

تحلیل محتوای اسناد مقالات پزشکی دسترسی آزاد و غیر آزاد به روش عقیده‌کاوی بافتار اسناد

خدیجه جوکار^۱، مریم یقطین^۲، هاجر ستوده^۳، مهدیه میرزاییگی^۳

مقاله پژوهشی

چکیده

مقدمه: اعتبارسنجی مقالات دسترسی آزاد، همواره از جمله دغدغه‌های جامعه علمی بوده است. از آن‌جا که تحلیل استنادی کمی در ارزیابی مقالات علمی با چالش روبه‌رو است، تحلیل استنادی محتوامحور از جمله عقیده‌کاوی بافتار اسناد، می‌تواند تصویر روشن‌تری از اعتبار آن‌ها ارائه دهد. هدف از انجام پژوهش حاضر، بررسی عقاید اسنادگران درباره مقالات پزشکی دسترسی آزاد و مقایسه آن‌ها با مقالات دسترسی غیر آزاد مشابه بود.

روش بررسی: این مطالعه به روش تحلیل محتوای کمی با رویکرد تحلیل استنادی عقیده‌کاوانه انجام شد. اطلاعات مقالات از پایگاه‌های PubMed، بافتارهای اسناد از پایگاه Colil و نمره عقیده‌های بافتارهای اسناد از SentiWords استخراج گردید. پس از پردازش زبان طبیعی متن چکیده‌ها، عناوین و بافتارهای اسناد، شباهت کسینوسی مقادیر TF-IDF (Term Frequency-Inverse Document Frequency) محاسبه و مقالات دسترسی آزاد و غیر آزاد مشابه زوج شدند. داده‌ها با استفاده از آزمون Friedman و همبستگی مورد تجزیه و تحلیل قرار گرفت.

یافته‌ها: اختلاف معنی‌داری در نمره میانگین عقیده مقالات دسترسی آزاد با مقالات دسترسی غیر آزاد مشابه وجود نداشت؛ اگرچه اختلاف استنادات آن‌ها معنی‌دار بود. همچنین، همبستگی معنی‌داری بین فاصله عقیده‌های مقالات دسترسی آزاد و غیر آزاد با شباهت متنی و بافتار اسناد آن‌ها مشاهده نشد.

نتیجه‌گیری: اگرچه مقالات دسترسی آزاد بیشتر از مقالات دسترسی غیر آزاد، اسناد دریافت کرده‌اند، اما عقاید اسنادگران درباره این دو گروه مقاله تفاوت معنی‌داری ندارد. همچنین، آثاری که به لحاظ موضوعی با هم شباهت دارند، لزوماً وضعیت عقیده‌های مشابهی را نشان نمی‌دهند. از این‌رو، لزوم تحلیل‌های محتوامحور برای تکمیل نتایج تحلیل استنادی کمی تأیید می‌شود.

واژه‌های کلیدی: بافتار اسناد؛ تحلیل استنادی؛ دسترسی آزاد به اطلاعات؛ عقیده‌کاوی

پیام کلیدی: به کارگیری عقیده‌کاوی بافتار اسناد به عنوان یک روش تحلیل استنادی محتوامحور در کنار تحلیل استنادی کمی می‌تواند دقت بیشتری را در ارزیابی منابع دسترسی آزاد به همراه داشته باشد.

دریافت مقاله: ۱۳۹۹/۵/۱۲

پذیرش مقاله: ۱۳۹۹/۸/۱۰

تاریخ انتشار: ۱۳۹۹/۹/۱۵

ارجاع: جوکار خدیجه، یقطین مریم، ستوده هاجر، میرزاییگی مهدیه. تحلیل محتوای اسناد مقالات پزشکی دسترسی آزاد و غیر آزاد به روش عقیده‌کاوی بافتار اسناد. مدیریت اطلاعات سلامت ۱۳۹۹؛ ۱۷ (۵): ۲۴۴-۲۵۰

مقدمه

تأثیر دسترسی آزاد بر کیفیت آثار علمی، از بحث‌های مهم پیرامون این الگوی انتشاراتی می‌باشد (۱). اعتبار مدل‌های دسترسی آزاد، از آغاز دستخوش چالش‌هایی بوده است. در الگوی دسترسی آزاد سبز (Green Model) که در آن نویسندگان مقاله خود را در وبسایت یا سپارهای (Repositories) سازمانی یا موضوعی آرشیو (Self-Archive) می‌کند، همه نسخه‌های یک مقاله (مانند نسخه‌های پیش‌چاپ یا داوری شده نهایی) اعتبار یکسانی ندارند. الگوی طلایی (Gold Model) نیز که ناظر بر انتشار مجلات دسترسی آزاد می‌باشد، از آغاز به گمان نبود داوری بی‌اعتبار دانسته می‌شد. به مرور و در پرتو روشنگری‌های طرفداران دسترسی آزاد، روشن شد که در مجلات دسترسی آزاد طلایی نیز فرایند داوری برای کنترل کیفیت محتوا اعمال می‌شود (۲) و تنها تفاوت، در تغییر محمل از چاپی به الکترونیکی است.

با این وجود، برخی بدرقتاری‌های علمی در انتشار مجلات دسترسی آزاد، بار دیگر به بحث اعتبار آن‌ها دامن زد. به ویژه از هنگامی که برخی ناشران سودجو (Predatory) (۳)، به انتشار مجلات طلایی بی‌کیفیت و دریافت وجوه هنگفت از

نویسندگان روی آورده‌اند، جامعه علمی بار دیگر اعتبار مجلات طلایی را به پرسش گرفته است. با وجود برخی نگرش‌ها و برداشت‌های مثبت (۴، ۵)، برخی شواهد پژوهشی نشان می‌دهد که محققان همه منابع دسترسی آزاد را معتبر نمی‌دانند یا اعتبار آن‌ها را به دیده تردید می‌نگرند (۶-۸).

مقاله حاصل پایان‌نامه کارشناسی ارشد می‌باشد که با حمایت دانشگاه شیراز انجام شده است.

- ۱- دانشجوی کارشناسی ارشد، علم اطلاعات و دانش‌شناسی، گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه شیراز، شیراز، ایران
- ۲- دانشجوی دکتری تخصصی، علم اطلاعات و دانش‌شناسی، گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه شیراز، شیراز، ایران
- ۳- دانشیار، علم اطلاعات و دانش‌شناسی، گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه شیراز، شیراز، ایران

نویسنده طرف مکاتبه: هاجر ستوده؛ دانشیار، علم اطلاعات و دانش‌شناسی، گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه شیراز، شیراز، ایران
Email: sotudeh@shirazu.ac.ir

روش بررسی

این پژوهش حاضر به روش تحلیل محتوای کمی با رویکردهای تحلیل استنادی کمی و محتوامحور انجام شد. تحلیل محتوا به روش پردازش زبان طبیعی و در دو سطح عینی و ذهنی صورت گرفت. در سطح عینی، واژه‌های بافتار استناد، عنوان و چکیده مقالات بدون توجه به بار عقیده‌ای هر واژه تحلیل گردید. در سطح ذهنی، از تکنیک عقیده‌کاوی برای تعیین نمره عقیده هر واژه در بافتار استناد استفاده شد.

مقایسه مقالات دسترسی آزاد و غیر آزاد، نیازمند مجموعه‌ای از مجلات ترکیبی حاوی هر دو گروه مقاله بود. بدین منظور، نمونه هدفمندی از مقالات در ۱۸ عنوان مجله دسترسی آزاد ترکیبی Elsevier در حوزه پزشکی طی سال‌های ۲۰۱۳ تا ۲۰۱۵ که توسط عرب‌زاده (۲۵) شناسایی شده بودند، انتخاب گردید. داده‌های کتاب‌شناختی و چکیده‌ها از پایگاه PubMed و بافتارهای استناد از پایگاه Colil استخراج گردید.

نظر به پراکندگی موضوعی، نمونه اولیه به گونه‌ای پالایش شد که به مجموعه‌ای از مقالات با موضوعات مشابه و مهم (به لحاظ توزیع کلمات در مدرک و مجموعه) محدود شود. چگونگی اجرای مطالعه و چگونگی هدفمندسازی نمونه در ادامه آمده است.

به منظور سنجش شباهت متنی مقالات دسترسی آزاد و غیر آزاد، ابتدا با تحلیل مدارک در نرم‌افزار KNIME، مقالاتی با حداقل یک کلمه مهم (Salient Word) از میان ۱۰ هزار کلمه دارای بالاترین مقدار TF-IDF (Term Frequency-Inverse Document Frequency) شناسایی و به نمونه وارد شدند. KNIME یک پلتفرم داده‌کاوی متن‌باز است که می‌تواند مراحل داده‌کاوی، استخراج، انتقال و بارگذاری داده‌ها را انجام دهد. سپس با تحلیل شباهت، شبیه‌ترین مدارک انتخاب گردید. بدین منظور، نخست با ساخت یک Workflow متشکل از Module‌های مناسب در KNIME، پیش‌پردازش زبان‌شناختی بر عناوین و چکیده‌ها به عنوان بازنمون‌های اصلی و اثربخش مقالات (۲۶، ۲۷) انجام گرفت. این فرایند شامل حذف نشانه‌های سجاوندی و ناواژه‌ها (Stop Words)، تبدیل حروف بزرگ به کوچک و ریشه‌یابی با الگوریتم Snowball بود. سپس با محاسبه مقادیر TF-IDF، اهمیت واژه‌ها بر اساس فراوانی هر واژه در مدرک (TF) و معکوس فراوانی آن در کل مدارک (IDF) تعیین شد. در محاسبه TF، از فراوانی نسبی واژه (Relative TF) یعنی نسبت میان TF بر تعداد کل کلمات مدرک استفاده گردید تا تأثیر بلندی مدرک خنثی شود. در ادامه به کمک سنجش شباهت کسینوسی، شباهت مدارک دسترسی آزاد و غیر آزاد محاسبه شد. به کمک تکنیک k-NN (k-Nearest Neighbors) (k = ۱۰)، نزدیک‌ترین همسایه‌ها برای هر مقاله دسترسی آزاد شناسایی گردید. در این مرحله، ۵۲۴ مقاله شامل ۲۷۹ مقاله دسترسی آزاد و ۲۴۵ مقاله دسترسی غیر آزاد با حداقل ۵۰ درصد شباهت انتخاب شد.

به منظور استخراج و سنجش شباهت بافتارهای استناد، ۳۶۳۹ بافتار استناد از پایگاه Colil استخراج گردید. این پایگاه با فراهم‌آوری امکان جستجو با PMID (PubMed identifier) یا کلید واژه، دسترسی به استنادها، بافتارهای استناد و محل رخداد استناد مقالات زیرمجموعه دسترسی آزاد PMC-OAS (PubMed Central Open Access Subset) را فراهم می‌کند. با حذف ۵۲ مقاله بی‌استناد، اندازه نمونه به ۴۷۲ مقاله کاهش یافت. بافتارهای استناد نیز به کمک نرم‌افزار KNIME مانند بازنمون‌های مقالات تحلیل و شباهت آن‌ها

برخی تحقیقات نشان دهنده برتری استنادی مجلات و مقالات دسترسی آزاد در همه الگوها اعم از سبز، طلایی و ترکیبی (Hybrid Model) است (۹-۱۲). این برتری لزوماً به معنای برتری کیفی آن‌ها نیست؛ چرا که ممکن است دسترسی آزاد با گسترش طیف مخاطبان در سراسر جهان، منجر به نمایانی، استفاده و در نتیجه، استناد بیشتر شود (۱۳). با این حال، احتمال برتری کیفی را نیز نمی‌توان منکر شد؛ چرا که ممکن است نویسندگان، با رویکردی گزیده‌گرایانه و به امید دریافت بازخوردهای زود هنگام، مقالاتی را برای انتشار آزاد برگزینند که اندیشه‌های نوآورانه و محتوایی تازه و باکیفیت داشته باشد (۱۴).

با توجه به این دوگانگی، واکنش جامعه علمی به مقالات دسترسی آزاد پرسش‌برانگیز است؛ از یک سو، برتری استنادی مقالات دسترسی آزاد بیان‌کننده اقبال جامعه علمی است و از سوی دیگر، انتشار آثار بی‌کیفیت زیر چتر دسترسی آزاد برای اهداف سودجویانه، بدبینی تازه‌ای را در جامعه علمی به دنبال داشته است. بنابراین، به طور قطع نمی‌توان برتری استنادی مقالات دسترسی آزاد را نشانگر تأیید اعتبار آن‌ها از سوی استنادگران دانست. از این رو، نیاز به روش‌ها و فونونی دقیق‌تر برای تحلیل آثار دسترسی آزاد احساس می‌شود. پیش‌تر پژوهش‌ها پیرامون اعتبار آثار دسترسی آزاد به طور عمده از تحلیل استنادی کمی و پیمایش بهره گرفته بودند (۹-۱۴). تحلیل استنادی محتوامحور، از جمله تحلیل بافتارهای استناد می‌باشد که می‌تواند به ارزیابی دقیق‌تر کمک کند (۱۵)؛ چرا که بافتار استناد در مقالات استنادگر، جایگاهی است که در آن مقاله استناد شده به بحث کشیده می‌شود، یافته‌های آن تبیین می‌شود و چه بسا عقاید و نظرات نویسنده استنادگر درباره آن طرح می‌گردد.

عقیده‌کاوی (Opinion Mining) یا تحلیل احساسات (Sentiment Analysis)، حوزه پژوهشی نوپایی است که می‌کوشد نظرات مصرف‌کنندگان و کاربران درباره کالا، خدمات، سیاست یا موضوعی خاص (۱۶) را با کدکاو متون ساخت‌نیافته‌ای مانند فرسته‌های (Posts) بلاگ‌ها، یادداشت‌ها، نقد و بررسی‌ها و توییت‌ها دریابد (۱۶-۲۰). عقیده‌کاوی می‌تواند نگرش نویسنده، حالات وی در زمان نوشتن یا احساسی که قصد داشته است به خواننده نشان دهد را ارزیابی نماید (۲۱). عقاید درباره مقالات علمی نیز در رسانه‌های علمی و اجتماعی انعکاس می‌یابد. شماری از مطالعات به سنجش کیفیت استنادها (۲۲)، رابطه عقاید استنادی و ربط مقالات (۲۳)، ارزیابی کمی و کیفی تأثیر پژوهشی (۲۴) و رتبه‌بندی استنادها بر اساس قطبیت عقاید (۱۹) پرداخته‌اند. با این حال، تحقیقی به روش عقیده‌کاوی استنادی درباره مقالات دسترسی آزاد انجام نشده است. از این رو، پژوهش حاضر به کمک عقیده‌کاوی بافتار استناد، به تحلیل و مقایسه محتوای استناد مقالات پزشکی دسترسی آزاد و غیر آزاد مشابه آن‌ها پرداخت. بدین منظور، ضروری است مقالات دسترسی آزاد و غیر آزاد به لحاظ شباهت محتوایی با هم زوج شوند و با تحلیل فاصله عقیده‌ای و فاصله استنادی آن‌ها روشن شود که آیا برتری استنادی مقالات دسترسی آزاد با عقاید مثبت‌تر یا دستکم برابر همراه بوده است. همچنین، تحلیل همبستگی میان فاصله عقیده‌ای مقالات دسترسی آزاد و غیر آزاد با شباهت متنی و شباهت بافتار استناد آن‌ها برای شناخت بهتر الگوی عقیده‌ای آن‌ها انجام شد.

با این که دسترسی آزاد دستمایه مطالعات بسیاری به روش پیمایش و تحلیل استنادی کمی بوده است، تحقیق حاضر برای بررسی اختصاصی دیدگاه استنادگران درباره مقالات دسترسی آزاد به روش عقیده‌کاوی با رویکردی تطبیقی انجام گردید.

محاسبه گردید.

کمک همبستگی Spearman در نرم افزار SPSS نسخه ۲۲ (version 22, IBM Corporation, Armonk, NY) مورد تجزیه و تحلیل قرار گرفت.

یافته‌ها

در هر دو گروه، فراوانی مقالات در قطبیت مثبت بیش از قطبیت منفی بود. اگرچه میانگین نمره عقیده هر دو گروه مثبت شد، اما مقدار عددی آن برای مقالات دسترسی آزاد بالاتر به دست آمد (۵/۱۱ در برابر ۳/۰۲). میانگین استاندارد گروه دسترسی آزاد (۱۱/۸۳) نیز بالاتر از گروه مقابل (۶/۷۵) گزارش گردید (جدول ۱). نتایج آزمون Friedman نشان داد که اختلاف معنی‌داری بین میانگین رتبه دو گروه مقالات دسترسی آزاد و غیر آزاد در نمره عقیده وجود نداشت؛ در حالی که میانگین رتبه مقالات دسترسی آزاد به لحاظ استانداردها بالاتر از گروه مقابل بود (جدول ۲).

جدول ۲: مقایسه مقالات دسترسی آزاد و غیر آزاد به لحاظ

نمره عقیده و استاندارد

شاخص	گروه	میانگین رتبه	مقدار χ^2	مقدار P
نمره عقیده	آزاد	۱/۵۱	۰/۰۱۳	۰/۹۰۸
	غیر آزاد	۱/۴۹		
استاندارد	آزاد	۱/۶۴	۵/۸۸	۰/۰۱۵
	غیر آزاد	۱/۳۶		

به منظور بررسی این که نبود تفاوت معنی‌دار میان نمرات عقیده مقالات دسترسی آزاد و غیر آزاد متأثر از میزان شباهت موضوعی آن‌ها است یا خیر، همبستگی میان فاصله عقیده‌ای زوج مقالات با شباهت متنی و شباهت بافتار استاندارد آن‌ها به کمک تحلیل همبستگی Spearman بررسی گردید (جدول ۳).

جدول ۳: همبستگی فاصله عقیده‌ای زوج مقالات دسترسی

آزاد و غیر آزاد با شباهت متنی و بافتار استاندارد آن‌ها

شاخص	ضریب همبستگی	مقدار P
فاصله عقیده‌ای	-۰/۷۷۰	۰/۵۱۳
شباهت بافتار استاندارد	-۰/۱۹۰	۰/۸۷۱

بر اساس داده‌های جدول ۳، فاصله عقیده‌ای مقالات دسترسی آزاد و غیر آزاد با شباهت متنی یا شباهت بافتار استاندارد آن‌ها همبستگی معنی‌داری نداشت. به عبارت دیگر، فاصله عقیده‌ای یا تفاوت عقیده مقالات استاندارد درباره زوج مقالات آزاد- غیر آزاد لزوماً تابعی از کاهش یا افزایش شباهت آن‌ها- چه در سطح چکیده- عنوان و چه در سطح بافتار استاندارد- نبود.

جدول ۱: ویژگی‌های نمونه‌ها به لحاظ مؤلفه‌های مورد بررسی

گروه	فراوانی مقالات در			میانگین نمره عقیده	میانگین استاندارد
	قطبیت				
	کل	مثبت	منفی		
مقالات دسترسی آزاد	۷۵	۶۳	۰	۵/۱۱	۱۱/۸۳
مقالات دسترسی غیر آزاد	۱۲۳	۱۱۱	۰	۳/۰۲	۶/۷۵

در بخش عقیده‌کاوی بافتارهای استاندارد، نمره عقیده از ابزار SentiWords برگرفته از SentiWordNet استخراج شد. این لغتنامه، مجموعه بزرگی از ۱۵۵ هزار واژه را در برمی‌گیرد و هر واژه را بر اساس نقش‌های آن به دسته‌های اسم، صفت، فعل و قید دسته‌بندی می‌کند و نمره عقیده هر یک را به تفکیک ارائه می‌دهد.

پس از استخراج ۱۳۰۰۰۲ جمله موجود در بافتارهای استاندارد، برچسب‌گذاری (PoS Tagging) به کلمات افزوده و بن‌واژه‌ها (Lemmas) به دست آمد. منفی‌سازها (Negation Words) قطبیت عقیده را عوض می‌کنند و باید کنترل و خنثی شوند. فهرست منفی‌سازها شامل «no, not, n't, never, neither, nor, none, nobody, nowhere, nothing, cannot, can not, without, no one, no way» از پژوهش Athar (۲۸) به دست آمد و با حالت کوتاه آن‌ها (Contraction) تکمیل گردید. برای خنثی‌سازی منفی‌سازها، پنجره‌های واژه‌ای از ۲ تا ۱۵+ (۲ تا ۱۵ کلمه بعد از منفی‌ساز) توصیه شده است. با این که تفاوت معنی‌داری میان این پنجره‌ها مشاهده نشد (۲۸)، برای اطمینان، از پنجره بلند ۱۵+ استفاده گردید. فرایند خنثی‌سازی با افزودن NOT_ به انتهای این ۱۵ کلمه و حذف منفی‌ساز اصلی انجام گرفت. خنثی‌سازی با رسیدن به انتهای جمله اصلی (نشانه نقطه) یا انتهای جمله پایه یا پیرو (نشانه کاما یا کامانقطه) متوقف شد.

عقیده‌کاوی به صفت‌ها و قیدها که بیش از دیگر ادات سخن حامل عقیده هستند (۲۹)، محدود گردید. واژه‌هایی که در فرایند برچسب‌گذاری ادات سخن، برچسب صفت یا قید دریافت کرده بودند، در SentiWords جستجو و نمره عقیده‌ای آن‌ها استخراج شد. به منظور محاسبه نمره عقیده‌ای هر بافتار استاندارد، مجموع نمرات عقیده‌ای واژه‌های آن محاسبه گردید. برای مقاله دارای چندین بافتار استاندارد، میانگین نمرات عقیده‌ای بافتارهای استاندارد آن محاسبه شد (۲۳). از ۴۷۲ مقاله دارای بافتار استاندارد، ۱۸۹ عنوان در بافتارهای استاندارد خود هیچ‌گونه صفت یا قیدی نداشتند. همچنین، واژه‌های عقیده‌ای ۱۵ مقاله در SentiWords نبود. بنابراین، اندازه نمونه به ۲۶۸ مدرک (شامل ۱۱۱ مقاله دسترسی آزاد و ۱۵۷ مقاله دسترسی غیر آزاد) رسید که ۴۵۳ زوج دسترسی آزاد- غیر آزاد مشابه را تشکیل دادند. برخی مقالات دسترسی آزاد، با بیش از یک مقاله دسترسی غیر آزاد زوج شدند. بنابراین، شمار زوج‌ها از شمار مطلق مدارک فراتر رفت. از آن‌جا که شباهت محتوایی مقالات در پژوهش حاضر اهمیت کانونی داشت، آستانه شباهت ۵۰ درصد برای ورود به نمونه نهایی انتخاب گردید. به این ترتیب، تعداد زوج‌های قابل بررسی به ۲۳۱ زوج (شامل ۷۵ مقاله دسترسی آزاد و ۱۲۳ مقاله دسترسی غیر آزاد) رسید. در نهایت، فاصله عقیده‌ای زوج مقالات دسترسی آزاد- غیر آزاد به دست آمد. بدین منظور، تقاض نمره عقیده استاندارد درباره مقالات دسترسی آزاد و غیر آزاد که میزان شباهت یا تفاوت عقیده آن‌ها را نشان می‌دهد، محاسبه گردید.

با توجه به نرمال نبودن توزیع داده‌ها، نمره عقیده‌ای مقالات دسترسی آزاد و غیر آزاد با استفاده از آزمون ناپارامتریک Friedman مقایسه گردید. همبستگی بین فاصله عقیده‌ای مقالات با شباهت متنی و شباهت بافتار استاندارد آن‌ها به

روایت یافته‌های عینی جهت‌دار صادق می‌بود، انتظار می‌رفت شباهت موضوعی با فاصله عقیده‌ای مدارک همبسته باشد؛ به گونه‌ای که فاصله عقیده‌ای مقالات به نسبت شباهت موضوعی کم یا زیاد شود؛ در حالی که چنین رابطه‌ای مشاهده نشد. بنابراین، شاید بتوان این یافته را تأییدی بر آن دانست که در مقالات مورد بررسی، نظرات و موضع‌گیری نویسندگان استنادگر درباره مقالات استناد شده، در قطبیت بافتارهای استناد مؤثر بوده است. با این حال، قضاوت قطعی در این باره نیازمند انجام پژوهش‌های بیشتر است.

با این که تمایز نظرات منفی از روایت عینی یافته‌های منفی برای ارزیابی مطالعه ضروری است، عقیده‌کاوی واژه‌نامه‌ای توان جداسازی این دو را ندارد. محدودیت دیگر تحقیق حاضر، تأثیرپذیری استناد از اعتبار مجلات و ناشران بود که احتیاط در تفسیر و تعمیم نتایج به مجلات و ناشرانی با درجه اعتبار متفاوت را می‌طلبد. وابستگی قطبیت و نمره عقیده برخی واژه‌ها به حوزه تخصصی، نیازمند ابزارهای معنایی است که بر خلاف SentiWords و SentiWordNet به حوزه موضوعی وابسته باشند. همچنین، شماری از صفات و قیدها در این ابزارها نبودند. با وجود فراوانی اندک این واژه‌ها و تخصصی بودن آن‌ها که احتمال خنثی بودن آن‌ها را پیش می‌آورد، ابزارهای کامل‌تر می‌تواند دقت بیشتری را به دنبال داشته باشد. محدودیت ابزارهای به کار رفته برای سنجش عقیده و تنها مبتنی بر واژه‌نامه بودن روش عقیده‌کاوی از جمله محدودیت‌های پژوهش حاضر بود.

نتیجه‌گیری

با وجود جایگاه مستحکم تحلیل‌های استنادی در اعتبارسنجی مجلات، از آغاز، ضعف استناد در انعکاس کیفیت و تأثیر واقعی مقالات آشکار بود. در تحلیل استنادی کمی، همه استنادها ارزش برابری دارند و به منزله رأی مثبت به اعتبار مقاله استناد شده تلقی می‌شوند؛ در حالی که نویسندگان لزوماً اعتبار اثر مورد استناد را تأیید نمی‌کنند، بلکه ممکن است آن را به نقد بکشند یا بی‌طرفانه روایت کنند (۱۹).

نتایج مطالعه حاضر با تأیید برتری استنادی مقالات دسترسی آزاد بر مقالات دسترسی غیر آزاد در عین برابری آن‌ها در عقاید دریافتی، نشان داد که تحلیل استنادی کمی و عقیده‌کاوی استناد به نتایج لزوماً یکسانی نمی‌رسند و بار دیگر ضرورت تقویت تحلیل استنادی کمی با تحلیل‌های استنادی محتوامحور را آشکار ساخت.

در میان منابع علمی، اعتبارسنجی منابع دسترسی آزاد اهمیتی دوچندانی دارند؛ چرا که انبوه منابع آزاد در وب، تمایز آثار معتبر از نامعتبر را دشوار می‌کند. دانش‌پژوهان که بازیگران اصلی در استفاده از منابع دسترسی آزاد هستند (۳۲)، هم در مقام خواننده و هم در مقام نویسنده، نیازمند آگاهی از محتوای مقالات و نظرات دیگر نویسندگان درباره خودشان هستند. از این‌رو، در سامانه‌های بازیابی استنادی می‌توان در کنار سنجه‌های استنادی، نتایج عقیده‌کاوی به ویژه در شکل تطبیقی آن که در تحقیق حاضر به کار رفت را ارایه کرد و به این ترتیب، با تصحیح تأثیر سنجه‌های استنادی، جستجوگران را در دستیابی به بینشی برای اعتبارسنجی مقالات بازیابی شده یاری رساند. همچنین، علم‌سنجان، مدیران پژوهشی و کتابداران می‌توانند با تلفیق این دو نوع تحلیل استنادی در ارزیابی پژوهش، کمیت‌گرایی را با افزودن عناصر محتوایی تعدیل نمایند.

بحث

نتایج پژوهش حاضر نشان دهنده برتری استنادی مقالات دسترسی آزاد بر مقالات دسترسی غیر آزاد مشابه بود. این یافته با نتایج مطالعات پیشین (۹-۱۱) همخوانی داشت. تحقیقات مذکور نشان داد که حتی پس از کنترل تأثیر موضوع مقالات که نقش بسزایی در استناد دارد، برتری استنادی مقالات دسترسی آزاد همچنان برقرار است (۹-۱۱). با این حال، این دو گروه به لحاظ نمره عقیده‌ای که از استنادگران دریافت داشته‌اند، فاصله معنی‌داری نداشتند. به عبارت دیگر، نمره عقیده‌ای مقالات دسترسی آزاد به لحاظ آماری با مقالات دسترسی غیر آزاد برابر بوده است. این یافته نشان می‌دهد اگرچه مقالات دسترسی آزاد نسبت به گروه همتای خود استناد بیشتری دریافت کردند، اما عقیده استنادگران درباره این دو گروه مقالات کم و بیش یکسان بود.

بافتارهای استناد مجالی فراهم می‌کند که نویسندگان استنادگر به بحث پیرامون مقاله یا ویژگی‌های آن پردازند. احتمال می‌رود که استنادگران در فرایند استناد، گزینشی آگاهانه و نوعی داوری را روا دارند. بنابراین، بافتار استناد، ظرفیتی را برای نقل عینی یافته‌های مقالات پیشین یا نقد و نظر درباره آن‌ها فراهم می‌کند. به این ترتیب، آنچه به عنوان عقیده در این بافتارها ظاهر می‌شود یا ذهنی است و نشانگر عقیده و یا یافته‌های عینی جهت‌دار را بیان می‌کند. در حالت نخست، روش عقیده‌کاوی مبتنی بر واژه‌نامه به کار رفته در پژوهش حاضر، نمی‌تواند به صراحت سه نوع بافتار عقیده‌ای مثبت، منفی یا خنثی را متمایز کند، اما بافتار عقیده‌ای مثبت- خواه نظر نویسنده استنادگر باشد و خواه روایت عینی وی از یافته‌های مثبت پژوهش‌های پیشین- را می‌توان تأییدی هر چند غیر مستقیم بر یافته‌های پیشین دانست (۳۰). بنابراین، این یافته را می‌توان به سه شیوه تفسیر کرد؛ الف. عقاید استنادگران درباره مقالات دسترسی آزاد و غیر آزاد متفاوت نیست، ب. یافته‌های مطالعات استنادگر به یک اندازه با هر دو گروه مقالات دسترسی آزاد و غیر آزاد همسو است و ج. یافته‌های مقالات دسترسی آزاد و غیر آزاد به یک اندازه همسو است.

از آن‌جا که پیش‌تر تحقیق عقیده‌کاوانه‌ای به مقایسه مقالات دسترسی آزاد و غیر آزاد نپرداخته است، نمی‌توان این یافته را در پرتو شواهد پیشین تبیین کرد، اما می‌توان دریافت که بی‌اعتمادی یا تردیدها نسبت به کیفیت مقالات دسترسی آزاد حداقل در مجلات ترکیبی معتبر به چشم نمی‌خورد یا بیش از مقالات دسترسی غیر آزاد مشابه آن‌ها نیست. این امر می‌تواند به دلیل اعتبار مجلات ترکیبی ناشران بزرگ (۳۱) که پیش‌تر شهرت و جایگاهی مستحکم در مجامع علمی برای خود ساخته‌اند، باشد.

همبستگی معنی‌داری بین فاصله عقیده‌ای مقالات با شباهت متنی و شباهت بافتارهای استناد آن‌ها وجود ندارد. از آن‌جا که شباهت مقالات در متن و در بافتار استناد بیان‌کننده شباهت موضوعی آن‌ها است، این یافته نشان می‌دهد شباهت موضوعی به طور قطع با افزایش یا کاهش شباهت عقیده‌ای همراه نیست. به عبارت دیگر، آثار دارای محتوای مشابه، لزوماً الگوی عقیده‌ای مشابه یا متفاوتی ندارند.

پیش‌تر گفته شد که در عقیده‌کاوی مبتنی بر واژه‌نامه، آنچه تحلیل می‌شود یا ذهنی و ناظر بر آرای نویسندگان استنادگر درباره مقاله است و یا عینی و ناظر بر سمت و سوی یافته‌ها (برای مثال گزارش تأثیر منفی یا مثبت متغیری بر آزمودنی‌ها) می‌باشد. از این‌رو، چنانچه در مقالات مورد بررسی، حالت دوم یعنی

تشکر و قدردانی

بدین وسیله از کلیه افرادی که در انجام این مطالعه همکاری نمودند، تشکر و قدردانی به عمل می‌آید.

تضاد منافع

در انجام پژوهش حاضر، نویسندگان هیچ‌گونه تضاد منافی نداشته‌اند.

پیشنهادها

تعمیم‌پذیری نتایج نیازمند انجام مطالعات بیشتر بر گستره‌ای از ناشران و مجلات با درجات اعتبار متفاوت است. همچنین، محدودیت ابزارهای به کار رفته برای سنجش عقیده، ضرورت تکرار تحقیق با ابزارهایی دقیق‌تر و تخصصی‌تر را برای اطمینان از دقت یافته‌ها پیش می‌آورد. با توجه به محدودیت روش عقیده‌کاوی مبتنی بر واژه‌نامه، تکرار پژوهش با روش‌های یادگیری ماشینی به منظور تمایز دو نوع بافتار استاد حاوی نظرات منفی و روایت عینی یافته‌های منفی ضروری به نظر می‌رسد.

References

- Grozanick SE. The impact factor: Implications of open access on quality. *null* 2010; 6(4): 389-99.
- Beall J. Predatory publishing is just one of the consequences of gold open access. *Learn Publ* 2013; 26(2): 79-84.
- Beall J. Beall's List of Predatory Publishers 2013 [Online]. [cited 2012 Dec 4]; Available from: URL: https://mail.pide.org.pk/pdf/not_recognized.pdf
- Kaba A, Said R. Open access awareness, use, and perception: A case study of AAU faculty members. *New Library World* 2015; 116(1-2): 94-103.
- Togia A, Korobili S. Attitudes towards open access: A meta-synthesis of the empirical literature. *Inf Serv Use* 2014; 34(3-4): 221-31.
- Mischo WH, Schlembach MC. Open access issues and engineering faculty attitudes and practices. *J Libr Adm* 2011; 51(5-6): 432-54.
- Rodriguez JE. Awareness and attitudes about open access publishing: A glance at generational differences. *J Acad Libr* 2014; 40(6): 604-10.
- Peekhaus W, Proferes N. An examination of North American Library and Information Studies faculty perceptions of and experience with open-access scholarly publishing. *Libr Inf Sci Res* 2016; 38(1): 18-29.
- Piwovar HA, Vision TJ. Data reuse and the open data citation advantage. *Peer J* 2013; 1: e175.
- Sotudeh H, Ghasempour Z, Yaghtin M. The citation advantage of author-pays model: The case of Springer and Elsevier OA journals. *Scientometrics* 2015; 104(2): 581-608.
- Xia J, Myers R, Wilhoite S. Multiple open access availability and citation impact. *J Inf Sci* 2011; 37(1): 19-28.
- Wang X, Liu C, Mao W, Fang Z. The open access advantage considering citation, article usage and social media attention. *Scientometrics* 2015; 103(2): 555-64.
- Lewis DW. The inevitability of open access. *Coll Res Libr* 2012; 73(5): 493-506.
- Sotudeh H, Arabzadeh H, Mirzabeigi M. How do self-archiving and Author-pays models associate and contribute to OA citation advantage within hybrid journals. *J Acad Libr* 2019; 45(4): 377-85.
- Hernandez-Alvarez M, Gomez JM. Survey about citation context analysis: Tasks, techniques, and resources. *Nat Lang Eng* 2016; 22(3): 327-49.
- Liu B. Sentiment analysis and opinion mining. *Synth Lect Hum Lang Technol* 2012; 5(1): 1-167.
- Khan K, Baharudin B, Khan A, Malik F. Mining opinion from text documents: A survey. *Digital Ecosystems and Technologies. Proceedings of the 3rd IEEE International Conference on Digital Ecosystems and Technologies, DEST '09; 2009 Jun 1-3; Istanbul, Turkey.*
- Feldman R. Techniques and applications for sentiment analysis. *Commun ACM* 2013; 56(4): 82-9.
- Parthasarathy G, Tomar DC. Sentiment analyzer: Analysis of journal citations from citation databases. 2014; p. 923-8.
- Parthasarathy G, Tomar DC. A Survey of sentiment analysis for journal citation. *Indian J Sci Technol* 2015; 8(35): 1-8.
- Kaur A, Gupta V. A survey on sentiment analysis and opinion mining techniques. *J Emerg Technol Web Intell* 2013; 5(4): 367-71.
- Sendhilkumar S, Elakkiya E, Mahalakshmi GS. Citation semantic based approaches to identify article quality. *Proceedings of the 3rd International Conference on Computer Science, Engineering and Applications, ICCSEA 2013; 2013 May 24-26; New Delhi, India.*
- Cavalcanti DC, Prudencio BC, Pradhan SS, Shah JY, Petrobon RS. Good to be bad? Distinguishing between positive and negative citations in scientific impact. *Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2011; 2011 Nov 7-9; Boca Raton, Florida USA.*
- Abu-Jbara A, Ezra J, Radev D. Purpose and polarity of citation: Towards NLP-based bibliometrics. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2013 Jun 9-14; Atlanta, Georgia, USA.*
- Arabzadeh H. A study of the role of green open-access model in the persistence of author-pays model: the case of Elsevier's hybrid open access journals in 2013-2015. [MSc Thesis]. Shiraz, Iran: Shiraz University; 2018. [In Persian].
- Hartley J, Kostoff RN. How useful are 'Key Words' in scientific journals? *J Inf Sci* 2003; 29(5): 433-8.
- Jenuwine ES, Floyd JA. Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on

- sleep in healthy individuals. *J Med Libr Assoc* 2004; 92(3): 349-53.
28. Athar A. Sentiment analysis of scientific citations [PhD Thesis]. Cambridge, UK: University of Cambridge; 2014.
 29. Esuli A, Sebastiani F. SentiWordNet: A publicly available lexical resource for opinion mining. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06); 2006 May 22-28; Genoa, Italy.
 30. Jia M. Citation function and polarity classification in biomedical papers [MSc Thesis]. London, ON: University of Western Ontario; 2018.
 31. Hines T. Viewpoint: Open is an opportunity and a choice. *J Biocommun* 2016; 40(1): 1-6.
 32. Fuchs C, Sandoval M. The diamond model of open access publishing: Why policy makers, scholars, universities, libraries, labour unions and the publishing world need to take non-commercial, non-profit open access serious. *TripleC* 2013; 11(2): 428-43.

Content-Based Citation Analysis of Open Access and Non-Open Access Medical Articles Using Opinion Mining of Citances

Khadijeh Jokar¹, Maryam Yaghtin², Hajar Sotudeh³, Mahdiah Mirzabeigi³

Original Article

Abstract

Introduction: Scientific communities have always been concerned about validity of open-access articles. Given the challenges of quantitative citation analysis in evaluating scientific articles, content-based citation analysis, including opinion mining of citances, can bring about more transparent results about their validity. In view of this, the present study compared the opinions contained in citances about open-access and non-open-access articles.

Methods: We used a quantitative content analysis method with citation and opinion analysis approaches. The citances, bibliographic, and bibliometric data were extracted from Colil and PubMed databases. Opinion scores were assigned to the citances through SentiWords. After processing the titles, abstracts, and citances, Cosine similarity of Term Frequency-Inverse Document Frequency (TF-IDF) values were calculated. The open-access and non-open-access articles were then paired by their similarities in abstracts, titles, and citances. The data were analyzed using Friedman test and Spearman correlation.

Results: There was no significant difference between the open-access and non-open-access articles in terms of their opinion scores, despite a significant difference in citation advantages. The pairs' citance and textual similarities had no significant correlation with their opinion distance.

Conclusion: Although the open-access studies had citation advantage over their similar non-open-access peers, they showed no significant opinion distance. Besides, similar texts did not necessarily follow the same opinion patterns. Consequently, to complete the results of quantitative citation analysis, the content-based citation analysis is emphasized.

Keywords: Citances; Citation Analysis; Open Access to Information; Opinion Mining

Received: 02 Aug., 2020

Accepted: 31 Oct., 2020

Published: 05 Dec., 2020

Citation: Jokar K, Yaghtin M, Sotudeh H, Mirzabeigi M. **Content-Based Citation Analysis of Open Access and Non-Open Access Medical Articles Using Opinion Mining of Citances.** Health Inf Manage 2020; 17(5): 244-50.

Article resulted from MSc thesis funded by Shiraz University.

1- MSc Student, Knowledge and Information Science, Department of Knowledge and Information Science, School of Education and Psychology, Shiraz University, Shiraz, Iran

2- PhD Student, Knowledge and Information Science, Department of Knowledge and Information Science, School of Education and Psychology, Shiraz University, Shiraz, Iran

3- Associate Professor, Knowledge and Information Science, Department of Knowledge and Information Science, School of Education and Psychology, Shiraz University, Shiraz, Iran

Address for correspondence: Hajar Sotudeh; Associate Professor, Knowledge and Information Science, Department of Knowledge and Information Science, School of Education and Psychology, Shiraz University, Shiraz, Iran; Email: sotudeh@shirazu.ac.ir