Related Records Retrieval and Pennant Retrieval: An Exploratory Case Study

Müge Akbulut
Department of Information Management, Faculty of Humanities and Social Sciences
Ankara Yıldırım Beyazıt University, 06760 Ankara, Turkey

Yaşar Tonta
Department of Information Management, Faculty of Letters
Hacettepe University, 06800 Beytepe, Ankara, Turkey

Howard D. White
College of Computing and Informatics
Drexel University, Philadelphia, PA 19104, USA

Abstract

The Related Records feature in the Web of Science retrieves records that share at least one item in their reference lists with the references of a seed record. This search method, known as bibliographic coupling, does not always yield topically relevant results. Our exploratory case study asks: How do retrievals of the type used in pennant diagrams compare with retrievals through Related Records? Pennants are two-dimensional visualizations of documents co-cited with a seed paper. In them, the well-known $tf*idf$ (term frequency*inverse document frequency) formula is used to weight the co-citation counts. The weights have psychological interpretations from relevance theory; given the seed, tf predicts a co-cited document's cognitive effects on the user, and idf predicts the user's relative ease in relating its title to the seed's title. We chose two seed papers from information science, one with only two references and the other with 20, and used them to retrieve 50 documents per method in WoS for each of our two seeds. We illustrate with pennant diagrams. Pennant retrieval indeed produced more relevant documents, especially for the paper with only two references, and it produced mostly different ones. Related Records performed almost as well on the paper with the longer reference list, improving remarkably as the coupling units between the seed and other papers increased. We argue that relevance rankings based on co-citation, with pennant-style weighting as an option, would be a desirable addition to WoS and similar databases.

Keywords Bibliographic coupling · co-citation analysis · relevance theory · tf*idf · Web of Science

Article Highlights:

Using two classic papers as seeds, one with few references and the other with many, this case study tests two retrieval metrics in the Web of Science.

Pennant retrieval, based on co-citation, is shown to compare favorably with WoS Related Records retrieval, based on bibliographic coupling, for both seeds.

The interpretability of the co-citation set is further enhanced by $tf*idf$ weighting as used in pennant diagrams.

Related Records Retrieval and Pennant Retrieval: An Exploratory Case Study

Introduction

This is an exploratory case study in bibliometric retrieval of related documents. Specifically, we retrieve and evaluate documents that two citation-based similarity metrics relate to the same two seed documents. Both metrics rank retrieved documents by predicted closeness to the seed. The first is bibliographic coupling (Kessler 1963), as implemented in the Web of Science (WoS) under the name Related Records. The second, based on co-citation counts and yet to be implemented, is the one used in creating pennant diagrams (White 2007a,b, 2009, 2010, 2015, 2018a; White and Mayr 2013). These diagrams, shaped like triangular flags, position the works co-cited with a seed on two dimensions—namely, by their predicted relevance to it and by how specifically their titles relate to its title. Hence, we call our second method "pennant retrieval".

Searching for records of related documents is one of the most important capabilities that online retrieval services offer their users. Terms used to indicate or retrieve such records vary from system to system (e.g., "related records," "related documents," "find similar," "more like this"), but the overall goal is clear: "Given a document that the user has indicated interest in, the system task is to retrieve other documents that the user may also want to examine" (Lin and Wilbur 2007: 3). In this context, "relatedness" is generally an indicator of the degree of topical similarity between two documents, as evidenced by shared properties such as descriptors, keywords, references, or co-citations. For instance, the higher the number of items shared in the reference lists of two documents—that is, the stronger their bibliographic coupling—the more likely it is that both are on the same topic (Belter 2017; Horsley, Dingwall, and Sampson 2011). Small's (1973) co-citation metric—the frequency with which any two earlier documents are cited by later ones—is comparable: the higher the frequency, the stronger the presumption that citers regard the earlier pair as topically similar (Beel, Gipp, Langer, and Breitinger 2016: 320). In this study, one member of the co-cited pair is always the seed, and it is the similarity of other documents to the seed that is measured. The following considerations set our research problem. When documents are coupled to a seed by only one or two references, their topical similarity to it is frequently low because of "topic drift" (Huang, Xue, Zhang, Chen, Yu, and Ma 2004). Yet bibliographic coupling counts above some threshold (e.g., three) usually produce reasonably good retrievals. The same is true of co-citation counts. White has claimed, most recently in 2018b, that *all* the works in a pennant retrieval are relevant to the seed in varying degrees by empirical co-citation evidence. Such evidence comes from multiple co-citing authors repeatedly relating the seed as relevant, thus creating a stronger topical tie between the co-cited documents. Pennant retrieval, however, has never been tested against any other IR method.

Earlier work comparing the retrieval performance of co-citation and bibliographic coupling found the former more effective (Bichteler and Eaton 1980; Zarrinkalam and Kahani 2012). Accordingly, our main motive was to compare pennant retrieval with Related Records retrieval as ways of finding potentially useful documents. We further wished to compare the two methods for (1) a seed with few items in its reference list as against (2) a seed with relatively many. Our first seed was thus Maron & Kuhns (1960), a seminal paper on information retrieval (IR) with only two references. Our hunch was that the bibliographic coupling method would probably fail to produce satisfactory results for a relatively old seed paper. Thus, we wished to find out if the pennant retrieval can be used to complement the shortcomings of Related Records retrieval for such old papers with no or scanty reference lists. Our second seed was Cooper's (1988) "Getting beyond Boole" with 20 items in its reference list. We predicted that Related Records retrieval would probably produce much better results for the second seed than the first one. Despite this, we thought it would be worthwhile to see if the pennant retrieval could still be of help in discovering further relevant documents for papers with many items in their reference lists. We entered these two documents into WoS as separate queries and judged the ranked lists of records they retrieved for relevance to the seeds. For both seeds, we conjectured that pennant retrievals would produce more relevant items in WoS than lists based on Related Records, and that is what we found. We are of course aware that our results are not generalizable, but our in-depth analysis may suggest testable hypotheses in bibliometrically-enhanced IR.

We chose our seeds so that retrievals generated by them would be meaningful to readers in information science. The Maron & Kuhns paper—hereafter M&K—replaced the "two-valued thinking about IR with probabilistic notions" (Maron 2008: 971), thereby paving the way, almost 60 years ago, for Google-like systems that we take for granted today (Bensman 2013). Cooper's paper, another classic of IR, discussed the general shortcomings of Boolean IR systems and proposed alternative models.

We also present pennant diagrams for both M&K and Cooper that visualize their relationships with other papers, suggesting their impact on, e.g., the formation of major IR models. In general, visually rich pennants offer ways of

exploring the intellectual structure around both classic and non-classic works. They can be published as permanent displays to help researchers interpret the complex intellectual history of specific domains, or they can be generated online as disposable aids to help users browse an existing corpus of documents (White 2015, 2018a). Preliminary work on interpreting M&K through pennant diagrams appeared in Akbulut (2016a, b), an independent replication of White's methodology. The CiteSpace system has added pennant diagrams to its suite of visualization software.

Our retrievals resemble those produced by document recommender systems, in that, like them, ours are documents algorithmically ranked by degrees of relevance to a seed. However, our approach differs somewhat in emphasis from those typically found in the recommender system literature. That is:

- Since our retrievals are based on seed-related bibliometric counts, we assume that researchers may be interested in the seed's history as an intellectual contribution. In other words, they enter it as a search term because they want an overview of the other publications associated with it in the literature. Given a seed such as M&K, for example, they may search for documents bibliographically coupled with it or co-cited with it simply to see what these methods turn up as the seed's near or distant neighbors. A bibliometric motive of this kind is not usually addressed in the recommender system literature.

- We use topical similarity to the seed to judge retrievals for relevance, but we do not assume that high topical similarity is necessarily the *only* goal of the retrievals. As in bibliometric mapping, shades of similarity, including marked dissimilarity, may also be of interest—for example, to the seed's author(s) or to researchers well acquainted with the field the seed represents. Such persons would presumably bring relatively high levels of sophistication and curiosity to interpreting retrievals.

- Rather than testing the relevance of retrievals against baseline documents in a standard experimental set and reporting the results abstractly, we show, with commentary, the top 50 documents actually returned by each method for each seed, so that readers themselves may judge their attributes.

Related Studies

Comparative studies on related records search systems are not that many. Studies that evaluate Related Records results are even scarcer. The following review is limited to systems using bibliographic coupling or co-citation linkages (or both) as relatedness measures. These include some recommender systems, which are reviewed in general in Beel et al. (2016). Details of pennant creation and interpretation are also given.

Bibliographic Coupling

Similarity measures based on linguistic features of documents such as keywords reflect natural language with all its "inherent complexities and ambiguities" (Zarrinkalam and Kahani 2012: 100). Bibliographic coupling is a similarity measure that is "independent of words and language," thereby "avoid[ing] all the difficulties of language, syntax and word habits" and requiring "[n]o expert reading or judgment" (Kessler 1963: 11). Or as Clarivate Analytics (2017a) puts it, Related Records is implemented in WoS to find similar documents "regardless of whether their titles, abstracts, or keywords contain the same terms". Kessler defined each reference common to two documents as a "coupling unit". For two documents to be bibliographically coupled, at least one of their references must be shared. The more items a document has in its reference list, the more documents it is somehow related to, and the more likely it is to appear in Related Records searches.

The number of shared references signifies what is called "cognitive overlap" or "intellectual overlap" between two documents, "with higher number of overlaps indicating greater relevance" (Belter 2017: 733; Colavizza et al. 2018: 604). More formally, intellectual overlap is "the proportion of references that a pair of publications have in common", and this formula is used to calculate it:

$$N_{qd} / min (N_q, N_d), \hspace{4cm} (1)$$

where N_{qd} denotes the number of overlapping references of publications, and N_q and N_d denote the total number of references in publications *q* & *d* respectively.

The intellectual overlap equals one if all references of the publication with the shorter reference list are also cited by the other publication (Colavizza et al. 2018: 604-605). For a given publication *q*, the calculation is repeated for all the publications (*d_i*'s) in the collection and those that overlap are then ranked.

The main issue with the overlap formula is that it does not take into account the total items, including non-overlapping ones in the reference lists of bibliographically coupled documents. This is well illustrated in a Venn

diagram which has a pair of articles (say, *q* & *d*), one (*q*) with 40 items and the other (*d*) with 90 items, of which 10 are common to both (i.e., N_{qd} = 10) in their reference lists (Smith, Georges and Nguyen, 2015: 1667). Suppose we have another pair of articles with half the number of items in their reference lists including the ones in the intersection (i.e., *q* = 20, *d* = 45, and N_{qd} = 5). Formula (1) produces the same outcome for both pairs of articles, but it is not clear that these two pairs of articles have the same degree of overlap in terms of topical similarity.

Further, suppose that for the first pair of articles *q* & *d*, all 40 items in the reference list of *q* overlap with those of *d* (i.e., the items in the reference list of *q* are a subset of that of *d*). This means, by definition, a perfect intellectual overlap. But what about the remaining 50 items in the reference list of *d*? Could they be about a somewhat related but nevertheless a slightly different subject? What if *d* had 100 non-overlapping items instead of 50? Would the subset have equal intellectual overlap with the whole in both cases? As the subset (or intersection) gets smaller, a given article (*q*) would be related to a higher number of non-overlapping references of probably less relevant articles, thereby causing topic drift.

Even so, systems based on bibliographic coupling offer a popular searching and browsing capability. They enable users to start their searches with a "known-item" relevant record and then find possibly similar items. In Kessler's day they were not easy to implement because of the limited power and storage capacities of the computers that produced citation indexes. Eugene Garfield, the founder of the Institute for Scientific Information (ISI), took 25 years to apply the idea of bibliographic coupling to the citation indexes that are now in the Web of Science. It was finally implemented as the Related Records capability "in the CD-ROM version of SCI and SSCI in 1988" (Garfield 2001: 2). Since then, this capability has become available in two other citation indexes, Scopus and Microsoft Academic. It is also seen in CiteSeer[X] and PubMed (Giles et al. 1998; Lin and Wilbur 2007; Peterson and Graves 2009). More recently, a refined method that uses *tf*idf* formula to compute weighted bibliographic coupling strengths was introduced (Shen, Zhu, Rousseau, Su and Wang 2019).

Co-citations

Small (1973) linked bibliometrics with IR and proposed a new measure of document similarity based on co-citation of documents (Åström 2007). Co-citation counts "focus on the broader relatedness of publications" by reflecting how many later works cite pairs of earlier works (Colavizza et al. 2018: 601). Co-citation analysis became a powerful tool for mapping the structure of scientific fields (Garfield 2001: 2).

Co-citation counts are dynamic, whereas counts of the bibliographic coupling links are static (Sugimoto and Larivière 2018: 67). Bichteler and Eaton (1980: 279) pointed out that "[u]nlike bibliographic coupling, the strength of co-citation can increase over time as new papers that cite previous papers are written". Or in the words of Garfield (2001: 3): "Bibliographic coupling is retrospective whereas co-citation is essentially a forward-looking perspective". The dynamism of co-citation emerges in two ways: the count for any pair of co-cited documents can grow indefinitely over time, or authors can continually cite a given document with other documents in new combinations. Bibliographic coupling thus "fails to evolve with the field" while "co-citation similarity evolves and changes as fields change" (Wesley-Smith et al. 2016: 1-2).

Nevertheless, WoS does not offer retrieval based on co-citations. DialogClassic offered it in the ISI citation databases until 2013, but that service no longer exists (White 2018b). Nor is it possible to retrieve records on the basis of co-citations in Scopus or Google Scholar, although they offer their own versions of interrelated records searching. For instance, Scopus displays, by default, the top three documents based on the combination of shared references, keywords and authors. Scopus also allows users to browse each list separately (Scopus 2018).

Citation-based measures are used in some recommender systems to rank papers (Beel et al. 2016; He et al. 2010; Huang et al. 2004; Liang et al. 2011). For example, Zarrinkalam and Kahani (2012) proposed a metric of relatedness between documents to improve the performance of recommender systems. They defined six types of relations (including bibliographic coupling and co-citation) between two documents, each weighted differently. They designed experiments to test the performance of this metric on a subset (*c*. 30 thousand papers) of the CiteSeer[X] Digital Library, using recall, probability of documents being co-cited, and normalized discounted cumulative gain (NDCG). Direct citations, co-citations, and bibliographic coupling had the most positive impact on performance.

Zarrinkalam and Kahani's paper is one of the few that compared the retrieval performance of co-citation and bibliographic coupling. They found that co-cited papers were more likely to be cited than the bibliographically coupled ones and that they had the highest recall of recommended documents (Zarrinkalam and Kahani 2012: 105-106). In Bichteler and Eaton's (1980) small-scale experiment, co-citation-based retrieval modestly improved performance (13%), suggesting that "knowing the cocitations of two papers helps predict the subject similarity of the papers more

accurately than is possible simply from knowing their bibliographic coupling strength" (p. 282). However, Haruna et al. (2018) found that a recommendation algorithm based on co-citation counts alone performed worst of three in finding the latent associations among papers, presumably because of the topic drift mentioned earlier.

More recent papers have explored various methods to improve the retrieval performance of co-citation-based search and recommendation systems. For instance, Eto (2013) found that using co-citation contexts has positive effects on co-citation searching. Ahmad and Afzal (2017) experimented with combining co-citation relevance scores with metadata to recommend more related papers and reported 25% improvement in retrieval performance.

Pennant Diagrams

White combined co-citation analysis with IR and relevance theory (RT) to predict the relevance of documents to a seed author or work (White 2007a, b). From RT, a subfield of linguistic pragmatics, he borrowed Sperber and Wilson's (1995: 261) psychological definition of relevance as a property of inputs to individual minds:

$$\textit{Relevance = cognitive effects / processing effort} \qquad\qquad (2)$$

where the relevance of an input varies directly with its cognitive effects on a person, but inversely with the effort the input costs the person to process (Wilson and Sperber 2002). In terms of RT, a person has a set of assumptions — that is, a context of thought—about the seed. The inputs to this context are the co-cited documents. When considered in relation to the seed, a co-cited document may strengthen an assumption, eliminate it, or combine with it to yield new conclusions. These are RT's three main cognitive effects. The greater an input's cognitive effects, the greater its relevance. At the same time, the greater the effort of processing the input, the *less* its relevance, as will be discussed below. Both variables are inherently ordinal (not ratio) scales. Users would usually seek documents with the greatest cognitive effects, since they do not like to waste their "mental energy entertaining uncertainties" (Soll et al. 2015).

White mapped the Sperber and Wilson formula for relevance onto the *tf*idf* (term frequency*inverse document frequency) formula from IR. In the original IR formula, the *tf* and *df* values are used to weight documents for their relevance to a query; *tf* is a query term's frequency in a document, while *df* is its frequency in all the documents in the collection being searched (Salton and Yang 1973; Hiemstra, 2000). As a term's *df* count increases, the more common it is in the collection, and presumably the less discriminating it becomes. Inverting this frequency as *idf* elevates terms that are less common and presumably more query-specific. Sparck Jones (1972) in fact introduced the *idf* factor as "statistical specificity".

The seed document is the query in White's studies, and *tf* is the number of times a document has been co-cited with a seed. As the document's *tf* increases, its relevance to the seed increases. By contrast, *df* is the number of citations each co-cited document has received in the database. As a document's *df* count increases, the more it is cited across a variety of other works, presumably because its title and contents suggest a wider applicability than the seed's. Inverting *df* as *idf* elevates documents that are *less* widely cited and presumably more specifically and obviously related to the seed when their titles (and possibly other bibliographic details) are compared (White 2007a: 538-542). Hence, White calls *idf* in this case "ease of processing" rather than "processing effort"; higher *idf* means higher ease.

In pennant retrieval, a relevance score for each document is obtained by multiplying *tf* and *idf* as in Manning and Schütze (2000: 541-542):

$$\textit{Relevance} = 1 + \log (\textit{tf}) * \log (N/\textit{df}) \qquad\qquad (3)$$

where *tf* and *df* are defined as above, *N* is the estimated number of documents in the database, and the base-10 logs are a damping factor. Appendices 1b and 2b have 100 examples of scores from *(3)* being used to rank documents co-cited with a seed. Other examples will be found in White (2018b), where they are called "bag of works" retrievals, and in White (2010). These examples tend to show works obviously relevant to the seed ranked higher than those not obviously relevant to it, in keeping with the ease-of-processing factor.

By contrast, in pennant diagrams the *tf* and *idf* values from *(3)* are plotted separately, not multiplied to yield a single weight. The two values determine the coordinates of each co-cited document on two axes (see, e.g., White 2007a, b; Tonta and Özkan Çelik 2013: 39). The seed's point is at the tip of the pennant. A document's placement relative to the seed predicts its cognitive effects and the ease of processing it. Its *tf* value (cognitive effects) is plotted on the pennant's *x*-axis; its *idf* value (ease of processing) is plotted on the *y*-axis. Higher *tf* scores pull documents closer to the seed. Their *idf* scores predict the ease of relating their titles to the seed's title. "Easier" documents on the *y*-axis tend to have title terms identical to, or synonymous with, or frequently associated with the seed's title terms; "harder" documents on the *y*-axis do not. (A "harder" document can still be highly relevant to the seed on the *x*-

axis.) Since pennants have limited space for labeling, they need to be accompanied by fuller bibliographic data for each point. They are intended to show what *tf*idf* weighting does to numerous actual titles, thereby illustrating an RT-based explanation of the formula's popularity.

Data, Methods, and Techniques

The first of our two seeds, Maron & Kuhns (1960), presented what was then a new approach to IR, and, as sometimes happens with pioneering works, it cited only two items:

- Shannon and Weaver's (1949) famous book *The Mathematical Theory of Communication*
- Yule's (1912) foundational paper on measuring association between attributes

In a sense, using M&K in a test of bibliographic coupling is less than fair: not only are its references few, but both are very broad in their connotations, neither obviously resembles the seed, and neither obviously resembles the other. One might therefore expect many Related Records retrievals to be markedly different from it in their subject matter, and that is indeed what we found. Our relevance judgments on the Related Records retrieval and the pennant retrieval for M&K were thus very much in the latter's favor.

By contrast, Cooper (1988) cited 20 items, and its Related Records retrievals were dramatically better. It describes several problems with the "conventional" Boolean model of IR (e.g., unfriendly Boolean formulas, too little or too much output, inability to emphasize different facets of the search). For each problem, it offers a design solution that was non-conventional in 1988 (e.g., probabilistic methods, term-weighting, advanced statistical techniques). Most of its references are keyed to these alternatives. Nine are on probabilistic IR, eight are on IR systems in general, and three are on use of the maximum entropy principle in IR experiments.

Through separate Related Records searches of the WoS core collection (consisting, at the time, of documents from 1945 to 2017), we obtained the bibliographic coupling data on our two seeds in December 2017. When a seed's full bibliographic record is displayed in WoS, a link labeled "View Related Records" returns all documents that share at least one reference with the seed, ranked by how many are shared. Our search identified 9,803 works coupled with M&K and 2,919 works coupled with Cooper. After retrieving these works, we exported them in *.txt* format to MS Excel and used a macro to check the number of references each record shared with each of our seeds.

The co-citation data retrievals were conducted in WoS in March 2018. We performed Cited Reference (CR) searches in WoS for each seed paper separately. M&K had been cited in 333 articles and Cooper's paper in 43 articles in WoS-indexed journals. We downloaded the bibliographic records of all articles citing these two papers, including their reference lists. The latter were processed offline to count the frequencies with which our seeds were cited jointly with other references. The 333 direct citations to M&K yielded a total of 4,176 unique co-citations, while the 43 direct citations to Cooper yielded a total of 711. We wrote several macros[1] to clean, match, count, merge, cluster, and visualize these data.

To count the frequency of each unique citation in the reference lists, we wrote a script to match citations to the same works in the reference lists of the 333 and 43 papers. Data cleaning and merging were necessary because WoS source publications use different citation styles (e.g., APA, MLA, Chicago), and names are not cited consistently, which fragments the associated counts. For instance, the string "Maron ME" was cited in 19 slightly different ways in the database. The journal in which M&K's paper was published (*Journal of the Association for Computing Machinery*) was abbreviated in eight different ways (e.g., *ACM J*, *J ACM* and *J ASS COMP MACH*). This fragmentation caused citations to some works to appear fewer than they in fact were, making it difficult to determine *tf* correctly. To overcome this problem, we ran a similarity algorithm based on the bag of words (not bag of works) technique and set the similarity threshold at 80 percent to increase the accuracy of matching. We examined the output visually and checked the sources before merging.

Figure 1 illustrates our WoS record structure with specimen data. Works 1, 2, 3 and 4 have publication years (PY) 2010-2012. Their unique cited references (CR) are A, B, C, D, E, and F.[2] The seed paper is represented by A, included in the reference lists of all four works. Due to the nature of our data set, the total frequency count for each reference, as seen in Table 1, actually gives its co-citation count with the seed. In our actual data, the reference lists

---

[1] Macros are available at: http://www.mugeakbulut.com/YL_Tez/veriler_makrolar/makrolar/
[2] We needed only PY and CR from the 60 field tags in the WoS standard file to calculate the frequencies. For all tags and their definitions, see Clarivate Analytics (2018a).

of citing works were linked through their citing years.

To calculate the *idf* values, we exported the cited references by checking each item on WoS "marked lists" and created citation reports from these lists. The *tf* and *idf* values were calculated as in formula (3):

$$tf = 1 + \log (tf) \qquad (4)$$

$$idf = \log (N/df) \qquad (5)$$

where the *tf* was based on each work's co-citations with the seed, and *idf* was based on its total citations in the database.
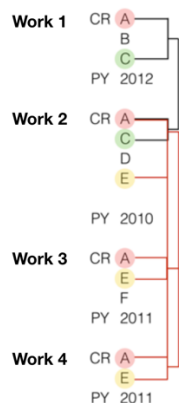

Table 1. Co-citation frequencies by years

References	2010	2011	2012	Total
A	1	2	1	4
B	0	0	1	1
C	1	0	1	2
D	1	0	0	1
E	1	2	0	3
F	0	1	0	1

Fig. 1. A specimen of WoS data structure

We used a macro to count the co-citations and total citations for each record and computed the *tf*idf* weightings to come up with a ranked list. *N*, the total items in the collection, was assumed to be five million.

We then found the distribution of WoS-assigned research areas for all records related to the two seeds. "Research area" denotes a subject categorization scheme used in all WoS databases (Clarivate Analytics 2017b, 2018b). The scheme comprises 252 subject categories in the sciences, social sciences, arts, and humanities. Some works are assigned to more than one area. For each seed, we listed the 10 areas most frequently assigned to its bibliographically coupled papers. We thus could compare the effect of M&K's two references with Cooper's 20 on the broad topicality of retrievals.

Next, we assessed the topical relevance of individual papers to our seeds. In classical IR tests, need-based queries are collected from users, and the same users judge the results for relevance. In effect, that is what two of us (MA and YT) did here on a very small scale. It is not unreasonable to think the same could be done on a much larger scale in TREC-style evaluations. The two seeds represented real interests of ours, the retrievals were from a real system, and our judgments were straightforward reactions to scientific papers. Although we had conjectured that pennant retrievals would produce superior results, we proceeded as we would in any literature search and evaluated records using Swanson's (1986: 397) "topic-oriented relevance" to identify objective relationships with the seeds—or the absence of them. This was deemed appropriate because "[t]he testing of an information system . . . must operate with objectified requests and objective relevance of the responding documents" (Swanson 1986: 396).

As our "responding documents", we chose the top 50 records related to our two seeds by our two metrics—200 in all. We examined the full bibliographic details for each (e.g., author, title, assigned keywords), and read its abstract. A few items (e.g., book chapters) lacked abstracts, in which case we read their introductory sections. In judging whether a record was indeed related in topic to its respective seed, we sought consensus between the two researchers. Our binary judgments are shown in the Appendixes: relevant (R) or not relevant (NR).

Finally, we calculated Bollmann's (1983) generalized normalized recall measure (R_{norm}) for bibliographically coupled or co-cited items at various cut-off points (the first 5, 10, 25, and 50 items in the ranked retrievals):

$$R_{norm} (\Delta) = 0.5 * [1 + (R^+ - R^-) / R^+_{max}], \qquad (6)$$

where $R_{norm} (\Delta)$ is the retrieval output; R^+ is the number of document pairs in which a relevant document ranked higher than a non-relevant one (agreeing pairs); R^- is the number of document pairs in which a non-relevant

document ranked higher than a relevant one (contradictory pairs); and R^+_{max} is the maximum number of R^+'s (agreeing pairs). R_{norm} measures the effectiveness of an IR system based on the ranking of the retrieval outputs (Yao 1995: 142). It penalizes systems that do not reject non-relevant works successfully. We compared each list of records related to M&K or Cooper to find out which method displayed the relevant papers at higher ranks of retrieved outputs. Mean normalized recall ratios measure if search systems display relevant works in the top ranks of the retrieval outputs. Retrieval output Δ_1 is better than output Δ_2 if Δ_1 has fewer non-relevant works at top ranks than Δ_2 (Bitirim, Tonta and Sever 2002).

It is clearly very labor-intensive to clean, match and merge references, as well as to obtain the corresponding *tf* and *idf* values for each of them. There is, as yet, no way of automatically creating pennants, and almost all of the existing ones were made possible by the now defunct DialogClassic system and its unique instant formatting of the necessary data (White 2018b).

Findings and Discussion

Maron and Kuhns

M&K's subject matter can be briefly suggested. Its authors formulated a model based on probabilistic indexing: "If a user comes to the system with a query containing a specific term, what is the probability that this user would find a particular document containing that term relevant?" They set up an experimental library with a small collection of articles and estimated each article's probability of relevance to a given search query. Then they used probabilistic weighting factors based on distance/closeness measures between index terms in the document/request space. Relevant articles were retrieved for more than two-thirds of the queries.

Records Bibliographically Coupled with M&K

Table 2 reveals the top 10 research areas of the 9,803 works bibliographically coupled with M&K. Almost half occur in the top four categories. Most have nothing to do with M&K's topics.

Table 2. Distribution of Related Records for M&K in the top 10 research areas

Rank	Research Area	N
1	Environmental Sciences Ecology	1628
2	Computer Science	1458
3	Engineering	985
4	Marine Freshwater Biology	812
5	Physics	772
6	Psychology	770
7	Business Economics	546
8	Agriculture	493
9	Plant Sciences	487
10	Mathematics	468

Since M&K's paper refers only to Shannon and Weaver's book and Yule's paper, Related Records retrieved mostly non-relevant items. For instance, Prevedelli et al., the first paper in Appendix 1(a), is titled "Relationship of non-specific commensalism in the colonization of the deep layers of sediment". Among its 23 references, it cited both items that M&K also cited, which presumably made it the paper most topically related to M&K. By formula (1), there was 100% intellectual overlap between M&K and Prevedelli et al., but the latter is on freshwater biology.

The rest of the top 50 papers listed in Appendix 1(a) shared only one reference with M&K's paper. (When multiple bibliographic coupling counts are all tied, the WoS ranking principle is not clear, but it ranks them nonetheless.) Some of the papers in Appendix 1(a) seem indirectly related with M&K (see 4, 6, 25, 30, 32, and 36), probably because both they and M&K cited Yule's paper on association measures. Yet the topics of the rest of the 50 are very diverse, as the global categories in Table 2 suggest.

Overall, we judged 15 of the top 50 records to be relevant to M&K, as marked in Appendix 1(a). Their distribution at cut-off points is: one in the top five (20%), four in the top ten (40%), and eight in the top 25 (32%). The mean normalized recall ratios were respectively 25%, 38%, 51%, and 54% at cut-off points 5, 10, 25 and 50.

Thus, more than two-thirds of the records were not relevant to the seed paper. This less-than-satisfactory outcome is

due to the way Related Records works for papers with few items in their reference lists. All but one paper in the top 50 shared one reference with M&K—i.e., 50% intellectual overlap. Though 50% may not seem low, formula (1) does not take into account the total items in the reference lists of bibliographically coupled papers, as discussed earlier. WoS therefore produced many false drops, along with the extreme diversity of research areas seen in Table 2. In other words, the Related Records algorithm quickly drifted away from M&K's topic (Huang et al. 2004).

Pennant Retrieval with M&K

The performance of the pennant algorithm with M&K was quite impressive. We scored all 50 items in Appendix 1b as relevant to the seed. The normalized recall values (R_{norm}) were thus perfect (100%) at all cut-off points. Most are obviously so at title level. In a few cases, such as Bush's "As we may think" (rank 29) or Kleinberg's "Authoritative sources in a hyperlinked environment" (rank 45), the ties are not obvious in title language, but we know from our own backgrounds that citers have linked them to M&K because they bear strongly on IR techniques. (The authors are, to some extent, already familiar with the IR literature and in the past have read the full texts of several papers that were retrieved.)

The top 50 works co-cited with M&K appear as points in the pennant diagram in Fig. 2. They are labeled with their rank numbers from Appendix 1(b), their first or sole author, and their publication year. The points are divided (intuitively, not algorithmically) into sectors labeled A, B, and C to facilitate interpretation. In pennants, all works are relevant to the seed by historical co-citation evidence, but they vary in their degree of relevance and in how easily that relevance may be inferred from their titles. To repeat, the *tf* factor in formula (4) pushes items more relevant to the seed rightward on the *x*-axis, while the *idf* factor in formula (5) pushes items more specifically related to the seed upward on the *y*-axis (White 2018a: 762).
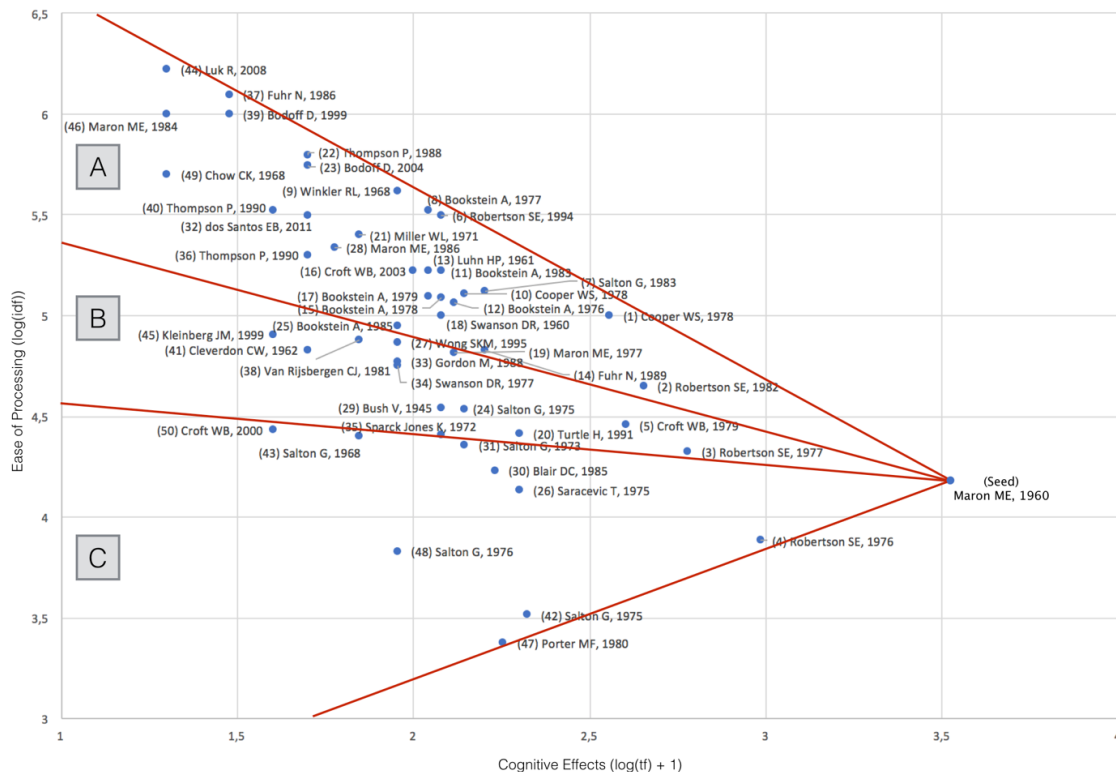

Fig. 2. Pennant diagram of works co-cited with Maron and Kuhns' seed paper (top 50 records)

Works with the highest *tf* values appear closest to the seed in all sectors (Fig. 2). They are predicted to be most relevant to M&K's paper in terms of cognitive effects when considered or read with it. But they differ in the ease of relating their titles to the seed's title. For example, among papers closest to the seed, Robertson [Maron and Cooper] (1982), Robertson (1977), and Croft [and Harper] (1979) are higher on the *y*-axis than Robertson [and Sparck Jones] (1976). The three higher papers' titles seem quite close to the seed's, while the last-named paper's title (italicized) is a bit broader in scope—i.e., less specific:

Seed: On relevance, probabilistic indexing and information retrieval

- Robertson SE (1982): Probability of relevance: a unification of two competing models for document retrieval
- Croft WB (1979) Using probabilistic models of document retrieval without relevance information.
- Robertson SE (1977): The probability ranking principle in IR
- *Robertson SE (1976): Relevance weighting of search terms*

In sector A, the total citations to works are low relative to their co-citation counts with the seed. Recall that the inverse factor that puts them higher on the *y*-axis is statistical specificity; their titles are predicted to be more specifically related to M&K's title and thus easier for a user to associate with it. In sector A, for example, we find Cooper [and Maron] (1978) on probabilistic and utility-theoretic indexing, along with Fuhr (1989), which builds on and advances M&K's Model 1. Also in sector A are papers by Thompson (1990a, b) and Maron (1977) that discuss the concept of "aboutness" as a factor contributing to relevance. These papers appear semantically close to M&K. Figure 3 highlights some papers in sector A for specificity analysis. More than three quarters of co-cited works with "probabilistic" in their titles are located in sector A. They thus all seem directly related to the seed's topic.
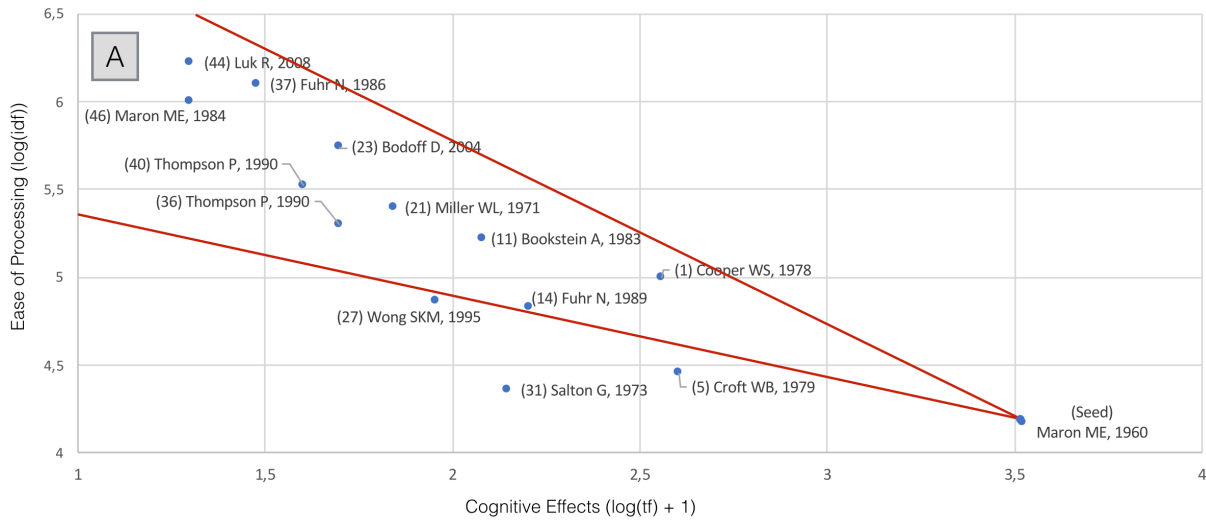

Fig. 3. Works with "probabilistic" in their titles

Papers pushed lower on the *y*-axis in sectors B and C tend to be less specifically related to the seed. The works in sector B mostly do not refer to probabilistic indexing in their titles; they are on other aspects of retrieval. Likewise, most of the works in sector C are less relevant to M&K's paper, in the sense of being harder to associate with it semantically (e.g., Salton's discussions of automatic indexing, Porter's suffix stripping). However, by co-citation evidence they are a meaningful part of the display, not simply instances of topic drift. Their histories of use by citers show them to be weakly relevant to M&K, even though their titles do not resemble M&K's.

We cannot provide the distribution of co-cited papers by WoS research areas (like those in Table 2), since the Cited Reference search in WoS does not offer this option, and it was not possible to identify research areas automatically. However, the titles of documents grouped by the pennant algorithm often imply intra- and interdisciplinary relations that are not otherwise easily observable. In Figure 4 we have applied *tf**idf weighting to all 4,176 papers co-cited with M&K—not just the top-ranked 50. Each of them is represented with a dot. (Numerous papers with low *tf* values are on the same dots.) We then labeled clusters of dots to suggest broad topical similarities in their titles, and these convey some major research areas associated with M&K, indicating its history of intellectual associations.
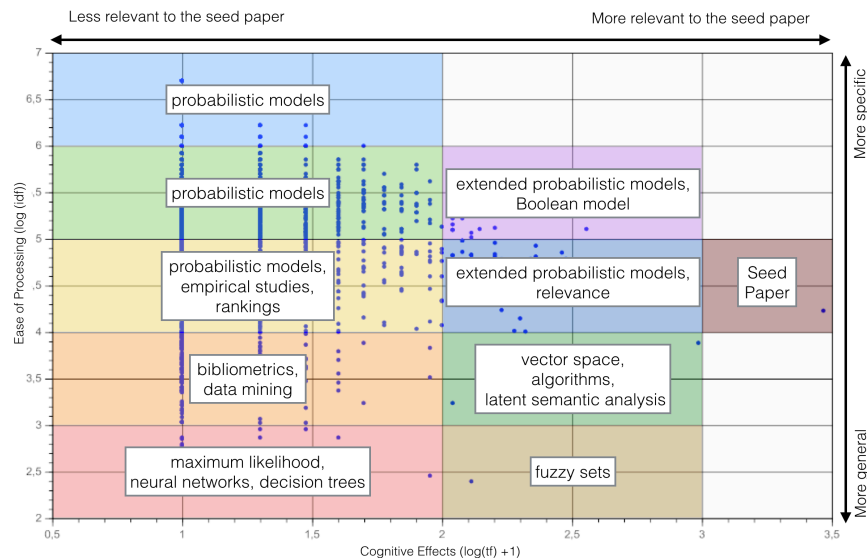
Fig. 4. Distribution of related papers by broad topics

The papers closest to the seed on the *x*-axis deal with probabilistic, Boolean, and vector space IR models and also with relevance as a concept. They are relatively few in number and are predicted to have greater cognitive effects when read with the seed than the less specifically related papers further back. The latter deal with topics such as fuzzy sets, bibliometrics, data mining, and neural networks. A researcher studying probabilistic IR models can easily locate the broad areas in which M&K's paper has had direct or indirect influence.

Comparison of Relevance Ranking Lists for M&K

We conjectured that pennant retrieval would retrieve more relevant items than Related Records for papers with both few and many references. Our findings for M&K strongly support the first part of our hypothesis. Of course, they are based on a single seed paper. Yet it is not uncommon for papers, especially older ones, to have brief reference lists. For instance, Tonta and Özkan Çelik (2013) created a pennant diagram for a 1941 paper by a famous mathematician, Cahit Arf, that had only one reference. It is likely that the retrieval performance of Related Records would be similar for many such older papers.

The retrieval sets for M&K in Appendix 1 permit detailed comparisons of the two methods we studied. The sets have zero overlap, indicating that our two methods in this case functioned differently. Related Records retrieved mostly non-relevant items. In contrast, the top 50 records produced by pennant retrieval could be used as a reading list for courses on IR models and IR history. Paul Thompson, a former student of Maron's, discussed M&K's contribution to the IR literature in a special issue of *Information Processing & Management* (Thompson, 2007). Of the 51 works he cited in his paper, 28 are in the pennant diagram, whereas none of them appears in the Related Records list.

Cooper

With Cooper, neither bibliographic coupling nor co-citation produced retrievals as glaringly off-topic as the many we marked "not relevant" in the case of M&K. In fact, all 100 items retrieved by the two metrics belong to the literature of information science, and most are on formal techniques of IR in the same sense as our two seeds. However, we did encounter a few items whose topics seemed somewhat peripheral to Cooper's and that would have required extra effort to consider in relation to it. That makes them, technically, *less* relevant to the seed, but we scored them as not relevant, as if making a first pass at prioritizing items from a recommender system. All titles appear in Appendix 2.

Records Bibliographically Coupled with Cooper

The retrieval performance of Related Records appears to improve tremendously as the number of shared references with the seed paper increases. Its performance for Cooper is impressive: 24 of the first 25 records were relevant. Altogether there were only six non-relevant records in the first 50. The mean normalized recall ratios at cut-off points of 5 and 10 are 100%. The overall normalized recall ratio for the top 50 records is 90%. Of the six non-

relevant records, five were on OPACs—online public access catalogs—in particular libraries (records 20, 35, 43, 47, and 48 in Appendix 2(a)).

Table 3 gives the top 10 research areas of the 2,919 works that Cooper retrieved through Related Records. They are much more similar to the seed in broad subject matter than those retrieved by M&K in Table 2. Some 90% of them came from Computer Science (compare 15% for M&K). In addition, 36% were assigned to the area WoS calls "Information Science Library Science" (some records were assigned to both top areas). The bibliographic coupling algorithm apparently works better for works citing relatively many references.

Table 3. Distribution of Related Records for Cooper in the top 10 research areas

Rank	Research Area	N
1	Computer Science	2617
2	Information Science Library Science	1064
3	Engineering	374
4	Mathematics	75
5	Operations Research Management Science	62
6	Telecommunications	62
7	Medical Informatics	50
8	Imaging Science Photographic Technology	48
9	Psychology	44
10	Automation Control Systems	38

The topics of the retrieved papers can be gleaned from their keywords: Boolean (10 papers), fuzzy (8), OPACs (6), general IR systems (5), term weighting (4), probabilistic IR (3), and ranking (3). Nearest-neighbor searching and natural language processing in IR are represented with two papers each.

Records Co-cited with Cooper

As we did for M&K in Fig. 2, we ranked the 711 papers co-cited with Cooper by their *tf*idf* weights and plotted the top 50 in the pennant diagram in Fig. 5. Cooper is rightmost. Parenthesized numbers refer to ranks of papers in Appendix 2(b).
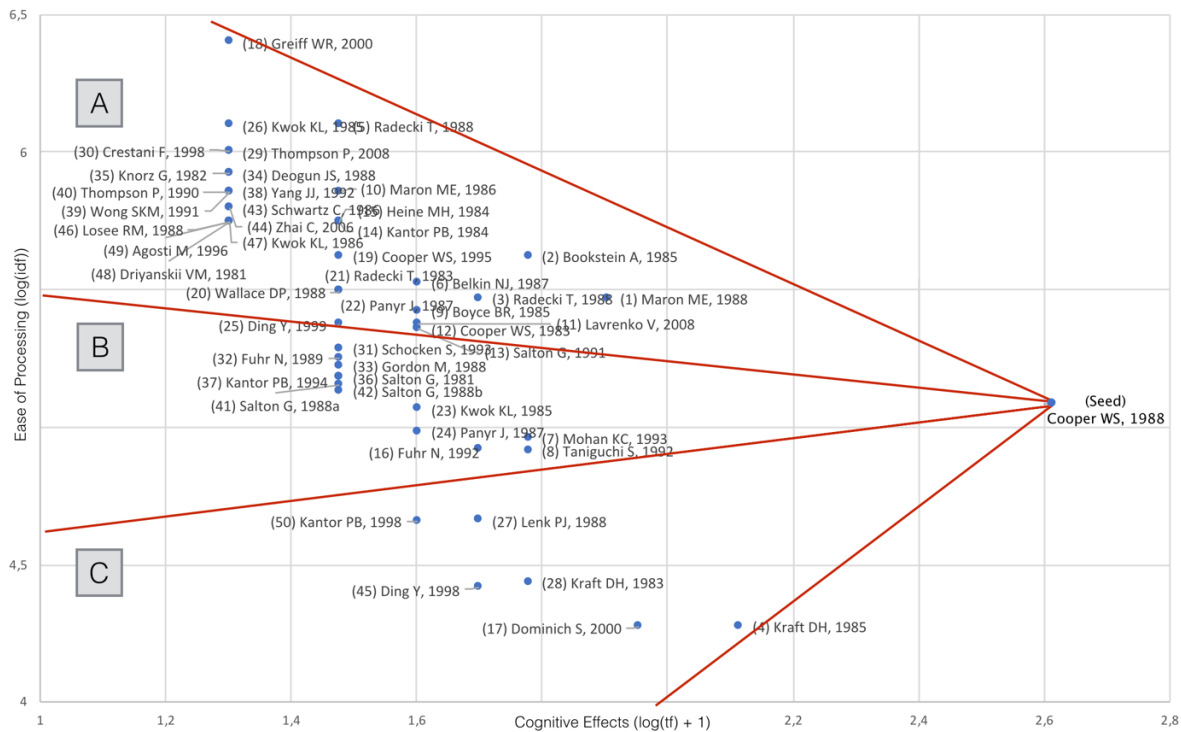

Fig. 5. Pennant diagram of works co-cited with Cooper's seed paper (top 50 records)

12

Of the papers co-cited with Cooper, 20 concentrate on probabilistic IR (including probabilistic indexing, fuzzy sets, and relevance) followed by 11 on IR in general, seven on vector space model (including automatic indexing and parallel text search), and four on the maximum entropy principle.

The three papers predicted to be the most specifically relevant to Cooper's are Maron ME (1988), Bookstein A (1985), and Radecki T (1988). Their *tf* and *idf* values are both high, placing them in sector A. Below, their titles may be compared with that of the italicized fourth-ranked one, Kraft DH (1985), which has a much higher *tf* value than the first three but a much lower *idf* value, placing it in sector C. Kraft is there because it is a more general work than the three above it—a highly cited literature review—and its relation to Cooper is less obvious than theirs:

> *Seed: Getting beyond Boole*
>
> - Maron ME (1988): Probabilistic design principles for conventional and full-text retrieval systems
> - Bookstein A (1985) Probability and fuzzy-set applications to information retrieval.
> - Radecki T (1988): Probabilistic methods for ranking output documents in conventional Boolean retrieval systems
> - *Kraft DH (1985) Advances in information retrieval: where is that /#*&@¢ record?*[3]

Regardless of where the co-cited papers are placed in the Fig. 5 pennant, we marked 46 of the first 50 as relevant to Cooper. The four non-relevant papers (ranks 20, 25, 43 and 45 in Appendix 2(b)) are Wallace DP (1988) on effective display size in online IR systems; Ding Y (1998, 1999), which are author co-citation analyses of IR as a field; and Schwartz C (1986), an ARIST review of subject analysis.

Comparison of Relevance Ranking Lists for Cooper

By our scoring, pennant retrieval produced 46 relevant papers in the first 50, and Related Records produced 44. The difference, while negligible, shows pennant retrieval working slightly better than Related Records for a seed paper with many items in its reference list. Moreover, as shown in Table 4, the two lists of 50 had only five records in common. The pennant retrieval thus contributed an additional 41 relevant items in this particular retrieval. It would be of great interest if low overlap were typical of retrievals when both methods use the same seed.

Pennant retrieval had 100% mean normalized recall ratios at cut-off points for 5 and 10 records, 91% for 25 records and 67% for 50 records. The comparable Related Records scores were, respectively, 100%, 100%, 79% and 81%.

Table 4. Titles included in both ranked lists

Author(s)	Title	Pennant retrieval rank	Related Records rank
1. Salton, G. & Buckley, C.	Term-weighting approaches in automatic text retrieval	42	1
2. Radecki, T.	Trends in research on information-retrieval - the potential for improvements in conventional Boolean retrieval-systems	5	3
3. Belkin, N.J. & Croft, W.B.	Retrieval techniques	6	16
4. Maron, M.E.	Probabilistic design principles for conventional and full-text retrieval-systems	1	14
5. Kraft, D.H. & Buell, D.A.	Fuzzy-sets and generalized Boolean retrieval-systems	28	24

General Comparison

Taking all 200 items into account, the pennant algorithm yielded 96 relevant papers, while Related Records yielded 59, thereby supporting our conjecture. Put another way, Related Records performed much worse than pennant retrieval for a paper with two references, and no better for a paper with relatively many.

---

[3] For those unfamiliar with the convention, nonsensical characters are an old-fashioned way of comically indicating an unprintable curse-word.

Figure 6 charts the mean normalized recall ratios attained for M&K and Cooper. The papers co-cited with M&K were all judged specifically or generally relevant to it in topic, giving pennant retrieval perfect scores. Related Records was much less successful. Its ratios ranged between 25% and 54%, the average being 42%. In the case of Cooper, pennant retrieval was not quite as good as it was with M&K, but there was little difference in how it compared with Related Records. Both methods averaged 90% over the four cut-off points. The greater falloff of pennant retrieval at right happens because, as noted, the ratios penalize non-relevant papers that are relatively high-ranked. These outcomes evoke possible studies that use many more seeds to compare the two methods.
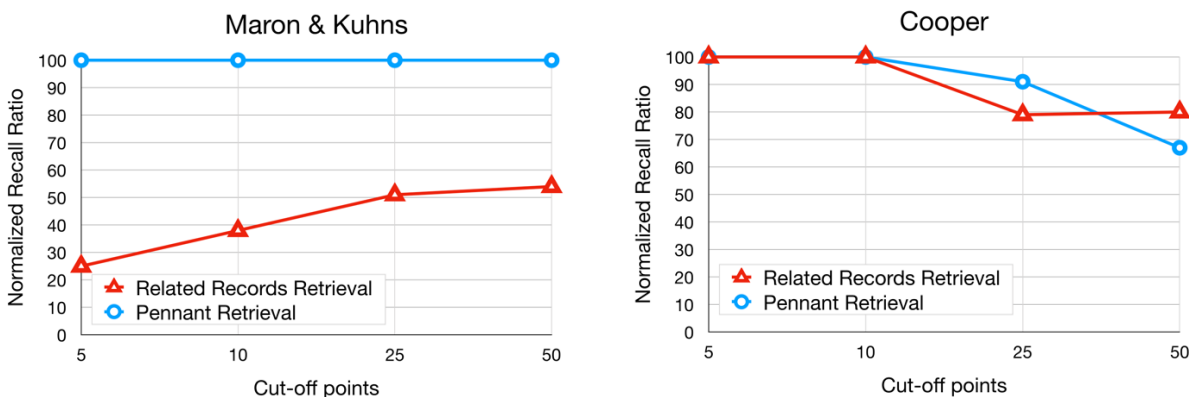

Fig. 6. Mean normalized recall ratios for the two seeds with Related Records retrieval and pennant retrieval

The WoS research areas of the papers retrieved by Related Records are more homogenous for Cooper (Table 3) than for M&K (Table 2). In other words, Cooper's 20 references brought greater topical focus to bibliographically coupled works. No comparable data exist for the pennant retrievals, but they generally appear to be well-focused for M&K in both the broad areas and the particular specialties we identified (Figure 4).

Our two methods retrieved relevant document sets with little overlap from the same collection. If full-scale tests confirm this effect, it would suggest the desirability of offering both bibliographic coupling and co-citation as retrieval methods in important databases such as WoS.

It further suggests that co-citation or pennant-style retrieval might be a way of adding to a pool of relevant records in IR experiments. For instance, in TREC experiments the top *n* items retrieved by different IR systems are aggregated in a pool so as to benefit from each system's different strengths. The assumption is that each system would position the greatest number of relevant results at the top of its output list. Hence, aggregating diverse output lists with their top *n* retrievals would increase the likelihood of creating a more complete collection for relevance judgments (Clough and Sanderson 2013).

Conclusions and Further Research

This is the first study to compare the pennant algorithm with the Related Records algorithm. Our main contribution is to suggest that pennant retrievals can be used successfully in search systems. For papers with few items in their reference lists, pennant retrievals using co-citation data may find many useful items. Tonta and Özkan Çelik (2013) replicated White's approach for a relatively old paper (1941) with only one reference and obtained promising results. Pennant retrieval moreover compared favorably with Related Records retrieval for a seed with numerous bibliographic coupling links to other papers.

The limitations of our study are plain: the quantitative findings cannot be used to estimate parameters, pennant retrieval is highly labor-intensive, and we have not considered the practicality and cost of incorporating it in operational IR systems such as WoS. Relevance assessments carried out by researchers rather than users with real information needs can be seen a limitation as well. However, as implied earlier, the classic IR testing method could be used if large sets of users supplied seed documents and then judged retrievals based on different bibliometric relationships, presumably using IR systems with interactive capabilities (Borlund and Ingwersen 1997).

It is likely that the choice of the first seed document in our exploratory case study has had an influence on the results because it had only two items in its reference list. Be that as it may, it should be pointed out that for the two seeds with both a few and many items in their reference lists, pennant retrieval seems to have provided comparable, if not

better, retrieval performance in both cases and found additional relevant records complementing those of Related Records retrieval. This suggests that pennant retrieval with its capability of finding different but potentially relevant documents can be used to improve Related Records retrieval, especially for older papers with sparse reference lists. Nonetheless, further research is needed to better understand how co-citation retrieval performs against bibliographic coupling retrieval when seed papers have both few and many items in their reference lists.

Our findings may provoke interest in integrating co-citation retrieval capabilities into current citation databases (e.g., WoS, Scopus, Google Scholar) and recommender systems (Carevic and Mayr 2014; Carevic and Schaer 2014; White 2018b). DialogClassic offered co-citation retrieval until a few years ago, and WoS has offered Cited Reference searching as one of its main search options for a long time. A new goal would be to complement retrievals based on bibliographic coupling with retrievals based on co-citation in WoS's huge network (see Clarivate Analytics (2019) for current coverage). Records could be ranked on the basis of their degree of co-citation with seed documents, thereby identifying and perhaps visualizing associations among papers in the database. Pennant-style weighting could be an option. The methods by which pennant diagrams might be incorporated in existing systems should be studied further, along with the scalability issues.

Contributions

Akbulut and Tonta: Conceptualization, methodology, software, formal analysis, and visualizations; original and
　　　　revised drafts
White: Review, editing and revisions, final draft

References

Ahmad, S., & Afzal, M. T. (2017). Combining co-citation and metadata for recommending more related papers. In 15th International Conference on Frontiers of Information Technology (FIT) (pp. 218-222). Islamabad, Pakistan: IEEE.

Akbulut, M. (2016a). *Atıf klasiklerinin etkisinin ve ilgililik sıralamalarının pennant diyagramları ile analizi [The analysis of the impact of citation classics and relevance rankings using pennant diagrams].* Yayımlanmamış yüksek lisans tezi, Hacettepe Üniversitesi, Ankara [Unpublished master's thesis, Hacettepe University, Ankara]. http://www.mugeakbulut.com/yayinlar/Muge_Akbulut_YL_Tez.pdf.

Akbulut, M. (2016b). Extended abstract: The analysis of the impact of citation classics and relevance rankings using pennant diagrams. http://www.mugeakbulut.com/yayinlar/tez_extended_abstract.pdf.

Åström, F. (2007). Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. *Journal of the American Society for Information Science and Technology*, *58*(7): 947–957.

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). Research-paper recommender systems: A literature survey. *International Journal on Digital Libraries*, *17*(4): 305-338.

Belter, C. W. (2017). A relevance ranking method for citation-based search results. *Scientometrics, 112*(2): 731–746.

Bensman, S. J. (2013, December 13). Eugene Garfield, Francis Narin, and PageRank: The theoretical bases of the Google Search Engine. https://arxiv.org/pdf/1312.3872.pdf.

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

Bitirim, Y., Tonta, Y., & Sever, H. (2002). Information retrieval effectiveness of Turkish search engines, *Lecture Notes in Computer Science*, 2457: 93-103.

Bollmann, P. (1983). The normalized recall and related measures. In *Proceedings of the 6th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '83)* (pp. 122-128). New York: ACM Press. https://doi.org/10.1145/511793.511811.

Borlund, P, & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3): 225-250.

Carevic, Z., & Mayr, P. (2014). Recommender systems using pennant diagrams in digital libraries. *(arXiv:1407.7276)*. NKOS Workshop London. http://arxiv.org/pdf/1407.7276.pdf.

Carevic, Z., & Schaer, P. (2014). On the connection between citation-based and topical relevance ranking: Results of a pretest using iSearch. In *Proceedings of the First Workshop on Bibliometric-enhanced Information Retrieval* (pp. 37-44). Amsterdam, The Netherlands. http://ceur-ws.org/Vol-1143/paper5.pdf.

Clarivate Analytics (2017a). Related Records. https://images.webofknowledge.com/images/help/WOS/hp_related_records.html.

Clarivate Analytics (2017b). Research Areas (Categories/Classification).
https://images.webofknowledge.com/images/help/WOS/hp_research_areas_easca.html.

Clarivate Analytics (2018a). Advanced Search Examples.
https://images.webofknowledge.com/images/help/WOS/hp_advanced_examples.html

Clarivate Analytics (2018b). Research Area Schemes.
http://help.incites.clarivate.com/inCites2Live/filterValuesGroup/researchAreaSchema.html.

Clarivate Analytics (2019). Web of Science platform: Web of Science: Summary of Coverage
https://clarivate.libguides.com/webofscienceplatform/coverage

Clough, P., & Sanderson, M. (2013). Evaluating the performance of information retrieval systems using test collections. *Information Research*, 18(2) paper 582. http://InformationR.net/ir/18-2/paper582.html

Colavizza, G., Boyack, K. W., Van Eck, N. J., & Waltman, L. (2018). The closer the better: Similarity of publication pairs at different cocitation levels. *Journal of the Association for Information Science & Technology*, *69*(4): 600-609.

Cooper, W. S. (1988). Getting beyond Boole. *Information Processing & Management*, *24*(3), 243-248.

Cooper, W. S., & Maron, M. E. (1978). Foundations of probabilistic and utility-theoretic indexing. *Journal of the ACM, 25*(1): 67–80.

Croft, W. B., & Harper, D. J. (1979). Using probabilistic models of document retrieval without relevance information. *Journal of Documentation, 35*(4): 285-295.

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

Fuhr, N. (1989). Models for retrieval with probabilistic indexing. *Information Processing & Management, 22*(1): 55–72.

Garfield, E. (2001). From bibliographic coupling to co-citation analysis via algorithmic historio-bibliography. A citationist's tribute to Belver C. Griffith. Paper presented at Drexel University, Philadelphia, PA. http://garfield.library.upenn.edu/papers/drexelbelvergriffith92001.pdf.

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). CiteSeer: An automatic citation indexing system. In *Digital Libraries – Third ACM Conference on Digital Libraries* (pp. 89-98). Ed. by I. Witten et al. New York: ACM Press. https://clgiles.ist.psu.edu/papers/DL-1998-citeseer.pdf.

Haruna, K., Ismail, M. A., Bichi, A. B., Chang, V., Wibawa, S., & Herawan, T. (2018). A citation-based recommender system for scholarly paper recommendation. In O. Gervasi et al. (Eds.). *International Conference on Computational Science and Its Applications, ICCSA 2018*, LNCS 10960 pp 514-525).

He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, L. (2010). Context-aware citation recommendation. In *Proceedings of the 19th International Conference on World Wide Web (Raleigh, North Carolina, USA, April 26 - 30, 2010. WWW '10. ACM, New York, NY*, pp. 421-430. http://www.cse.psu.edu/~duk17/papers/citationrecommendation.pdf.

Hiemstra, D. (2000). A probabilistic justification for using tf×idf term weighting in information retrieval. *International Journal on Digital Libraries*, *3*(2): 131-139.

Horsley, T., Dingwall, O., & Sampson, M. (2011). Checking reference lists to find additional studies for systematic reviews. *Cochrane Database Systems Review*, *1*(6). 10.1002/14651858.MR000026.pub2.

Huang, S., Xue, G-R., Zhang, B-Y., Chen, Z., Yu, Y., & Ma, W-Y. (2004). TSSP: A reinforcement algorithm to find related papers. In: WI 2004, Washington, DC, USA (pp. 117–123). IEEE Computer Society, Los Alamitos. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1410792&tag=1

Kraft, D. H. (1985). Advances in information retrieval: Where is that /#*&@¢ record? In Yovits, M. C. (Ed.)*, Advances in Computers*, (pp. 277-318).

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1): 10-25.

Liang, Y., Li, Q., & Qian, T. (2011). Finding relevant papers based on citation relations. In *Proceedings of the International Conference on Web-Age Information Management* (pp. 403–414). Berlin: Springer.

Lin, J., & Wilbur, W. J. (2007). PubMed related articles: A probabilistic topic-based model for content similarity. *BMC Bioinformatics*, *8*: 423 10.1186/1471-2105-8-423.

Manning, C., & Schütze, H. (2000). *Foundations of statistical natural language processing.* (2nd printing). Cambridge, MA: MIT Press. http://ics.upjs.sk/~pero/web/documents/pillar/Manning_Schuetze_StatisticalNLP.pdf.

Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science, 28*(1): 38–43.

Maron, M.E. (1988). Probabilistic design principles for conventional and full-text retrieval systems. *Information Processing and Management*, *24*(3), 249-255.

Maron, M. E. (2008). An historical note on the origins of probabilistic indexing. *Information Processing & Management*, *44*(2): 971-972.

Maron, M. E., & Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. *Journal of the ACM, 7*(3): 216-244.

Peterson, G., & Graves, R. S. (2009). How similar is similar? An evaluation of "related records" applications among health literature portals. *Proceedings of the Association for Information Science & Technology*, *46*(1): 1-3.

Prevedelli, D., Simonini, R., & Ansaloni, I. (2001). Relationship of non-specific commensalism in the colonization of the deep layers of sediment. *Journal of the Marine Biological Association of the United Kingdom*, *81*(6): 897-901. 10.1017/S0025315401004817.

Robertson, S. E. (1977). The probability ranking principle in IR. *Journal of Documentation*, *33*(4): 294–304. http://parnec.nuaa.edu.cn/xtan/IIR/readings/jdRobertson1977.pdf.

Robertson, S. E., & Sparck Jones, K. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science*, *27*(3): 129-146.

Robertson, S. E., Maron, M.E., & Cooper, W.S. (1982). Probability of relevance: A unification of two competing models for document retrieval. *Information Technology: Research and Development, 1*(1)*:* 1–21.

Salton, G., & Yang, C. S. (1973). On the specification of term values in automatic indexing. *Journal of Documentation*, *29*(4): 351-372.

Scopus. (2018). References and related documents. https://service.elsevier.com/app/answers/detail/a_id/14190/supporthub/scopus/ Accessed April, 25, 2019.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, Ill: University of Illinois Press.

Shen, S., Zhu, D., Rousseau, R., Su, X., & Wang, D. (2019). A refined method for computing bibliographic coupling strengths. *Journal of Informetrics*, 13: 605-615.

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.

Smith, C. H., Georges, P., & Nguyen, N. (2015). Statistical tests for 'related records' search results. *Scientometrics*, *105*(3): 1665-1677.

Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). Outsmart your own biases. *Harvard Business Review*, *93*(5): 64-71.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application to retrieval. *Journal of Documentation*, 28(1), 11-21.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. (2d ed.) Oxford: Blackwell.

Sugimoto, C. R., & Larivière, V. (2018). *Measuring research: What everyone needs to know,* New York: Oxford University Press.

Swanson, D.R. (1986). Subjective versus objective relevance in bibliographic retrieval systems. *The Library Quarterly*, 56(4): 389-398.

Thompson, P. (1990a). A combination of expert opinion approach to probabilistic information retrieval, Part 1: The conceptual model. *Information Processing & Management, 26*(3): 371–382.

Thompson, P. (1990b). A combination of expert opinion approach to probabilistic information retrieval, Part 2: Mathematical treatment of CEO model 3. *Information Processing & Management*, *26*(3): 383–394.

Thompson, P. (2007). Looking back: On relevance, probabilistic indexing and information retrieval. *Information Processing & Management*, *44*(2): 963-970.

Tonta, Y., & Özkan Çelik, A. E. (2013). Cahit Arf: Exploring his scientific influence using social network analysis, author co-citation maps and single publication h index. *Journal of Scientometric Research*, *2*(1): 37-51.

Wesley-Smith, I., Bergstrom, C. T., & West, J. D. (2016). Static ranking of scholarly papers using article-level eigenfactor (ALEF). *The 9th ACM International Conference on Web Search and Data Mining (WSDM). February 22–25, 2016, San Francisco, CA, USA.* http://octavia.zoology.washington.edu/publications/WesleySmithEtAl16.pdf.

White, H. D. (2007a). Combining bibliometrics, information retrieval, and relevance theory. Part 1: First examples of a synthesis. *Journal of the American Society for Information Science and Technology, 58*(4): 536-559.

White, H. D. (2007b). Combining bibliometrics, information retrieval, and relevance theory. Part 2: Some implications for information science. *Journal of the American Society for Information Science and Technology, 58*(4): 583-605.

White, H. D. (2009). Pennants for Strindberg and Persson. Celebrating scholarly communication studies: A festschrift for Olle Persson at his 60th birthday. *ISSI Newsletter*, Vol. 5-S. pp. 71-83. http://portal.research.lu.se/portal/files/5902071/1458992.pdf.

White, H. D. (2010). Some new tests of relevance theory in information science. *Scientometrics*, *83*(3): 653–667.

White, H. D. (2015). Co-cited author retrieval and relevance theory: Examples from the humanities. *Scientometrics*, *102*(3): 2275-2299.

White, H. D. (2018a). Pennants for Garfield: Bibliometrics and document retrieval. *Scientometrics, 114*(2): 757-778.

White, H. D. (2018b). Bag of works retrieval: TF*IDF weighting of works co-cited with a seed. *International Journal of Digital Libraries, 19*(2-3): 139-149.

White, H. D., & Mayr, P. (2013). Pennants for descriptors. In NKOS Workshop 2013. Valletta, Malta. http://arxiv.org/abs/1310.3808.

Wilson, D., & Sperber, D. (2002). Relevance theory. In Ward, G., & Horn, L. (Eds.) *Handbook of pragmatics.* Oxford: Blackwell. https://www.dan.sperber.fr/?p=93

Yao, Y. (1995). Measuring retrieval effectiveness based on user preference of documents. *Journal of the American Society for Information Science*, *46*(2): 133–145.

Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, *75*(6): 579-652.

Zarrinkalam, F., & Kahani, M. (2012). A new metric for measuring relatedness of scientific papers based on non-textual features, *Intelligent Information Management*, *4*(4): 99-107. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.6934&rep=rep1&type=pdf.

Appendix 1. Top 50 (a) Related Records and (b) pennant retrievals for Maron and Kuhns*

Rank	(a) Relevance Rankings Based on WoS Related Records	R** or NR
Seed	Maron, M. E., & Kuhns, J.L. (1960). On relevance, probabilistic indexing and information retrieval. *JACM*	
1	Prevedelli, D., et al. (2001). Relationship of non-specific commensalism in the colonization of the deep layers of sediment. *JMBA, UK*	NR
2	Cui, Y., et al. (2013). A novel mobile device user interface with integrated social networking services. *Int J Human-Computer Studies*	NR
3	Zhang, W., et al. (2017). Significant prognostic features and patterns of somatic TP53 mutations in human cancers. *Cancer Informatics*	NR
4	Todeschini, R., et al. (2012). Similarity coefficients for binary chemoinformatics data… *JCIM*	R
5	Cicchetti, D.V., et al. (1991). Establishing the reliability and validity of neuropsychological disorders with low base rates… *JCEN*	NR
6	Gangemi, G. (1978). Reviewing statistical measures of association for triple dichotomy. *Quality & Quantity*	R
7	Wilson, E.B. (1964). Comparative experiment and observed association, II. *PNAS, USA*	R
8	Immer, F.R. (1931). The efficiency of the correlation coefficient for estimating linkage intensities. *American Naturalist*	R
9	Hagenaars, J.A. (2015). Methodological issues in categorical data analysis. *Methodology*	NR
10	Baraniuk, S., et al. (2012). Comparison of global statistical test and composite outcome for secondary analyses of...heart disease outcomes. *Prog Cardiovasc Dis*	NR
11	Rousson, V. (2007). The gamma coefficient revisited. *Stat & Probab Lett*	R
12	McKinlay, S.M. (1975). Design and analysis of observational study. *JASA*	NR
13	Celotto, E. (2017). Visualizing the behavior and some symmetry properties of Bayesian confirmation measures. *Data Min Knowl Discov*	NR
14	Liao, S.G., et al. (2014). Missing value imputation in high-dimensional phenomic data… *BMC Bioinformatics*	NR
15	Malik, S., et al. (2013). An improved estimator using two auxiliary attributes. *App Math and Comp*	NR
16	Hong, C.S., et al. (2011). Mutual information and redundancy for categorical data. *Stat Pap*	NR
17	Greenstreet, J.E.S., et al. (1957). The existence of differences between the infra-red absorption spectra of bacteria. *Spectrochimica Acta*	NR
18	Miklós-Thal, J. (2012). Linking reputations through umbrella branding. *Quant Mark & Econ*	NR
19	Ke, Y., et al. (2008). Efficient correlation search from graph databases. *IEEE TKDE*	R
20	Bonett, D.G., et al. (2007). Statistical inference for generalized Yule coefficients in 2*2 contingency tables. *SMR*	NR
21	Walter, S.D. (2001). Hoehler's adjusted kappa is equivalent to Yule's Y. *J of Clinical Epidemiology*	NR
22	Goodman, L.A. (1996). A single general method for the analysis of cross-classified data… *JASA*	R
23	Sebena, R., et al. (2011). Selected sociodemographic factors and related differences in patterns of alcohol use among university students in Slovakia. *BMC Public Health*	NR
24	Mihulka, S., et al. (2006). Invasiveness of Oenothera congeners alien to Europe… *PPEES*	NR
25	Tan, P.N., et al. (2004). Selecting the right objective measure for association analysis. *Inf Syst*	R
26	O'Gorman, T.W., et al. (1990). The stability of several measures of association in small contingency tables. *Communications in Statistics—Theory and Methods*	R
27	Saxén, I. (1975). The association between maternal influenza, drug consumption and oral clefts. *Acta Odontologica Scandinavica*	NR
28	Blömeke, S., et al. (2017). Process mediates structure: The relation between preschool teacher education and preschool teachers' knowledge. *J Educ Psyc*	NR
29	Bruckner, C.T., et al. (2006). Interpreting kappa in observational research: Baserate matters. *AJMR*	NR
30	Jackson, D.A., et al. (1989). Similarity coefficients: Measures of co-occurrence and association or simply measures of occurrence? *American Naturalist*	R
31	Schwab, M.R. (1976). Interpersonal compatibility and student-faculty evaluations: The effects of need complementarity in the classroom. *Sociological Focus*	NR
32	Lucia, L., et al. (2014). Extended comprehensive study of association measures for fault localization. *J Software: Evolution and Proc*	R
33	Gilbert, G.H., et al. (2001). Twenty-four month incidence of root caries among a diverse group of adults. *Caries Res*	NR
34	Li, R., et al. (1996). Reproducibility of extracranial carotid atherosclerotic lesions assessed by B-mode ultrasound… *Ultrasound in Med & Biology*	NR
35	Holland, P.W., et al. (1987). Dependence function for continuous bivariate densities. *Comm in Statistics—Theory and Methods*	NR
36	Goodman, L.A., et al. (1954). Measures of association for cross classifications. *JASA*	R

37	Clements, F. (1931). Plains Indian tribal correlations with sun dance data. *American Anthropologist*	NR
38	Egozcue, J.J., et al. (2015). Independence in contingency tables using simplicial geometry. *Comm in Statistics—Theory and Methods*	NR
39	Ke, Y., et al. (2007). Correlation search in graph databases. *Proc 13th Conf ACM SIGKDD*	R
40	MacKenzie, D. (1999). The science wars and the past's quiet voices. *Soc Stud of Sci*	NR
41	Shi, G.R. (1993). Multivariate data analysis in palaeoecology and palaeobiogeography—A review. *Palaeo3*	NR
42	Cicchetti, D.V., et al. (1990). High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epid*	NR
43	Koch, G.G. (1969). The effect of non-sampling errors on measures of association in 2×2 contingency tables. *JASA*	NR
44	Lerman, I.C. (2016). Comparing attributes by a probabilistic and statistical association II. *Foundations and Methods in Combinatorial and Statistical Data Analysis and Clustering*	NR
45	Albatineh, A.N., et al. (2012). On the equivalence of some indices of similarity: Implication for binary presence/absence data. *ANZJS*	R
46	Duffy, K.R., et al. (2012). Activation-induced B cell fates are selected by intracellular stochastic competition. *Science*	NR
47	House, A.E., et al. (1981). Measures of interobserver agreement: Calculation formulas and distribution effects. *JPBA*	NR
48	McKinlay, S.M. (1978). The effect of nonzero second-order interaction on combined estimators of the odds ratio. *Biometrika*	NR
49	Dible, J.H. (1921). The enterococcus and the faecal streptococci: Their properties and relations. *J Pathology and Bacteriology*	NR
50	Tushkanova, O. (2015). Comparative analysis of the numerical measures for mining associative and causal relationships in big data. *Tech & Data Sci: First Conf Proc*	R

Rank	(b) Relevance Rankings Based on Pennant Retrievals	TF*IDF	R or NR
Seed	Maron, M. E., & Kuhns, J.L. (1960). On relevance, probabilistic indexing and information retrieval. *JACM*	14.7	
1	Cooper, et al. (1978). Foundations of probabilistic and utility-theoretic indexing. *JACM*	12.8	R
2	Robertson, et al. (1982). Probability of relevance: A unification of two competing models for document retrieval. *IT: Res and Dev*	12.3	R
3	Robertson, S.E. (1977). The probability ranking principle in IR. *JDoc*	12.0	R
4	Robertson, S.E., et al. (1976). Relevance weighting of search terms. *JASIS*	11.6	R
5	Croft, W.B., et al. (1979). Using probabilistic models of document retrieval without relevance information. *JDoc*	11.6	R
6	Robertson, S.E. (1994). Documentation note: Query-document symmetry and dual models. *JDoc*	11.4	R
7	Salton, G., et al. (1983). Extended Boolean information retrieval. *CACM*	11.3	R
8	Bookstein, A., et al. (1977). Operations research applied to document indexing and retrieval decisions. *JACM*	11.3	R
9	Winkler, R.L. (1968). The consensus of subjective probability distributions. *Management Science*	11.0	R
10	Cooper, W.S. (1978). Indexing documents by gedanken experimentation. *JASIS*	11.0	R
11	Bookstein, A. (1983). Outline of a general probabilistic retrieval model. *JDoc*	10.9	R
12	Bookstein, A., et al. (1976). A general mathematical model for information retrieval systems. *LQ*	10.7	R
13	Luhn, H.P. (1997). The automatic derivation of information retrieval encodements from machine-readable texts. *Readings in Information Retrieval*	10.7	R
14	Fuhr, N. (1989). Models for retrieval with probabilistic indexing. *IPM*	10.7	R
15	Bookstein, A. (1978). On the perils of merging Boolean and weighted retrieval systems. *JASIS*	10.6	R
16	Croft, B., et al. (Eds.). (2013). *Language modeling for information retrieval*	10.4	R
17	Bookstein, A. (1979). Relevance. *JASIS*	10.4	R
18	Swanson, D.R. (1960). Searching natural language text by computer. *Science*	10.4	R
19	Maron, M.E. (1977). On indexing, retrieval and the meaning of about. *JASIS*	10.2	R
20	Turtle, H., et al. (1991). Evaluation of an inference network-based retrieval model. *ACM TOIS*	10.2	R
21	Miller, W.L. (1971). A probabilistic search strategy for MEDLARS. *JDoc*	10.0	R
22	Thompson, P. (1988). Subjective probability and information retrieval: A review of the psychological literature. *JDoc*	9.9	R
23	Bodoff, D., et al. (2004). A new unified probabilistic model. *JASIST*	9.8	R
24	Salton, G., et al. (1975). A theory of term importance in automatic text analysis. *JASIS*	9.7	R
25	Bookstein, A. (1985). Probability and fuzzy-set applications to information retrieval. *ARIST*	9.7	R
26	Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *JASIS*	9.5	R

27	Wong, S.K.M., et al. (1995). On modeling information retrieval with probabilistic inference. *ACM TOIS*	9.5	R
28	Maron, M.E., et al. (1986). An inductive search system: Theory, design, and implementation. *IEEE TSMC*	9.5	R
29	Bush, V. (1945). As we may think. *Atlantic Monthly*	9.5	R
30	Blair, D.C., et al. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. *CACM*	9.4	R
31	Salton, G., et al. (1973). On the specification of term values in automatic indexing. *JDoc*	9.4	R
32	dos Santos, E.B., et al. (2011). Bayesian network classifiers: Beyond classification accuracy. *Intell Data Anal*	9.3	R
33	Gordon, M. (1988). Probabilistic and genetic algorithms in document retrieval. *CACM*	9.3	R
34	Swanson, D.R. (1977). Information retrieval as a trial-and-error process. *LQ*	9.3	R
35	Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *JDoc*	9.2	R
36	Thompson, P. (1990). A combination of expert opinion approach to probabilistic information retrieval, part 1: The conceptual model. *IPM*	9.0	R
37	Fuhr, N. (1986). Two models of retrieval with probabilistic indexing. *Proc 9th ACM SIGIR Conf on Res and Dev in IR*	9.0	R
38	Van Rijsbergen, et al. (1981). The selection of good search terms. *IPM*	9.0	R
39	Bodoff, D. (1997). A re-unification of two competing models for document retrieval. *JASIS*	8.9	R
40	Thompson, P. (1990). A combination of expert opinion approach to probabilistic information retrieval, part 2: Mathematical treatment of CEO model 3. *IPM*	8.9	R
41	Cleverdon, C.W. (1962). *Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems.*	8.2	R
42	Salton, G., et al. (1975). A vector space model for automatic indexing. *CACM*	8.2	R
43	Salton, G., et al. (1968). Computer evaluation of indexing and text processing. *JACM*	8.1	R
44	Luk, R.W. (2008). On event space and rank equivalence between probabilistic retrieval models. *Inf Ret*	8.1	R
45	Kleinberg, J.M. (1999). Authoritative sources in a hyperlinked environment. *JACM*	7.9	R
46	Maron, M.E. (1984). Probabilistic retrieval models. *Prog in Comm Sci*	7.8	R
47	Porter, M.F. (1980). An algorithm for suffix stripping. *Program*	7.6	R
48	Salton, G., et al. (1976). Automatic indexing using term discrimination and term precision measurements. *IPM*	7.5	R
49	Chow, C., et al. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Trans Inf Theory*	7.4	R
50	Croft, W.B. (2002). Combining approaches to information retrieval. *Advances in IR*	7.1	R

*Full bibliographic records are available at https://goo.gl/fBzvS1
** R for "relevant", NR for "non-relevant"

Appendix 2. Top 50 (a) Related Records and (b) pennant retrievals for Cooper*

Rank	(a) Relevance Rankings Based on WoS Related Records	R** or NR
Seed	Cooper, W.S. (1988). Getting beyond Boole. *IPM*	
1	Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *IPM***	R
2	Buell, D.A., & Kraft, D.H. (1981). Threshold values and Boolean retrieval systems. *IPM*	R
3	Radecki, T. (1988). Trends in research on information retrieval—The potential for improvements in conventional Boolean retrieval systems. *IPM*	R
4	Salton, G., Fox, E.A., & Wu, H. (1983). Extended Boolean information retrieval. *CACM*	R
5	Chen, S.J. (2011). Fuzzy information retrieval based on a new similarity measure of generalized fuzzy numbers. *Intel Automat & Soft Comput*	R
6	Fox, E.A., & Koll, M.B. (1988). Practical enhanced Boolean retrieval: Experiences with the SMART and SIRE systems. *IPM*	R
7	Waller, W.G., & Kraft, D.H. (1979). A mathematical model of a weighted Boolean retrieval system. *IPM*	R
8	Bookstein, A. (1980). Fuzzy requests: An approach to weighted Boolean searches. *JASIST*	R
9	Kerre, E.E., Zenner, R.B., & De Caluwe, R.M. (1986). The use of fuzzy set theory in information retrieval and databases: A survey. *JASIST*	R
10	Kwok, K.L. (1988). On the use of bibliographically related titles for the enhancement of document representations. *IPM*	R
11	Doszkocs, T.E. (1986). Natural language processing in information retrieval. *JASIST*	R
12	Perry, S.A., & Willett, P. (1983). A review of the use of inverted files for best match searching in information retrieval systems. *JIS*	R
13	Biswas, G., Bezdek, J.C., Subramanian, V., & Marques, M. (1987). Knowledge-assisted document-retrieval: II. The retrieval process. *JASIST*	R
14	Maron, M.E. (1988). Probabilistic design principles for conventional and full-text retrieval systems. *IPM*	R
15	Bookstein, A. (1978). On the perils of merging Boolean and weighted retrieval systems. *JASIST*	R
16	Belkin, N.J., & Croft, W.B. (1987). Retrieval techniques. *ARIST*	R
17	Wong, S.K.M., Ziarko, W., Raghavan, V.V., & Wong, P.C.N. (1989). Extended Boolean query processing in the generalized vector space model. *Inf Systems*	R
18	Bolc, L., Kowalski, A., Kozlowska, M., & Strzalkowski, T. (1985). A natural language information retrieval system... *Int J Man Mach Stud*	R
19	Buell, D.A., & Kraft, D.H. (1981). A model for a weighted retrieval system. *JASIST*	R
20	Van Pulis, N., & Ludy, L.E. (1988). Subject searching in an online catalog with authority control. *College & Research Libraries*	NR
21	Markey, K. (1988). Integrating the machine-readable LCSH into online catalogs. *ITL*	R
22	Cater, S.C., & Kraft, D.H. (1989). A generalization and clarification of the Waller-Kraft wish list. *IPM*	R
23	Radecki, T. (1979). Fuzzy set theoretical approach to document retrieval. IPM	R
24	Kraft, D.H., & Buell, D.A. (1983). Fuzzy sets and generalized Boolean retrieval systems. *Int J Man Mach Stud*	R
25	Noreault, T., Koll, M., & McGill, M.J. (1977). Automatic ranked output from Boolean searches in SIRE. *JASIST*	R
26	Negoiță, C.V. (1973). On the application of the fuzzy sets separation theorem for automatic classification in information retrieval systems. *Information Sciences*	R
27	Salton, G. (1986). Another look at automatic text-retrieval systems. *CACM*	R
28	Turtle, H., & Croft, W.B. (1991). Evaluation of an inference network-based retrieval model. ACM *TOIS*	R
29	McCarthy, C. (1986). The reliability factor in subject access.	R
30	Al-Hawamdeh, S., de Vere, R., Smith, G., & Willett, P. (1991). Using nearest-neighbour searching techniques to access full-text documents. *Online Review*	R
31	Byrne, A., & Micco, M. (1988). Improving OPAC subject access: The ADFA experiment. *College & Research Libraries*	R
32	Blackshaw, L., & Fischhoff, B. (1988). Decision making in online searching. *JASIST*	R
33	Croft, W.B., & Savino, P. (1988). Implementing ranking strategies using text signatures. *ACM TOIS)*	R
34	Losee, R. (1987). Probabilistic retrieval and coordination level matching. *JASIST*	R
35	Lipetz, B.A., & Paulson, P. J. (1987). A study of the impact of introducing an online subject catalog at the New York State Library.	NR
36	Robertson, S.E. (1986). On relevance weight estimation and query expansion. *JDoc*	<u>R</u>
37	Lucarella, D. (1988). A document retrieval system based on nearest neighbour searching. JIS	R

38	Saracevic, T. (2008). Effects of inconsistent relevance judgments on information retrieval test results: A historical perspective. *Library Trends*	NR
39	Robertson, S.E. (1978). On the nature of fuzz: A diatribe. *JASIST*	R
40	Kantor, P. B. (1981). Logic of weighted queries. *IEEE Trans Syst, Man, & Cybern*	R
41	Tahani, V. (1976). A fuzzy model of document retrieval systems. *IPM*	R
42	Mischo, W.H. & Jounghyoun, L. (1987). End-user searching of bibliographic databases. *ARIST*	R
43	Larson, R.R., & Graham, V. (1983). Monitoring and evaluating MELVYL. *Inf Tech & Lib*	NR
44	Robertson, S.E. (1977). The probability ranking principle in IR. *JDoc*	R
45	Wong, W.Y.P., & Lee, D. L. (1990). Signature file methods for implementing a ranking strategy. *IPM*	R
46	Kochtanek, T.R. (1982). Bibliographic compilation using reference and citation links. *IPM*	R
47	Kaske, N.K. (1988). A comparative study of subject searching in an OPAC among branch libraries of a university library system. *Inf Tech & Lib*	NR
48	Blazek, R., & Bilal, D. (1988). Problems with OPAC: a case study of an academic research library. *RQ*	NR
49	Cohen, P.R., & Kjeldsen, R. (1987). Information retrieval by constrained spreading activation in semantic networks. *IPM*	R
50	Bordogna, G., Carrara, P., & Pasi, G. (1991). Query term weights as constraints in fuzzy information retrieval. *IPM*	R

Rank	(b) Relevance Rankings Based on Pennant Retrievals	TF*IDF	R or NR
Seed	Cooper, W.S. (1988). Getting beyond Boole. *IPM*	13.2	
1	Maron, M.E. (1988). Probabilistic design principles for conventional and full-text retrieval systems. *IPM***	13.2	R
2	Bookstein, A. (1985). Probability and fuzzy-set applications to information retrieval. *ARIST*	9.9	R
3	Radecki, T. (1988). Probabilistic methods for ranking output documents in conventional Boolean retrieval systems. *IPM*	9.2	R
4	Kraft, D.H. (1985). Advances in information-retrieval: Where is that /#*&@¢ record? *Advances in Computers*	9.0	R
5	Radecki, T. (1988). Trends in research on information retrieval—The potential for improvements in conventional Boolean retrieval systems. *IPM*	9.0	R
6	Belkin, N.J., & Croft, W.B. (1987). Retrieval techniques. *ARIST*	8.8	R
7	Mohan, K.C. (1993). Free-text retrieval-systems—Research and development in information retrieval. *JSIR*	8.8	R
8	Taniguchi, S. (1992). Progress of the research of information retrieval models in the 80's: A critical review. *LIS*	8.7	R
9	Boyce, B.R., & Kraft, D. H. (1985). Principles and theories in information science. *ARIST*	8.6	R
10	Maron, M.E., Curry, S., & Thompson, P. (1986). An inductive search system: Theory, design, and implementation. *IEEE Trans Cybern*	8.6	R
11	Lavrenko, V. (2008). *A generative theory of relevance* (Vol. 26). Springer Science & Business Media.	8.6	R
12	Cooper, W.S. (1983). Exploiting the maximum entropy principle to increase retrieval effectiveness. *JASIST*	8.5	R
13	Salton, G. (1991). Developments in automatic text retrieval. S*cience*	8.5	R
14	Kantor, P.B. (1984). Maximum entropy and the optimal design of automated information retrieval systems. *NITRD*	8.4	R
15	Heine, M.H. (1984). Information retrieval from classical databases from a signal-detection standpoint—A review. *Inform Technol R&D*	8.4	R
16	Fuhr, N. (1992). Probabilistic models in information retrieval. *Comput J*	8.3	R
17	Dominich, S. (2000). A unified mathematical definition of classical information retrieval. *JASIST*	8.3	R
18	Greiff, W.R., & Ponte, J.M. (2000). The maximum entropy approach and probabilistic IR models. *ACM TOIS*	8.3	R
19	Cooper, W.S. (1995). Some inconsistencies and misidentified modeling assumptions in probabilistic information retrieval. *ACM TOIS*	8.3	R
20	Wallace, D.P., Boyce, B.R., & Kraft, D.H. (1988). Estimating effective display size in online retrieval	8.1	NR

	systems. *IPM*		
21	Radecki, T. (1983). Incorporation of relevance feedback into Boolean retrieval systems. In *Proceedings of the 5th annual ACM SIGIR conference*	8.1	R
22	Panyr, J. (1987). Vector-space model and cluster-analysis in information retrieval systems. *Nachrichten für Dokumentation*	8.1	R
23	Kwok, K.L. (1985). A probabilistic theory of indexing using author-provided relevance information. *ASZS'85: Proceedings of the 48th ASIS annual meeting*	8.1	R
24	Panyr, J. (1987). Interactive retrieval strategies—Relevance feedback. *Nachrichten für Dokumentation*	7.9	R
25	Ding, Y., Chowdhury, G., & Foo, S. (1999). Mapping the intellectual structure of information retrieval studies: An author co-citation analysis, 1987-1997. *JIS*	7.9	NR
26	Kwok, K.L. (1985). A probabilistic theory of indexing and similarity measure based on cited and citing documents. *JASIST*	7.9	R
27	Lenk, P.J., & Floyd, B.D. (1988). Dynamically updating relevance judgements in probabilistic information systems via users' feedback. *Management Science*	7.9	R
28	Kraft, D.H., & Buell, D.A. (1983). Fuzzy sets and generalized Boolean retrieval systems. *Int J Man Mach Stud*	7.8	R
29	Thompson, P. (2008). Looking back: On relevance, probabilistic indexing and information retrieval. *IPM*	7.8	R
30	Crestani, F., Lalmas, M., Van Rijsbergen, C.J., & Campbell, I. (1998). "Is this document relevant?…" *ACM CSUR*	7.8	R
31	Schocken, S., & Hummel, R.A. (1993). On the use of the Dempster- Shafer model in information indexing and retrieval applications. *Int J Man Mach Stud*	7.8	R
32	Fuhr, N., & Hüther, H. (1989). Optimum probability estimation from empirical distributions. *IPM*	7.7	R
33	Gordon, M. (1988). Probabilistic and genetic algorithms in document retrieval. *CACM*	7.7	R
34	Deogun, J.S., & Raghavan, V.V. (1988). Integration of information retrieval and database management systems. *IPM*	7.7	R
35	Knorz, G. (1982). A decision theory approach to optimal automatic indexing. In *Proceedings of the 5th annual ACM SIGIR conference*	7.7	R
36	Salton, G., Wu, H., & Yu, C.T. (1981). The measurement of term importance in automatic indexing. *JASIST*	7.6	<u>R</u>
37	Kantor, P.B. (1994). Information retrieval techniques. *ARIST*	7.6	R
38	Yang, J.J., & Korfhage, R.R. (1992). Query modification using genetic algorithms in vector space models.	7.6	R
39	Wong, S.K.M., & Yao, Y.Y. (1991). A probabilistic inference model for information retrieval. *Information Systems*	7.6	R
40	Thompson, P. (1990). A combination of expert opinion approach to probabilistic information retrieval, part 1: The conceptual model. *IPM*	7.6	R
41	Salton, G., & Buckley, C. (1988). Parallel text search methods. *CACM*	7.6	R
42	Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *IPM*	7.6	R
43	Schwartz, C., & Eisenmann, L.M. (1986). Subject analysis. *ARIST*	7.5	NR
44	Zhai, C., & Lafferty, J. (2006). A risk minimization framework for information retrieval. *IPM*	7.5	R
45	Ding, Y. (1998). Visualization of intellectual structure in information retrieval: Author cocitation analysis. *Int Forum Inf & Docum*	7.5	NR
46	Losee, R.M. (1988). Parameter estimation for probabilistic document-retrieval models. *JASIST*	7.4	R
47	Kwok, K.L. (1986). The concept of document components for probabilistic indexing. *Proc of the 49th Annual Meeting of ASIS*	7.4	R
48	Driyanskii, V.M. (1981). Retrieval models in on-line documentary information systems: An analytic review. *Cybern Syst Analysis*	7.4	R
49	Agosti, M., Crestani, F., & Melucci, M. (1996). Design and implementation of a tool for the automatic construction of hypertexts for information retrieval. *IPM*	7.4	R
50	Kantor, P.B., & Lee, J.J. (1998). Testing the maximum entropy principle for information retrieval. *JASIST*	7.4	R

* Full bibliographic records are available at https://goo.gl/fBzvS1
** R for "relevant", NR for "non-relevant"
*** Cells in gray represent overlapping records in the two lists in Appendix 2(a) and 2(b).