To be published at iConference 2017

Evaluating the CC-IDF citation-weighting scheme: How effectively can 'Inverse Document Frequency' (IDF) be applied to references?

Joeran Beel[1, 2], Corinna Breitinger[3], Stefan Langer[4]

[1] National Institute of Informatics Tokyo, Digital Content and Media Sciences Division, Japan
[2] Trinity College Dublin, School of Computer Science & Statistics, ADAPT Centre, Ireland
[3] University of Konstanz, Department of Computer and Information Science, Germany
[4] Otto-von-Guericke University Magdeburg, Department of Computer Science, Germany

Abstract
In the domain of academic search engines and research-paper recommender systems, CC-IDF is a common citation-weighting scheme that is used to calculate semantic relatedness between documents. CC-IDF adopts the principles of the popular term-weighting scheme TF-IDF and assumes that if a rare academic citation is shared by two documents then this occurrence should receive a higher weight than if the citation is shared among a large number of documents. Although CC-IDF is in common use, we found no empirical evaluation and comparison of CC-IDF with plain citation weight (CC-Only). Therefore, we conducted such an evaluation and present the results in this paper. The evaluation was conducted with real users of the recommender system *Docear*. The effectiveness of CC-IDF and CC-Only was measured using click-through rate (CTR). For 238,681 delivered recommendations, CC-IDF had about the same effectiveness as CC-Only (CTR of 6.15% vs. 6.23%). In other words, CC-IDF was not more effective than CC-Only, which is a surprising result. We provide a number of potential reasons and suggest to conduct further research to understand the principles of CC-IDF in more detail.

Keywords: recommender systems; cc-idf; digital libraries; weighting schemes; tf-idf; related document search
Citation: Editor will add citation

Acknowledgements: This work was supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD). Additional support came in the form of a Doctoral Stipend of the Carl-Zeiss Foundation. This publication also has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 13/RC/2106.
Contact: joeran.beel@adaptcentre.ie / http://beel.org

1 Introduction

The citation-weighting scheme CC-IDF[1] was introduced in 1998 in the digital library and citation-indexing system *CiteSeer* (Bollacker, Lawrence, & Giles, 1998; Giles, Bollacker, & Lawrence, 1998). CiteSeer offered a link for retrieving a list of "related documents" beside each search result, and the list of related documents was calculated, among others, using CC-IDF. CC-IDF stands for "Common Citation-Inverse Document Frequency" and it consists namely of the common citation frequency (CC) for a citation and the inverse frequency of documents in a corpus containing that citation[2] (IDF). Using IDF to weight citations was a novel concept at that time and was inspired by TF-IDF, one of the most popular text-weighting schemes in information retrieval (Jones, 1972; Salton, Wong, & Yang, 1975)[3]. The assumption of IDF when applied to citations is that "if a very uncommon citation is shared by two documents, this should be weighted more highly than a citation made by a large number of documents" (Giles et al., 1998). However, there is a difference between TF-IDF and the traditional CC-IDF measure. In TF-IDF, the term frequency TF expresses how often a term occurs in a particular document. In contrast, CC is a binary measure, which only specifies if a document contains (1) or does not contain (0) a reference.

Figure 1 illustrates the rationale underlying CC-IDF. For a given input document d_i, a list of related documents must be identified. All documents that share at least one reference with d_i are considered potentially related, a concept also known as bibliographic coupling (BC). In the example, the bibliographically coupled documents are d_{BC}^1, d_{BC}^2, d_{BC}^3, and d_{BC}^4. According to CC-IDF, d_{BC}^1 and d_{BC}^2 are the least related documents to d_i, because they each share only one reference (d_{cited}^1) with d_i and this

[1] Also called CCIDF, CCxIDF, CC*IDF, CC–IDF, and CC×IDF
[2] Note that we will use the terms "citation" and "reference" interchangeably in this paper.
[3] We assume the reader to be familiar with the concept of TF-IDF and do not explain it in this paper.

reference is cited in total by three documents in the corpus (d_{BC}^1, d_{BC}^2 and d_{BC}^3)[4]. Hence, for d_{BC}^1 and d_{BC}^2 CC-IDF calculates as $CC \times IDF(d_i, d_{BC}^{1|2}) = {}^1/_3$. In contrast, d_{BC}^4 also shares a single reference (d_{cited}^2) with d_i, but this reference is only cited twice in the corpus (namely by d_{BC}^3 and d_{BC}^4). Hence, $CC \times IDF(d_i, d_{BC}^4) = {}^1/_2$ and d_{BC}^4 is regarded as more closely related to d_i than d_{BC}^1 and d_{BC}^2. In Figure 1, for all documents in the collection, document d_{BC}^3 is the most closely related to the input document d_i, because they share the two references d_{cited}^1 and d_{cited}^2. CC-IDF sums up the individual relatedness values, hence $CC \times IDF(d_i, d_{BC}^3) = {}^1/_2 + {}^1/_3 = {}^5/_6$.

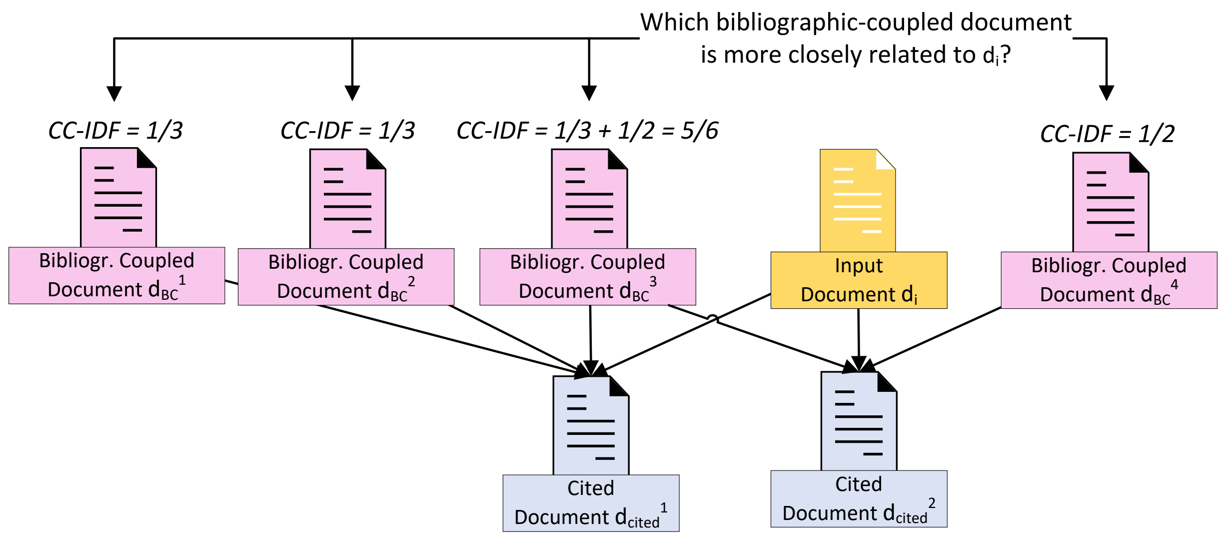


Figure 1: Illustration of CC-IDF

Since 1998, CC-IDF has been used in several recommender systems, and served as a baseline in many evaluations. Furthermore, CC-IDF is mentioned by researchers as a standard approach for calculating document relatedness using citations (Chakraborty, Modani, Narayanam, & Nagar, 2015; Ekstrand et al., 2010; Huynh & Hoang, 2012; Huynh et al., 2012; Küçüktunç, Saule, Kaya, & Çatalyürek, 2013; Liang, Li, & Qian, 2011; Narwekar, 2016; Pan, Dai, Huang, & Chen, 2015; Zhang, Li, Zhang, & Wang, 2012). However, there are ambiguous reports regarding the effectiveness of CC-IDF. For instance, sometimes, CC-IDF was found to perform better and other times worse than simple bibliographic coupling and co-citation strength (Küçüktunç, Saule, Kaya, & Çatalyürek, 2012; Küçüktunç et al., 2013; Liang et al., 2011; Pan et al., 2015; Zhang et al., 2012). Compared to more advanced approaches such as HITS, PaperRank, and Katz, CC-IDF performs usually poorly (Küçüktunç et al., 2012; Pan et al., 2015).

To the best of our knowledge, CC-IDF has never been compared to CC-Only, i.e. a simple citation weighting scheme based only on the CC component and ignoring IDF. This means, the basic assumption underlying CC-IDF – namely that “if a very uncommon citation is shared by two documents, this should be weighted more highly than a citation made by a large number of documents” – has never been evaluated for its effectiveness. Of course, the assumption seems plausible, and for terms the effectiveness of IDF has been shown multiple times (Robertson, 2004). However, the absence of empirical evidence on the rationale of IDF motivated us to assess its suitability when applied to references[5].

2 Related Work

To find related documents for a given input document using citations, four assumptions are generally made (cf. Figure 2). First, documents that cite an input document can be considered related. Second, documents that are being cited by an input document can be considered related. Third, documents that are co-cited can be considered related, i.e. documents being cited in the same documents that cite the input document. Finally, documents that cite the same documents as the input document can be considered related, i.e. documents containing the same entries in their bibliography as the input document (bibliographic coupling).

[4] If the input document was considered to be part of the corpus, the number of documents would be four instead of three. However, for calculating document relatedness using CC-IDF it does not matter if the input document is counted or not.
[5] An evaluation of CC-IDF was previously conducted in the PhD thesis of Beel (2015); However, the current paper represents the first peer-reviewed publication and the first detailed discussion of the evaluation.

Beyond literature search and recommender systems, a third practical application of calculating document relatedness based on citations lies in the field of academic plagiarism detection (Gipp, Meuschke, & Breitinger, 2014).

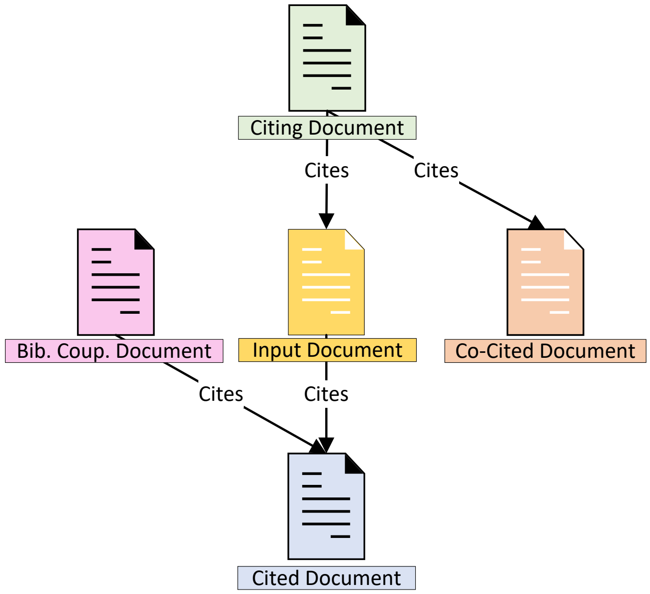


Figure 2: Types of document relations in citation analysis

Naturally, absolute citation counts are the simplest measure for calculating document relatedness. For instance, the more references two documents share in their bibliography, the higher their "bibliographic coupling strength", and thus their relatedness. Similarly, the more frequently two documents are co-cited together in other documents, the stronger their co-citation strength. However, there are more sophisticated relatedness measures, several of which we will briefly present in the following sections.

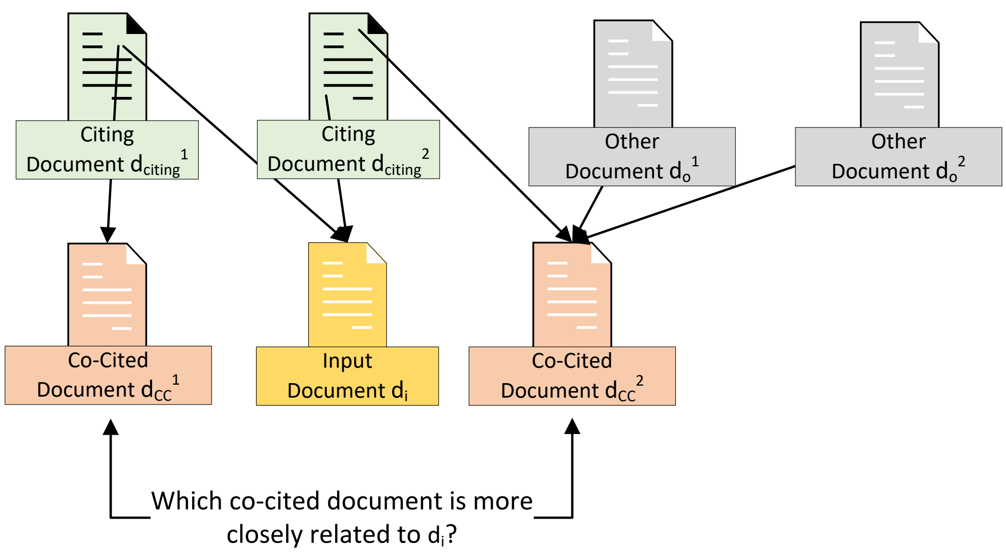


Figure 3: Document relatedness using co-citation

2.1 Relatedness using Co-Citations

Assume that an input document d_i is cited by two documents d_{citing}^1 and d_{citing}^2 (cf. Figure 3) Each of the two documents also cites one more document, namely d_{CC}^1 and d_{CC}^2. The co-citation strength of d_{CC}^1 and d_i as well as of d_{CC}^2 and d_i is 1 because they are each co-cited one time. The question that arises is which of the two documents is more closely related to d_i. There are various approaches to answer this question. Among the oldest is “relative co-citation strength”, which was introduced by Small (1973). The relative co-citation strength divides the absolute co-citation strength by the number of all cited papers. The relative co-citation strength of d_i and d_{CC}^1 in Figure 3 is $1/_1$, because d_i and d_{CC}^1 are co-cited once, and in total the co-cited document d_{CC}^1 is cited only once in the document corpus[6]. In comparison, for d_i and d_{CC}^2 the relative co-citation strength is $1/_3$ because d_{CC}^2 is cited in total three times by the documents of the corpus. This concept of relative co-citation strength corresponds to the idea of IDF.

A more recently proposed alternative to relative co-citation strength is co-citation proximity analysis (CPA), which uses a co-citation proximity index (Gipp & Beel, 2009). The index expresses the proximity at which two documents are cited within a paper. Figure 3 illustrate how d_i and d_{CC}^1 are cited by d_{citing}^1 in close proximity, i.e. in the same sentence. Hence, d_i and d_{CC}^1 are considered closely related. In contrast, d_i and d_{CC}^2 are cited by d_{citing}^2 in less close proximity, i.e. in different paragraphs. Hence, d_i and d_{CC}^2 are considered less closely related. Variants of the CPA approach, and an overview of additional citation-based measures are described by Gipp (2014, p. 47). Beyond academic citations alone, co-citation proximity analysis has also been demonstrated as suitable when applied to links, for example, to generate literature recommendations for related Wikipedia articles (Schwarzer et al., 2016).

2.2 Relatedness using “Cited” relations

Assume that an input document d_i cites two documents d_{cited}^1 and d_{cited}^2 (cf. Figure 4). To calculate document relatedness between d_i and the cited documents, the frequency of in-text citations can be used as a weight (Gipp, Beel, & Hentschel, 2009). In Figure 4, d_{cited}^1 is cited three times in the body-text of d_i, while d_{cited}^2 is cited only once. Hence, d_{cited}^1 is considered more related to d_i than d_{cited}^2. Another approach includes considering how often a document is cited overall, and to then decrease the weight of highly cited papers. In the example, d_{cited}^1 is only cited by d_i, while d_{cited}^2 is also cited by two other documents $d_o^{1,2}$. Hence, d_{cited}^1 is assumed to be more closely related to d_i than d_{cited}^2.

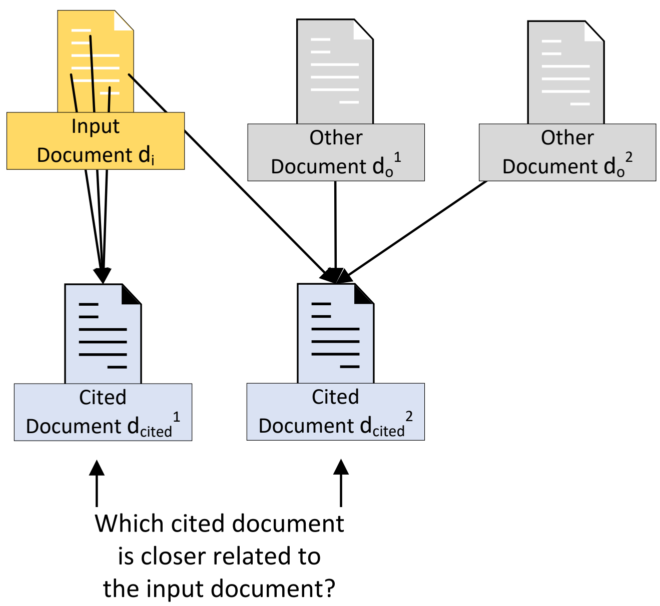


Figure 4: Document relatedness using “cited” relations

2.3 Relatedness using Bibliographic Coupling

We explained bibliographic coupling in the introduction and in Figure 1. However, there are additional variations. In Figure 5, all four documents $d_{BC}^{1 \ldots 4}$ share one reference with d_i. Hence, the absolute bibliographic coupling strength between $d_{BC}^{1 \ldots 4}$ and d_i is always 1. One option for calculating a relative bibliographic coupling strength is to analyze what percentage of the bibliographies of two documents overlap. In the example in Figure 4, d_i and d_{BC}^4 have one reference in common (d_{cited}^2), but d_{BC}^4 cites two additional documents (d_o^1 and d_o^2). This means, d_i shares only 1/3 of the references with d_{BC}^4. In contrast,

[6] We regard the input document as external to the document corpus. If it was part of the document corpus, all counts would increase by one.

the documents $d_{BC}^{1...3}$ all cite only a single document (d_{cited}^{1}). This means, d_i shares 100% of its references with $d_{BC}^{1...3}$. Consequently, according to relative bibliographic coupling strength, $d_{BC}^{1...3}$ could be considered more related to d_i than d_{BC}^{4}.

We would like to emphasize that this type of relative bibliographic strength may lead to different results for document-relatedness than CC-IDF. With CC-IDF, $d_{BC}^{1...3}$ would be considered less related to d_i than d_{BC}^{4}, because d_{BC}^{4} and d_i share a rarely cited reference (d_{cited}^{2} is cited only once), while $d_{BC}^{1...3}$ and d_i share the reference d_{cited}^{1}, which is cited three times.

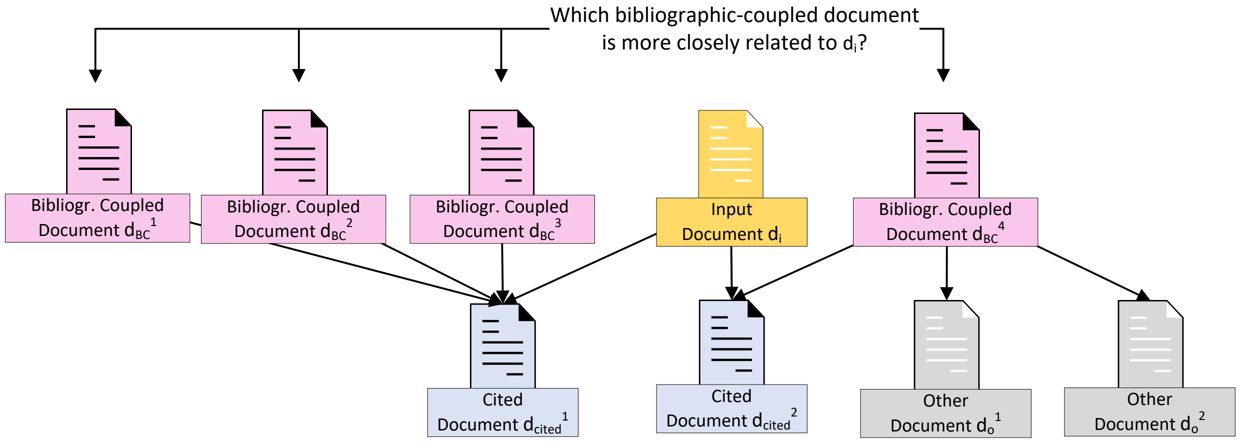


Figure 5: Document relatedness using bibliographic coupling

3 Methodology

To evaluate the effectiveness of IDF applied to citations, we compared the effectiveness of CC-IDF with CC-Only. The evaluation was conducted using the recommender system of the reference-management software *Docear* (Beel, Gipp, Langer, & Genzmehr, 2011; Beel, Gipp, & Mueller, 2009; Beel, Langer, Gipp, & Nürnberger, 2014; Beel, Langer, Genzmehr, & Nürnberger, 2013). Docear is comparable to the tools JabRef, Zotero and Mendeley, which enable users to organize their references and PDF files (typically research articles, and occasionally other resources, such as websites). A unique feature of Docear is that the collections are not simply lists of references and PDF files, but are structured as mind-maps into which users can insert references or link PDF files (Figure 6). For our current research, this distinction is not of importance, since we only require a large number of users, each of whom has one or multiple collections (i.e. mind-maps) with a number of references and PDF files.

Compared to the original CC-IDF approach, we implemented some changes to make the approach applicable to our scenario. In the original CC-IDF approach, there is one input document for which a list of related documents is wanted, and related documents are found via bibliographic coupling with CC-IDF weighting. We utilized a user's collection of mind-maps as input (instead of a single research paper), and we interpreted the link to, or reference of, a paper in a user's collection as a citation of that paper[7]. In addition, the original CC-IDF approach uses a binary weight for the CC component. We calculated CC as the frequency for how often a reference or link to a paper occurred in a user's collection. The identification and matching of papers was done only by comparing titles. In the case of PDF files, titles were extracted with Docear's PDF Inspector (Beel, Gipp, Shaker, & Friedrich, 2010; Beel, Langer, Genzmehr, & Müller, 2013).

Figure 7 illustrates the recommendation process. Similar to an input document d_i that references documents d_1 and d_2, a user has documents d_1, d_2, and many other documents in his or her collection. In the example (cf. Figure 7), the two most recently added documents, i.e. d_1, and d_2, are used to build the user's user model. The user model *um* equals a "joined" document that contains all the references from the selected documents, in this case, the user's collections of mind maps. The recommendations are displayed in Docear (Figure 8). Users were automatically shown new recommendations every few days and they could additionally request recommendations explicitly. For more details on Docear's recommender system please refer to Beel, Langer, Kapitsaki, Breitinger, & Gipp (2015), Beel (2015), Beel et al. (2014) and Langer & Beel (2014).

[7] More precisely, our recommender system only utilized a subset of the user's most recently added documents.

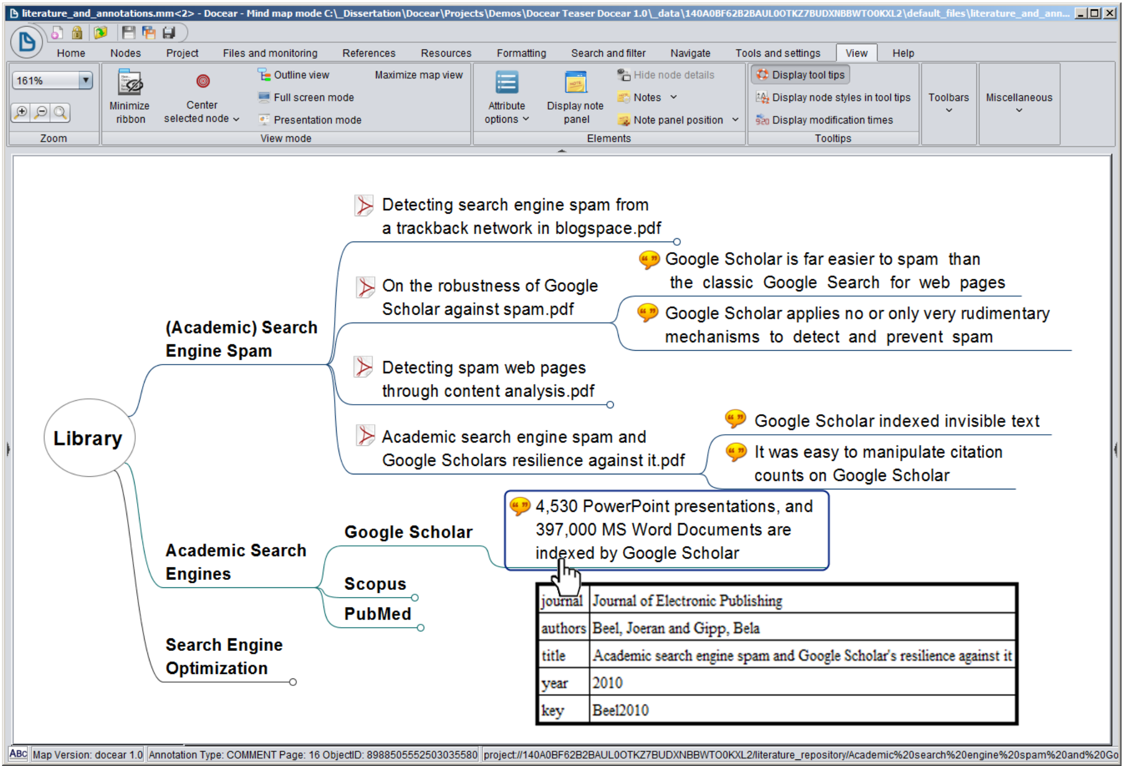


Figure 6: Screenshot of Docear

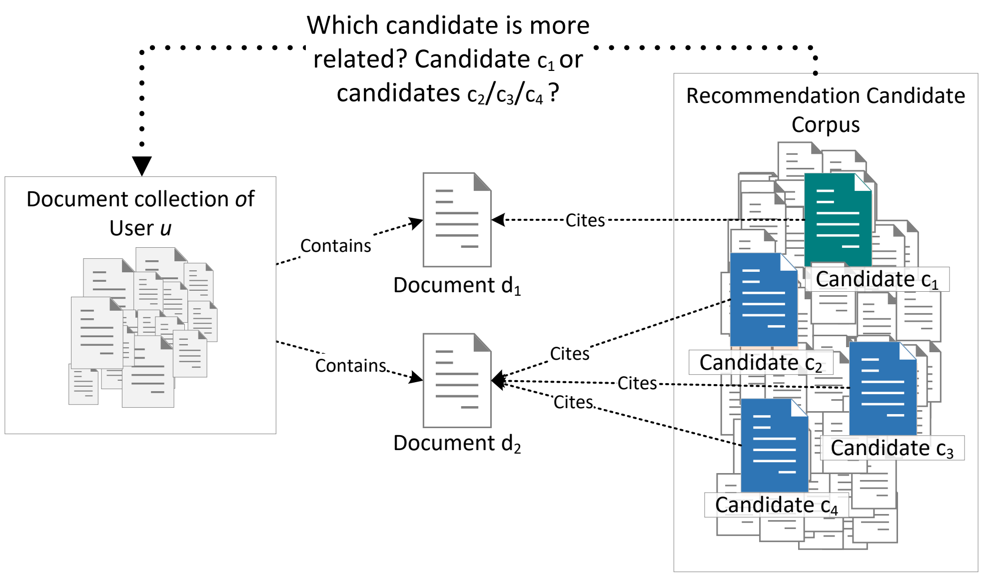


Figure 7: CC-IDF in the context of user modelling

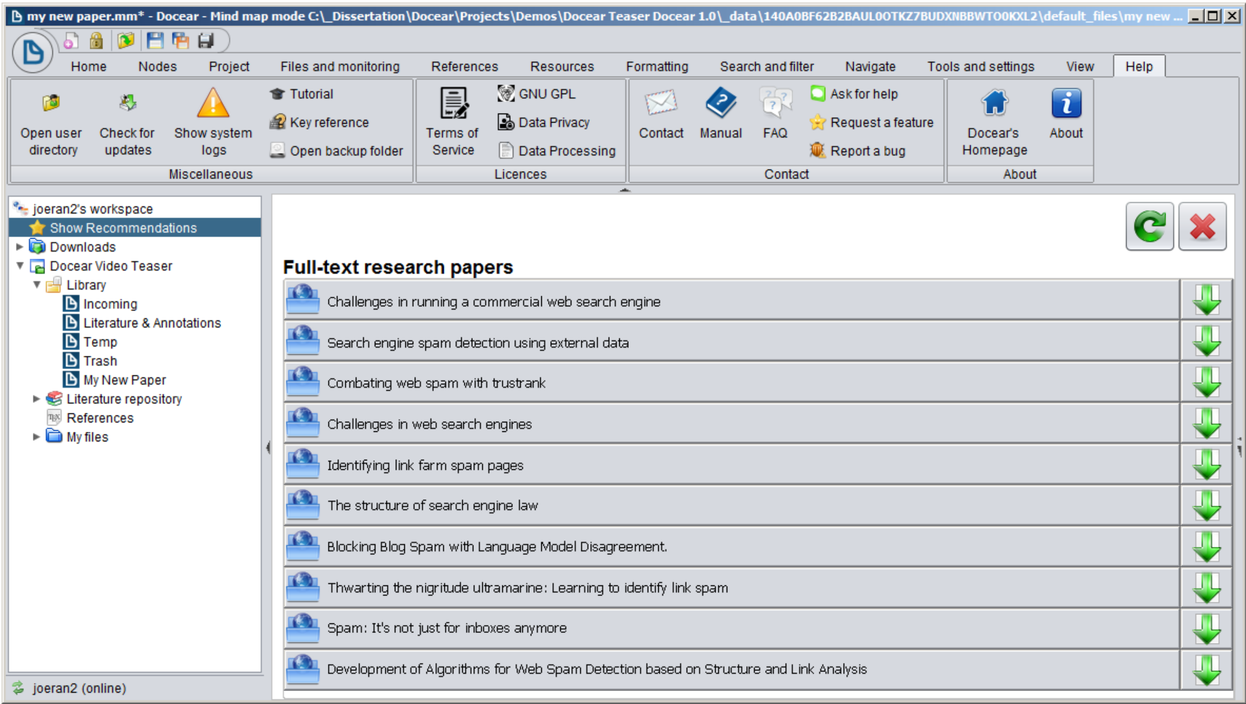


Figure 8: Recommendations in Docear

We evaluated the effectiveness of CC-IDF and CC-Only with an A/B Test. Whenever recommendations were generated, one of the two weighting schemes was randomly chosen, and the click-through rate was recorded (CTR). CTR describes the ratio of displayed recommendations to clicked recommendations. For instance, when 10,000 recommendations using CC-IDF were made and 500 of these recommendations were clicked, the average CTR of CC-IDF would be $\frac{500}{10,000} = 5\%$. The assumption is that the higher the CTR, the more effective the weighting scheme. There is some discussion to what extend CTR is appropriate for measuring recommendation effectiveness, but we found CTR to be well suitable for our scenario, because we found that it correlates well with user ratings (Beel, Breitinger, Langer, Lommatzsch, & Gipp, 2016; Beel & Langer, 2015). As an additional baseline, we measured the effectiveness of classic TF-IDF and TF-only. In this assessment, the terms from a user's document collection were utilized instead of the references. Between January 2014 and September 2014, 238,681 recommendations were delivered to 3,561 users. Unless stated otherwise, all results are statistically significant based on a two-tailed t-test ($p<0.05$).

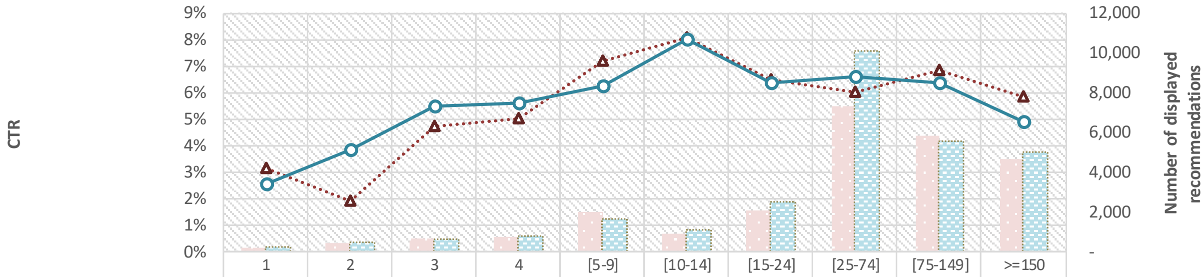
4 Results & Discussion

As expected, TF-IDF (CTR = 5.09%) performed significantly better than TF-Only (4.06%) (Table 1). This confirms the well-known finding that TF-IDF is superior over TF-only as a weighting scheme.

However, there was no statistically significant difference between CC-Only (CTR = 6.23%) and CC-IDF (6.15%) (Table 1). The result remains the same when looking at different numbers of references being utilized (Figure 9). The effectiveness of CC-IDF and CC-Only is about the same. For instance, when a user model contained 15 to 24 references, CTR for CC-Only was 6.50% and for CC-IDF 6.35%.

	CC-Only	CC-IDF	TF-Only	TF-IDF
Delivered	24,821	27,986	139,474	46,400
Clicks	1,546	1,721	5,665	2,361
CTR	6.23%	6.15%	4.06%	5.09%

Table 1. Number of delivered recommendations, clicks, and CTR for the different weighting schemes



	1	2	3	4	[5-9]	[10-14]	[15-24]	[25-74]	[75-149]	>=150
CC-Only (Dspld Recs)	180	451	667	713	1,964	896	2,098	7,329	5,848	4,675
CC-IDF (Dspld Recs)	252	496	600	770	1,660	1,075	2,484	10,114	5,555	4,979
CC-Only (CTR)	3.13%	1.93%	4.74%	5.00%	7.21%	8.08%	6.50%	6.00%	6.83%	5.87%
CC-IDF (CTR)	2.58%	3.86%	5.51%	5.62%	6.28%	8.03%	6.35%	6.64%	6.38%	4.92%

Number of utilized references

Figure 9: CTR for CC-IDF and CC-Only based on the number of utilized references

From the observed results, we would conclude that CC-IDF and CC-Only are equally effective, i.e. calculating IDF does not increase effectiveness compared to using CC-Only. Consequently, there would be little reason to use CC-IDF, because it is more complex to calculate than CC-Only. However, it is too early to draw such general conclusions from our results for the following reasons:

1. CC-IDF is usually applied in the context of related-document search. We applied it in the context of user-modelling. Although, we believe that this should not make a significant difference, we suggest to conduct additional research in a classic related-document scenario.
2. The document corpus of Docear is rather small (2 million documents). We could imagine that CC-IDF performs better on larger corpora. Consequently, we suggest to research the effectiveness of CC-IDF on a larger corpus.
3. Many users of Docear have only few references in their collection. It might be interesting to analyze how CC-IDF performs with users who have larger document collections with many references.
4. We used ParsCit to extract references from the recommendation candidates (Councill, Giles, & Kan, 2008). ParsCit has a reasonable, but not an outstanding accuracy. Hence, our reference data might be noisy and of mediocre suitability for calculating IDF values. We suggest performing further evaluations with reference data of higher quality.
5. We did not use a binary weighting for the CC component. Although we believe that this should not significantly affect the effectiveness of IDF, it might be sensible to nonetheless repeat our experiment with a binary CC component.

Despite the limitations of our research, there are a number of reasons why CC-IDF might indeed not be a significant improvement over CC-Only. Please note that the following hypotheses are still speculative, and that more research will be required in order to confirm or reject each assumption.

1. Research papers usually contain thousands of unique terms. Consequently, it is important to identify the most descriptive terms. In contrast, a research paper usually contains few citations (maybe 5 or 10 for conference papers, or 30 for journal article, although this number can differ widely depending on the discipline). Consequently, the need – and the potential benefit – of identifying the most important citations is lower, because likely almost all references in an article will have some significance.
2. In a large corpus, some terms occur in millions of documents. In contrast, even the world's most frequently occurring reference occurs only in 305,000 citing documents[8]; and the vast majority of references occurs only in few documents, because typically research papers receive few citations (or none at all). Consequently, IDF values for citations will be within a

[8] http://www.nature.com/news/the-top-100-papers-1.16224

smaller range than term-based IDF values. Therefore, we would expect IDF when applied to references to be less effective than IDF when applied to terms.

3. Older papers have more time to accumulate citations, while recently published papers typically have few or no citations. CC-IDF does not account for this, which could bias IDF calculations[9]. For instance, consider the previous example of bibliographic coupling and CC-IDF (cf. section 2.3), but this time assume that d_{cited}^1 was published in 1982, and d_{cited}^2 was published in 2016 (Figure 10). CC-IDF would be 1/3 for $d_{BC}^{1\ldots3}$ and 1 for d_{BC}^4. However, given the publication years, it would be expected that d_{cited}^1 has more citations than d_{cited}^2, and we intuitively would not believe, for instance, that d_{BC}^3 is less related to d_i than d_{BC}^4. We therefore suggest to analyze how CC-IDF performs when normalized by the documents' publication years.
4. CC-IDF does not normalize for the number of entries in a bibliography and may provide different recommendations than a classic relative bibliographic-coupling strength (see section 2.3). In future research, we suggest comparing CC-IDF with relative bibliographic coupling strength and also to evaluate the effectiveness of a CC-IDF measure that normalizes for the number of entries in a bibliography.
5. CC-IDF favors recommendation candidates that reference rarely cited papers over candidates that reference highly cited-papers. Maybe, papers that reference rarely cited papers tend to be of a different type than papers that reference highly cited papers, and maybe the latter type is more suitable for recommendation. For instance, we could imagine that papers with few citations might have a higher proportion of self-citations or citations from co-authors than highly cited papers (again, this is a speculative assumption to be examined). However, recommending a paper to a user, which the user or a co-author authored is probably not suitable, because the user already knows this paper. If this assumption were to be true, it would be interesting to analyze the performance of CC-IDF when self-citations were ignored in the calculations.

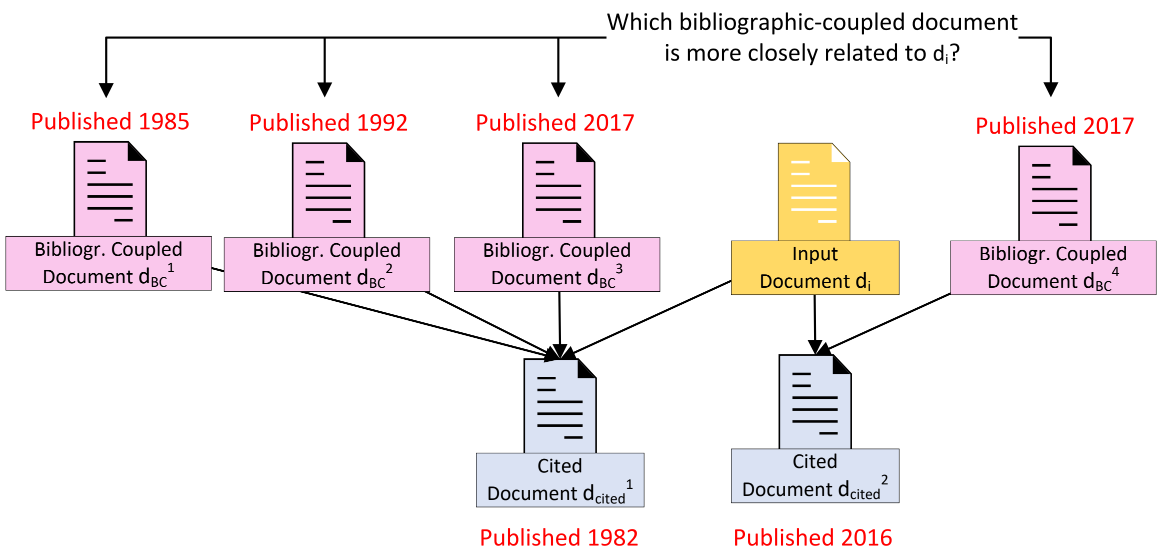


Figure 10: Illustration of a normalized CC-IDF measure

In summary, we were surprised to discover an equal performance of CC-IDF and CC-Only in our evaluation. Although we provided some arguments why CC-IDF might not be more effective than CC-Only, we are still supportive of the underlying assumption behind CC-IDF and believe that there must at least be some scenarios in which CC-IDF is more effective than CC-Only. We would also like to emphasize that the performance of CC-IDF varied strongly in experiments of other researchers who compared CC-IDF to e.g. bibliographic coupling (cf. section 1). Therefore, we suggest to conduct further research to gain insights on whether, and in which cases, CC-IDF is a suitable weighting scheme.

[9] To some extent, the same might be true for terms, but we assume the effect to be much stronger for citations.

5 References

Beel, J. (2015). Towards Effective Research-Paper Recommender Systems and User Modeling based on Mind Maps. *PhD Thesis. Otto-von-Guericke Universität Magdeburg*.

Beel, J., Breitinger, C., Langer, S., Lommatzsch, A., & Gipp, B. (2016). Towards Reproducibility in Recommender-Systems Research. *User Modeling and User-Adapted Interaction (UMUAI)*, *26*(1), 69–101. doi:10.1007/s11257-016-9174-x

Beel, J., Gipp, B., Langer, S., & Genzmehr, M. (2011). Docear: An Academic Literature Suite for Searching, Organizing and Creating Academic Literature. *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, JCDL '11 (pp. 465–466). ACM. doi:10.1145/1998076.1998188

Beel, J., Gipp, B., & Mueller, C. (2009). SciPlore MindMapping' - A Tool for Creating Mind Maps Combined with PDF and Reference Management. *D-Lib Magazine*, *15*(11). doi:10.1045/november2009-inbrief

Beel, J., Gipp, B., Shaker, A., & Friedrich, N. (2010). SciPlore Xtract: Extracting Titles from Scientific PDF Documents by Analyzing Style Information (Font Size). In M. Lalmas, J. Jose, A. Rauber, F. Sebastiani, & I. Frommholz (Eds.), *Research and Advanced Technology for Digital Libraries, Proceedings of the 14th European Conference on Digital Libraries (ECDL'10)*, Lecture Notes of Computer Science (LNCS) (Vol. 6273, pp. 413–416). Glasgow (UK): Springer.

Beel, J., & Langer, S. (2015). A Comparison of Offline Evaluations, Online Evaluations, and User Studies in the Context of Research-Paper Recommender Systems. In S. Kapidakis, C. Mazurek, & M. Werla (Eds.), *Proceedings of the 19th International Conference on Theory and Practice of Digital Libraries (TPDL)*, Lecture Notes in Computer Science (Vol. 9316, pp. 153–168). doi:10.1007/978-3-319-24592-8_12

Beel, J., Langer, S., Genzmehr, M., & Müller, C. (2013). Docears PDF Inspector: Title Extraction from PDF files. *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'13)* (pp. 443–444). ACM. doi:10.1145/2467696.2467789

Beel, J., Langer, S., Genzmehr, M., & Nürnberger, A. (2013). Introducing Docear's Research Paper Recommender System. *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'13)* (pp. 459–460). ACM. doi:10.1145/2467696.2467786

Beel, J., Langer, S., Gipp, B., & Nürnberger, A. (2014). The Architecture and Datasets of Docear's Research Paper Recommender System. *D-Lib Magazine*, *20*(11/12). doi:10.1045/november14-beel

Beel, J., Langer, S., Kapitsaki, G. M., Breitinger, C., & Gipp, B. (2015). Exploring the Potential of User Modeling based on Mind Maps. In F. Ricci, K. Bontcheva, O. Conlan, & S. Lawless (Eds.), *Proceedings of the 23rd Conference on User Modelling, Adaptation and Personalization (UMAP)*, Lecture Notes of Computer Science (Vol. 9146, pp. 3–17). Springer. doi:10.1007/978-3-319-20267-9_1

Bollacker, K. D., Lawrence, S., & Giles, C. L. (1998). CiteSeer: An autonomous web agent for automatic retrieval and identification of interesting publications. *Proceedings of the 2nd international conference on Autonomous agents* (pp. 116–123). ACM.

Chakraborty, T., Modani, N., Narayanam, R., & Nagar, S. (2015). Discern: a diversified citation recommendation system for scientific queries. *2015 IEEE 31st International Conference on Data Engineering* (pp. 555–566). IEEE.

Councill, I. G., Giles, C. L., & Kan, M. Y. (2008). ParsCit: An open-source CRF reference string parsing package. *Proceedings of LREC* (Vol. 2008, pp. 661–667). European Language Resources Association (ELRA).

Ekstrand, M. D., Kannan, P., Stemper, J. A., Butler, J. T., Konstan, J. A., & Riedl, J. T. (2010). Automatically building research reading lists. *Proceedings of the fourth ACM conference on Recommender systems* (pp. 159–166). ACM.

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). CiteSeer: An automatic citation indexing system. *Proceedings of the 3rd ACM conference on Digital libraries* (pp. 89–98). ACM.

Gipp, B. (2014). *Citation-based Plagiarism Detection - Detecting Disguised and Cross-language Plagiarism using Citation Pattern Analysis* (p. 350). Springer Vieweg Research. doi:0.1007/978-3-658-06394-8

Gipp, B., & Beel, J. (2009). Citation Proximity Analysis (CPA) - A new approach for identifying related work based on Co-Citation Analysis. In B. Larsen & J. Leta (Eds.), *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.

Gipp, B., Beel, J., & Hentschel, C. (2009). Scienstein: A Research Paper Recommender System. *Proceedings of the International Conference on Emerging Trends in Computing (ICETiC'09)* (pp. 309–315). Virudhunagar (India): IEEE.

Gipp, B., Meuschke, N., & Breitinger, C. (2014). Citation-based Plagiarism Detection: Practicability on a Large-scale Scientific Corpus. *Journal of the American Society for Information Science and Technology (JASIST)*, *65*(2), 1527–1540. doi:10.1002/asi.23228

Huynh, T., & Hoang, K. (2012). Modeling collaborative knowledge of publishing activities for research recommendation. *International Conference on Computational Collective Intelligence* (pp. 41–50). Springer.

Huynh, T., Hoang, K., Do, L., Tran, H., Luong, H., & Gauch, S. (2012). Scientific publication recommendations based on collaborative citation networks. *Collaboration Technologies and Systems (CTS), 2012 International Conference on* (pp. 316–321). IEEE.

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, *28*(1), 11–21.

Küçüktunç, O., Saule, E., Kaya, K., & Çatalyürek, Ü. V. (2012). *Recommendation on Academic Networks using Direction Aware Citation Analysis*. *arXiv preprint arXiv:1205.1143*.

Küçüktunç, O., Saule, E., Kaya, K., & Çatalyürek, Ü. V. (2013). Towards a personalized, scalable, and exploratory academic recommendation service. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 636–641). ACM.

Langer, S., & Beel, J. (2014). The Comparability of Recommender System Evaluations and Characteristics of Docear's Users. *Proceedings of the Workshop on Recommender Systems Evaluation: Dimensions and Design (REDD) at the 2014 ACM Conference Series on Recommender Systems (RecSys)* (pp. 1–6). CEUR-WS.

Liang, Y., Li, Q., & Qian, T. (2011). Finding relevant papers based on citation relations. *Proceedings of the 12th international conference on Web-age information management* (pp. 403–414). Springer.

Narwekar, A. A. (2016). *An Academic Search Engine and Problems in Citation Networks. PhD Thesis*. Indian Institute of Technology Madras.

Pan, L., Dai, X., Huang, S., & Chen, J. (2015). Academic Paper Recommendation Based on Heterogeneous Graph. *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data* (pp. 381–392). Springer.

Robertson, S. (2004). Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of Documentation*, *60*(5), 503–520.

Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*, *18*(11), 613–620.

Schwarzer, M., Schubotz, M., Meuschke, N., Breitinger, C., Markl, V., & Gipp, B. (2016). Evaluating Link-based Recommendations for Wikipedia. *Proceedings of the 16th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, JCDL '16 (pp. 191–200). Newark, New Jersey, USA: ACM. doi:10.1145/2910896.2910908

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*, 265–269.

Zhang, Q., Li, J., Zhang, Z., & Wang, L. (2012). Relation regularized subspace recommending for related scientific articles. *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 2503–2506). ACM.