International Journal of Information Science and Management
Vol. 17, No. 2, 2019, 19-31

A Correlation Study of Co-opinion and Co-citation Similarity Measures

Maryam Yaghtin
Ph.D. candidate in Knowledge & Information Sciences, Department of Knowledge & Information Sciences, Faculty of Education & Psychology, Shiraz, Iran
yaghtin.maryam@gmail.com

Hajar Sotudeh
Associate Prof. Department of Knowledge & Information Sciences, Faculty of Education & Psychology, Shiraz University, Shiraz, Iran
Corresponding Author: sotudeh@shirazu.ac.ir

Mehdi Mohammadi
Associate Prof. Department of Educational Management and Planning, Faculty of Education & Psychology,
Shiraz University, Shiraz, Iran
m48r52@gmail.com

Mahdieh Mirzabeigi
Assistant Prof. Department of Knowledge & Information Sciences, Faculty of Education & Psychology, Shiraz University, Shiraz, Iran
mmirzabeigi@gmail.com

Seyed Mostafa Fakhrahmad
Assistant Prof. Department of Computer Science & Engineering
School of Electrical and Computer Engineering
Shiraz University, Shiraz University, Shiraz, Iran
mfakhrahmad@yahoo.com

Abstract

Co-citation forms a relational document network. Co-citation-based measures are found to be effective in retrieving relevant documents. However, they are far from ideal and need further enhancements. Co-opinion concept was proposed and tested in previous research and found to be effective in retrieving relevant documents. The present study endeavors to explore the correlation between opinion (dis)similarity measures and the traditional co-citation-based ones including Citation Proximity Index (CPI), co-citedness and co-citation context similarity. The results show significant, though weak to medium, correlations between the variables. The correlations are direct for co-opinion measure, while being inverse for the opinion distance. Accordingly, the two groups of measures are revealed to represent some similar aspects of the document relation. Moreover, the weakness of the correlations implies that there are different dimensions represented by the two groups.

Keywords: Opinion, Co-opinion, Co-citation, Correlation, Citation Proximity Index, Similarity.

Introduction

Citation is believed to convey the impact of cited papers on the citing ones as regards their utility and impact (De Bliss, 2009; MacRoberts & MacRobers, 2010). It is also a proxy of socially determined quality (Cole & Cole, 1967; Lundberg, 2006). Furthermore, there exist some kinds of subject relatedness between cited and citing papers (Smith, 1981; Egghe & Rousseau, 1990). Citations are, thus, widely accepted and applied in evaluating and retrieving documents (Zhao, 2015; Gurulingappa et al., 2010; Ritchie, 2009; Rousseau, 2007). Co-citation is among the citation-based measures successfully used in retrieving scientific papers (Janssens and Gwinn, 2015; Yoon, Kim, Park, 2016; Eto, 2013; Zhao, 2014; Egghe & Rousseau, 1990; Bichteler & Eaton, 1980; Badran, 1984). It is defined as the relation between two documents cited together within a citing paper. The co-citedness, or the frequency of co-citation is defined as "the frequency with which two documents are cited together" (Small, 1973). The higher the frequency, the tighter the relation of the co-cited articles.

Given the weaknesses of quantity-based citation measures, it is strongly suggested to use content-based citation analysis (Small, 1982). "Syntactic analysis" or "citation proximity analysis (CPA)" (Gipp & Beel, 2009) measures the proximity of co-cited articles within their co-citing articles. The similarity between citation contexts (CC) and co-citation contexts (Co-CC), as a special case of citation relation, is also used as a representative of the subject similarity of co-cited papers (Callahan, Hockema, & Eysenbach, 2010; Jeong, Song, Ding, Zhang, Chambers, Song, Wang & Zhai, 2014). The merit of Co-CCs lies in the fact that they carry not only the co-citers' objective reports of different features of their co-cited articles, but also their opinions about them (Sendhilkumar, Elakkiya & Mahalakshmi, 2013; Ritchie, Robertson & Teufel, 2008; Agarwal Choubey & Yu, 2010; Doslu & Bingol, 2016). Opinion mining or sentiment analysis tries to extract opinions by analyzing their written texts (Shuy 2005; Liu and Zhang, 2012; Liu et al., 2012). People's opinion reflects their views, judgments and attitudes towards something (including individuals, concepts, things, etc.). Opinion mining techniques are helpful in extracting co-citers' viewpoints characterized by their opinion strengths and polarities (as being positive, negative or neutral) (Teufel, Siddharthan, Tidhar, 2006; Schafer & Spurk, 2010; Small, 2011; Athar & Teufel, 2012).

Based on the co-citation assumptions, we recently proposed a new measure called "co-opinion" of co-cited papers. It is defined as similarity of co-cited papers in terms of the opinions they receive from their co-citing papers. After analyzing the Co-CCs texts, we measured the opinions based on the opinion scores extracted from Senti-Wordnet. Using two dimensions of opinion strengths and polarities, the co-opinion concept was measured based on the inverse of opinion distance between two Co-CCs. The results showed that the co-cited papers were strongly co-opinionated whether in terms of the opinion polarities or strengths. We, also, proved the effectiveness of the measure in improving co-citation-based information retrieval (Yaghtin, Sotudeh, Mirzabeigi et al., under review).

Theoretically, the co-opinion concept is supported by the assumptions that authors always refer to articles relevant to their works (Bichteler & Eaton, 1980; Egghe & Rousseau, 1990). Consequently, co-cited articles are usually similar in their contents (Eto, 2015; White, 2016; Small, 1973; Hamedani , Kim & Kim, 2016). Besides, authors widely tend to take positive positions (Mahalakshmi, Siva, Sendhilkumar, 2015) and avoid negative credits when citing previous articles (MacRoberts & MacRoberts, 1984; Athar & Teufel, 2012; Mahalakshmi et al., 2015; Sendhilkumar et al., 2013). Given the fact that there are several traditional co-citation measures, a question arises as to whether the co-opinion and co-citation measures are in accordance with each other.

To answer the question, we conducted a research on the sample of co-cited medical papers studied in our previous work to test the correlation between the co-opinion concept, on the one hand and three traditional co-citation-based measure, including co-citedness, Citation Proximity Index (CPI) and Co-Citation context (Co-CC) textual similarity. Another new feature of the present study is that we calculated the co-opinion concept in a new way. “Opinion intersection” indicates the portion of the opinions shared by two co-polar co-citing articles. In other words, it reveals the extent to which the co-citing articles concurrently reject or confirm their co-cited articles. It is obvious that two anti-polar contexts, i.e. one with positive and another with negative attitude towards a co-cited article, have no opinion in common. The “opinion distance” can be, thus, considered as the other side of the coin, which measures the absolute difference between two opinions. The measures quantify two related concepts: while the former calculates the opinion similarity by calculating the minimum opinion shared by the co-citing articles, the latter indicates the dissimilarities between the opinion degrees.

To clarify the concept, let’s consider the documents A and B cociting C and D. Let’s first consider the polarity of the opinions the cocited pairs (C and D) receive from their cociting papers (A and B). If A and B both give positive credits to C and D, the cocited pairs (C and D) can be considered positively co-opinionated. If the cociting documents both reject C and D, the cocited pairs are negatively co-opinionated. If A rejects, but B confirms the pair C and D, the cocited papers can be considered as anti-polar, i.e. showing no similarity in the opinions they receive.

To further clarify the concept, let’s consider opinion strengths. If C and D receive +3 and +4 degrees of confirmation from their cociting papers, they have 3 degrees of opinion similarity, i.e. co-opinion. If they are rejected with the same degrees of opinions, they are still co-opinionated with the same strength. If C is rejected and D is confirmed with -3 and +4 degrees, respectively, they have no similarity in the opinions, but 7 degrees of opinion distance.

Research Questions

1- Is there any significant correlation between co-cited papers in terms of their opinion (dis)similarity and co-citedness?

2- Is there any significant correlation between co-cited papers in terms of their opinion (dis)similarity and CPI?

3- Is there any significant correlation between co-cited papers in terms of their opinion (dis)similarity and Co-CC textual similarity?

Literature Review

Document relations have been widely believed to be effective in evaluating and retrieving documents. Citation is the primary relation between documents forming the building block of the citation index proposed by Garfield to improve information retrieval and later to evaluate documents. Citing a document signifies its relation in content to the cited document (Smith, 1981; Egghe & Rousseau, 1990), its importance and utility (MacRoberts & MacRobers, 2010), its cognitive impact (De Bellis, 2009), and the recognition of its merits by scientific community leading to a socially controlled quality (Cole & Cole, 1967; Lundberg, 2006). Although, these assumptions have been challenged (Verbeek, Debackere, Luwel & Zimmermann, 2002; MacRoberts & MacRobers, 1989; 2010), no more objective method has been devised so far. In fact, citers constitute a global expert community to analyze and evaluate document contents. It is obvious that no promotion and tenure committee or indexing team is known to be truly as wide as the world as is the citation realm.

Co-citation and bibliographical coupling are among the well-known document networks derived from citations. They were found to be effective in information retrieval (Ritchie, 2009; Bichteler & Eaton, 1980; Nanba, Kando. Okumura, 2000; Zhao, 2014). However they, too, were found to suffer from some defects (Gipp & Beel, 2009; Eto, 2013; Martin, 1964; Dabrowska & Larsen, 2015; Boyack, Small & Klavans, 2013). In response to the deficiencies of quantitative citation analysis, content-based citation analysis is proposed. The content-based techniques can be considered on a continuum depending on the extent of their use of content elements. As instance, while CPA, or the Citation Proximity Analysis, uses the minimum content elements, i.e. the position of the citation within the citing document (Gipp & Beel, 2009), the citation context similarity can use a wide range of textual clues around the citation (Callahan, Hockema & Eysenbach, 2010).

CC opinion mining or sentiment analysis is among those innovative techniques based on content analysis of CCs. Negative and positive citation opinions can be used to evaluate cited papers and delve their actual impact (Abu-Jbara, Ezra, & Radev, 2013; Hernandez-Alvarez & Gomez, 2016; Yu, 2013; Athar & Teufel, 2012; Small, 2011). Negative citations help to uncover criticisms about the cited papers (Athar & Teufel, 2012). As instance, opinion and quantitative citation analyses differently rank cited papers (Cavalcanti, Prudêncio, Pradhan, Shah & Pietrobon, 2011; Sendhilkumar et al., 2013). According to Amadi (2014) highly cited papers do not necessarily attract conformational citations. Besides, the polarity of the opinions depends on the position where the citation occur within the citing documents. Although. there is not yet enough standard infrastructures to ensure the reliability of the results of opinion mining of scientific papers (Hernandez-Alvarez & Gomez, 2016), a growing number of studies have employed the techniques (Teufel, Siddharthan & Tidhar, 2006; Athar, 2011; 2014; Jochim & Schütze, 2012; Abu-Jbara et al., 2013; Athar & Teufel, 2012). As instance, Athar (2014) classified the opinions extracted from CCs using support vector machine classifiers and confirmed its effectiveness over naïve Bayesian classifiers.

Moreover, previous studies approved the correlation between citation and opinion measures. As instance, according to Abu-Jbara et al. (2013), total citation quantity exhibit stronger associations with positive citations compared to negative ones. Correlation between citation and opinion polarity were also confirmed by Teufel et al. (2006; 2009), Dong & Schäfer (2011) and Jochim & Schütze (2012). Given the novelty of the co-opinion measure (Yaghtin et al., under review), it is necessary to explore its correlation to such well-known, effective co-citation measures as co-citedness, CPI, and Co-CC textual similarity.

Methodology

Using a content analysis method with citation analysis approach, the present study analyzes CCs to investigate opinions expressed by co-citing papers about their co-cited works. To calculate Co-CC textual similarity, all words, except for stop words, were analyzed. However, the opinion analyses were carried out on nouns, adjectives, adverbs and verbs that are believed to carry opinion loads in a sentence (Esuli & Sebastiani, 2007; Mahalakshmi et al., 2015).

Research Sample

Using CITREC test collection, we built a sub-collection. It consisted of 30 randomly selected seed documents that served as queries. They were selected from those papers having at least 100 co-citations to ensure the inclusion of a high number of co-cited papers in the sample. 5394 co-cited papers for the 30 seed documents were identified.

In the next step, after identifying 9620 co-citing papers, their CCs were extracted from CoLiL database[1] (Fujiwara & Yamamoto, 2015).

Co-CC textual and Opinion Similarity

The CC contents extracted from CoLiL were processed by KNIME (the Konstanz Information Miner). Alongside the linguistic preprocessing, Parts-of-Speech (PoS) tagging and lemmatization were also applied. Furthermore, the data were purified in terms of trivial words and characters (i.e. stop words, numbers, and punctuations). At last, cosine similarity of the seed documents to their co-cited papers was calculated.

In the next step, opinion similarity was calculated in two ways. First, the sentiment or opinion scores were extracted from SentiWordNet (SWN). SWN uses Synsets to reflect different senses of a given word and assign an opinion score to each. The sum of positive, negative and neutral scores equals one within every Synset. Besides, each word may have different scores, depending on its PoS. To yield a single opinion score for each word, we calculated the average of its opinion scores after subtracting its negative opinion scores from the positive ones as proposed by Cavalcanti et al. (2011). To be doubly sure, we also used the opinion scores calculated by Sentiwords (SWs). SWs is a tool developed by Gatti, Guerini, Turchi (2016) based on SWN to manage the multi-meaning problem and proved to have a higher precision compared to other approaches.

Of the 38178 CCs extracted, there were some words not covered by SWN. They were mostly consisted of formulas, chemical constituents, abbreviations and other technical words that are not likely to have opinion loads. Also, a scarce number (1.3 %) of the identified words did not gain opinion scores. Negation words identified from Wilson, Hoffmann, Somasundaran, Kessler, Wiebe, Choi & Patwardhan (2005) consisted less than one percent of the words and were not managed, given their scarcity as well as their insignificant impact on identifying CCs opinions (Athar, 2014).

The opinion score of each CC was, then, calculated based on the average opinion score of its words.

Data analysis method

Calculating opinion measures: CCs' co-opinion degrees were analyzed using the two newly-introduced measures including "opinion intersection" and "opinion distance". "Opinion intersection" represents the co-opinion concept, while opinion distance reflects the cociters' disagreement about their cocited papers. As mentioned before, the former measures the least opinion score shared by a pair of co-cited papers. The latter is calculated by the absolute value of subtraction of their opinion scores. Besides, the co-opinion concept was measured for co-polar co-cited papers. Given the similarity of the polarities of two opinions, their intersection can be calculated by:

$$\text{co} - \text{opinion} = Min\ (|Opinion_1|, |Opinion_2|)$$

As instance two co-polar Co-CCs, with +3 and +4 or -3 and -4 opinion scores, have 3 degrees of co-opinion.

The opinion distance between two contexts was calculated by:

$$opinion\ distance = \ |Opinion_1 - Opinion_2|$$

As instance the opinion distance between two anti-polar contexts with -3 and +4 opinion scores equals 7 and it is the same as that of two other co-polar contexts with +1 and +8 opinion scores.

Calculating co-citation measures

The Co-CC textual similarity was calculated using cosine similarity in KNIME. The co-citedness and CPI values were, then, calculated using Citrec's sim_cocit and sim_cpa_simple measures, respectively.

The data were analyzed using SPSS 22. Given the non-normality of the data distributions, non-parametric partial correlation was used to analyze the association between the opinion and co-citation measures. The use of partial correlation helped to control for the effect of different polarities that were not taken into consideration in calculating opinion (dis)similarity.

Findings

As previously mentioned, partial correlation was used to test the correlation between the opinion (dis)similarity and co-citation similarity measures. It analyzes the correlation between variables at two order. The "zero order" and "control order" show the result of correlations, before and after controlling for the effects of the confounding variable, respectively.

In the present study, the opinion polarity stratification of cocited pairs are considered as confounding variable. As mentioned above, the polarity of the opinions was not taken into consideration when calculating their (dis)similarity. The verification of the polarities of the pairs indicated that there are co-polar, anti-polar, and mixed-polarity cocited pairs. As the polarity (dis)similarity groups may not be necessarily homogenous in the opinion strengths, their effects were controlled when studying the correlation between co-citation and co-opinion measures.

Correlation between co-citedness and opinion (dis)similarity measures

As seen in Table 1, co-citedness shows not to be significantly correlated to the opinion distance as measured by SWN (r= 0.018, Sig.= 0.209), while being significantly, directly correlated to the opinion intersection (r= 0.181, Sig.= 0.000). By controlling for the effects of the polarity similarity grouping, the correlation between the co-citedness and the opinion intersection tends to be stronger (r= 0.567, Sig.= 0.000). Besides, at the control level, the correlation between the co-citedness and the opinion distance turns to be significantly inverse (r= -0.202, Sig.= 0.000). The results also hold for the opinion (dis)similarity measures as calculated by SWs. The finding implies that the stronger the co-citations, the more similar the co-cited documents are in their opinions and vice versa. All of the relationships are estimated to be weak to medium.

Table 1
The partial correlation between co-citedness and the opinion (dis)similarity measures

Opinion tool	Order	Variables	Correlation	Sig
SWN	Zero	Opinion intersection	0.181	0.000
		Opinion distance	0.018	0.209
	Control	Opinion intersection	0.567	0.000
		Opinion distance	-0.202	0.000
SWs	Zero	Opinion intersection	0.051	0.000
		Opinion distance	-0.012	0.414
	Control	Opinion intersection	0.170	0.000
		Opinion distance	-0.081	0.000

Correlation between CPI and the opinion (dis)similarity measures

There are significant, though weak, correlations between CPI and opinion intersection (r= 0.099) and (r= 0.210) before and after controlling for the polarity similarity grouping, respectively. The correlation, which is significant, indirect for opinion distance before controlling the effect of the grouping (r= -0.090), is strengthened after applying the control (r= -0.163). It signifies that with increase in CPI, the opinion similarity increases, while the opinion distance decreases and vice versa. The same results are observed for the opinion measures calculated by SWs. All the correlations are estimated to be weak (Table 2).

Table 2
The partial correlation between CPI and the opinion (dis)similarity measures

Opinion tool	Order	Variables	Correlation	Sig
SWN	Zero	Opinion intersection	0.099	0.000
		Opinion distance	-0.090	0.000
	Control	Opinion intersection	0.210	0.000
		Opinion distance	-0.163	0.000
SWs	Zero	Opinion intersection	0.057	0.000
		Opinion distance	-0.122	0.000
	Control	Opinion intersection	0.103	0.000
		Opinion distance	-0.153	0.000

Correlation between Co-CC cosine similarity and the opinion (dis)similarity measures

The Co-CC cosine similarity shows to have significant association with the proposed co-opinion measure, so that the more similar the co-cited articles in Co-CCs, the higher is their opinion intersection, whether before (r= 0.135) or after (r= 0.299) controlling for the effect of polarity grouping. Their inverse association to the opinion distance also confirms the fact. As seen in table 3, the opinion distance significantly reduces, if the similarity of the co-cited articles in their Co-CCs increases and vice versa (r= -0.049 and r= -0.149 before and after the control, respectively). All the correlations are weak. The same results are also observed for the opinion measures calculated by SWs (Table 3).

Table 3
The partial correlation between Co-CC cosine similarity and the opinion (dis)similarity measures

Opinion tool	Order	Variables	Correlation	Sig
SWN	Zero	Opinion intersection	0.135	0.000
		Opinion distance	-0.049	0.001
	Control	Opinion intersection	0.299	0.000
		Opinion distance	-0.149	0.000
SWs	Zero	Opinion intersection	0.090	0.000
		Opinion distance	-0.116	0.000
	Control	Opinion intersection	0.135	0.000
		Opinion distance	-0.143	0.000

Discussion and conclusion

Co-citedness is among quantitative co-citation-based measures that has been found to be effective in retrieving relevant documents (Egghe & Rousseau, 1990; Bichteler & Eaton, 1980; Badran, 1984; Zhao, 2014). However, due to its purely quantitative nature, its effectiveness is estimated to be relatively low and needs to be improved by content-based elements (Gipp & Beel, 2009; Eto, 2013; Jeong et al., 2014; Elkiss et al., 2008; Liu et al., 2014). CPI is a content-based co-citation measure useful in retrieving relevant documents (Gipp & Beel, 2009; Callahan et al., 2010; Eto, 2012; 2013; 2014; 2015; Boyack et al., 2013; Liu, Chen, Ding, Wang, Xu, & Lin, 2014). Textual similarity of Co-CCs has been also proposed to delve the content similarity of co-cited documents (Jeong et al., 2014). Opinion mining of CCs has been found helpful in improving retrieval and evaluation results (Parthasarathy & Tomar 2014; Athar & Teufel, 2012; Small, 2011; Cavalcanti et al., 2011; Sendhilkumar et al., 2013; Amadi, 2014; Piao, Ananiadou, Tsuruoka, Sasaki & McNaught, 2007). Because, it can discriminate between papers based on the opinion they receive from their citers (Abu-Jbara et al., 2013; Parthasarathy & Tomar, 2014; 2015; Amadi, 2014). The co-cited articles are expected to show some degrees of similarity in the opinions they receive. Because, authors tend to positively cite other articles (Mahalakshmi et al., 2015). Given the prevalence of the conformational behavior, they are expected to show similar attitudes towards the papers they cite. On this basis, the co-opinion concept has been proposed and shown to be effective in information retrieval (Yaghtin et al., under review).

The results of the present study confirm the existence of significant, though not strong, correlations between traditional co-citation indicators and the newly proposed co-opinion measure. Co-opinion was measured by opinion intersection, i.e. the least opinion degree shared by the co-cited articles. Obviously, it is not expected that an absolute co-opinion dominates the co-citation realm. As a result, it is also expected to observe some kind of disagreement among the co-citers reflected in their opinions in the Co-CCs. Opinion distance was, thus, proposed and tested as a proxy of authors' disagreement. The results showed that, the two measures are significantly correlated to the well-known co-citation measures. However, the co-opinion exhibits a positive correlation, while the opinion distance indicates a negative association. The traditional co-citation metrics measure the similarity of the co-cited papers in their contents (measured by cosine similarity of Co-CCs), utility (measured by CPI) and importance (measured by co-citedness). Consequently, one may conclude that the more similar the co-cited articles are in their contents, utility and importance to co-citers, the more similar opinions they receive. In line with our previous findings (Yaghtin et al., Under review), the result confirms the existence of co-opinion phenomenon that inversely interacts with opinion distance in the co-citation environment. The measures can be, therefore, used to further discriminate between relevant papers by clustering them into similar and dissimilar opinion groups.

The significance of the correlations implies that the co-citation and co-opinion measures represent some similar dimensions of relevance and importance of documents from the viewpoints of their co-citers. However, the weakness of the correlations reveals that their similarity is weak. In other words, the traditional and the co-opinion measures widely differ in the dimensions they represent. Further studies are required to discover the dimensions represented by each of the measures. Previous research highlights the effectiveness of opinions explicitly or implicitly stated via CCs (Piao et al., 2007; Yu, 2013; Athar & Teufel, 2012; Abu-Jbara et al., 2013; Hernandez-Alvarez & Gomez, 2016; Parthasarathy & Tomar, 2014; 2015; Athar, 2014). Distinguishing negative and positive citations helps readers to compare different features of scholarly outputs (e.g. research results, designs, methodologies, arguments, etc.) and judge their scientific robustness (Eto, 2012). Co-opinion can be useful in further discriminating those documents that absorb similar or different opinions. Although, authors largely tend to give positive credits to their cited articles (MacRoberts & MacRoberts, 1984; Athar & Teufel, 2012; Mahalakshmi et al., 2015), negative opinions are also important for users to contrast the viewpoints and judge their strengths and weaknesses. Consequently, further analyses are required to explore how negatively co-polar, positively co-polar and anti-polar Co-CCs are interacting within the co-citation sphere.

Endnote

1. http://colil.dbcls.jp/browse/papers/

References

Abu-Jbara, A., Ezra, J., & Radev, D. R. (2013). Purpose and Polarity of Citation: Towards NLP-based Bibliometrics. In *HLT-NAACL* (pp. 596-606).

Agarwal, S., Choubey, L., & Yu, H. (2010, November). Automatically classifying the role of citations in biomedical articles. In *Proceedings of American Medical Informatics Association Fall Symposium (AMIA), Washington, DC* (pp. 11-15).

Amadi, U. P. (2014). *Exploiting the role of polarity in citation analysis*. University of Maryland, Baltimore County.

Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. In *Proceedings of the ACL 2011 student session* (pp. 81-87). Association for Computational Linguistics.

Athar, A. (2014). Sentiment analysis of scientific citations. *Technical Report, University of Cambridge, Computer Laboratory, (UCAM-CL-TR-856).*

Athar, A., & Teufel, S. (2012). Context-enhanced citation sentiment detection. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for*

Computational Linguistics: Human Language Technologies (pp. 597-601). Association for Computational Linguistics.

Badran, O. M. (1984, January). An alternative search strategy to improve information retrieval. In *Proceedings of the 47th ASIS annual meeting* (pp. 137-140).

Bichteler, J., & Eaton, E. A. (1980). The combined use of bibliographic coupling and co-citation for document retrieval. *Journal of the American Society for Information Science, 31*(4), 278-282.

Boyack, K. W., Small, H., & Klavans, R. (2013). Improving the accuracy of co-citation clustering using full text. *Journal of the American Society for Information Science and Technology*, 64(9), 1759-1767.

Callahan, A., Hockema, S., & Eysenbach, G. (2010). Contextual co-citation: Augmenting co-citation analysis and its applications. *Journal of the American Society for Information Science and Technology*, *61*(6), 1130-1143.

Cavalcanti, D. C., Prudêncio, R. B., Pradhan, S. S., Shah, J. Y., & Pietrobon, R. S. (2011,). Good to be bad? Distinguishing between positive and negative citations in scientific impact. In *Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on* (pp. 156-162). IEEE.

Cole, S., & Cole, J. R. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American sociological review*, 32(3), 377-390.

Dabrowska, A., & Larsen, B. (2015). Exploiting citation contexts for physics retrieval. In *Second Workshop on Bibliometric-enhanced Information Retrieval* (pp. 14-21).

De Bellis, N. (2009). Bibliometrics and citation analysis from the science citation index to cybermetrics. Lanham, Md: Scarecrow Press.

Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the association for information science and technology*, *65*(9), 1820-1833.

Dong, C., & Schäfer, U. (2011). Ensemble-style Self-training on Citation Classification. In *IJCNLP* (pp. 623-631).

Doslu, M., & Bingol, H. O. (2016). Context sensitive article ranking with citation context analysis. *Scientometrics*, 108(2), 653–671.

Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Elsevier, North-Holland, Amsterdam, The Netherlands

Elkiss, A. et al. (2008). Blind men and elephants: What do citation summaries tell us about a research article?. *Journal of the American Society for Information Science and Technology*, *59*(1), 51-62.

Esuli, A., & Sebastiani, F. (2007). SWN: a high-coverage lexical resource for opinion mining. *Evaluation*, 1-26.

Eto, M. (2012). Spread co-citation relationship as a measure for document retrieval. In *Proceedings of the fifth ACM workshop on Research advances in large digital book repositories and complementary media* (pp. 7-8). ACM.

Eto, M. (2013). Evaluations of context-based co-citation searching. *Scientometrics*, *94*(2), 651-673.

Eto, M. (2014). Document retrieval method using random walk with restart on weighted co-citation network. *Proceedings of the American Society for Information Science and*

Technology, *51*(1), 1-4.

Eto, M. (2015). Combination Effects of Word-based and Extended Co-citation Search Algorithms. In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 245-246). ACM.

Fujiwara, T., & Yamamoto, Y. (2015). Colil: a database and search service for citation contexts in the life sciences domain. *Journal of biomedical semantics*, 6(1), 38.

Gatti, L., Guerini, M., & Turchi, M. (2016). SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis. *IEEE Transactions on Affective Computing*, *7*(4), 409-421.

Gipp, B., & Beel, J. (2009, July). Citation Proximity Analysis (CPA)-A new approach for identifying related work based on Co-Citation Analysis. In *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)* (Vol. 2, pp. 571-575). Rio de Janeiro (Brazil): International Society for Scientometrics and Informetrics.

Gurulingappa, H., Müller, B., Klinger, R., Mevissen, H. T., Hofmann-Apitius, M., Friedrich, C. M., & Fluck, J. (2010). Prior Art Search in Chemistry Patents Based On Semantic Concepts and Co-Citation Analysis. In TREC.

Hamedani, M. R., Kim, S. W., & Kim, D. J. (2016). SimCC: A novel method to consider both content and citations for computing similarity of scientific papers. *Information Sciences*, *334*, 273-292.

Hernández-alvarez, M. Y. R. I. A. M., & GOMEZ, J. M. (2016). Survey about citation context analysis: Tasks, techniques, and resources. *Natural Language Engineering*, *22*(03), 327-349.

Janssens, A. C. J., & Gwinn, M. (2015). Novel citation-based search method for scientific literature: application to meta-analyses. *BMC medical research methodology*, *15*(1), 84.

Jeong, Y. K., Song, M., & Ding, Y. (2014). Content-based author co-citation analysis. *Journal of Informetrics*, *8*(1), 197-211.

Jochim, C., & Schütze, H. (2012). Towards a generic and flexible citation classifier based on a faceted classification scheme. Proceedings of COLING 2012, 1343-1358.

Liu, S., Chen, C., Ding, K., Wang, B., Xu, K., & Lin, Y. (2014). Literature retrieval based on citation context. *Scientometrics*, 101(2), 1293-1307.

Liu B., Zhang L. (2012) A Survey of Opinion Mining and Sentiment Analysis. In: Aggarwal C., Zhai C. (eds) Mining Text Data. Springer, Boston, MA

Lundberg, J. (2006). Bibliometrics as a research assessment tool: impact beyond the impact factor. Institutionen för lärande, informatik, management och etik, LIME/Department of Learning, Informatics, Management and Ethics (Lime).

MacRoberts, M.H., & MacRoberts, B.R. (2010). Problems of Citation Analysis: A Study of Uncited and Seldom-Cited Influences. Journal of the American Society for Information Science and Technology, 21(1), 1–13.

MacRoberts, M. H., & MacRoberts, B. R. (1984). The negational reference: Or the art of dissembling. *Social Studies of Science*, *14*(1), 91-94.

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for information Science*, *40*(5), 342-349.

Mahalakshmi, G. S., Siva, R., & Sendhilkumar, S. (2015). Context Based Retrieval of Scientific Publications via Reader Lens. In *Computational Intelligence in Data Mining-Volume 3* (pp. 583-596). Springer India.

Martyn, J. (1964). Bibliographic coupling. *Journal of Documentation*, *20*(4), 236-236.

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: towards automatic review article generation. In *The American Society for Information Science/the 11th SIC Classification Research Workshop* (pp. 117-134).

Parthasarathy, G., & Tomar, D. C. (2014). Sentiment analyzer: Analysis of journal citations from citation databases. In *Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference-* (pp. 923-928). IEEE.

Parthasarathy, G., & Tomar, D. C. (2015). A Survey of Sentiment Analysis for Journal Citation. *Indian Journal of Science and Technology*, *8*(35), 1-8.

Piao, S., Ananiadou, S., Tsuruoka, Y., Sasaki, Y., & McNaught, J. (2007, January). Mining opinion polarity relations of citations. In *International Workshop on Computational Semantics (IWCS)* (pp. 366-371).

Ritchie, A., Robertson, S., & Teufel, S. (2008). Comparing citation contexts for information retrieval. In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 213-222). ACM.

Ritchie, A. (2009). *Citation context analysis for information retrieval*. Doctoral dissertation, University of Cambridge.

Rousseau, R. (2007). Publication and citation analysis as a tool for information retrieval. *Social Information Retrieval Systems. Emerging technologies and applications for searching the web effectively (Goh D & Foo S, eds.). Hershey (PA)*, 252-267.

Schafer, U., & Spurk, C. (2010). TAKE scientist's workbench: semantic search and citation-based visual navigation in scholar papers. In *Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on* (pp. 317-324). IEEE.

Sendhilkumar, S., Elakkiya, E., & Mahalakshmi, G. S. (2013). Citation semantic based approaches to identify article quality. In *Proceedings of International conference ICCSEA* (pp. 411-420).

Shuy, R. W. (2005). Discourse analysis in the legal context. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The handbook of discourse analysis*. Oxford: Blackwell

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science*, 24(4), 265-269.

Small, H. (1982). Citation context analysis. *Progress in communication sciences*, 3, 287-310.

Small, H. (2011). Interpreting maps of science using citation context sentiments: a preliminary investigation. *Scientometrics*, *87*(2), 373-388.

Smith, L. C. (1981). Citation Analysis. *Library Trends*, 30(1), 83-106.

Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation function. In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 103-110). Association for Computational Linguistics.

Teufel, S., Siddharthan, A., & Tidhar, D. (2009). An annotation scheme for citation function. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue* (pp. 80-87). Association for Computational Linguistics.

Verbeek, A., Debackere, K., Luwel, M., & Zimmermann, E. (2002). Measuring progress and evolution in science and technology–I: The multiple uses of bibliometric indicators. *international Journal of management reviews*, *4*(2), 179-211.

White, H. D. (2016). Bag of Works Retrieval: TF* IDF Weighting of Co-cited Works. In *BIR@ ECIR* (pp. 63-72).

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y. & Patwardhan, S. (2005, October). OpinionFinder: A system for subjectivity analysis. In *Proceedings of hlt/emnlp on interactive demonstrations* (pp. 34-35). Association for Computational Linguistics.

Yoon, S. H., Kim, S. W., & Park, S. (2016). C-Rank: A link-based similarity measure for scientific literature databases. *Information Sciences*, 326, 25-40.

Yaghtin, M., Sotudeh, H., Mirzabeigi, M., Fakhrahmad, S.M., Mohammadi, M. (Under review). In quest of new document relations: evaluating co-opinion relations between co-citations and its impact on Information Retrieval effectiveness. *Scientometrics*. (under review)

Yu, B. (2013). Automated citation sentiment analysis: what can we learn from biomedical researchers? Proceedings *of the American Society for Information Science and Technology*, *50*(1), 1-9.

Zhao, H. (2014). Sharding for literature search via cutting citation graphs. In *Big Data (Big Data), 2014 IEEE International Conference on* (pp. 77-79). IEEE.

Zhao, H. (2015). *The Role of Document Structure and Citation Analysis in Literature Information Retrieval*. DREXEL UNIVERSITY.