۳۷۷) در پایان‌نامه «بررسی وضعیت نمایه‌سازی در کتابخانه‌ها و مراکز اطلاع‌رسانی مستقر در شهر تهران» وضع نمایه‌سازی را از لحاظ شیوه و زبان نمایه‌سازی، تخصیص نیروی انسانی، بودجه تخصیص‌یافته برای این بخش و استفاده از رایانه در این بخش در کتابخانه‌ها و مراکز اطلاع‌رسانی مستقر در شهر تهران بررسی می‌کند؛

نیاکان (۱۳۸۲) در پژوهش «امکان‌سنجی نمایه‌سازی ماشینی مدارک زبان فارسی در مرکز اطلاع‌رسانی جهاد کشاورزی» فرایند نمایه‌سازی ماشینی و نتایج امکانات لازم را در مرکز اطلاع‌رسانی جهاد کشاورزی مورد بررسی قرار داده تا الگویی مناسب برای استفاده از نمایه‌سازی ماشینی به زبان فارسی در ایران ارائه شود.

*** در زمینه همخوانی و انطباق میان توصیف‌گرهای نمایه‌سازی و کلیدواژه‌های عنوانی**

داور پناه (۱۳۷۵) در پژوهشی، میزان همخوانی عناوین مقالات فارسی با محتوای آنها را در گروه‌های مختلف علمی بررسی کرد؛ بزرگی (۱۳۷۷) نیز پژوهشی با عنوان «بررسی میزان سازگاری عنوان‌ها و موضوع‌های کتاب‌های منتشر شده در ایران در فاصله سال‌های ۷۵-۱۳۶۵» انجام داده است؛

انصاری (۱۳۷۹) در پایان‌نامه خود، «بررسی انطباق میان توصیف‌گرهای نمایه‌سازی و کلیدواژه‌های عنوان پایان‌نامه‌های دکترای تخصصی کودکان، زنان، قلب و عروق، روانپزشکی» درباره میزان انطباق میان توصیف‌گرهای نمایه‌سازی و کلیدواژه‌های عنوانی بحث کرده است؛

ایرانشاهی (۱۳۸۲) نیز به تحقیق درباره میزان همخوانی کلیدواژه‌های عنوان و توصیف‌گرهای نمایه‌سازان در پایگاه چکیده پایان‌نامه‌های ایران پرداخت.

نتایج مقدماتی تحلیل آماری بر اساس اصطلاح مدرک «امکان‌سنجی رده‌بندی ماشینی مفاهیم در رشته‌ای خاص را با استفاده از تحلیل‌های آماری بسامد واژگانی بر مبنای قانون زیف بررسی کردند؛

نظام اسمارت در دانشگاه هاروارد از سال ۱۹۶۱ تا سالیان متمادی ادامه داشت. سالتون که به مدت بیست سال مسئول این نظام بود، با استفاده از طرح‌های گوناگون ارزش‌گذاری اصطلاح‌ها، راهبردهای جست‌وجو، فرمول‌بندی خودکار پرسش جست‌وجو بر مبنای بازخوردهای ربط و جست‌وجوهای بسط‌یافته بولی، دست به بررسی‌های گسترده‌ای زد. در این مطالعات و بررسی‌ها از شیوه‌های پیشرفته تحلیل معنایی و نحوی استفاده شد؛

کریستوفر شوارز (۱۹۹۰) در پژوهشی با عنوان «تحلیل نحوی متن آزاد به‌صورت خودکار» به مسائل و مشکلات موجود در تحلیل نحوی متن آزاد پرداخته است؛

پژوهش دیگر درباره تحلیل متن خودکار، پایان‌نامه دکترای جینگز (۱۹۹۷) با عنوان «حل مسئله عدم تطابق واژه از طریق تحلیل متن خودکار» است؛

مارتین گرومان (۲۰۰۰) در زمینه امکان‌سنجی نمایه‌سازی ماشینی در پژوهشی به بررسی کاربردهای نمایه‌سازی ماشینی در کتابخانه‌های تخصصی پرداخته و با آزمایش بازیابی مدارک، پیاده‌سازی آن را در کتابخانه‌های عمومی بررسی کرد و به این پرسش پاسخ داد که آیا نمایه‌سازی ماشینی، تولید فهرست‌های رایانه‌ای را بهبود می‌بخشد یا خیر؟ - باتل و پرایبیش (۱۹۷۰) در پژوهشی به بررسی سودمندی نمایه‌سازی کوئیک در مقالات روان‌شناسی پرداخته و دریافتند در چکیده‌نامه روان‌شناسی ۶۰ درصد مقالات با استفاده از کلیدواژه‌های عنوان قابل بازیابی است؛

انانسون (۱۹۸۷) در پژوهشی به ارزیابی جست‌وجوی کلیدواژه‌های عنوان پرداخت؛

شفرود (۱۹۸۷) میزان همخوانی کلیدواژه‌های عنوانی و توصیف‌گرهای موضوعی را در پایگاه اطلاعاتی OON مورد بررسی قرار داد.

پیشینه پژوهش در داخل ایران

*** در زمینه تحلیل متن بر اساس بسامد واژگانی**

اگرچه پژوهشی درباره تحلیل متن بر اساس بسامد واژگانی در حوزه نمایه‌سازی و بازیابی اطلاعات در ایران یافت نشد، در حوزه زبان‌شناسی و تحلیل واژگان متن می‌توان به پایان‌نامه دکترای فریدون بدره‌ای (۱۳۵۰) با عنوان «وسعت واژگان کودکان دبستانی ایران بر پایه نوشته‌های آنان» اشاره کرد؛

پژوهش دیگر درباره واژگان پایه فارسی، رساله کارشناسی ارشد مهدی ضرغامیان (۱۳۷۱) در رشته زبان‌شناسی همگانی با عنوان «واژگان خوانداری کودکان دبستانی ایران» است که در آن مجموعه‌ای از واژگان خوانداری کودکان دبستانی ایران در دهه شصت بررسی شده است؛

فرامرز صفرزاده (۱۳۷۶) در پایان‌نامه خود، «واژگان گفتاری پایه کودکان دبستانی ایران در مدارس تهران»، پژوهشی بر روی شصت دانش‌آموز دختر و پسر (پایه‌های سوم، چهارم و پنجم) انجام داده است؛

پایان‌نامه دیگر پژوهش علی محمد شستیری (۱۳۷۶) با عنوان «واژگان

نوع و روش پژوهش

این پژوهش از نوع کاربردی است و با روش تحلیل محتوا صورت گرفته است.

جامعه آماری، حجم نمونه و روش نمونه‌گیری

جامعه آماری این پژوهش، تمام مقالات چاپ شده در آخرین شماره‌های مجلات معتبر کتابداری و اطلاع‌رسانی در سال ۱۳۸۳ است و شامل ۱۳ مقاله از فصلنامه کتاب، هفت مقاله از فصلنامه کتابداری و اطلاع‌رسانی آستان قدس رضوی و هفت مقاله از مجله کتابداری کتابخانه مرکزی دانشگاه تهران است. از آنجا که در این پژوهش تحلیل واژگان مورد نظر بوده، از این رو با توجه به حجم زیاد کار، با مشورت استادان راهنما و مشاور ده مقاله با روش نمونه‌گیری تصادفی ساده انتخاب شده‌اند.

عنوان مقالات مورد بررسی عبارت‌اند از: ۱. پست‌مدرنیته و علوم کتابداری و اطلاع‌رسانی؛ ۲. آرشیوهای رقومی؛ ۳. تنش در کتابخانه؛ ۴. مدیریت مجلات الکترونیکی در کتابخانه‌ها؛ ۵. بررسی نظر دانشجویان دانشکده علوم تربیتی و روان‌شناسی دانشگاه فردوسی مشهد درباره خواندن به قصد یادگیری، خواندن به قصد لذت، خواندن به قصد تصمیم‌گیری؛ ۶. مطالعه‌ای در مورد نحوه ارائه خدمات به دانشجویان نابینا در کتابخانه مرکزی و مرکز اسناد دانشگاه تهران؛ ۷. خدمات تحویل مدرک: مفهومی ناآشنا در کتابداری ایران؛ ۸. وب‌نگاشت‌ها و نقش آنها در استفاده از منابع علمی اینترنت؛ ۹. خدمات مرجع دیجیتال در کتابخانه‌های دانشگاهی؛ و ۱۰. کتابداران و تجارت الکترونیک: وادار کنید تجارت الکترونیک در خدمت شما باشد.

ابزار و روش گردآوری داده‌ها

داده‌های مورد نیاز پژوهش طی مراحل زیر گردآوری شده است: ۱. تهیه متن الکترونیکی مقالات؛ ۲. تفکیک واژگان (برای تفکیک واژگان از دو روش «صرفاً ماشینی بدون دخالت عامل انسانی» و تفکیک واژگان به صورت «ماشینی با دخالت عامل انسانی» استفاده شده است؛ به علاوه معیار تفکیک واژگان «تعیین حدود واژه» در پژوهش حاضر در شیوه «صرفاً ماشینی بدون دخالت عامل انسانی» مبتنی بر برداشت ساختاری و در شیوه «ماشینی با دخالت عامل انسانی» مبتنی بر برداشت نشانه‌مند از واژه‌هاست؛ ۳. شمارش بسامد واژگان؛ و ۴. ابزار تحلیل گر متن و تفکیک واژگان.

محدودیت‌های پژوهش

* مشکل زمانی برای انجام این تحقیق

پژوهشگر بیشترین محدودیت این پژوهش را کمبود زمان بیان کرده است. زیرا کار با واژگان بسیار دقیق و وقت‌گیر و نیازمند صرف وقت فراوان بوده است. کار تحلیل متن و شمارش واژگان و تهیه جداول زیف و تعیین واژه نمایه‌ها در مرحله اول و سپس تجزیه و تحلیل آماری برای مقایسه همخوانی بین سه نوع کلیدواژه مورد نظر در این تحقیق (بسامد واژگانی، کلیدواژه عنوانی و توصیف‌گر موضوعی) در مرحله دوم به‌ویژه اینکه تمام این مراحل در دو شیوه متمایز «صرفاً ماشینی» و «ماشینی با دخالت عامل انسانی» انجام گرفته همگی مستلزم کار مستمر و

توان فرسا با رایانه بودند.

* مشکلات مربوط به تهیه متن الکترونیکی مقالات

درباره تهیه متن الکترونیکی مقالات فقط مجله کتابداری آستان قدس در قالب word دریافت شد و فصلنامه کتاب در قالب pdf و کتابداری دانشگاه تهران در قالب زرنگار دریافت شد. از آنجا که امکان تحلیل واژگان در قالب اخیر وجود نداشت، لذا باید به قالب word تبدیل می‌شدند. بدین ترتیب به‌منظور رعایت یکدستی هنگام تفکیک واژگان، مقالات این دو مجله بر اساس الگوی تایپ مقالات مجله کتابداری آستان قدس رضوی در نرم افزار word تایپ شدند.

* مشکلات مربوط به تعیین حدود واژگان در شیوه «ماشینی با دخالت عامل انسانی»

در این شیوه برای تعیین مرز واژگان، نیاز بود از اصول و معیار استاندارد استفاده شود. بدین منظور باید تعریف واژه از دیدگاه‌های مختلف مورد بررسی و مقایسه قرار گیرد که کار بسیار پرزحمت و وقت‌گیری بود.

نتایج فرضیه‌های پژوهش

- فرضیه اول مبتنی بر اینکه واژه‌های با بسامد بالای مقالات مورد بررسی، از فرمول پیش‌بینی‌پذیر زیف تبعیت می‌کند، مورد تأیید قرار گرفت؛

- فرضیه دوم مبتنی بر اینکه نسبت واژه‌های با بسامد پایین در مقالات مورد بررسی، بر مبنای فرمول‌های زیف و بوث نسبتی پیش‌بینی‌پذیر است، مورد تأیید قرار گرفت؛

- فرضیه سوم پژوهش مبتنی بر اینکه واژه‌هایی که بین دو قطب واژه‌های با بسامد بالا و واژه‌های با بسامد پایین در محتوای متن مقالات می‌آید، احتمالاً بیشترین توصیف را از محتوای آن مقاله به‌دست می‌دهد و نمایه‌سازان می‌توانند به این واژگان در هنگام نمایه‌سازی توجه کنند، مورد تأیید قرار گرفت. همچنین نتیجه این بررسی‌ها نشان داد که واژه‌های زیر پرسامدترین واژگان بازدارنده در متون فارسی کتابداری و اطلاع‌رسانی هستند؛ و در / به / که / از / را / است / این / برای / یا / یک / باید / آن / آنها / نیز / بر / مانند / می‌شود / هر / بین / می‌کند / دارد / یا / روی .

- در مورد فرضیه چهارم مبنی بر اینکه نتایج حاصل از بررسی فرضیه سوم با کلیدواژه‌های عنوانی و توصیف‌گرهای موضوعی مقالات همخوانی دارد، نتیجه‌ای که از بررسی‌ها می‌توان گرفت این است که در شیوه تفکیک واژگانی صرفاً ماشینی ۲۷/۶ درصد توصیف‌گرهای موضوعی و حدود ۱۲/۵ درصد کلیدواژه‌های عنوانی مقالات مورد بررسی با واژه نمایه‌های حاصل از بسامد واژگانی همخوانی (کامل و نسبی) دارد. در حالی که در شیوه تفکیک واژگانی با دخالت عامل انسانی ۶۳/۱۸ درصد توصیف‌گرهای موضوعی و حدود ۶۱ درصد از کلیدواژه‌های عنوانی با واژه‌نمایه‌های حاصل از بسامد واژگانی همخوانی (کامل و نسبی) دارد. بنابراین، تفکیک واژگانی به شیوه «ماشینی با دخالت عامل انسانی» نسبت به شیوه «صرفاً ماشینی» نتیجه بهتری به‌دست می‌دهد. به عبارت دیگر تفکیک واژگان به شیوه «ماشینی با دخالت عامل انسانی» موجب جلوگیری از ذخیره‌سازی غلط، هماری و بازیابی کاذب در نظام ذخیره

و بازیابی می‌شود.

در ادامه بررسی این فرضیه، این سؤال مطرح شد: چه تفاوتی بین تعداد متوسط کلیدواژه‌های عنوانی، توصیف‌گرهای موضوعی و واژه‌نمایه‌ها وجود دارد؟

آزمون‌ها و بررسی‌ها نشان داد که واژه‌نمایه‌های حاصل از بسامد واژگانی توصیف عمیق‌تری از مدرک به‌دست می‌دهند و دارای جامعیت بیشتری‌اند. به‌عبارت دیگر می‌توان گفت در کلیدواژه‌های عنوانی و کلیدواژه‌های موضوعی اختصاص داده شده به مقالات، برخی مفاهیم موضوعی موجود در مدرک نادیده گرفته شده و پنهان باقی مانده‌اند. با استفاده از بسامد واژگانی می‌توان مفاهیم پنهان مدرک را شناسایی و با اختصاص واژه‌نمایه‌ها نقاط دسترسی بیشتری برای مدرک فراهم آورد. بی‌تردید هر چه نقاط دسترسی بیشتر شود، میزان دسترس‌پذیری آن مدرک نیز افزایش می‌یابد.

نتایج سؤال‌های پژوهش

در پاسخ به سؤال اول، بررسی‌ها نشان داد که در هر دو شیوه تفکیک «صرفاً ماشینی بدون دخالت عامل انسانی» و «ماشینی با دخالت عامل انسانی» نسبت پیش‌بینی شده به نتایج محاسبه شده بوث نزدیک‌تر است. و قاعده بوث در مقایسه با معادله زیف پیش‌بینی بهتری را ارائه می‌کند.

در پاسخ به سؤال دوم، بررسی‌ها نشان داد به‌رغم تفاوت در طول مقالات مورد بررسی، الگوی واژگانی آنها اختلاف چندانی ندارد و به‌طور کلی در مقالات با طول متفاوت (تعداد واژه‌ها)، الگوی مشابهی بر فراوانی واژه‌ها حاکم است.

در پاسخ به سؤال سوم، بررسی‌ها نشان داد که الگوی حاکم بر فراوانی واژه‌ها در چکیده مقالات مشابه با الگوی حاکم بر فراوانی واژه‌ها در متن مقالات است. به‌گونه‌ای که واژگان بازدارنده در صدر سیاهه و واژه‌نمایه‌ها در میانه سیاهه و واژگان کم تکرار در انتهای سیاهه قرار دارند. بنابراین توزیع واژگانی چکیده‌ها از فرمول پیش‌بینی‌پذیر زیف تبعیت می‌کند. همچنین حضور تمام یا برخی از کلیدواژه‌های عنوانی، توصیف‌گرهای موضوعی و واژه‌نمایه‌ها در همه چکیده‌ها نشان‌دهنده این الگوی واحد و پیش‌بینی‌پذیر است.

در پاسخ به سؤال چهارم، معلوم شد همخوانی میان دو شیوه نمایه‌سازی ذهنی (توصیف‌گرهای موضوعی) و نمایه‌سازی عینی (بسامد واژگانی) در دو شیوه تفکیک واژگانی مورد استفاده در این پژوهش دارای تفاوت قابل ملاحظه‌ای است. به‌گونه‌ای که در شیوه تفکیک «ماشینی با دخالت عامل انسانی» همخوانی دو شیوه نمایه‌سازی تقریباً سه برابر بیش از شیوه تفکیک «صرفاً ماشینی» است.

پیشنهادهای اجرایی

با توجه به شیوه‌های مختلف نگارش و در نتیجه پراکندگی و عدم یکدستی در انتخاب واژگان، پیشنهاد می‌شود، شیوه واحد و استاندارد برای حروفچینی مقالات فارسی اتخاذ شود.

معیار تفکیک واژه و کلمات مرکب زیر نظر یک مرکز خاص (مانند فرهنگستان زبان فارسی) به‌صورت یکسان و هماهنگ تعریف شود تا پژوهش‌های بعدی یکدستر و قابل مقایسه‌تر شوند.

یافته‌های پژوهش نشان داد، در میان واژه‌نمایه‌ها (بسامد واژگانی) به‌دست آمده از طریق معادله ۴ گافمن، کلیدواژه‌های مناسب و مبین محتوای موضوعی مدارک وجود دارند که در میان توصیف‌گرهای موضوعی و کلیدواژه‌های عنوانی نیامده‌اند. بنابراین پیشنهاد می‌شود با اعمال معیارهای بسامدی در نمایه‌سازی خودکار، این واژگان مورد توجه قرار گرفته و بدین ترتیب نقاط دسترسی به منابع مرتبط افزایش یابد.

یافته‌های پژوهش نشان داد که میزان انطباق و همخوانی میان واژه‌نمایه‌ها (بسامد واژگانی) و توصیف‌گرهای موضوعی در شیوه تفکیک «ماشینی با دخالت عامل انسانی» بیشتر است. این امر به‌دلیل اعمال برخی اصول و معیارهای برشمرده و ویرایش مقالات قبل از تفکیک واژگان بود. از این‌رو پیشنهاد می‌شود با اعمال نوعی نمایه‌سازی به کمک رایانه و به‌منظور دستیابی به نتایج بهتر در انتخاب واژه‌نمایه‌های مطلوب و کاهش ریزش کاذب، بهبود اثربخشی بازیابی اطلاعات را افزایش دهیم.

پیشنهادهایی برای پژوهش‌های آینده

در این پژوهش یک بررسی مقدماتی به‌منظور شناسایی الگوی حاکم بر توزیع واژگانی و چگونگی بسامد واژگانی در مقالات کتابداری و اطلاع‌رسانی صورت گرفت. پیشنهاد می‌شود چنین پژوهشی در سایر حوزه‌های موضوعی نیز صورت پذیرد.

پیشنهاد می‌شود درباره بسامد و رتبه واژه‌ها در متون مقالات فارسی تحقیق و پژوهش شود و بر اساس یافته‌های آن، توصیف‌گرها و ارجاع‌های موجود در اصطلاح‌نامه‌های فارسی بازنگری شود.

یافته‌های این پژوهش نشان داد که واژگان بازدارنده، بالاترین بسامد را به خود اختصاص داده و با حذف این واژگان به میزان قابل توجهی از حجم واژگان مدارک کاسته می‌شود. لذا پیشنهاد می‌شود پژوهشی برای شناسایی واژگان بازدارنده در متون حوزه‌ها و رشته‌های مختلف علمی صورت گیرد تا به‌عنوان فهرست استاندارد هنگام نمایه‌سازی در اختیار مراکز اطلاع‌رسانی قرار گیرد.

بر مبنای مدل زیف امکان بهینه‌سازی نمایه‌سازی خودکار متون فارسی بررسی می‌شود.

نقد و نظر

پژوهش کاملاً کاربردی است و در زبان فارسی همان‌طور که خود پژوهشگر بیان کرده، نیاز به چنین پژوهشی احساس می‌شده است.

بعضی از موارد از جمله آنهایی که در «تعاریف مفهومی و عملیاتی» آمده‌اند به اندازه‌ای رسا نیستند که کاربران عادی بتوانند به‌راحتی از آن استفاده کنند.

در بررسی سوابق پژوهش، پژوهشگر دامنه وسیعی از پژوهش‌های پیشین را ذکر کرده که سبب می‌شود که موضوع بهتر مطرح شود و بر اعتبار پژوهش می‌افزاید.

جامعه آماری پژوهش مجلات پرمخاطب کتابداری و اطلاع‌رسانی است که این موجب شده، پژوهش برای کاربران این حوزه ملموس‌تر شود.

با توجه به محدودیت‌هایی که پژوهشگر با آنها مواجه بوده، کوشیده است نتایج پژوهش به اندازه کافی مشخص و کامل ارائه شود.

پیشنهادهای اجرایی در شرایط کنونی کاملاً قابل اعمال هستند.