Measuring Similarities between Open Peer Review Comments and Contents of Scientific Articles: a Natural Language Processing Technique Inquiry

K. Rashidi Sharifabad[1] | H. Sotodeh[2]
M. Mirzabeigi[3] | S. Mostafa Fakhrahmad[4]

Purpose: The social web provides a platform for publicizing open peer review reports. In this sphere, journal readers, authors, editors, and reviewers can involve in multilateral discussions on the reviewed papers and share their comments and viewpoints on the merits and probable pitfalls of papers. Open peer review comments may, hence, reflect the features of their mother articles. To identify this potential, the present study investigates to what extent similar comments accurately predict similar papers.
Methodology: Applying natural language processing techniques, it analyzes the contents of a sample of papers in medicine and life sciences and the comments received by them. To do so, a test collection is built from the papers openly published on F1000Research, an open access publishing platform that adheres to an open peer reviewing process by transparently providing the public with peer review reports, authors' responses, and users' comments. The test collection consists of 2212 papers and their comments. 100 papers are randomly selected as seed documents that serve as queries. The similarities between the comments and the contents of the papers are calculated using Cosine similarity of TF-IDF values. The TF-IDF values are calculated for both unigrams and bigrams extracted from the contents and comments. The correlation between the content and comment similarities is analyzed using Spearman correlation, given the non-normality of the data distributions. The accuracy of prediction of the papers' content similarity by the similarity of their comments is tested using Receiver Operating Characteristic (ROC) curves.
Findings: The results of the Spearman correlation revealed a significant correlation between the content and comment similarities. This signifies that similar papers are more likely to receive similar comments and vice versa. The ROC curves show that similar comments can significantly identify similar papers, either at unigram or bigram level. The prediction is highly accurate.
Conclusion: Similar comments are effective in representing similar papers. In other words, similar comments are expected to present similar papers. This finding has implications for interactive information retrieval systems, where users are interested in reading experts' comments on a given paper before viewing or downloading the paper itself. The findings also may pave the path towards new studies about the application of the comments in such spheres as information retrieval, evaluation or classification, where content similarity is of importance.

Keywords:
Comments, Peer reviewing, Natural language processing, Similarity, ROC curve analysis

Received: 5, Apr. 2020
Accepted: 27, Apr. 2020

DOI: 10.30484/NASTINFO.2020.2480.1937

1. PhD Candidate, Knowledge and Information Science, Shiraz University, Shiraz, Iran, kia_rsh85@yahoo.com
2. Associate Professor, Knowledge and Information Science, Shiraz University, Shiraz, Iran (Corresponding author), sotudeh@shirazu.ac.ir
3. Associate Professor, Knowledge and Information Science, Shiraz University, Shiraz, Iran, mmirzabeigi@gmail.com
4. Assistant Professor, Computer Science and Engineering and Information Technology, Shiraz University, Shiraz, Iran, fakhrahmad@shirazu.ac.ir

سنجش شباهت نظرات داوری آزاد و محتوای مقالات علمی به‌روش پردازش زبان طبیعی

کیانوش رشیدی شریف‌آباد[۱] | هاجر ستوده[۲]
مهدیه میرزابیگی[۳] | سیدمصطفی فخراحمد[۴]

هدف: شناسایی قابلیت داوری‌های آزاد در بازشناخت مقالات پزشکی براساس شباهت آنها به مقالات مربوط.

روش‌شناسی: آزمونی متشکل از ۲۲۱۲ مقاله اف‌هزار ریسرچ و نظرات داوری آنها ساخته شد. ۱۰۰ مقاله به‌عنوان مدرک پایه به‌صورت تصادفی انتخاب شد. شباهت نظرات داوری و محتواهای مدارک براساس سنجهٔ شباهت کسینوسی مقادیر TF-IDF در سطح تک‌واژه‌ها و دوواژه‌ها محاسبه شد. شباهت محتوا و نظرات با تحلیل همبستگی اسپیرمن تحلیل شد. صحت پیش‌بینی شباهت محتوای مقالات براساس شباهت نظرات دریافت‌شده به‌کمک منحنی مشخصه عملکرد سامانه آزمون شد.

یافته‌ها: توان نظرات داوران در بازشناخت مقالات مشابه تأیید شد. میان محتوا و نظرات، همبستگی معنادار وجود دارد. منحنی‌های تحلیل عملکرد سامانه نیز نشان داد شباهت نظرات داوری، خواه در سطح تک‌واژه‌ها و خواه دوواژه‌ای‌ها توانایی شناسایی مقالات با محتوای مشابه را دارد.

نتیجه‌گیری: اعتبار نظرات داوران ریشه در توان تخصصی و شناختی آنان دارد. بنابراین، نظرات می‌توانند در شبکه مدارک، در زمره منابع مرتبط اثربخش در بازشناخت مدارک به‌شمار آیند. این یافته راه را برای پژوهش در کاربرد نظرات کاربران در حوزه‌های بازیابی، ارزیابی، یا طبقه‌بندی متون هموار می‌کند که شباهت محتوایی در آنها اهمیت دارد.

کلیدواژه‌ها
نظرات کاربران، داوری آزاد، پردازش زبان طبیعی، شباهت، منحنی تحلیل عملکرد سامانه

دریافت: ۹۹/۰۱/۱۶ پذیرش: ۹۹/۰۲/۰۷

۱. دانشجوی دکترای علم اطلاعات و دانش‌شناسی، واحد بین‌الملل دانشگاه شیراز، شیراز، ایران kia_rsh85@yahoo.com
۲. دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه شیراز، شیراز، ایران (نویسنده مسئول) sotudeh@shirazu.ac.ir
۳. دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه شیراز، شیراز، ایران mmirzabeigi@gmail.com
۴. استادیار گروه مهندسی و علوم کامپیوتر و فناوری اطلاعات، دانشگاه شیراز، شیراز، ایران fakhrahmad@shirazu.ac.ir

مطالعات ملی کتابداری و سازماندهی اطلاعات، دوره سی‌ویکم، شماره دوم، تابستان ۱۳۹۹، ص ۸۶-۱۰۳.
DOI: 10.30484/NASTINFO.2020.2480.1937

مقدمه

تحلیل محتوای مقالات نه تنها توسط متخصصان، بلکه توسط ماشین و به‌کمک فنون پیشرفته پردازش زبان طبیعی میسر است. زبان طبیعی، در کنار زبان مهارشده دو رویکرد مهم در بازنمون مدارک به‌شمار می‌آیند (Lewis & Jones, 1996). ازآنجاکه استفاده از تک‌واژه‌ها[1] به‌عنوان واحد بازنمون مدرک می‌تواند به ازدست‌رفتن برخی ویژگی‌های معنایی متن منجر شود (Brennan, 2012; Thomas, Vinod, & Dhanya, 2014)، استفاده از چندواژه‌ای[2]ها، یعنی تک‌واژه‌های پیاپی توصیه می‌شود. هرچه تعداد چندواژه‌ای‌های مشترک بین دو متن بیشتر باشد شباهت آن دو بیشتر است (Golcher & Reznicek, 2011; Constantin, Du Mouza, Faget, & Rigaux, 2011). به‌طور معمول، در مطالعات کنونی از تک‌واژه تا سه واژه استفاده می‌شود (Cui, Mittal, & Datar, 2006; Cao, Li, Liu, Li, & Ji, 2015). از این‌رو، با توجه به آنکه تک‌واژه‌ها به روابط بین واژگان توجهی ندارند در این پژوهش، علاوه بر تک‌واژه‌ها، از دوواژه‌ای‌ها[3] نیز بهره گرفته شد؛ زیرا دوواژه‌ای‌ها به‌طور معمول در پژوهش‌ها به‌خدمت گرفته شده است.

از سوی دیگر، تحلیل محتوا توسط متخصصان به تولید متون جدیدی می‌انجامد. برای نمونه، نویسندگان استنادگر در زمره متخصصانی هستند که تحلیل آنها از محتوای آثار در بافتار استناد، یعنی متن پیرامونی یک استناد در متن استنادگر بازتاب می‌یابد و نقشی به‌سزا در شناساندن محتوای اثر و درنتیجه ربط موضوعی آن با دیگر آثار دارد (Ritchie, Teufel, & Robertson, 2006; Agarwal, Choubey, & Yu, 2010; Doslu & Bingol, 2016). تجارب تازه‌تر در شبکه‌های اجتماعی و در پیوند میان مدرک با متون اجتماعی در حال شکل‌گیری است و درباره آن نگاشته شده است (Bennett, El-Arini, Joachims, & Svore, 2012). مشارکت کاربران در تولید محتوا در رسانه‌های اجتماعی، در گستره‌ای از فعالیت‌های اجتماعی نظیر نظرگذاشتن، توییت‌کردن، «برچسب‌گذاری اشتراکی»[4]، و رأی‌دهی شکل می‌گیرد (O'Reilly, 2005; Voss, 2007; Abbasi & Liu, 2013; Manikonda, Meduri, & Kambhampati, 2016, Shu, Wang, Tang, Zafarani, & Liu, 2017; Manikonda, Meduri, & Kambhampati, 2016). نظرات کاربران از ابزارهای مهم برای تبادل نظرات نه تنها در رسانه‌های عمومی، بلکه شبکه‌های علمی مانند کوپرنیکوس، مث‌سای‌نت[5]، آمازون، و سایت‌های مجلات علمی به‌شمار می‌آیند. نظرات خوانندگان و پست‌ها درباره آثار علمی منبعی از اطلاعات ارزشمند، بالقوه مفید، و غنی در اختیار پژوهشگران قرار می‌دهند (Wang, Bian, Chang, & Tseng, 2012; Potthast, Stein, Loose & Becker, 2012; Yamamoto, Masuda, Ohira, &

1. Single words
2. Word N-grams
3. Unigrams and bigrams
4. Collaborative tagging
5. Mathscinet

(Nagao, 2008, Hu, Sun, & Lim, 2008; Yee, Yates, Liu, & Frieder, 2009). داوران آزاد گروهی دیگر از متخصصان هستند که نتایج تحلیل محتوا و داوری مقاله را در نظرات آزاد گزارش می‌دهند (Ford, 2013; Ross-Hellauer, 2017). «داوری آزاد»[1] دستاورد وب ۲ است که امکان مشارکت و تعامل کاربران را فراهم کرده است (Ford, 2013).

شکل‌گیری داوری آزاد در پاسخ به داوری تخصصی دوسرکور، در مجلات شکل گرفت. این شیوه که بر پنهان‌سازی هویت نویسندگان، داوارن، یا هر دو استوار است شفافیت و آزادی دسترسی خوانندگان به نظرات داور/ان و پاسخ نویسنده/گان میسر نیست و نقدها و نگرانی‌های جدی درباره عملکرد و کیفیت آن به‌وجود آورده است (Wang, Rath, Deike, & Wu, 2016). بی‌اطمینانی و بی‌ثباتی ناشی از قضاوت انسانی؛ پاسخ‌گونبودن؛ فسادپذیری؛ هدررفت اطلاعات بالقوه ارزشمند در نظرات داوری؛ و سوگیری‌های اجتماعی براساس جنسیت، ملیت، وابستگی سازمانی، زبان، و رشته از مهم‌ترین دل‌نگرانی‌های داوری بسته است (Peters & Ceci, 1982; Gillespie, Chubin, & Kurzon, 1985; Travis & Collins, 1991; Daniel, 1993; Link, 1998; Tregenza, 2002; Dall'Aglio, 2006; Ross et al., 2006; Cronin, 2009; Budden et al., 2008; Ross-Hellauer, 2017). در مقابل، داوری تخصصی آزاد می‌تواند امکان بحث و گفتگو میان داوران، نویسندگان، و جامعه (Nentwich, 2005)؛ کوتاه‌شدن چرخهٔ داوری و تسریع در فرایند داوری (Boldt, 2011)؛ افزایش تعداد داوران (Ross-Hellauer, 2017)؛ کیفیت بررسی‌ها و داوری‌های مقالات (Nicholson & Alperin, 2016)؛ و ایجاد انصاف و «عدالت اجتماعی»[2] (Ford, 2013) را در انتشارات علمی فراهم کند.

در بیشتر بسترهایی که برای داوری تخصصی آزاد ایجاد شده است گزارش‌های داوری و پاسخ‌های نویسنده و نظرات بازدیدکننده‌های سایت و علاقه‌مندان در اختیار خوانندگان قرار می‌گیرد. یکی از این بسترها، «اف‌هزار ریسرچ»[3] در علوم زیستی و پزشکی است. در اف‌هزار ریسرچ ابتدا گروه دبیران داخلی با دقت مقالات را برای اطمینان از شایستگی‌های نویسنده، مطلوب‌بودن متن به‌لحاظ محتوا، قابل‌فهم‌بودن، کیفیت، لحن و فرمت، کامل‌بودن، پایبندی به استانداردهای اخلاقی، و نبود شواهد سرقت علمی بررسی می‌کنند. سپس داوری تخصصی رسمی پس از انتشار الکترونیکی مقالات انجام می‌شود. نظرات داوران در این پایگاه منتشر می‌شود. ازآنجاکه داوری تخصصی پس از انتشار انجام می‌شود گزارش داوران اغلب شامل پیشنهادهایی برای بهبود یا تذکر نقاط ضعف مقاله است.

1. Open Peer Review
2. Social Justice
3. F1000 Research

نظرات یا توصیه‌های داوران رسمی یا کاربران اجتماعی می‌تواند تحلیل آنها از محتوا و نظرات‌شان از جنبه‌های مختلف اثر باشد. ارزش نظرات کاربران شبکه‌های اجتماعی، کتاب‌ها، مقالات خبری، منابع صوتی تصویری، و بلاگ‌ها با اندازه‌گیری بازیابی، رتبه‌بندی، طراحی سامانه‌های توصیه‌گر، و سامانه‌های خلاصه‌ساز مشخص شده است. ازجمله آنها نظرات کاربران در حوزهٔ بازیابی متون، تصاویر، ویدئو، منابع چندرسانه‌ای و وبلاگ‌هاست. چنانکه میزان شباهت نظرات به موضوع شیء اصلی و جایگزینی نظرات با متن نظردهی‌شده به‌منظور رتبه‌بندی بررسی و نشان داده شده است ۱۰ نظر برای محاسبه شباهت بالا میان متن و نظرات کافی است. از طرفی، ۱۰۰ الی ۵۰۰ نظر برای جایگزین‌کردن با متن نظردهی‌شده به‌منظور رتبه‌بندی مناسب است (Potthast, 2009). همچنین تأثیر نظرات کاربران بر نمایه‌سازی و بازیابی محتواهای صوتی تصویری (Eickhoff, Li, & Vries, 2013)، مفیدبودن نظرات کاربران وب برای جستجو (Yee et al., 2009)، تأثیر نظرات کاربران در محیط وب بر بازیابی منابع بین‌رسانه‌ای[1] (Potthast, Stein, & Becker, 2010)، اثربخشی نظرات کاربران وب بر رتبه‌بندی تصاویر (San Pedro, Yeh, & Oliver, 2012)، و تأثیر بازخوردهای اجتماعی بر رتبه‌بندی ویدئوهای یوتیوب (Chelaru, Orellana-Rodriguez, & Altingovde, 2014) تأیید و نشان داده شده است.

در حوزه کتاب‌ها، تأثیر نظر خوانندگان کتاب‌ها در بازیابی نشان داده شده است (Chaa, Nouali, & Bellot, 2018) و سامانه جستجوی کتابی‌ای ارائه شده است که نه تنها رتبه‌بندی را براساس محتوای کتاب‌ها و ابرداده‌ها امکان‌پذیر می‌کند، بلکه اطلاعات اجتماعی (نظرات، تگ‌ها، و درجه‌بندی) را در جستجوی کتاب‌ها درنظر گرفته است (Yin et al., 2016). در پژوهشی نظرات و یادداشت‌ها برای پیشنهاد ابردادهٔ واقعی افزوده برای مجموعه مدارک مفید دانسته شده است (Stvilia & Jörgensen, 2010). در پژوهشی دیگر نشان داده شد نظرات خوانندگان بخش مهمی از بلاگ‌ها هستند و استفاده از نظرات، بازیابی بلاگ‌ها را بهبود می‌بخشد. همچنین اگر نظرات در جستجوی کلیدواژه‌ای مد نظر قرار گیرد دقت تا ۱۵ درصد افزایش می‌یابد (Mishne & Glance, 2006).

برخی اهمیت نظرات کاربران در سامانه‌های توصیه‌گر را بررسی کرده‌اند؛ ازجمله پژوهشی که سامانه پیشنهاددهندهٔ جدیدی در بسترهای رسانه‌های اجتماعی ارائه داده است که در آن از نظرات کاربران، به‌عنوان پالاینده برای پیشنهاد خبر استفاده می‌شود (Li, Wang, Chen, & Lin, 2010). پژوهش‌های دیگر، خلاصه‌سازی بلاگ‌ها

1. Cross-media

و مدارک وبی را براساس نظرات کاربران مد نظر قرار دادند (ازجمله Hu, Sun, & Lim, 2007; 2008).

با آنکه پژوهش‌ها دربارهٔ نظرات کاربران رو به افزایش است، تاکنون پژوهش وسیعی دربارهٔ ارزش اطلاعاتی نظرات خوانندگان مقالات علمی و کاربرد آنها انجام نشده است. جنبه‌ای از ارزش اطلاعاتی نظرات خوانندگان را می‌توان در شباهت میان نظرات دید. درصورتی‌که مقالات مشابه نظرات داوری مشابه داشتند، می‌توان از نظرات در رتبه‌بندی مجدد نتایج بازیابی و بازیابی تعاملی بهره گرفت؛ زیرا در بسیاری اوقات در بازیابی تعاملی و هدفمند، صرفِ دراختیارداشتن مقالات با محتوای مشابه برای کاربر کافی نیست؛ چرا که ممکن است او به‌دنبال آگاهی از بازخوردهای خوانندگان پیشین باشد.

نظرات دربارهٔ ابعاد عینی مقاله می‌تواند به درک بهتر محتوا بیانجامد و اعتبار و کیفیت آن را آشکار سازد. کاربران چه‌بسا بخواهند مقالات با بازخوردهای مشابه داوران یا منتقدان را در کنار هم ببینند. پیش‌شرط چنین سامانه‌ای آن است که شرح درباره مقاله، بتواند موضوع آن را نیز منعکس کند. بنابراین، به سنجش شباهت نظرات در مقالات مشابه نیاز است.

در پژوهش حاضر ما کوشیده‌ایم میزان مشابهت نظرات داوری مقالاتی را مشخص کنیم که در محتوا به‌هم شبیه‌اند تا با تأیید غنای اطلاعاتی نظرات، بتوان از پتانسیل نظرات مقالات علمی در ارزیابی، بازیابی، توصیه، یا خلاصه‌سازی استفاده کرد. بدین منظور، به دو پرسش زیر پاسخ می‌دهیم:

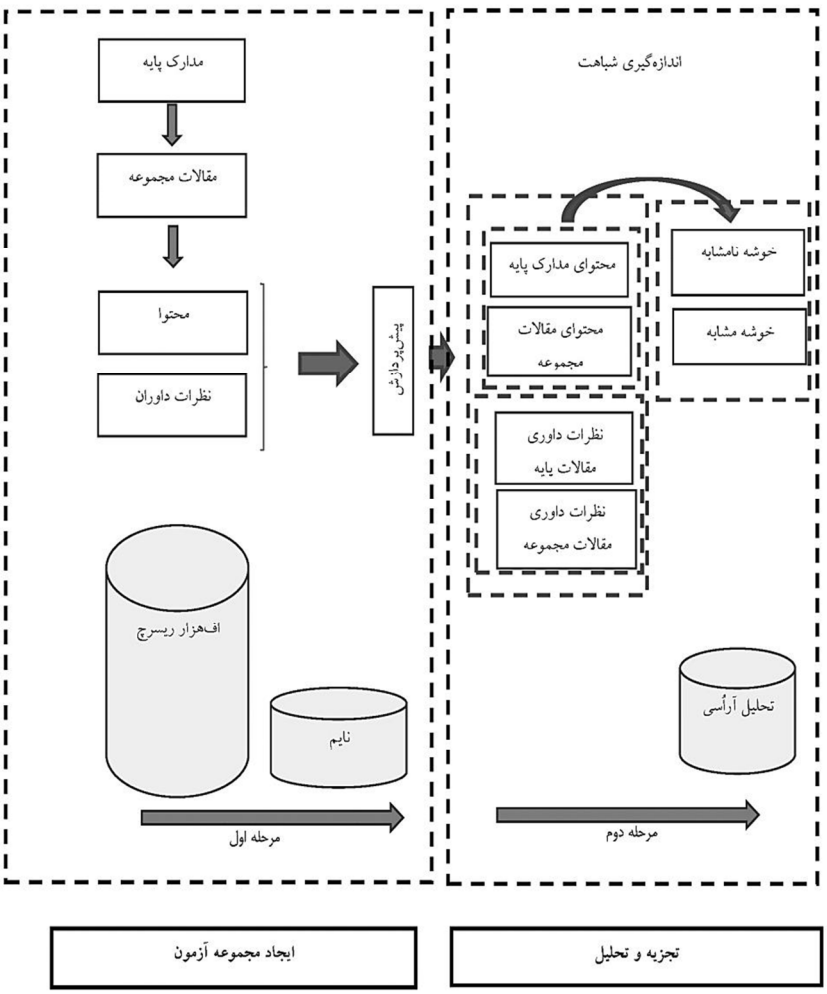
- آیا با افزایش شباهت متنی مدارک، شباهت نظرات داوری آنها هم افزایش می یابد؟
- آیا نظرات داوری مشابه، پتانسیل تشخیص صحیح مقالات مشابه را دارند؟

روش‌شناسی

ما نمونه‌ای از مقالات علوم زیستی و پزشکی در پایگاه اف‌هزار ریسرچ را برگرفته‌ایم و محتوا و نظرات داوران آنها را به‌کمک فنون پردازش زبان طبیعی تحلیل و شباهت‌های آنها را محاسبه کرده‌ایم. واحد تحلیل، تک‌واژه و عبارت دوواژه‌ای است. تحلیل محتوا به‌روش پردازش زبان طبیعی بر عنوان و چکیده و کلیدواژه انجام شده

است؛ زیرا این سه، بازنمون‌های مهم مدارک هستند (Hartley & Kostoff, 2003) که اثربخشی آنها در بازنمایی متن آثار تأیید شده است (Jenuwine & Floyd, 2004).

نظرات گزارش‌های داوری آزاد است. در زمان گردآوری داده‌ها (خرداد ماه ۹۷)، ۲۲۱۲ مقاله بر بستر اف‌هزار ریسرچ منتشر شده بود. ۱۵۹ مقاله هنوز نظر داوری دریافت نکرده بودند. به‌منظور بارگذاری مقالات و نظرات داوران، برنامه‌ای به زبان C# نوشته شد. از مقالات و نظرات داوران این پایگاه برای ساخت مجموعه آزمون‌ها استفاده کردیم (شکل ۱).



شکل ۱. نمودار مراحل پژوهش

ایجاد مجموعه آزمون: در ابتدا ۱۰۰ مقاله را به‌عنوان «مدارک پایه»[1] به‌طور تصادفی برگزیدیم و با ابزار داده‌کاوی نایم[2] متن آنها را با تمام ۲۰۶۱ مقالهٔ موجود در مجموعه مقایسه کردیم و میزان شباهت آنها را براساس سنجهٔ شباهت کسینوسی بین مقادیر TF-IDF[3] واژگان‌شان مشخص کردیم. ازآنجاکه ۱۰۰ مدرک پایه و ۲۰۶۱ مقاله مجموعه به‌ترتیب ۹۲ و ۲۱۶۱ نظرات داوری داشتند، فرایند سنجش شباهت به ساخت ۱۸۰۴۱۲ زوج مدرک منجر شد.

محاسبه شباهت: به‌منظور سنجش شباهت متنی مقالات به‌لحاظ محتوا و نظرات، از فنون پردازش زبان طبیعی استفاده کردیم. در ابتدا با ساخت یک workflow، پیش‌پردازش زبان‌شناختی بر عناوین، چکیده‌ها، و کلیدواژه‌ها انجام گرفت. بدین ترتیب که علائم سجاوندی و کلمات بازدارنده[4] حذف و همهٔ حروف بزرگ به کوچک تبدیل شد و ریشه‌یابی با استفاده از الگوریتم snowball (پرتر۲[5]) انجام شد. در تحلیل شباهت متنی، تمامی واژه‌های مندرج در عنوان و چکیده و کلیدواژهٔ مقالات منهای کلمات بازدارنده تحلیل شد.

به‌منظور محاسبهٔ شباهت مدارک پایه به دیگر مقالات موجود در مجموعه به‌لحاظ محتوا و نظرات، از سنجه شباهت کسینوسی (TF-IDF) استفاده کردیم و با محاسبهٔ مقادیر فراوانی نسبی واژه در مدرک و فراوانی مدارک حاوی واژه، اهمیت و ارزش کلمات در یک مدرک را تعیین کردیم.

روش تجزیه و تحلیل: به‌منظور آماده‌سازی داده‌ها، محاسبه شاخص‌ها، و تحلیل داده‌ها از ابزار داده‌کاوی نایم، برنامه Excel، و نرم‌افزار SPSS 26 استفاده کردیم. برای پاسخ به پرسش نخست، شباهت متن مقالات و شباهت نظرات داوری آنها تعیین شد. با توجه به نرمال‌نبودن داده‌ها و بالابودن شمار مقالات، از آزمون اسپیرمن با فن بوت‌استرپ[6] استفاده کردیم.

به‌منظور پاسخ به پرسش دوم، از منحنی «مشخصه عملکرد سامانه»[7] استفاده کردیم. این تحلیل برای ارزیابی عملکرد سامانه مستقل در هر آستانه مناسب است (Zuva & Zuva, 2012). بدین منظور، نخست به‌کمک شباهت کسینوسی شاخص TF-IDF، شباهت متنی بین هر مدرک پایه و تک‌تک مدارک مجموعه محاسبه شد. سپس شباهت بین نظرات داوری هر مدرک پایه و تک‌تک مدارک مجموعه را سنجیدیم. ازآنجاکه کاربرد منحنی ROC در سامانه‌های طبقه‌بندی دودویی است نخست مدارک براساس شباهت متن آنها با استفاده از فن خوشه‌بندی k-means به دو

1. Seed documents
2. KNIME
۳. فراوانی واژه در مدرک/ فراوانی مدارک حاوی واژه
4. Stop Words
5. Porter2
6. Bootstrap
7. Receiver Operating Characteristics (ROC)

گروه مشابه و متفاوت رده‌بندی شد. این تحلیل به شناسایی دو خوشه منجر شد. در سطح تک‌واژه‌ها، خوشهٔ نامشابه (۹۷/۱۸ درصد) ۱۷۵۳۳۷ زوج مدرک و خوشه مشابه (۲/۸۱ درصد) ۵۰۷۵ زوج مدرک بود. در سطح دوواژه‌ها، خوشه نامشابه ۱۸۰۲۷۸ (۹۹/۹۲ درصد) و خوشه مشابه ۱۳۴ (۰/۰۷ درصد) زوج مدرک بود. خوشه‌بندی معیار درنظر گرفته شد؛ به این معنی که مدارکی که هم به‌لحاظ محتوا و هم به‌لحاظ نظرات با مدارک پایه مشابه بودند، به‌عنوان مثبت‌های حقیقی در خوشه محتوای مشابه قرار گرفتند. مدارکی که محتوای مشابه، اما نظرات متفاوت با مدارک پایه داشتند به‌عنوان مثبت کاذب در خوشه محتوای نامشابه قرار گرفتند. اگر مدرکی در خوشه محتوای متفاوت، به‌لحاظ نظرات مشابه بود، به‌عنوان منفی کاذب و اگر به‌لحاظ نظرات هم متفاوت بود، به‌عنوان منفی حقیقی تعریف می‌شد (جدول ۱).

جدول ۱. دسته‌بندی دودویی مدارک براساس شباهت آنها به پرسش‌ها

	نظرات مشابه	نظرات متفاوت
محتوای مشابه	مثبت حقیقی	مثبت کاذب
محتوای متفاوت	منفی کاذب	منفی حقیقی

در منحنی ROC نرخ مثبت حقیقی همان بازیافت[1] یا حساسیت[2] است که در محور عمودی (Y) قرار گرفت. در محور افقی (X) مثبت کاذب، یعنی همان «ریزش کاذب»[3] قرار گرفت که از تفاضل «منفی‌های حقیقی»[4] یا ویژگی[5] از یک به‌دست می‌آید.

منفی‌های حقیقی− ۱ = ویژگی − ۱ = ریزش کاذب

بر همین اساس، منحنی دقت- بازیافت نیز ترسیم می‌شود. در محور افقی، وقتی بازیافت به‌سوی یک، میل کند، یعنی تمامی مدارک مرتبط در مجموعه به‌درستی تشخیص داده شده و در طبقه مربوط قرار داده شده است. در محور عمودی دقت قرار دارد. وقتی دقت برابر با یک باشد، یعنی تمامی آثار بازیابی‌شده به‌درستی تشخیص داده شده و در طبقه درست قرار گرفته‌اند.

1. Recall
2. Sensitivity
3. Fall out rate
4. True Negative
5. Specificity

یافته‌ها

- **همبستگی میان مقالات مشابه و نظرات داوری آنها**

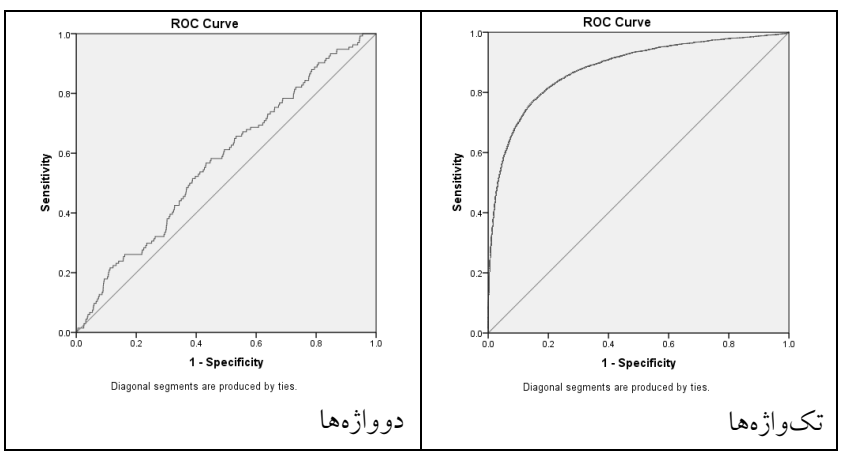
جدول ۲ نتیجه تحلیل همبستگی اسپیرمن میان شباهت محتوایی و نظرات داوری مقالات را با استفاده از فن بوت‌استرپ به‌نمایش می‌گذارد. همان‌گونه که مشاهده می‌شود همبستگی مثبت و معناداری میان میزان شباهت محتوایی و نظرات داوری مقالات وجود دارد. بدین معنا که هرچه میزان شباهت مقالات پایه به مقالات مجموعه افزایش یابد بر میزان شباهت نظرات آنها افزوده می‌شود.

جدول ۲. همبستگی بین شباهت محتوایی و شباهت نظرات داوری مدارک

ضریب همبستگی	سطح معناداری	سطح اطمینان ۹۵ درصد	
		بالاترین حد	پایین ترین حد
۰/۳۰۸	۰/۰۰۰	۰/۴۰۷	۰/۲۰۷

- **صحت پیش‌بینی شباهت محتوایی مقالات براساس شباهت نظرات داوری آنها**

شکل ۲ نتایج تحلیل منحنی ROC را به‌منظور آزمون صحت پیش‌بینی شباهت محتوایی مقالات براساس شباهت نظرات داوری آنها در سطح تک‌واژه و دوواژه نشان می‌دهد.



شکل ۲. صحت پیش‌بینی شباهت محتوایی براساس شباهت نظرات داوری مقالات

خط مرجع نشانگر نقطه ۰/۵ و ۰/۵ است و حد متوسط ریزش کاذب و بازیافت نشان می‌دهد هرچه عملکرد سامانه از حد متوسط فراتر رفته است سامانه عملکرد مطلوب‌تری داشته و منحنی از خط مرجع بالاتر رفته است. بالاترین سطح ذیل منحنی

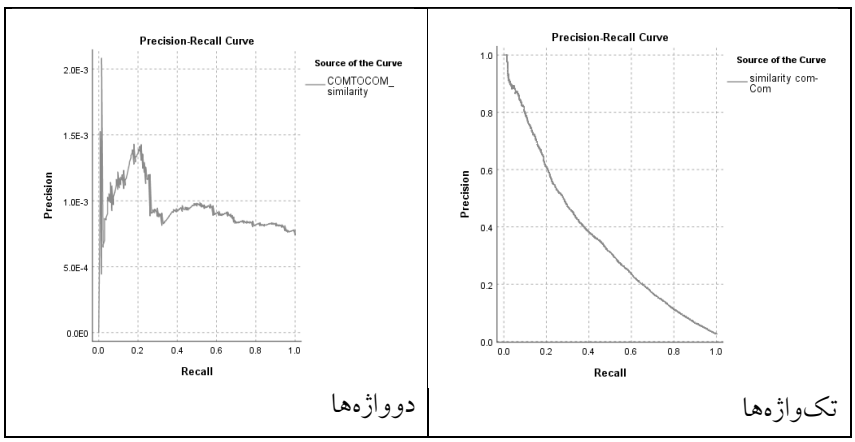
۰/۸۸۷ و کمترین سطح ذیل منحنی ۰/۸۷۷ است. سطح ذیل منحنی معنادار شده است (AUC=.882, Std. Error=0.003, Asymptotic Sig.=0.000). این بدین معناست که احتمال اینکه نظرات مشابه بتوانند محتوای مشابه را به‌درستی تشخیص دهند ۰/۸۸۲ است. به‌عبارت دیگر، سامانهٔ مبتنی بر نظرات مشابه می‌تواند با احتمال ۸۸/۲ درصد مقالات مشابه را به‌درستی شناسایی کند.

سطح ذیل منحنی به‌دست‌آمده برای دوواژه‌ای‌ها نیز فراتر از خط مرجع (۰/۵ و ۰/۵) است و ارزش سطح ذیل منحنی در دامنه ۰/۶۲۲ در بالاترین مقدار و ۰/۵۲۸ در کمترین مقدار قرار دارد و ارزش سطح ذیل منحنی آنها معنادار شده است (AUC=0.575, Std. Error=0.024, Asymptotic Sig.=0.003). با این حال، صحت پیش‌بینی برای دوواژه‌ای‌ها کمتر از تک‌واژه‌ای‌هاست.

شکل ۳، منحنی‌های بازیافت را برحسب دقت برای تحلیل تک‌واژه‌ها و دوواژه‌ها نشان می‌دهد. آشکار است زمانی‌که سطح زیر منحنی در تحلیل ROC بزرگ باشد مقادیر هردو سنجه نشان‌دهندهٔ دقت و بازیافت بالاست. بنابراین، در سامانه آرمانی هردو سنجه به طرف یک میل می‌کنند. اما، دستیابی به چنین سامانه‌ای ممکن نیست؛ زیرا دو سنجه دقت و بازیافت در موازنه[1] هستند و افزایش یکی به کاهش دیگری منجر می‌شود. بنابراین، در تحلیل این منحنی‌ها باید به نقطه‌ای توجه کرد که عملکرد نسبی مطلوب را به‌لحاظ هردو شاخص به‌دست دهد. همان‌گونه که ملاحظه می‌شود چنان‌که انتظار می‌رود در منحنی دقت- بازیافت تک‌واژه‌ها، هرچه بازیافت افزایش می‌یابد از دقت کاسته می‌شود. برای نمونه، وقتی بازیافت برابر با ۰/۲ است، یعنی ۲۰ درصد از مقالات مشابه موجود در مجموعه به‌درستی تشخیص داده شده‌اند، دقت تقریباً برابر با ۰/۶ است، یعنی ۶۰ درصد از این منابع بازیابی‌شده مربوط بوده‌اند. بنابراین، اگر انتظار رود که سامانه به دقتی برابر با ۶۰ درصد برسد، نرخ بازیافت آن برابر با ۲۰ درصد خواهد بود. با این حال، اگر نرخ بازیافت بالاتری برای مثال ۴۰ درصد مد نظر باشد، میزان دقت به ۴۰ درصد کاهش می‌یابد. وضعیت برای دوواژه‌ای‌ها تاحدی متفاوت است؛ به‌نحوی‌که برخلاف منحنی تک‌واژه‌ای، در این منحنی پس از نوساناتی وضعیت تقریباً باثبات به‌نظر می‌رسد؛ به این معنا که با افزایش بازیافت، دقت کاهش نمی‌یابد، بلکه تقریباً بی‌تغییر باقی می‌ماند. این لزوماً به‌معنای عملکرد بهتر آن نیست؛ زیرا مقدار دقت در منحنی دوواژه‌ای‌ها از اساس بسیار کمتر از مقادیر مربوط به تک‌واژه‌ای‌هاست. برای نمونه، تحلیل تک‌واژه‌ای‌ها در نقطه برش بازیافت برابر با ۰/۴ دقتی در همین حد را نشان می‌دهد؛ حال آنکه در

1. Tradeoff

نمودار مربوط به دوواژه‌ای‌ها، در این نقطه بازیابی، دقت کمتر از ۰/۰۰۱ است. بر این اساس، می‌توان استنباط کرد که در نقطه بازیافت ۴۰ درصد هر دو سامانه به دقت نسبتاً مقبولی در حد ۴۰ درصد دست می‌یابند.



شکل ۳. نمودار دقت- بازیافت

نتیجه‌گیری

با آنکه واژگان طبیعی به‌کاررفته در مدارک علمی می‌توانند به بازنمون محتوای آنها کمک کنند متن مدارک به‌دلیل ایستایی و پوشش ناکافیِ روابط معنایی به منابعی مکمل نیاز دارند که هم بتوانند به غنای واژگانی متن بیافزایند و هم پویایی روابط علمی بین مدارک را بازتاب دهند. نقش متون اجتماعی در بازنمایی رسانه‌های گوناگون در پژوهش‌هایی چند بررسی شده است؛ با این حال، پژوهشگران به ارزش این متون برای بازنمایی مقالات علمی کمتر توجه کرده‌اند. در فضای داوری اجتماعی، گفتگوی چندجانبه در میان خوانندگان، نویسندگان، دبیران مجلات، و داوران درباره شایستگی‌ها یا کاستی‌های برون‌دادهای علمی درمی‌گیرد (Ford, 2013; Wilson, Boe, Sala, Puttaswamy, & Zhao, 2009; Hille & Bakker, 2014). از این‌رو، انتظار می‌رود که این متون که نشانگر «هوش اشتراکی اجتماعی»[1] (Hsu, Khabiri, & Caverlee, 2009; Ito & Shimizu, 2012) و دانش و ادراک خوانندگان هستند منابع غنی از اطلاعات در اختیار بگذارند.

نتایج این پژوهش مؤید این انتظار بود. نتایج نشان داد مقالات مشابه، نظرات داوری مشابه دارند. از یک‌سو، همبستگی بین شباهت محتوایی و شباهت نظرات

1. Social collective intelligence

داوری مقالات معنادار و مثبت است و از سوی دیگر، منحنی‌های تحلیل عملکرد سامانه نشان داد نظرات داوران در تشخیص مقالاتی که محتوای مشابه دارند به‌درستی عمل می‌کنند. این نتیجه هم برای تک‌واژه‌ها و هم برای دوواژه‌ها صدق می‌کند.

همان‌گونه که پیش‌تر گفتیم در زمینه تأثیر نظرات خوانندگان بر بازیابی مقالات علمی پژوهش انجام نشده است. از این‌رو، امکان مقایسه نتایج این پژوهش با پژوهش‌های مرتبط وجود ندارد. با این حال، چنانچه پژوهش‌هایی مد نظر قرار گیرد که درباره ارزش نظرات خوانندگان در محیط‌های دیگر است پژوهش حاضر هم‌راستا با آنهاست. ازجمله پژوهشی که نشان داد ۱۰۰ الی ۵۰۰ نظر خوانندگان برای جایگزین‌کردن با متن نظردهی‌شده به‌منظور رتبه‌بندی مناسب است (Potthast, 2009). پژوهشی دیگر نشان داد از نظرات واردشده بر آیتم‌های وبی می‌توان برای بازیابی منابع بین‌رسانه‌ای استفاده کرد (Potthast et al., 2010). پژوهشگرانی نیز معتقدند «نظرات، ادراک یا بازخورد خوانندگان» محتوای مدرک هستند و قاعدتاً با مقالات مادر وجوه اشتراک بسیار دارند. با این حال، انتظار اینکه تصویری کامل و بی‌نقص از مقاله ارائه دهند نیز منطقی نیست؛ زیرا نظرات خوانندگان آزادانه و در فضایی پویا، جهانی، و متکثر خلق می‌شوند (Hu et al., 2008).

نتایج این پژوهش، با تأیید شباهت بین نظرات داوران و مقالات علمی مربوط ارزش اطلاعاتی آنها را تأیید کرد. این شباهت از دو جهت اهمیت دارد: از یک سو، این نظرات می‌توانند بازنمون مقالات به‌شمار آیند؛ از سوی دیگر، محدودبودن فضای نظرات و الزام داوران به بازتاب نظرات خود در محدودهٔ مشخص از کلمات باعث می‌شود چگالی موضوعی نظرات در مقایسه با خود مقالات افزایش یابد. این امر سبب می‌شود نظرات نه تنها به‌لحاظ اثربخشی، بلکه به‌لحاظ کارایی نیز ارزشمند شود. بدین ترتیب، نظرات داوران مقالات علمی را می‌توان در بهبود کارهای مبتنی بر پردازش زبان طبیعی مانند بازیابی، خلاصه‌سازی، ارزیابی، یا رتبه‌بندی مد نظر قرار داد. با این حال، پژوهش ما نخستین از این نوع بر نمونه‌ای محدود اجرا شده است. قضاوت قطعی درباره ارزش نظرات داوران در بازنمون مقالات علمی نیاز به پژوهش‌های گسترده‌تر دارد.

مآخذ

Abbasi, M. A., & Liu, H. (2013). Measuring user credibility in social media. *In International Conference on Social Computing, Behavioral-Cultural*

Modeling, and Prediction, April 2-5, (pp. 441-448). Berlin, Heidelberg: Springer

Agarwal, S., Choubey, L., & Yu, H. (2010). Automatically classifying the role of citations in biomedical articles. In *Proceedings of American Medical Informatics Association Fall Symposium (AMIA),* (pp. 11-15). Retrieved June 6, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041379/

Bennett, P. N., El-Arini, K., Joachims, T., & Svore, K. M. (2012). Enriching information retrieval. *ACM Sigir Forum, 45* (2), 60-65. Retrieved June 6, 2020, from https://www.cs.cornell.edu/people/tj/publications/bennett_ etal_ 11a.pdf

Boldt, A. (2011). Extending ArXiv. org to achieve open peer review and publishing. *Journal of Scholarly Publishing*, *42* (2), 238-242.

Brennan, P. (2012). Uses of computational stylometry to determine demographics for online reputation management. *Language Engineering for Online Reputation Management*, 15-16. Retrieved June 6, 2020, from http://www.lrec-conf.org/proceedings/lrec2012/workshops/15.LREC% 202012%20Online%20Reputation%20Proceedings.pdf

Budden, A. E., Tregenza, T., Aarssen, L. W., Koricheva, J., Leimu, R., & Lortie, C. J. (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution*, *23* (1), 4-6.

Cao, Z., Li, S., Liu, Y., Li, W., & Ji, H. (2015). A novel neural topic model and its supervised extension. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. Retrieved June 6, 2020, from https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9303/9544

Chaa, M., Nouali, O., & Bellot, P. (2018). Combining Tags and Reviews to Improve Social Book Search Performance. In *International Conference of the Cross-Language Evaluation Forum for European Languages, September 10-14,* (pp. 64-75). Cham: Springer.

Chelaru, S., Orellana-Rodriguez, C., & Altingovde, I. S. (2014). How useful is social feedback for learning to rank YouTube videos? *World Wide Web*, *17* (5), 997-1025.

Constantin, C., Du Mouza, C., Faget, Z., & Rigaux, P. (2011). The melodic signature index for fast ontent-based retrieval of symbolic scores. In *12th International Society for Music Information Retrieval Conference (ISMIR), October 24-28*, (pp. 363-368). Retrieved June 6, 2020, from http://ismir2011.ismir.net/papers/PS3-2.pdf

Cronin, B. (2009). Vernacular and vehicular language. *Journal of the American Society for Information Science and Technology*, *60* (3), 433-433.

Cui, H., Mittal, V., & Datar, M. (2006). Comparative experiments on sentiment classification for online product reviews. In *The Twenty-First National Conference on Artificial Intelligence (AAAI-06)*, July 16–20, (pp. 1265-1270). Retrieved June 6, 2020, from https://static.googleusercontent.com/ media/ research.google.com/en/ir/pubs/archive/4.pdf

Dall'Aglio, P. (2006). Peer review and journal models. *arXiv preprint physics/0608307*. Retrieved June 6, 2020, from https://arxiv.org/pdf/ physics/0608307.pdf

Daniel, H. D. (1993). An evaluation of the peer review process at Angewandte Chemie. *Angewandte Chemie International Edition in English*, *32* (2), 234-238.

Doslu, M., & Bingol, H. O. (2016). Context sensitive article ranking with citation context analysis. *Scientometrics*, *108*, 653-671.

Eickhoff, C., Li, W., & De Vries, A. P. (2013). Exploiting user comments for audio-visual content indexing and retrieval. In *European Conference on Information Retrieval, March 24-27*, (pp. 38-49). Berlin, Heidelberg: Springer.

Ford, E. (2013). Defining and characterizing open peer review: a review of the literature. *Journal of Scholarly Publishing*, *44* (4), 311-326.

Gillespie, G. W., Chubin, D. E., & Kurzon, G. M. (1985). Experience with NIH peer review: Researchers' cynicism and desire for change. *Science, Technology, & Human Values*, *10* (3), 44-54.

Golcher, F., & Reznicek, M. (2011). Stylometry and the interplay of topic and L1 in the different annotation layers in the FALKO corpus. Retrieved June 6, 2020, from https://edoc.hu-berlin.de/bitstream/handle/18452/2022/golcher.pdf?sequence=1

Hartley, J., & Kostoff, R. N. (2003). How useful are key words' in scientific journals? *Journal of Information Science, 29* (5), 433-438.

Hille, S., & Bakker, P. (2014). Engaging the social news user: Comments on news sites and Facebook. *Journalism Practice*, *8* (5), 563-572.

Hsu, C. F., Khabiri, E., & Caverlee, J. (2009). Ranking comments on the social web. In *International Conference on Computational Science and Engineering (Vol. 4), August 29-31,* (pp. 90-97). Calif., Los Alamitos: IEEE Computer Society.

Hu, M., Sun, A., & Lim, E. P. (2007). Comments-oriented blog summarization by sentence extraction. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 901-904). Retrieved June 9, 2020, from https://dl.acm .org/doi/10.1145/1321440. 1321571

Hu, M., Sun, A., & Lim, E. P. (2008). Comments-oriented document summarization: Understanding documents with readers' feedback. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 20-24*, (pp. 291-298). New York, NY: Association for Computing Machinery.

Ito, E., & Shimizu, K. (2012). Frequency and link analysis of online novels toward social contents ranking. In *2012 Second International Conference on Cloud and Green Computing, November 1-3,* (pp.531-536). IEEE. Retrieved June 6, 2020, from https://ieeexplore.ieee.org/document/6382867

Jenuwine, E. S., & Floyd, J. A. (2004). Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. *Journal of the Medical Library Association*, *92* (3), 349-353.

Lewis, D. D., & Jones, K. S. (1996). Natural language processing for information retrieval. *Communications of the ACM*, *39* (1), 92-101.

Li, Q., Wang, J., Chen, Y. P., & Lin, Z. (2010). User comments for news recommendation in forum-based social media. *Information Sciences*, *180* (24), 4929-4939.

Link, A. M. (1998). US and non-US submissions: an analysis of reviewer bias. *JAMA*, *280* (3), 246-247.

Manikonda, L., Meduri, V. V., & Kambhampati, S. (2016). Tweeting the mind and instagramming the heart: Exploring differentiated content sharing on social media. In *The Tenth International AAAI Conference on Web and Social Media, May 17-20*. Retrieved June 6, 2020, from https:// arxiv.org/pdf/1603.02718.pdf

Mishne, G., & Glance, N. (2006, May). Leave a reply: an analysis of weblog comments. In *Third Annual Workshop on the Weblogging Ecosystem*. Retrieved June 6, 2020, from http://leonidzhukov.net/hse/2011/seminar/ papers www2006-blogcomments.pdf

Nentwich, M. (2005). Quality control in academic publishing: Challenges in the age of cyberscience. *Poiesis & Praxis*, *3* (3), 181-198.

Nicholson, J., & Alperin, J. P. (2016). A brief survey on peer review in scholarly communication. *The Winnower*. Retrieved June 6, 2020, from https://thewinnower.com/papers/4659-a-brief-survey-on-peer-review-in-scholarly-communication#submit

O'reilly, T. (2005). What is web 2.0: Design Patterns and business models for the next generation of software. Retrieved June 6, 2020, from https://www. oreilly.com/pub/a/web2/archive/what-is-web-20.html

Peters, D. P., & Ceci, S. J. (1982). Peer review practices of psychological journals: the fate of published articles, submitted again. *Behavioral and Brain Sciences, 5* (2), 187-195.

Potthast, M. (2009). Measuring the descriptiveness of web comments. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, *July 19-23*, (pp. 724-725). New York, NY: ACM.

Potthast, M., Stein, B., & Becker, S. (2010). Towards comment-based cross-media retrieval. Retrieved June 6, 2020, from https://dl.acm.org/doi/pdf/ 10.1145/1772690.1772858

Potthast, M., Stein, B., Loose, F., & Becker, S. (2012). Information (2010). Towards comment-based cross-media retrieval in. In *Proceedings of the* comment sphere. *ACM Transactions19th international conference on Intelligent Systems and Technology (TIST)*, *3*(4), 68. *World wide web* (pp. 1169-1170). Retrieved June 9, 2020, from https://dl.acm.org/doi/10.1145/ 2337542.2337553

Ritchie, A., Teufel, S., & Robertson, S. (2006). How to find better index terms through citations. In *Proceedings of the Workshop on How Can Computational Linguistics Improve Information Retrieval?* (pp. 25-32). Retrieved June 6, 2020, from https://dl.acm.org/doi/pdf/10.5555/1629808. 1629813

Ross, J. S., Gross, C. P., Desai, M. M., Hong, Y., Grant, A. O., Daniels, S. R., et al. (2006). Effect of blinded peer review on abstract acceptance. *JAMA*, *295* (14), 1675-1680.

Ross-Hellauer, T. (2017). What is open peer review? a systematic review. *F1000Research*, *6*. Retrieved June 6, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5437951/

San Pedro, J., Yeh, T., & Oliver, N. (2012). Leveraging user comments for aesthetic aware image search re-ranking. In *Proceedings of the 21st international conference on World Wide Web, April 16- 20,* (pp. 439-448). New York, NY: ACM.

Shu, K., Wang, S., Tang, J., Zafarani, R., & Liu, H. (2017). User identity linkage across online social networks: a review. *Acm Sigkdd Explorations Newsletter*, *18* (2), 5-17.

Stvilia, B., & Jörgensen, C. (2010). Member activities and quality of tags in a collection of historical photographs in Flickr. *Journal of the American Society for Information Science and Technology*, *61* (12), 2477-2489.

Thomas, B., Vinod, P., & Dhanya, K. A. (2014). Multiclass emotion extraction from sentences. *International Journal of Scientific & Engineering Research*, *5* (2), 12-15.

Travis, G. D. L., & Collins, H. M. (1991). New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology, & Human Values*, *16* (3), 322-341.

Tregenza, T. (2002). Gender bias in the refereeing process? *Trends in Ecology & Evolution, 17* (8), 349-350.

Voss, J. (2007). Tagging, folksonomy & co-renaissance of manual indexing? *arXiv:cs/0701072*. Retrieved June 6, 2020, from https://arxiv.org/pdf/cs/0701072.pdf

Wang, P., Rath, M., Deike, M., & Wu, Q. (2016). Open peer review: an innovation in scientific publishing. Retrieved June 6, 2020, from https://core.ac.uk/download/pdf/158312603.pdf

Wang, X., Bian, J., Chang, Y., & Tseng, B. (2012). Model news relatedness through user comments. In *Proceedings of the 21st International Conference on World Wide Web, April 16-20,* (pp. 629-630). Retrieved June 6, 2020, from https://dl.acm.org/doi/pdf/10.1145/2187980.2188161

Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P., & Zhao, B. Y. (2009). User interactions in social networks and their implications. In *Proceedings of the 4th ACM European Conference on Computer Systems* (pp. 205-218). New York, NY: ACM. Retrieved June 9, 2020, from https://sites.cs.ucsb.edu/~ ravenben/publications/pdf/interaction-eurosys09.pdf

Yamamoto, D., Masuda, T., Ohira, S., & Nagao, K. (2008). Video scene annotation based on web social activities. *IEEE Multimedia*, *15* (3), 22-32.

Yee, W. G., Yates, A., Liu, S., & Frieder, O. (2009). Are web user comments useful for search? Retrieved June 6, 2020, from http://ceur-ws.org/Vol-480/paper7.pdf

Yin, X. C., Zhang, B. W., Cui, X. P., Qu, J., Geng, B., Zhou, F., et al. (2016). ISART: a generic framework for searching books with social information. *PloS one*, *11* (2), e0148479. Retrieved June 6, 2020, from https://doi.org/10.1371/journal.pone.0148479

Zuva, K., & Zuva, T. (2012). Evaluation of information retrieval systems. *International Journal of Computer Science & Information Technology*, *4* (3), 35-43

استناد به این مقاله:

رشیدی شریف‌آباد، کیانوش؛ ستوده، هاجر؛ میرزابیگی، مهدیه؛ و فخراحمد، سیدمصطفی (۱۳۹۹). سنجش شباهت نظرات داوری آزاد و محتوای مقالات علمی به‌روش پردازش زبان طبیعی. *مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۳۱ (۲)، ۸۶-۱۰۳.