

بررسی میزان همخوانی عبارات‌های جستجوی کاربران با اصطلاحات پیشنهادی مقالات در پایشینه‌های کتابشناختی پایگاه‌های اطلاعاتی غیرفارسی EBSCO و IEEE

اعظم صنعت‌جو (نویسنده مسئول)

استادیار گروه علم اطلاعات و دانش‌شناسی دانشگاه فردوسی مشهد
sanatjoo-a@ferdowsi.um.ac.ir

حمیده رجیبیان

کارشناس ارشد علم اطلاعات و دانش‌شناسی دانشگاه فردوسی مشهد
rajabeian@gmail.com
تاریخ دریافت: ۱۳۹۲/۰۲/۰۴؛ تاریخ پذیرش: ۱۳۹۴/۰۸/۰۳

چکیده

هدف: هدف پژوهش حاضر بررسی میزان همخوانی بین کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی پایگاه‌های اطلاعاتی است. پایگاه‌های اطلاعاتی محتوای موضوعی مدارک خود را با استفاده از واژگان زبان طبیعی یا زبان کنترل شده، در فیلدهای کتابشناختی مشخصی که همراه با سایر اطلاعات کتابشناختی مدرک همراه است، نمایان می‌کنند که اصطلاحات پیشنهادی مقالات نامیده می‌شوند.

روش: این پژوهش از نوع کاربردی است و به روش تحلیل محتوا انجام گرفته است. جامعه آماری این پژوهش، پایگاه‌های اطلاعاتی ایسکو و آی‌تریپل‌ای بودند که تراکنش‌های ثبت شده طی یک دوره سه ماهه مورد بررسی قرار گرفتند.

یافته‌ها: یافته‌ها نشان دادند که در هر دو پایگاه اطلاعاتی، همخوانی کامل کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی آزاد از اصطلاحات پیشنهادی کنترل شده پایگاه‌ها بیشتر بود. حال آنکه میزان همخوانی نسبی کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی کنترل شده بیش از اصطلاحات پیشنهادی آزاد بود. همچنین بین میزان همخوانی کلیدواژه‌های جستجوی کاربران با هر یک از انواع اصطلاحات پیشنهادی، تفاوت معناداری مشاهده شد. نتایج نشان داد که این اصطلاحات توانمندی بالایی برای به کارگیری در نمایه‌سازی منابع، انتخاب اصطلاحات اصطلاحنامه‌ها و منبع انتخاب کلیدواژه‌های جستجو در فرایند توسعه جستجو دارند.

اصالت/ارزش: ارزش این مقاله در نشان دادن ارزش کلیدواژه‌های آزاد در جستجوی متون از طریق وبگاه‌ها و لزوم توجه بیشتر به آنها در نمایه‌سازی خودکار و نیمه‌خودکار متون در پایگاه‌های اطلاعاتی است.

کلیدواژه‌ها: اصطلاحنامه، کلیدواژه‌های آزاد، کلیدواژه‌های کنترل شده، توصیفگر، نمایه‌سازی.

مقدمه

در پایگاه‌های اطلاعاتی، بازنمون اطلاعات برای کمک به بازیابی اطلاعات به شیوه‌های مختلف انجام می‌پذیرد. یکی از مهمترین شیوه‌های بازنمایی محتوا، نمایه‌سازی است که طی آن با استفاده از زبان طبیعی یا واژگان کنترل‌شده، محتوای مدرک به اصطلاحات نمایه‌ای تبدیل می‌شود. به منظور دسترس‌پذیر بودن هر چه بیشتر محتوای مدرک، پایگاه‌های اطلاعاتی محتوای موضوعی مدرک را در منطقه کتابشناختی مشخصی نمایش می‌دهند. به‌طور کلی هر نوع از این بازنمون‌های موضوعی مدرک، اصطلاحات پیشنهادی مقالات نامیده می‌شوند که غالباً در انتهای مقالات و همراه با سایر اطلاعات کتابشناختی مدرک می‌آیند. با توجه به گسترش انواع بازنمون‌های اطلاعات، قالب‌های متعددی از انواع اصطلاحات پیشنهادی نظیر اصطلاحات نویسندگان، اصطلاحات موضوعی و مانند آن در پایگاه‌های اطلاعاتی رایج شده است. اگر چه ایجاد هر کدام از این اصطلاحات پیشنهادی در پایگاه‌های اطلاعاتی بر منطقی استوار است اما مسئله پژوهش حاضر، میزان سودمندی هر یک از این اصطلاحات در تطابق با زبان جستجوی کاربران است. وجود منابع و پژوهش‌های متعدد منتشر شده در ارتباط با اصطلاحات پیشنهادی یا همان توصیفگرها و کلیدواژه‌های تعیین شده برای مقالات، حاکی از تمایل پژوهشگران به درک این اصطلاحات و تشخیص موقعیت آن‌ها نسبت به سایر توصیفگرها و کلیدواژه‌ها است. بنابراین، بیشتر پژوهش‌ها (انصاری، ۱۳۷۷؛ داورپناه و ایرانشاهی، ۱۳۸۴؛ فتاحی و ارسطوپور، ۱۳۸۵؛ فضیلت، ۱۳۸۸؛ نوکاریزی و دهقانی، ۱۳۹۰) به مقایسه این اصطلاحات با کلیدواژه‌های عنوان، سرعنوان‌های موضوعی، فهرست مندرجات و مانند آن پرداخته‌اند. در بررسی مرتبط‌ترین اثر، تنها نیک‌زمان (۱۳۸۹) به انجام پژوهشی پرداخت که در آن، زبان جستجوی موضوعی کاربران به لحاظ همخوانی با سرعنوان‌های موضوعی فارسی مورد بررسی قرار گرفت. در یک تحلیل کلی یافته‌های وی نشان داد که در جستجوهای موضوعی انجام گرفته در دانشکده کشاورزی دانشگاه فردوسی، میزان همخوانی کامل ۳۰/۹ درصد، میزان همخوانی نسبی ۱۹/۵ درصد و میزان عدم همخوانی ۴۹/۶ بود. در جستجوهای موضوعی انجام گرفته در دانشکده علوم تربیتی و روانشناسی دانشگاه فردوسی نیز میزان همخوانی کامل ۳۷/۷ درصد، میزان همخوانی نسبی ۲۱/۱ درصد و میزان عدم همخوانی ۴۱/۲ درصد بود. اما در این پژوهش

نیز، میزان همخوانی عبارت‌های جستجوی کاربران با انواع مختلف اصطلاحات پیشنهادی به ویژه اصطلاحات آزاد و کلیدواژه‌های نویسندگان بررسی نشده است. در خارج از کشور، پژوهش‌ها تنوع بیشتری را در این باره دارند. برای مثال ووربیج^۱ (۱۹۹۸)، به مقایسه بین کلیدواژه‌های عنوان و توصیفگرهای موضوعی رکوردها پرداخت. یا ناتسو^۲ (۱۹۹۸)، تعداد رخدادهای اصطلاحات نمایه‌ای و میزان تکرارپذیری آنها در مجلات مختلف را بررسی نمود. در انواع دیگری از پژوهش‌ها، انواع مختلف اصطلاحات پیشنهادی از جمله برجسب‌های کاربران، کلیدواژه‌های نویسندگان و توصیفگرها با یکدیگر مقایسه شده‌اند (گیل لیوا^۳ و آلونسو آرویو^۴، ۲۰۰۷). اما اینکه تا چه میزان این اصطلاحات پیشنهادی با اصطلاحات واقعی جستجو شده توسط کاربران نزدیک است و با آنها همخوانی دارد، تاکنون مورد بررسی قرار نگرفته است. بدین منظور این پژوهش به بررسی میزان همخوانی کلیدواژه‌های جستجوی کاربران با انواع اصطلاحات پیشنهادی موجود در پایگاه‌های اطلاعاتی ابسکو^۵ و آی‌تریپل‌ای^۶ می‌پردازد.

پرسش پژوهش

پرسش اصلی پژوهش: تا چه میزان کلیدواژه‌های جستجوی کاربران در هر یک از دو پایگاه اطلاعاتی لاتین با هر نوع از اصطلاحات پیشنهادی همان پایگاه اطلاعاتی از جمله اصطلاحات موضوعی، کلیدواژه‌های نویسندگان، اصطلاحات کنترل شده و اصطلاحات کنترل نشده همخوانی دارد؟

فرضیه‌ها

فرضیه یک: میزان همخوانی کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی نویسندگان در پایگاه عمومی ابسکو و پایگاه تخصصی آی‌تریپل‌ای با یکدیگر متفاوت است. فرضیه دو: بین میزان همخوانی کلیدواژه‌های جستجوی کاربران با هر یک از انواع اصطلاحات پیشنهادی (اصطلاحات موضوعی، کلیدواژه‌های نویسندگان، اصطلاحات کنترل شده و کنترل نشده) در هر پایگاه اطلاعاتی تفاوت معناداری وجود دارد.

1. Voorbij

2. Natsuo

3. Gil-Leiva

4. Alonso-Arroyo

5. EBSCO (Academic search premier)

6. IEEE (Institute of Electrical and electronics engineers)

روش پژوهش

پژوهش حاضر از نوع کاربردی است و با روش تحلیل محتوا انجام پذیرفت. جامعه آماری این پژوهش پایگاه‌های اطلاعاتی لاتین بودند که کتابخانه مرکزی دانشگاه فردوسی مشهد آنها را مشترک بود. از این پایگاه‌ها، دو پایگاه اطلاعاتی اسکوپ به‌عنوان پایگاه اطلاعاتی عمومی (پوشش‌دهنده مجلات و کتب علمی با حوزه‌های موضوعی متنوع) و پایگاه اطلاعاتی آی‌تریپل‌ای به‌عنوان پایگاه اطلاعاتی تخصصی (پوشش‌دهنده کتب و مجلات در حوزه برق و الکترونیک) انتخاب شدند. از دلایل انتخاب این دو پایگاه اطلاعاتی، در دسترس بودن آنها و ارائه طیف متنوعی از اصطلاحات پیشنهادی کنترل شده و آزاد توسط این پایگاه‌ها است که انجام بررسی و مقایسه را میسر می‌نمود. به‌منظور گردآوری داده‌ها، فایل‌های تراکش ثبت کلیدواژه‌های جستجوی کاربران در این دو پایگاه اطلاعاتی در بازه زمانی فروردین تا خرداد ۹۱ مورد تحلیل قرار گرفت و با اصطلاحات پیشنهادی مقالات در همان پایگاه اطلاعاتی مطابقت داده شد. در این میان، اسامی غیرموضوعی نظیر اسامی افراد و مجلات و همچنین عنوان مقالات از سیاهه کلیدواژه‌های جستجوی کاربران کنار گذاشته شد.

همچنین، گردآوری داده‌ها بر اساس دستور کار کدگذاری در سیاهه‌وارسی پژوهش انجام پذیرفت. در این شیوه، پس از استخراج کلیدواژه‌های جستجوی کاربران از فایل‌های تراکش مربوطه، انواع همخوانی‌ها بین کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی مقالات همان پایگاه اطلاعاتی از جمله همخوانی کامل و نسبی و یا عدم همخوانی مورد بررسی قرار گرفت. برای پاسخگویی دقیق به فرضیه‌ها از آزمون‌های ناپارامتری یو مان‌ویتنی^۱، رتبه‌های علامت‌دار ویلکاکسون^۲ و فریدمن^۳ استفاده گردید. برای تأیید پایایی سیاهه‌وارسی، دو بار کدگذاری (مطابقت‌دهی) صورت پذیرفت: یک بار توسط پژوهشگر و بار دیگر توسط یک متخصص از حوزه فنی و مهندسی. نزدیکی بسیار بالای هر دو کدگذاری حاکی از پایایی بالای این ابزار دارد. همچنین روایی سیاهه‌وارسی سنجش میزان همخوانی عبارت‌های جستجوی کاربران با توصیفگرهای پایگاه‌های اطلاعاتی، با نظر چند تن از متخصصان حوزه فنی و مهندسی و علوم انسانی مورد تأیید قرار گرفت.

1. Mann-Whitney U test

2. Wilcoxon signed-rank test

3. Friedman Test

یافته‌ها

۱. تا چه میزان کلیدواژه‌های جستجوی کاربران در هر پایگاه اطلاعاتی با انواع اصطلاحات پیشنهادی همان پایگاه اطلاعاتی (موضوعی، نویسندگان، کنترل شده، کنترل نشده) همخوانی دارد؟ اصطلاحات پیشنهادی موجود در پایگاه اطلاعاتی اِسکو که همراه با سایر اطلاعات کتابشناختی می‌آیند، در دو فیلد اصطلاحات موضوعی^۱ و اصطلاحات نویسندگان^۲ ارائه می‌شوند که میزان همخوانی هر کدام از آنها با کلیدواژه‌های جستجوی کاربران در جدول ۱ آمده است:

جدول ۱. میزان همخوانی کلیدواژه‌های جستجوی کاربران با انواع اصطلاحات پیشنهادی مقالات در پایگاه اطلاعاتی اِسکو

جمع کل		اصطلاحات موضوعی		اصطلاحات نویسندگان		انواع همخوانی
درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	
۴۶	۵۳۱	۳۱٫۹	۱۸۴	۶۰٫۲	۳۴۷	همخوانی کامل
۳۸	۴۳۸	۴۳٫۷	۲۵۲	۳۲٫۲	۱۸۶	همخوانی نسبی
۱۶	۱۸۵	۲۴٫۴	۱۴۱	۷٫۶	۴۴	عدم همخوانی
۱۰۰	۱۱۵۴	۱۰۰	۵۷۷	۱۰۰	۵۷۷	مجموع

همانطور که در جدول ۱ مشاهده می‌شود، بیشترین میزان همخوانی میان کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی موضوعی، از نوع همخوانی نسبی (۴۳٫۷ درصد) و با اصطلاحات پیشنهادی نویسندگان، از نوع همخوانی کامل (۶۰٫۲ درصد) بود. در مجموع بیشترین میزان همخوانی بین کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی مقالات پایگاه اطلاعاتی اِسکو از نوع همخوانی کامل (۴۶ درصد) بود. در مقایسه، نیک‌زمان در پژوهش خود نشان داد که بیشتر عبارت‌های موضوعی کاربران با سرعنوان‌های موضوعی فارسی همخوانی ندارند و نتیجه گرفت زبان سرعنوان‌های موضوعی فارسی از زبان جستجوی کاربران دور است. اما اصطلاحات نمایه‌ای پایگاه اطلاعاتی اِسکو که برگرفته از سرعنوان‌های

1. Subject terms

2. Author keywords

موضوعی کنگره است، با عبارت‌های جستجوی کاربران همخوانی نسبی و کامل خوبی داشتند و لذا به نظر، این سرعنوان موضوعی جستجو مدار تر است. می‌توان نتیجه گرفت که باید در افزایش نزدیکی زبان سرعنوان‌های موضوعی فارسی با زبان جستجوی کاربران کوشید. از سوی دیگر، اصطلاحات پیشنهادی موجود در پایگاه آی‌تریپل‌ای، از تنوع گسترده‌تری برخوردار است. در اکثر مدارک موجود در این پایگاه، می‌توان چهار نوع اصطلاح پیشنهادی ذیل را همراه با سایر اطلاعات کتابشناختی هر مدرک مشاهده نمود که عبارتند از: اصطلاحات آی‌تریپل‌ای^۱، اصطلاحات کنترل‌نشده اینسپک^۲، اصطلاحات کنترل‌نشده اینسپک^۳ و اصطلاحات نویسندگان^۴. میزان همخوانی عبارت‌های جستجوی کاربران را با هر نوع از این اصطلاحات می‌توان در جدول ۲ مشاهده نمود:

جدول ۲. میزان همخوانی کلیدواژه‌های جستجوی کاربران با انواع اصطلاحات پیشنهادی مقالات در پایگاه اطلاعاتی آی‌تریپل‌ای

جمع کل		اصطلاحات نویسندگان		اصطلاحات کنترل‌نشده اینسپک		اصطلاحات کنترل‌شده اینسپک		اصطلاحات آی‌تریپل‌ای		پایگاه‌ها انواع همخوانی
درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	
۲۹,۷	۳۸۴	۱۱,۸	۳۸	۹,۹	۳۲	۵۵,۹	۱۸۱	۴۱	۱۳۳	همخوانی کامل
۴۸,۹	۶۳۴	۵۱,۵	۱۶۷	۵۷,۷	۱۸۷	۳۸,۳	۱۲۴	۴۸,۲	۱۵۶	همخوانی نسبی
۲۱,۴	۲۷۸	۳۶,۷	۱۱۹	۳۲,۴	۱۰۵	۵,۸	۱۹	۱۰,۸	۳۵	عدم همخوانی
۱۰۰	۱۲۹۶	۱۰۰	۳۲۴	۱۰۰	۳۲۴	۱۰۰	۳۲۴	۱۰۰	۳۲۴	مجموع

بیشترین میزان همخوانی میان کلیدواژه‌های جستجوی کاربران با همه انواع اصطلاحات پیشنهادی در پایگاه آی‌تریپل‌ای (اصطلاحات خود آی‌تریپل‌ای، اصطلاحات کنترل‌شده اینسپک، اصطلاحات کنترل‌نشده اینسپک، و اصطلاحات پیشنهادی نویسندگان) از نوع همخوانی نسبی بود و درصد همخوانی نیز به ترتیب ۵۱/۵، ۵۷/۷ درصد، ۵۵/۹ درصد و ۴۸/۲ درصد به دست آمد.

1. IEEE Terms
3. INSPEC uncontrolled terms

2.. INSPEC controlled terms
4. Author Keywords

همچنین در نگاهی دیگر، با مقایسه انواع اصطلاحات پیشنهادی کنترل‌شده و آزاد در هر پایگاه اطلاعاتی، مشخص شد که در پایگاه اطلاعاتی ایسکو، بیش از ۷۵ درصد کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی از نوع کنترل‌شده (اصطلاحات موضوعی) همخوانی کامل و نسبی دارند. همچنین در پایگاه آی‌تریپل‌ای، بیش از ۶۰ درصد کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی کنترل‌شده (اصطلاحات آی‌تریپل‌ای و اصطلاحات کنترل‌شده اینسپک) همخوانی کامل و نسبی دارند اما تفاوت میان همخوانی نسبی با همخوانی کامل در اصطلاحات پیشنهادی کنترل‌شده پایگاه آی‌تریپل‌ای به میزان قابل توجهی از پایگاه ایسکو بیشتر است. این بدان معناست که میزان همخوانی کامل و نسبی اصطلاحات پیشنهادی کنترل‌شده در پایگاه ایسکو به هم نزدیک است و اختلاف موجود قابل توجه نیست در حالیکه در پایگاه آی‌تریپل‌ای، میان میزان همخوانی نسبی و همخوانی کامل تفاوت بیشتری وجود دارد یعنی همخوانی کامل میان کلیدواژه‌های جستجوی کاربران با این نوع از اصطلاحات پیشنهادی بسیار کم و میزان همخوانی نسبی کلیدواژه‌های جستجوی کاربران با این اصطلاحات زیاد است. شاید یکی از دلایل این امر را بتوان به نوع زبان کنترل‌شده اصطلاحات پیشنهادی در هر پایگاه اطلاعاتی دانست بدین معنا که اصطلاحات پیشنهادی کنترل‌شده در پایگاه اطلاعاتی ایسکو از سرعنوان‌های موضوعی کتابخانه کنگره آمریکا انتخاب می‌شوند در حالیکه اصطلاحات پیشنهادی کنترل‌شده در پایگاه اطلاعاتی آی‌تریپل‌ای، برگرفته از اصطلاحنامه‌های تخصصی (اینسپک و آی‌تریپل‌ای) هستند. چون مداخل سرعنوان‌های موضوعی عمومی‌تر و اصطلاحنامه‌ها تخصصی‌ترند لذا اصطلاحنامه‌های تخصصی نسبت به سرعنوان‌های موضوعی از زبان جستجوی کاربران دورتر هستند.

میزان همخوانی نسبی و کامل اصطلاحات پیشنهادی از نوع زبان آزاد در پایگاه اطلاعاتی ایسکو (کلیدواژه‌های نویسندگان) و در پایگاه اطلاعاتی آی‌تریپل‌ای (اصطلاحات کنترل‌نشده اینسپک و کلیدواژه‌های نویسندگان) با اصطلاحات جستجو به میزان بسیار قابل توجهی (بیش از ۹۰ درصد) بالا است. اما تفاوت میان همخوانی کامل و همخوانی نسبی در پایگاه اطلاعاتی ایسکو بسیار بیشتر از این تفاوت در پایگاه اطلاعاتی آی‌تریپل‌ای است بدین معنا که در پایگاه اطلاعاتی آی‌تریپل‌ای میزان همخوانی کامل و نسبی بین کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی آزاد در یک حدود است اما در پایگاه اطلاعاتی ایسکو میزان همخوانی

کامل دو برابر همخوانی نسبی است. لذا می‌توان نتیجه گرفت که زبان جستجوی کاربران به اصطلاحات پیشنهادی آزاد در پایگاه اطلاعاتی اِسکو که همان کلیدواژه‌های نویسندگان است، نزدیک‌تر از اصطلاحات پیشنهادی آزاد پایگاه اطلاعاتی آی‌تریپل‌ای است.

در هر دو پایگاه اطلاعاتی مورد نظر، عدم همخوانی کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی کنترل‌شده بیشتر از اصطلاحات پیشنهادی آزاد است. به عبارت دیگر، کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی آزاد و طبیعی بیشتر همخوانی دارند و بنابراین می‌توان نتیجه گرفت که زبان جستجوی کاربران بیشتر از نوع زبان طبیعی و آزاد است و چون اصطلاحات پیشنهادی آزاد و طبیعی، اصطلاحاتی است که عمدتاً از سوی نویسندگان تعیین و انتخاب شده‌اند، بنابراین می‌توان چنین استنباط کرد که زبان جستجوی کاربران در این دو پایگاه اطلاعاتی، با زبان نویسندگان بیشتر همخوانی دارد.

فرضیه یک. میزان همخوانی کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی نویسندگان در پایگاه عمومی اِسکو و پایگاه تخصصی آی‌تریپل‌ای با یکدیگر متفاوت است.

اصطلاحات پیشنهادی نویسندگان، تنها نوع اصطلاحات پیشنهادی هستند که در هر دو پایگاه اطلاعاتی اِسکو و آی‌تریپل‌ای وجود دارد. میزان انواع همخوانی میان آنها با کلیدواژه‌های جستجوی کاربران، در جدول ۳ نمایش داده شده است:

جدول ۳. اصطلاحات نویسندگان در پایگاه اطلاعاتی اِسکو و آی‌تریپل‌ای

اِسکو		آی‌تریپل‌ای		پایگاه نوع همخوانی
فراوانی	درصد فراوانی	فراوانی	درصد فراوانی	
۳۴۷	۶۰٫۱	۱۳۳	۴۱	همخوانی کامل
۱۸۶	۳۲٫۳	۱۵۶	۴۸٫۱	همخوانی نسبی
۴۴	۷٫۶	۳۵	۱۰٫۸	عدم همخوانی
۵۷۷	۱۰۰	۳۲۴	۱۰۰	مجموع

با توجه به محتوای جدول ۳، میزان همخوانی کامل در پایگاه اِسکو ۶۰٫۱ درصد و در پایگاه آی‌تریپل‌ای ۴۱ درصد است اما میزان همخوانی نسبی در پایگاه اطلاعاتی آی‌تریپل‌ای از

پایگاه ابسکو ۱۵/۹ درصد بیشتر است. برای تعیین معنی‌داری این تفاوت، از آزمون یو مان‌ویتنی استفاده شد که داده‌های آن در جدول ۴ ارائه شده است. متغیر «نوع همخوانی» یک متغیر ترتیبی دارای سه مقدار است. از آنجا که مقادیر این متغیر ترتیبی در دو جامعه مستقل (دو پایگاه) با یکدیگر مقایسه شد، آزمون یو مان‌ویتنی به کار گرفته شده است. در این آزمون برای محاسبه آماره آزمون و بررسی معنی‌داری آن از میانگین و مجموع رتبه‌ها در دو گروه مورد بررسی استفاده می‌شود.

جدول ۴. آزمون یو مان‌ویتنی برای بررسی معنی‌داری تفاوت فراوانی کلیدواژه‌های کاربران با اصطلاحات پیشنهادی نویسندگان در دو پایگاه

P _{value}	Z	یو مان‌ویتنی	مجموع رتبه‌ها (Sum of Ranks)	میانگین رتبه (Mean Rank)	تعداد	آماره‌ها
						پایگاه‌ها
-	-	-	۲۴۲۵۶۰,۵۰	۴۲۰,۳۸	۵۷۷	ابسکو
۰,۰۰	-۵,۲۹۱	۷۵۸۰۷,۵۰۰	۱۶۳۷۹۰,۵۰	۵۰۵,۵۳	۳۲۴	آی‌تریپل‌ای
					۹۰۱	مجموع

فرضیه ۲. بین میزان همخوانی کلیدواژه‌های جستجوی کاربران با هر یک از انواع اصطلاحات پیشنهادی (اصطلاحات موضوعی، اصطلاحات نویسندگان، اصطلاحات کنترل‌شده و کنترل‌نشده) در هر پایگاه اطلاعاتی تفاوت معناداری وجود دارد.

برای تعیین وجود تفاوت معنادار بین همخوانی کلیدواژه‌های جستجوی کاربران با انواع اصطلاحات پیشنهادی در هر پایگاه اطلاعاتی، لازم است تا میانگین رتبه‌های هر یک از انواع اصطلاحات پیشنهادی در هر پایگاه با یکدیگر مقایسه شود. به دلیل اینکه متغیر مورد بررسی ترتیبی بود، از آزمون ناپارامتری استفاده شد و چون اصطلاحات کاربران با اصطلاحات خود پایگاه مقایسه می‌شد، در نتیجه با یک جامعه سروکار داشتیم که دوبار از آن اندازه‌گیری به عمل آمده بود و متغیرها از نظر آماری به هم وابسته بودند. در نتیجه آزمون رتبه‌های علامت‌دار ویلکاکسون استفاده شد.

همانگونه که در جدول (۵) مشاهده می‌شود، از ۵۷۷ واژه مورد بررسی، تنها در چهار مورد میزان همخوانی با اصطلاحات موضوعی بیشتر از همخوانی با کلیدواژه‌های نویسندگان است و در مقابل در ۲۰۸ مورد میزان همخوانی با اصطلاحات موضوعی کمتر از همخوانی با کلیدواژه‌های نویسندگان است و در ۳۶۵ مورد میزان همخوانی با اصطلاحات موضوعی و کلیدواژه‌های نویسندگان با یکدیگر برابر است، که این اختلاف در سطح ۰/۰۵ معنی‌دار تشخیص داده شده است، بدین معنا که در پایگاه اِسکو میزان همخوانی کامل و نسبی کلیدواژه‌های جستجوی کاربران با اصطلاحات موضوعی کمتر از همخوانی با کلیدواژه‌های نویسندگان است.

جدول ۵. آزمون رتبه‌های علامت‌دار ویلکاکسون برای سنجش تفاوت میان میزان همخوانی کلیدواژه‌های جستجوی کاربران با کلیدواژه‌های موضوعی و نویسندگان در پایگاه اِسکو

P _{value}	Z	آماره‌ها			
		مجموع رتبه‌ها	میانگین رتبه‌ها	تعداد	مقیاس‌ها
۰/۰۰	-۱۲/۹۶۲	۲۲۲۶۴/۰۰	۱۰۷/۰۴	a۲۰۸	رتبه‌های منفی
		۳۱۴/۰۰	۷۸/۵۰	b۴	رتبه‌های مثبت
				c۳۶۵	گره‌ها
				۵۷۷	مجموع

برای تعیین معنی‌داری تفاوت میان میزان همخوانی اصطلاحات آی‌تریپل‌ای، اصطلاحات کنترل‌شده اینسپیک و اصطلاحات کنترل‌نشده اینسپیک و کلیدواژه‌های نویسندگان با عبارت‌های جستجوی کاربران در پایگاه آی‌تریپل‌ای، از آزمون فریدمن استفاده شد و تفاوت در سطح ۰/۰۵ معنی‌دار تشخیص داده شد. آزمون فریدمن برای مقایسه یک متغیر ترتیبی در چند گروه وابسته استفاده می‌شود و یا زمانی که یک متغیر کیفی چندبار اندازه‌گیری شده و هدف مقایسه اختلاف میان اندازه‌گیری‌ها است. بر این اساس میزان همخوانی کلیدواژه‌های جستجوی کاربران با اصطلاحات پیشنهادی از نوع کنترل‌نشده اینسپیک از همه بیشتر بود. میزان همخوانی با کلیدواژه‌های نویسندگان در رتبه بعدی قرار گرفت و نهایتاً همخوانی کلیدواژه‌های جستجوی

کاربران با اصطلاحات پیشنهادی کنترل‌شده اینسپک و اصطلاحات آی تریپل ای به ترتیب در رتبه‌های بعدی قرار گرفتند.

جدول ۶. رتبه‌های آزمون فریدمن برای سنجش تفاوت میان میزان همخوانی کلیدواژه‌های جستجوی کاربران با اصطلاحات کنترل‌شده، کنترل‌نشده و نویسندگان در پایگاه آی تریپل ای

P value	df	کای اسکور	میانگین رتبه	آماره‌ها
				اصطلاحات
۰,۰۰	۳	۴۲۰,۷۶۸	۳,۰۲	اصطلاحات آی تریپل ای
			۲,۹۹	اصطلاحات کنترل‌شده اینسپک
			۱,۸۴	اصطلاحات کنترل‌نشده اینسپک
			۲,۱۶	کلیدواژه‌های نویسنده

نتیجه‌گیری

نتایج این پژوهش، به طور کلی حکایت از نزدیکی عبارات‌های جستجوی کاربران به اصطلاحات پیشنهادی آزاد و طبیعی، نسبت به اصطلاحات پیشنهادی کنترل‌شده دارد. این امر به لزوم بازنگری بیشتری در سیاست‌های نمایه‌سازی پایگاه‌های اطلاعاتی می‌انجامد. لذا پیشنهاد می‌گردد که از این نوع اصطلاحات پیشنهادی برای ایجاد یک اصطلاحنامه کاربرمدار در پایگاه اطلاعاتی استفاده شود. از سوی دیگر، اصطلاحات پیشنهادی کنترل‌شده که در واقع شکل‌های خاص، مرتبط و یا ساختاریافته کلیدواژه‌های جستجوی کاربران همانند مترادف‌ها، متضادها، اسامی عام، خاص و مرتبط هستند، ممکن است در بسط جستجو توسط کاربر به کار برده شوند. این دسته از اصطلاحات به‌عنوان منبعی ارزشمند برای شکل‌دهی درست و منطقی عبارات‌های جستجو در ذهن کاربر کمک خواهد کرد البته در صورتی که به تعداد کافی و متناسب در شکل‌های مختلف همراه موضوعات اصلی و اطلاعات کتابشناختی آثار بیایند. بررسی و اطمینان از وجود چنین تناسبی در میان خود اصطلاحات پیشنهادی در یک مقاله از حیثه فعالیت این پژوهش خارج شده است.

در این پژوهش مشخص شد که زبان جستجوی کاربران به اصطلاحات پیشنهادی کنترل‌شده در پایگاه‌های اطلاعاتی عمومی نسبت به پایگاه‌های اطلاعاتی تخصصی نزدیک‌تر

است. شاید بتوان دلیل این امر را تفاوت در نوع ابزار به کار گرفته شده در استخراج اصطلاحات پیشنهادی (تفاوت سرعنوان‌های موضوعی و اصطلاحنامه) دانست. لذا احتمالاً سرعنوان‌های موضوعی نسبت به اصطلاحنامه‌ها، جستجو مدارتر و به زبان کاربران نزدیک‌تر هستند. با توجه به همخوانی نسبی و کامل زیاد بین عبارت‌های جستجوی کاربران و اصطلاحات پیشنهادی کنترل‌شده پایگاه ابسکو که برگرفته از سرعنوان‌های موضوعی کتابخانه کنگره است و نتیجه پژوهش نیک‌زمان (۱۳۸۹) که نشان داد، بیشترین نوع همخوانی عبارت‌های جستجوی کاربران با مداخل سرعنوان‌های موضوعی فارسی بیشتر از نوع عدم همخوانی است، جستجو مدارتر بودن سرعنوان‌های موضوعی کنگره نسبت به سرعنوان‌های موضوعی فارسی نتیجه‌گیری می‌شود. البته شریف (۱۳۹۱) در پژوهش خود نشان داد سرعنوان‌های موضوعی فارسی با ۸۶/۱۷ درصد از زبان طبیعی پیروی می‌کنند و با توجه به اینکه بیشتر کاربران از زبان طبیعی برای جستجو استفاده می‌کنند، لذا شاید انتظار رود که این سرعنوان موضوعی به زبان جستجوی کاربران نزدیک باشد. البته ممکن است این نتیجه‌گیری تحت تأثیر نوع زبان جستجوی کاربران این پژوهش که زبان دوم آن‌ها بوده است، متغیر باشد. به هر حال به نظر می‌رسد پایگاه‌های اطلاعاتی قدرتمند همچون پایگاه آی‌تریپل‌ای، با توجه به درک مناسبی از انواع این کاستی‌ها (جستجو مداری کمتر اصطلاحنامه نسبت به سرعنوان موضوعی)، طیف متنوعی از انواع اصطلاحات پیشنهادی را در کنار یکدیگر به کار برده‌اند.

در نگاهی دیگر، با توجه به نتایج پژوهش گیل-لیوا و آلونسو-آرویو^۱ (۲۰۰۷) که نشان دادند ۲۵ درصد از کلیدواژه‌های نویسندگان همخوانی کامل با توصیفگرهای موضوعی و کنترل‌شده دارند، نتایج این پژوهش نشان داد، همخوانی کلیدواژه‌های جستجوی کاربران با اصطلاحات نویسندگان از جایگاه مناسب و ویژه‌ای برخوردار است و به زبان جستجوی کاربران بسیار نزدیک است و لذا این اصطلاحات در نمایه‌سازی متون، در انتخاب اصطلاحات اصطلاحنامه‌ها و پذیرفته شدن این اصطلاحات باید در اولویت قرار بگیرد. ضمن اینکه در نظام‌های توصیه‌گر، این اصطلاحات می‌توانند کلیدواژه‌های مناسبی برای توصیه به کاربران در بسط جستجو باشند.

1. Gil-Leiva & Alonso-Arroyo

پیشنهادها

به کارگیری انواع مختلف اصطلاحات پیشنهادی از جمله اصطلاحات کنترل‌شده و آزاد و یا اصطلاحات نویسندگان برای هر مدرک در پایگاه‌های اطلاعاتی لاتین، این اصطلاحات پیشنهادی را به یک ابزار قدرتمند برای انتخاب کلیدواژه جستجو، اصلاح فرایند جستجو و یا انتخاب اصطلاحات نمایه‌ای برای اصطلاحنامه‌های کاربرمدار تبدیل ساخته است. بدین منظور توصیه می‌شود در پایگاه‌های اطلاعاتی فارسی و از جمله پایگاه «سامان» متعلق به نهاد کتابخانه‌های عمومی کشور که به طیف گسترده‌ای از کاربران خدمت‌رسانی می‌کند، به در دسترس قرار دادن طیف متنوعی از اصطلاحات پیشنهادی به‌ویژه اصطلاحات پیشنهادی نویسندگان به‌منظور انجام جستجوی بهینه توجه شود. همچنین با توجه به این یافته که زبان جستجوی کاربران بیشتر از نوع طبیعی و آزاد است، بنابراین ایجاد اصطلاحنامه کاربرمدار یا جستجو در این پایگاه جهت هدایت آنان در انتخاب کلیدواژه‌های جستجو می‌تواند از دیگر اقدامات برای کمک به کاربران در بازیابی مدارک مرتبط باشد. توصیه می‌شود کتابداران در حین آموزش پایگاه‌های اطلاعاتی بر کاربری‌های این ابزار کمک کار جستجو تأکید بیشتری داشته باشند و از سوی دیگر پایگاه‌های اطلاعاتی با به کارگیری ترفندهایی نظیر استفاده از جلوه‌های گرافیکی، شخصی‌سازی، پیشنهاد استفاده به کاربران و غیره آن‌ها را به استفاده از این ابزار ترغیب و تشویق نمایند. همچنین پایگاه‌های اطلاعاتی فارسی، همانند پایگاه‌های اطلاعاتی لاتین به انتخاب، گردآوری و تدوین اصولی این اصطلاحات بر اساس سیاست‌های مناسب نمایه‌سازی مبادرت ورزند و با توجه به ارزش هر یک از این نوع اصطلاحات، به ارائه متناسبی از تعداد و انواع آنها به همراه اطلاعات کتابشناختی مقالات پردازند.

منابع

- ارسطویور، شعله (۱۳۸۵). بررسی میزان همخوانی سرعنوان‌های موضوعی فارسی با کلیدواژه‌های عنوان و فهرست مندرجات کتاب‌های فارسی (حوزه‌های علوم انسانی، علوم اجتماعی، علوم کاربردی و علوم محض). پایان‌نامه کارشناسی ارشد، دانشگاه فردوسی مشهد، مشهد.
- انصاری، مریم (۱۳۷۷). بررسی انطباق میان توصیفگرهای نمایه‌سازی و کلیدواژه‌های عنوان پایان‌نامه‌های دکترای تخصصی کودکان، زنان، قلب و عروق و روانشناسی. پایان‌نامه کارشناسی ارشد، دانشگاه علوم پزشکی و خدمات بهداشتی و درمانی ایران، تهران.

ایرانشاهی، محمد (۱۳۸۴). بررسی میزان همخوانی کلیدواژه‌های عنوان و توصیف‌گرهای نمایه‌سازان در پایگاه چکیده پایان‌نامه‌های ایران. پایان‌نامه کارشناسی ارشد، دانشگاه فردوسی مشهد، مشهد.

شریف، عاطفه (۱۳۹۱). شناسایی و تحلیل ابعاد ساختاری و مفهومی فهرست سرعنوان‌های موضوعی فارسی در مقایسه با فهرست سرعنوان‌های موضوعی کنگره. پایان‌نامه دکترا، دانشگاه فردوسی مشهد، مشهد.

فضیلت، مریم (۱۳۸۸). بررسی میزان همپوشانی و رابطه میان کلیدواژه‌های عنوان و موضوع مقالات پایگاه نمایه شماره ۸۴ (مهر ۸۵). اطلاع‌یابی و اطلاع‌رسانی، شماره ۱۸. بازیابی شده در تاریخ ۱۵ اسفند ۱۳۹۰ از: <http://ensani.ir/fa/content/213824/default.aspx>

نوکاریزی، محسن و دهقانی، کلثوم (۱۳۹۰). بررسی میزان همخوانی کلیدواژه‌های چکیده با توصیف‌گرهای نمایه‌سازان در پایگاه «چکیده پایان‌نامه‌های ایران». تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی، ۱۷ (۳)، ۱۵۳-۱۷۲.

References

- Dutta, B.; Majumder, K. & Sen, B. K. (2008). Classification of Keywords extracted from Research Articles Published in Science Journals. *Annals of library and information science*, 55(4), 317-333.
- Gil-Leiva, I. & Alonso-Arroyo, A. (2007). Keywords Given by Authors of Scientific Articles in Database Descriptors. *Journal of the American Society for Information Science and Technology*, 58 (8), 1175-1187.
- Natsuo, O. (1998). The 40th Anniversary Issue; For the Multidimensional Development of Information: Keyword Analysis of the Articles in JOHO KANRI in the Recent Decade. *Journal of Information Processing and Management*, 40 (10), 877-896.
- Strader, C. R. (2009). Author-Assigned Keywords Versus Library of Congress Subject Headings: Implications for the Cataloging of Electronic Theses and Dissertations. *Library Resources & Technical Services*, 53 (4), 243-250.
- Voorbij, H. J. (1998). Title Keywords and Subject Descriptors: A Comparison of Subject Search Entries of Books in the Humanities and Social Sciences. *Journal of Documentation*, 54 (4), 466-476.

به این مقاله این‌گونه استناد کنید:

صنعت‌جو، اعظم و رجیبیان، حمیده (۱۳۹۶). بررسی میزان همخوانی عبارت‌های جستجوی کاربران با اصطلاحات پیشنهادی مقالات در پیشینه‌های کتابشناختی پایگاه‌های اطلاعاتی غیرفارسی EBSCO و IEEE. تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی، ۲۳ (۲)، ۲۸۱-۲۹۴.

Correspondence between Users' Search Terms with the Proposed Terms of EBSCO and IEEE Databases

Azam Sanatjoo (corresponding author)
Assistant Professor, Dept. of KIS, Ferdowsi University of Mashhad, Iran
sanatjoo@um.ac.ir

Hamideh Rajabian
Master of KIS, Ferdowsi University of Mashhad, Iran
rajabeian@gmail.com

Received: 24th April 2013; Accepted: 25th October 2015

Abstract

Purpose: This study aims to investigate correspondence of users' queries with alternative terms of Latin databases namely IEEE and EBSCO. Databases display subjective content of their documents through natural or controlled language vocabularies in specified bibliographic fields along with other bibliographic information that are called papers alternative terms.

Methodology: We used content analysis method for implementation of this research. The statistical community includes EBSCO and IEEE databases in which the transaction logs of users have been investigated during 3 months.

Findings: The findings indicated that users' queries are correspond with free alternative terms rather than controlled alternative ones. However, the relative correspondence between users' queries and controlled alternative terms is more than free alternative ones. Also there was a meaningful difference between users queries correspondence with each kinds of alternative terms. Finally, results indicated that free alternative terms seems can be used in indexing and search engine thesauruses.

Originality/Value: This research seems valuable because of its concentration on the comparison of efficiency of free vocabularies vs. controlled vocabularies in search process and the need to more regard toward them in indexing processes.

Keywords: Alternative Terms, Indexing, Free vocabulary, Controlled Vocabulary.