Int J Digit Libr
DOI 10.1007/s00799-015-0156-0

Research-paper recommender systems: a literature survey

Joeran Beel[1] · **Bela Gipp**[2] · **Stefan Langer**[3] · **Corinna Breitinger**[4]

Received: 24 February 2014 / Revised: 10 June 2015 / Accepted: 22 June 2015

Abstract In the last 16 years, more than 200 research articles were published about *research-paper recommender systems*. We reviewed these articles and present some descriptive statistics in this paper, as well as a discussion about the major advancements and shortcomings and an overview of the most common recommendation concepts and approaches. We found that more than half of the recommendation approaches applied content-based filtering (55 %). Collaborative filtering was applied by only 18 % of the reviewed approaches, and graph-based recommendations by 16 %. Other recommendation concepts included stereotyping, item-centric recommendations, and hybrid recommendations. The content-based filtering approaches mainly utilized papers that the users had authored, tagged, browsed, or downloaded. TF-IDF was the most frequently applied weighting scheme. In addition to simple terms, n-grams, topics, and citations were utilized to model users' information needs. Our review revealed some shortcomings of the current research. First, it remains unclear which recommendation concepts and approaches are the most promising. For instance, researchers reported different results on the performance of content-based and collaborative filtering. Sometimes content-based filtering performed better than collaborative filtering and sometimes it performed worse. We identified three potential reasons for the ambiguity of the results. (A) Several evaluations had limitations. They were based on strongly pruned datasets, few participants in user studies, or did not use appropriate baselines. (B) Some authors provided little information about their algorithms, which makes it difficult to re-implement the approaches. Consequently, researchers use different implementations of the same recommendations approaches, which might lead to variations in the results. (C) We speculated that minor variations in datasets, algorithms, or user populations inevitably lead to strong variations in the performance of the approaches. Hence, finding the most promising approaches is a challenge. As a second limitation, we noted that many authors neglected to take into account factors other than accuracy, for example overall user satisfaction. In addition, most approaches (81 %) neglected the user-modeling process and did not infer information automatically but let users provide keywords, text snippets, or a single paper as input. Information on runtime was provided for 10 % of the approaches. Finally, few research papers had an impact on research-paper recommender systems in practice. We also identified a lack of authority and long-term research interest in the field: 73 % of the authors published no more than one paper on research-paper recommender systems, and there was little cooperation among different co-author groups. We concluded that several actions could improve the research landscape: developing a common evaluation framework, agreement on the information to include in research papers, a stronger focus on non-accuracy aspects and user modeling, a platform for researchers to exchange information, and an open-source framework that bundles the available recommendation approaches.

✉ Joeran Beel
beel@docear.org

Bela Gipp
bela.gipp@uni-konstanz.de

Corinna Breitinger
breitinger@docear.org

[1] Docear, Magdeburg, Germany

[2] University of Konstanz, Konstanz, Germany

[3] Otto-von-Guericke University, Magdeburg, Germany

[4] Linnaeus University, Kalmar, Sweden

Published online: 26 July 2015

Keywords Recommender system · User modeling · Research paper recommender systems · Content based filtering · Review · Survey

Contents

1 Introduction

In 1998, Giles et al. introduced the first research-paper recommender system as part of the *CiteSeer* project [1]. Since then, at least 216 articles relating to 120 research-paper recommendation approaches were published [2–217]. The amount of literature and approaches represents a problem for new researchers: they do not know which of the articles are most relevant, and which recommendation approaches are most promising. Even researchers familiar with research-paper recommender systems would find it difficult to keep track of the current developments, since the yearly number of articles steadily increases: 66 of the 217 articles (30 %) were published just in 2012 and 2013 alone (Fig. 1; Table 1). The few existing literature surveys in the field [186–188] cover just a fraction of the articles, or focus on selected aspects, such as recommender-system evaluation [190]. Thus, they do not provide an overview of the research field, or identify the most promising approaches.

We survey the field of research-paper recommender systems with the goal of enabling researchers and developers to (a) learn about the status-quo of research-paper recommender systems, (b) identify promising fields of research, and (c) motivate the community to solve the most urgent problems that currently hinder the effective use of research-paper recommender systems in practice. For clarity, we use the term "article" to refer to the reviewed journal articles, patents, websites, etc., and the term "paper" to refer to documents being recommended by research-paper recommender systems.[1] When referring to a large number of recommender systems with certain properties, we cite three exemplary articles. For instance, when we report how many recommender systems apply content-based filtering, we report the number and provide three references [7,58,80].

To identify relevant literature for our survey, we conducted a literature search on *Google Scholar*, *ACM Digital Library, Springer Link,* and *ScienceDirect*. We searched for *[paper* | *article* | *citation] [recommender* | *recommendation] [system* | *systems]* and downloaded all articles that had relevance for research-paper recommender systems. Our relevance judgment made use of the title and the abstract if the title alone did not indicate a recognizable relevance to research-paper recommender systems. We examined the bibliography of each article. If an entry in the bibliography pointed to a relevant article not yet downloaded, we downloaded that article. In addition, we checked on Google Scholar which articles cited the relevant article. If one of the citing articles seemed relevant, we also downloaded it. We expanded our search to websites, blogs, patents, and presentations on major academic recommender systems. These major academic services include the academic search engines *CiteSeer*(x),[2] *Google Scholar* (*Scholar Update*),[3] and *PubMed*;[4] the social network *ResearchGate*;[5] and the reference managers *CiteULike*,[6] *Docear*,[7] and *Mendeley*.[8] While these systems offer

[1] Some recommender systems also recommended "citations" but in our opinion, differences between recommending papers and citations are marginal, which is why we do not distinguish between these two terms in this paper.

[2] http://citeseerx.ist.psu.edu.

[3] http://scholar.google.com/scholar?sciupd=1&hl=en&as_sdt=0,5.

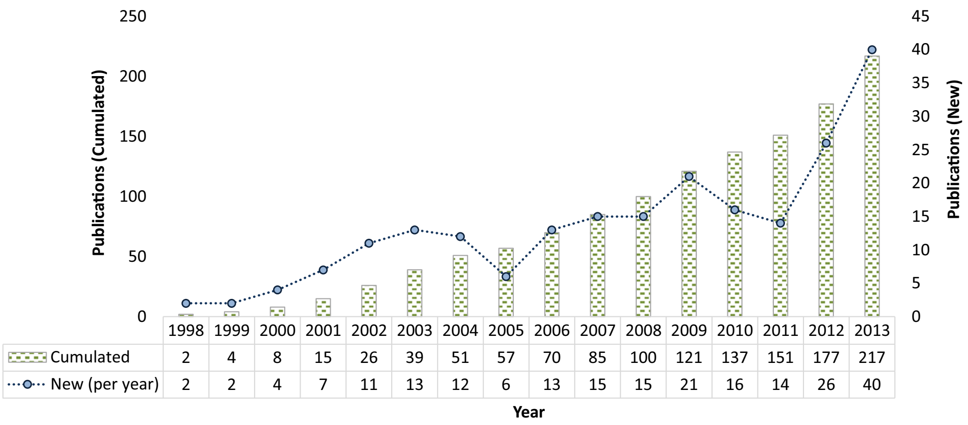
[4] http://www.ncbi.nlm.nih.gov/pubmed.

[5] http://www.researchgate.net/.

[6] http://www.citeulike.org/.

[7] http://www.docear.org.

[8] http://www.mendeley.com/.



	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
▨ Cumulated	2	4	8	15	26	39	51	57	70	85	100	121	137	151	177	217
···○··· New (per year)	2	2	4	7	11	13	12	6	13	15	15	21	16	14	26	40

Fig. 1 Annual publications in the field of research-paper recommender systems. Numbers are based on our literature search. Although, we believe our survey to be the most comprehensive survey about research-paper recommender systems, we may have missed a few articles. In addition, most likely, more than 40 papers were published in 2013 since we conducted the literature search in January 2014. Articles presented at conferences in late 2013 most likely had not been published in conferences proceedings by January 2014, and hence were not found through our search. Hence, the total number of papers published is probably higher than 217

recommender systems along with their main services, there are also a few stand-alone recommender systems, namely *BibTip*,[9] *bX*,[10] *RefSeer*,[11] *TheAdvisor*[12] and an experimental system called *Sarkanto*.[13]

The first literature search was conducted in June 2013 and found 188 relevant articles [1–188]. Three of the 188 articles were literature surveys [186–188], which were ignored in our survey. The remaining 185 articles consist of peer-reviewed conference articles (59 %), journal articles (16 %), pre-prints (5 %), and other formats such as Ph.D. theses, patents, presentations, and web pages (Table 2). Overall, the reviewed articles were comprehensive, with a median page count of eight. More than one-third (36 %) had 10 or more pages (Fig. 2). Another 23 % had eight or nine pages, while only 26 % of the articles had four or less pages.

Citation counts follow a typical power–law distribution: a few articles gained many citations (the maximum was 528 citations for [43]) and many articles had few citations, see Fig. 3. The mean citation count was 30, and median was seven. From the reviewed articles, 31 % had no citations. Citation counts were retrieved from Google Scholar in early 2014. Some researchers have reservations about using Google Scholar as source for citation counts [218–220], but the numbers can give a rough idea of a paper's popularity.

We reviewed the 185 articles, which contained information on 96 research-paper recommendation approaches [1–185]. In an initial review, we focused on the evaluation of the approaches. The review included an analysis of which evaluation methods were applied (e.g., user-studies or offline evaluations), which evaluation metrics were used (e.g., precision or recall), how many participants the user studies had, and how strongly datasets were pruned.

Of the 96 research-paper recommendation approaches presented in 185 articles, 62 approaches were chosen for an in-depth analysis, presented in 127 articles [1–127]. We chose these 62 approaches, because we classified the remaining 34 approaches, i.e., 58 articles, as not sufficiently related to 'research-paper recommender systems' [128–185]. We classified articles as insufficiently related if they provided no evaluation, or if their approach did not differ significantly from that of previous authors. We also excluded articles that could not be clearly interpreted due to grammar and language use, or when they were outside of the scope (even if the article's title suggested relevance to research-paper recommender systems). One example of an article outside of the research scope was 'Research Paper Recommender Systems—A Subspace Clustering Approach' [130]. The title appears relevant for this survey, but the article presents a collaborative filtering approach that is not intended for recommender systems for research papers. Instead, the paper used the *Movielens* dataset, which contains ratings of movies.

In January 2014, we conducted a second literature search and found 29 additional articles of relevance [189–217]. The goal of this search was to identify the overall number of articles published in 2013; see Fig. 1. However, time limitations prevented us from broadening the scope of the initially planned survey. Therefore, our review concentrates on the two subsets of the 217 articles: the 185 articles identified during the first round of the literature search, and the 127 articles that we chose for an in-depth review.

In the remaining paper, we present definitions (Sect. 2), followed by an introduction to related research fields (Sect. 3).

[9] http://www.bibtip.com/.

[10] http://www.exlibrisgroup.com/category/bXUsageBasedServices.

[11] http://refseer.ist.psu.edu/.

[12] http://theadvisor.osu.edu/.

[13] http://lab.cisti-icist.nrc-cnrc.gc.ca/Sarkanto/.

Table 1 List of reviewed articles by year

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
References	[1,43]	[80,155]	[106,123, 131,139]	[34,42, 82,91, 141,154, 174]	[31–33,57, 88,90,93, 102,103, 150,173]	[9,30,35–39,48, 140,157, 160,164, 165]	[27,55, 81,94, 105,107, 108,117, 162,163, 169,172]	[64,122, 130,133, 134,188]	[3,15,23, 46,53, 54,83,87, 104,129, 135,144, 195]	[28,63, 85,109, 110,114, 120,128, 143,147, 148,153, 170,180, 182]	[14,21, 47,52,78, 92,95,99, 100,127, 138,152, 166,176, 184]	[4,12,18, 19,24, 44,45,66, 89,96–98,124, 136,137, 146,156, 161,167, 168,175]	[13,22, 26,40,51, 62,65,79, 115,119, 121,145, 151,158, 181,185]	[10,11, 17,29,41, 50,77,84, 86,101, 111,116, 118,186]	[2,16,49, 56,58–61, 67,68,71–75,126, 132,142, 149,171, 177–179, 183,187]	[5–8,20, 25,69,70, 112,113, 125,159, 189–194, 196–217]

We then present the survey of the 96 approaches' evaluations (Sect. 4), followed by an analysis of the 62 approaches that we chose for the in-depth review (Sect. 5). Finally, we examine the field of research-paper recommender systems in general, and point out some shortcomings, such as the neglect of user modeling, a strong focus on accuracy alone at the expense of other aspects being ignored and scarce information on the details of the algorithms used (Sect. 6).

2 Definitions

We use the term "idea" to refer to a hypothesis about how recommendations could be effectively generated. To differentiate how specific the idea is, we distinguish between recommendation classes, approaches, algorithms, and implementations (Fig. 4).

We define a "recommendation class" as the least specific idea, namely a broad concept that broadly describes how recommendations might be given. For instance, the recommendation classes *collaborative filtering* (CF) and *content-based filtering* (CBF) fundamentally differ in their underlying ideas: the underlying idea of CBF is that users are interested in items that are similar to items the users previously liked. In contrast, the idea of CF is that users like items that the users' peers liked. However, these ideas are rather vague and leave room for different approaches.

A "recommendation approach" is a model of how to bring a recommendation class into practice. For instance, the idea behind CF can be realized with user-based CF [221], content-boosted CF [222], and various other approaches [223]. These approaches are quite different, but are each consistent with the central idea of CF. Nevertheless, these approaches to represent a concept are still vague and leave room for speculation on how recommendations are calculated.

A "recommendation algorithm" precisely specifies a recommendation approach. For instance, an algorithm of a CBF approach would specify whether terms were extracted from the title of a document or from the body of the text, and how terms are processed (e.g., stop-word removal or stemming) and weighted (e.g., TF-IDF). Algorithms are not necessarily complete. For instance, pseudo-code might contain only the most important information and ignore basics, such as weighting schemes. This means that for a particular recommendation approach there might be several algorithms.

Finally, the "implementation" is the actual source code of an algorithm that can be compiled and applied in a recommender system. It fully details how recommendations are generated and leaves no room for speculation. It is, therefore, the most specific idea about how recommendations might be generated.

A "recommender system" is a fully functional software system that applies at least one implementation to make rec-

Table 2 Percentage of different article types reviewed

Journal articles	Conference papers	Ph.D. Theses	Master's Theses	Patents	Pre-prints/ unpublished	Other
16 %	59 %	3 %	2 %	1 %	5 %	15 %

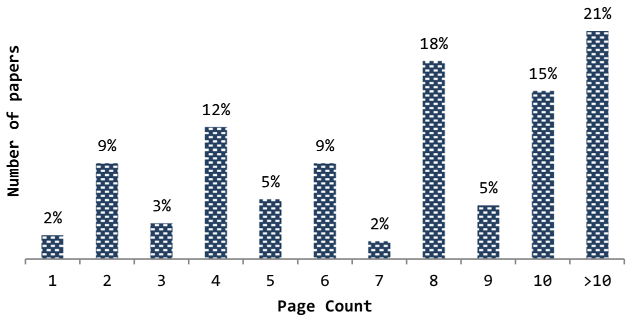


Fig. 2 Page count of reviewed articles

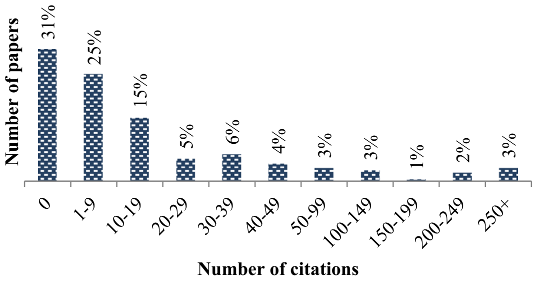


Fig. 3 Citation counts overview for reviewed articles

ommendations. In addition, recommender systems feature several other components, such as a user interface, a corpus of recommendation candidates, and an operator that owns/runs the system. Some recommender systems also use two or more recommendation approaches: CiteULike, a service for discovering and managing scholarly references, lets their users choose between two approaches [14,17], and Docear randomly selects one of three approaches each time users request recommendations [7].

The "recommendation scenario" describes the entire setting of a recommender system, including the recommender system and the recommendation environment, i.e., the domain and user characteristics.

By "effectiveness," we refer to the degree to which a recommender system achieves its objective. The objective of a recommender system from a broad perspective is to provide "good" [224] and "useful" [225] recommendations that make users "happy" [226] by satisfying user needs [87]. The needs of users vary. Consequently, some users might be interested in *novel* research-paper recommendations, while others might be interested in *authoritative* research-paper recommendations. Of course, users require recommendations specific to their fields of research [117]. When we use the term "effectiveness," we refer to the specific objective the evaluator wanted to measure. We use the terms "performance" and "effectiveness" interchangeably.

"Evaluation" describes any kind of assessment that measures the effectiveness or merit of a concrete idea or approach. More details about research paper recommender system evaluation methods follow in Sect. 4.

3 Related research fields

Several research fields are related to user modeling and (research-paper) recommender systems. Although we do not survey these fields, we introduce them so interested readers can broaden their research.

Research on *academic search engines* deals with calculating relevance between research papers and search queries [227–229]. The techniques are often similar to those used by research-paper recommender systems. In some cases, recommender systems and academic search engines are even identical. As described later, some recommender systems require their users to provide keywords that represent their interests. In these cases, research-paper recommender systems do not differ from academic search engines where users provide keywords to retrieve relevant papers. Consequently, these fields are highly related and most approaches for academic search engines are relevant for research-paper recommender systems.

The *reviewer assignment problem* targets using information-retrieval and information-filtering techniques to automate the assignment of conference papers to reviewers [230]. The differences from research-paper recommendations are minimal: in the reviewer assignment problem, a relatively small number of paper submissions *must* be assigned to a small number of users, i.e., reviewers; research-paper recommender systems recommend a few papers out of a large corpus to a relatively large number of users. However, the techniques are usually identical. The reviewer assignment problem was first addressed by *Dumais and Nielson* in 1992 [230]; 6 years before Giles et al. introduced the first research-paper recommender system. A good survey on the reviewer assignment problem was published by Wang et al. [231].

Scientometrics deals with analyzing the impact of researchers, research articles and the links between them. Scientometrics researchers use several techniques to calculate document relatedness or to rank a collection of articles. Some of the measures—*h*-index [232], co-citation strength [233] and bibliographic coupling strength [234]—have also been applied by research-paper recommender systems [13,123,

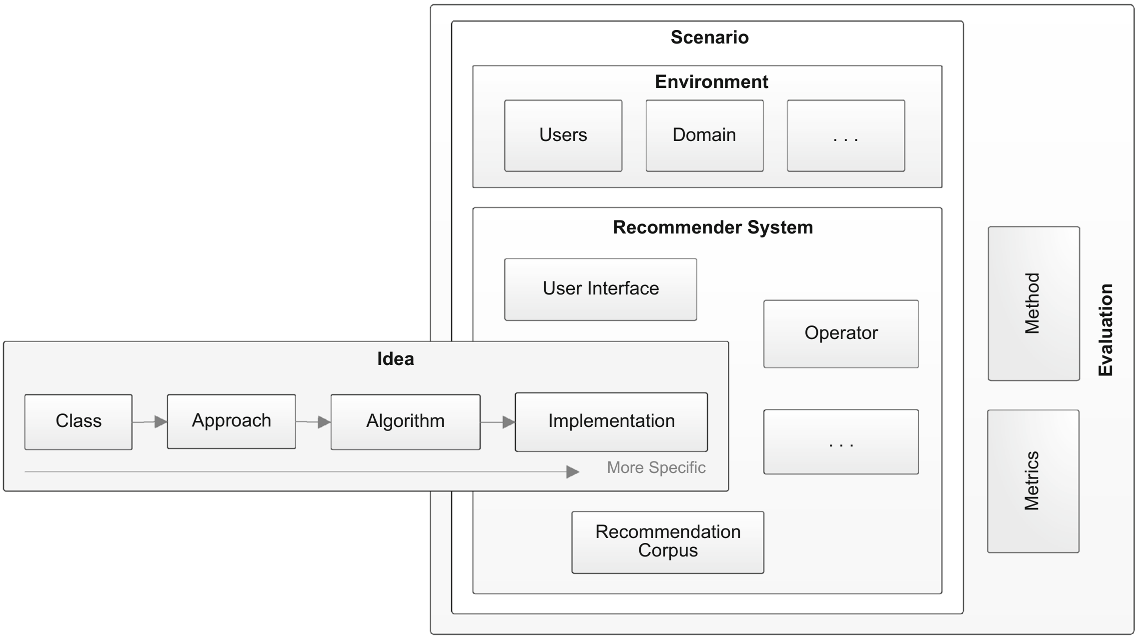


Fig. 4 Illustration of recommendation system terminology and concepts

126]. However, there are many more metrics in scientometrics that might be relevant for research-paper recommender systems [235].

User modeling evolved from the field of Human Computer Interaction. One thing user modeling focuses on is reducing users' information overload making use of users' current tasks and backgrounds [236]. User modeling shares this goal with recommender systems, and papers published at the major conferences in both fields (UMAP[14] and RecSys[15]) often overlap. User modeling is a central component of recommender systems because modeling the users' information needs is crucial for providing useful recommendations. For some comprehensive surveys about user modeling in the context of web personalization, refer to [237,238].

Other related research fields include book recommender systems [239], educational recommender systems [240], academic alerting services [241], expert search [242], automatic summarization of academic articles [243–245], academic news feed recommenders [246,247], academic event recommenders [248], venue recommendations [249], citation recommenders for patents [250], recommenders for academic datasets [251], and plagiarism detection. Plagiarism detection, like many research-paper recommenders, uses text and citation analysis to identify similar documents [252–254]. Additionally, research relating to crawling the web and analyzing academic articles can be useful for building research-paper recommender systems, for instance, author name extraction and disambiguation [255], title extraction [256–260], or citation extraction and matching [261]. Finally, most of the research on content-based [262] or collaborative filtering [263,264] from other domains, such as movies or news, can also be relevant for research-paper recommender systems.

4 Survey of the evaluations

Recommender-systems research heavily relies on evaluations to assess the effectiveness of recommendation approaches. Among the key prerequisites for thorough evaluations are appropriate evaluation methods, a sufficient number of study participants, and a comparison of the novel approach against one or more state-of-the-art approaches [265]. The novel approach and its evaluation must be clearly described. The soundness of the evaluation, the re-implementation of the approach, and the reproducibility and replicability of the results are guaranteed only if a clear description is given.

We reviewed the evaluation methods, metrics, and datasets used; the number of participants in the user studies; the baselines used for comparing the novel approaches; and several other factors to judge the appropriateness of the evaluations of the 96 approaches. Originally, our goal was to identify the approaches for which evaluations were thoroughly conducted. The further review would have then concentrated on these thoroughly evaluated approaches to identify the most promising approaches. However, as we will show in the following sections, the majority of evaluations contained

[14] http://www.um.org/conferences.

[15] http://recsys.acm.org/.

limitations, which made it impossible to determine a number of promising approaches.

4.1 Evaluation methods and their adequacy

Of the 96 reviewed recommendation approaches, 21 (22 %) were not evaluated by their authors [135,152,180]. In other cases, an evaluation was attempted, but the methods were questionable and were insufficiently described to be understandable or reproducible [137,176,181]. Of the remaining 75 evaluated approaches, 53 (71 %) were evaluated using offline evaluations, 25 (33 %) using quantitative user studies, two (3 %) using qualitative user studies, and five (7 %) using an online evaluation (Table 3). The different evaluation methods and their application in the field of research-paper recommender systems are introduced in the next sections.

4.1.1 User studies

User studies typically measure user satisfaction through explicit ratings. Users receive recommendations generated by different recommendation approaches, users rate the recommendations, and the approach with the highest average rating is considered most effective [263]. Study participants are typically asked to quantify their overall satisfaction with the recommendations. However, they might also be asked to rate individual aspects of a recommender system, for instance, how novel or authoritative recommendations are [117], or how suitable they are for non-experts [72]. A user study can also collect qualitative feedback, but qualitative feedback is rarely used in the field of (research-paper) recommender systems [156,159].

We distinguish between "lab" and "real-world" user studies. In lab studies, participants are aware that they are part of a user study, which together with other factors might affect user behavior and thereby the evaluation's results [266,267]. In real-world studies, participants are not aware of the study and rate recommendations for their own benefit, for instance because the recommender system improves recommendations based on user ratings (i.e., relevance feedback [262]), or user ratings are required to generate recommendations (i.e., collaborative filtering [221]). All reviewed user studies were lab-based.

Table 3 Evaluation methods used by reviewed recommendation approaches

	Offline	User study (Quant.)	User study (Qual.)	Online
Absolute	53	25	2	5
Relative	71 %	33 %	3 %	7 %

Some approaches were evaluated using several methods. As a result, percentages do not add up to 100 %

Often, user studies are considered the optimal evaluation method [268]. However, the outcome of user studies often depends on the questions asked. Cremonesi et al. found that it makes a difference if users are asked for the "perceived relevance" or the "global satisfaction" of recommendations [269]. Similarly, it made a difference whether users were asked to rate the *novelty* or the *relevance* of recommendations [270]. A large number of participants are also crucial to user study validity, which makes user studies relatively expensive to conduct. The number of required participants, to receive statistically significant results, depends on the number of approaches being evaluated, the number of recommendations being displayed, and the variations in the results [271,272]. However, as a rough estimate, at least a few dozen participants are required, often more.

Most participants in the reviewed user studies rated only a few recommendations and four studies (15 %) were conducted with fewer than five participants [62,123,171]; five studies (19 %) had five to ten participants [66,84,101]; three studies (12 %) had 11–15 participants [15,146,185]; and five studies (19 %) had 16–50 participants [44,118,121]. Six studies (23 %) were conducted with more than 50 participants [93,98,117]. Three studies (12 %) failed to mention the number of participants [55,61,149] (Table 4). Given these findings, we conclude that most user studies were not large enough to arrive at meaningful conclusions.

4.1.2 Online evaluations

Online evaluations were first used by the online advertising and e-commerce fields. They measure the acceptance rates of recommendations in real-world recommender systems. Acceptance rates are typically measured by click-through rates (CTR), i.e., the ratio of clicked recommendations to displayed recommendations. For instance, if a recommender system displays 10,000 recommendations and 120 are clicked, the CTR is 1.2 %. Other metrics include the ratio of downloaded or bought items to the number of items displayed. Acceptance rate is typically interpreted as an implicit measure for user satisfaction. The assumption is that when a user clicks, downloads, or buys a recommended item, the user liked the recommendation. Of course, this assumption is not always reliable because users might buy a book but rate it negatively after reading it. However, metrics such as CTR can be

Table 4 Number of participants in user studies of reviewed recommendation approaches

	Number of participants					
	n/a	<5	5–10	11–15	16–50	>50
Absolute	3	4	5	3	5	6
Relative	12 %	15 %	19 %	12 %	19 %	23 %

an *explicit* measures of effectiveness, namely when the operator receives money, e.g., for clicks on recommendations.

Online evaluations are not without drawbacks. Zheng et al. showed that CTR and relevance do not always correlate and concluded that "CTR may not be the optimal metric for online evaluation of recommender systems" and "CTR should be used with precaution" [273]. In addition, conducting online evaluations requires significantly more time than offline evaluations, they are more expensive, and they can only be conducted by researchers who have access to a real-world recommender system.

Among the 75 approaches that included some form of evaluation, only six (8 %) were evaluated using an online evaluation [7,92,94]. Despite the active experimentation in the field with a large number or evaluations being performed on research-paper recommender systems, we observed that many researchers have no access to real-world systems to evaluate their approaches. Interestingly, the researchers who *do* have access to real-world recommender systems often do not make use of this resource to conduct online evaluations, but rather perform offline evaluations or lab user studies. For instance, *Giles* and his co-authors, who are some of the largest contributors in the field, could have conducted online evaluations with their academic search engine CiteSeer. However, they chose primarily to use offline evaluations. The reason for this might be that offline evaluations are more convenient than conducting online evaluations or user studies. Results are available within minutes or hours and not within days or weeks as is the case for user studies and online evaluations. However, offline evaluations have a set of serious drawbacks, as shown in the next section.

4.1.3 Offline evaluations

Offline evaluations typically measure the *accuracy* of a recommender system based on a ground truth. To measure accuracy, precision at position *n* (P@n) is often used to express how many items of the ground truth are recommended within the top *n* recommendations. Other common evaluation metrics include recall, F-measure, mean reciprocal rank (MRR), normalized discounted cumulative gain (nDCG), mean absolute error, and root mean square error. Offline evaluations are also sometimes used to evaluate aspects such as novelty or serendipity of recommendations [226]. For a comprehensive overview of offline evaluations, refer to [274,275].

Offline evaluations were originally meant to identify a number of promising recommendation approaches [224,263,276,277]. These approaches should then be evaluated in detail with a user study or online evaluation to identify the most effective approaches. However, criticism has been raised on the assumption that offline evaluation could predict an algorithm's effectiveness in online evaluations or user studies. More precisely, several researchers have shown that results from offline evaluations do not necessarily correlate with results from user studies or online evaluations [93,269,270,278–281]. This means that approaches that are effective in offline evaluations are not necessarily effective in real-world recommender systems. McNee et al. observed that

> "the research community's dependence on offline experiments [has] created a disconnect between algorithms that score well on accuracy metrics and algorithms that users will find useful." [87]

Other researchers also voiced criticism of offline evaluations. Jannach et al. stated that "the results of offine [evaluations] may remain inconclusive or even misleading" and "real-world evaluations and, to some extent, lab studies represent probably the best methods to evaluate systems" [282]. Knijnenburg et al. reported that "the presumed link between algorithm accuracy […] and user experience […] is all, but evident" [283]. Said et al. consider "on-line evaluation [as] the only technique able to measure the true user satisfaction" [268]. Rashid et al. observed that biases in the offline datasets may cause bias in the evaluation [277]. The main reason for the criticism in the literature is that offline evaluations focus on accuracy yet ignore human factors; however, human factors strongly affect overall user satisfaction for recommendations. Despite the criticism, offline evaluations are the predominant evaluation method in the recommender community [284] and "surprisingly few studies [evaluate] algorithms in live experiments with real users" [283].

Our review indicates that the voiced criticism of offline evaluations also applies to the field of research-paper recommender systems. Some of the approaches were evaluated using both an offline evaluation and a user study. In two evaluations, results from the offline evaluations were indeed similar to results of the user studies [26,84]. However, the user studies had five and 19 participants, respectively, which led to statistically insignificant results. Three other studies reported contradicting results for offline evaluations and user studies (two of these studies had more than 100 participants) [57,93,117]. This means that offline evaluations could not reliably predict the effectiveness in the real-world use case. Interestingly, the three studies with the most participants were all conducted by the authors of TechLens [26,93,117], who are also the only authors in the field of research-paper recommender systems who discuss the potential shortcomings of offline evaluations [87]. It seems that other researchers in this field are not aware of—or chose not to address—problems associated with offline evaluations, although there has been quite a discussion outside the research-paper recommender-system community [93,269,270,278–281].

4.2 The operator's perspective

It is commonly assumed that the objective of a recommender system is to make users "happy" [226] by satisfying their needs [87]. However, there is another important stakeholder who is often ignored: the operator of a recommender system [224]. It is often assumed that operators of recommender systems are satisfied when their users are satisfied, but this is not always the case. Operators may also want to keep down costs of labor, disk storage, memory, computing power, and data transfer [263]. Therefore, for operators, an effective recommender system may be one that can be developed, operated, and maintained at a low cost. Operators may also want to generate a profit from the recommender system [224]. Such operators might prefer to recommend items with higher profit margins, even if user satisfaction is not optimal. For instance, publishers might be more interested in recommending papers the user must pay for than papers the user can freely download.

The operator's perspective has been widely ignored in the reviewed articles. Costs of *building* a recommender system, or implementing an approach were not reported in any article. Costs to run a recommender system were reported by Jack from Mendeley [59]. He stated that the costs on Amazon's S3 were $66 a month plus $30 to update the recommender system that served 20 requests per second generated by 2 million users.

Runtime information is crucial to estimate costs, and hence to estimate how feasible an approach will be to apply in practice. In one paper, the runtimes of two approaches differed by a factor of 600 [56]. For many operators, an approach that requires 600 times more computing power than another would probably not be an option. While this example is extreme, other runtime comparisons showed differences by a factor of five or more, which can also affect algorithm selection. However, information on runtime was provided only for 10 % of the approaches.

Reporting on computational complexity is also important. For operators who want to offer their system to a large number of users, computational complexity is important for estimating the long-term suitability of an approach. An approach may perform well enough for a few users, but it might not scale well. Approaches with exponentially increasing complexity most likely will not be applicable in practice. However, computational complexity was reported for even fewer approaches than runtime.

4.3 Coverage

Coverage describes how many papers of those in the recommender's database might potentially be recommended [285,286]. As such, coverage is an important metric to judge the usefulness of a recommender system. For text-based approaches, coverage is usually 100 %. For other approaches, coverage is typically lower. For instance, in collaborative filtering not all items are rated by users. Although the unrated items might be relevant, they cannot be recommended. High coverage is important because it increases the number of recommendations a user can receive. Of the reviewed articles, few considered coverage in their evaluations. He et al. judge the effectiveness of their approaches based on which approach provides the best tradeoff between accuracy and coverage [51]. The BibTip developers report that 80 % of all documents have been co-viewed and can be used for generating recommendations [92]. Pohl et al. report that co-download coverage on arXiv is close to 100 % while co-citation coverage is only around 30 % [110]. The TechLens authors report that all of their hybrid and CBF approaches have 100 % coverage, except pure CF which has a coverage of 93 % [117].

4.4 Baselines

Another important factor in evaluating recommender systems is the baseline against which an algorithm is compared. For instance, knowing that a certain approach has a particular CTR is not useful if the CTRs of alternative approaches are unknown. Therefore, novel approaches should be compared against a baseline representative of the state-of-the-art approaches. This way it is possible to quantify whether, and when, a novel approach is more effective than the state-of-the-art and by what margin.

Of the 75 evaluated approaches, 15 (20 %) were not compared against a baseline [27,86,185]. Another 53 (71 %) approaches were compared against trivial baselines, such as simple content-based filtering without any sophisticated adjustments. These trivial baselines do not represent the state-of-the-art and are not helpful for deciding whether a novel approach is promising. This is particularly troublesome since the reviewed approaches were not evaluated against the *same* trivial baselines. Even for a simple CBF baseline, there are many variables, such as whether stop words are filtered, which stemmer is applied, or from which document field the text is extracted. This means that almost all reviewed approaches were compared against different baselines, and results cannot be compared with each other. Seven approaches (9 %) were evaluated against approaches proposed by other researchers in the field. Only these evaluations allow drawing some conclusions on which approaches may be most effective.

It is interesting to note that in all evaluations, at least one of the novel approaches performed better than the baseline (if the approach was evaluated against a baseline). No article reported on a non-effective approach. We can just speculate about the reasons: First, authors may intentionally select baselines such that their approaches appear favorable. Sec-

Table 5 Evaluation metrics of reviewed recommendation approaches

	Precision	Recall	*F*-measure	nDCG	MRR	Other
Absolute	38	12	6	11	10	12
Relative	72 %	23 %	11 %	21 %	19 %	23 %

Some approaches' effectiveness was measured with multiple metrics; therefore, numbers do not add up to 100 %

ond, the simple baselines used in most evaluations achieve relatively poor results, so that any alternative easily performs better. Third, authors do not report failures. Lastly, journals and conferences might not accept publications that report on failures. Whatever the reasons are, we advocate that reporting failures is desirable since it could prevent other researchers from doing the same experiments, and hence wasting time.

4.5 Offline evaluation metrics

Precision was used as an evaluation metric in 38 offline evaluations (72 %) (Table 5). Recall was used in 12 evaluations (23 %); F-measure in 6 evaluations (11 %); nDCG in 11 evaluations (21 %); MRR in 10 evaluations (19 %); and other measures in 12 evaluations (23 %). Overall, results of the different measures highly correlated. That is, an algorithm that performed well measured by precision tended to perform well measured by nDCG, for instance. However, there were exceptions. Zarrinkalam and Kahani tested the effectiveness of abstract and title against abstract, title, and citation context [125]. When *co-citation probability* was used as an evaluation metric, title and abstract were most effective. Based on recall, the most effective field combination was abstract, title, and citation context. With the nDCG measure, results varied depending on how the candidate set was generated and which ranking approach was used.

4.6 Datasets and architectures

Researchers and developers in the field of recommender systems can benefit from publicly available architectures and datasets.[16] *Architectures* help with the understanding and building of recommender systems, and are available in various recommendation domains, such as e-commerce [287], marketing [288], and engineering [289]. *Datasets* enable the evaluation of recommender systems by allowing researchers to evaluate their systems with the same data. Datasets are available in several recommendation domains, including movies,[17] music,[18] and baby names.[19] Notable are also the various TREC datasets that facilitated and standardized evaluations in several domains.[20]

Architectures of research-paper recommender systems were published by few authors. The developers of *CiteSeer(x)* published an architecture that focused on crawling and searching academic PDFs [1,108]. This architecture has some relevance for recommender systems, since many tasks in academic search are related to recommender systems (e.g., crawling and indexing PDFs, and matching user models or search-queries with research papers). Bollen and van de Sompel published an architecture that later served as the foundation for the research-paper recommender system *bX* [15]. This architecture focuses on recording, processing, and exchanging scholarly usage data. The developers of *BibTiP* [33] also published an architecture that is similar to the architecture of bX (both bX and BibTip exploit usage data to generate recommendations).

Several academic services published datasets that eased the process of researching and developing research-paper recommender systems. *CiteULike*[21] and *Bibsonomy*[22] published datasets containing the social tags that their users added to research articles. The datasets were not originally intended for recommender-system research but are frequently used for this purpose [56,62,112]. *CiteSeer* made its corpus of research papers public,[23] as well as the citation graph of the articles, data for author name disambiguation, and the co-author network [290]. CiteSeer's dataset has been frequently used by researchers for evaluating research-paper recommender systems [20,24,51,56,65,106,112,117,125]. Jack et al. compiled a dataset based on the reference management software *Mendeley* [291]. The dataset includes 50,000 randomly selected personal libraries from 1.5 million users. These 50,000 libraries contain 4.4 million articles of which 3.6 million are unique. Due to privacy concerns, Jack et al. publish only the unique IDs of articles and no title or author names. Additionally, only those libraries with at least 20 articles were included in the dataset. *Sugiyama and Kan* released two small datasets,[24] which they created for their academic recommender system [115]. The datasets include

[16] Recommendation frameworks such as *LensKit or Mahout* may also be helpful for researchers and developers, but frameworks are not the topic of this paper.

[17] http://grouplens.org/datasets/movielens/.

[18] http://labrosa.ee.columbia.edu/millionsong/.

[19] http://www.kde.cs.uni-kassel.de/ws/dc13/.

[20] http://trec.nist.gov/data.html.

[21] http://www.citeulike.org/faq/data.adp.

[22] https://www.kde.cs.uni-kassel.de/bibsonomy/dumps/.

[23] http://csxstatic.ist.psu.edu/about/data.

[24] http://www.comp.nus.edu.sg/~sugiyama/SchPaperRecData.html.

Table 6 Source of datasets for reviewed recommendation approaches that performed offline evaluations

	CiteSeer	CiteUlike	ACM	DBLP	Others
Absolute	17	6	5	4	27
Relative	32 %	11 %	9 %	8 %	51 %

some research papers, and the interests of 50 researchers. The CORE project released a dataset[25] with enriched metadata and full-texts of academic articles that could be helpful in building a recommendation candidate corpus.

Of the 53 reviewed offline evaluations, 17 (32 %) were evaluated using data from CiteSeer and 6 (11 %) were evaluated using data from CiteULike (Table 6). Other data sources included ACM (9 %), DBLP (8 %), and a variety of others, often not publicly available datasets (51 %). Even when data originated from the same sources, it did not guarantee that the same datasets were used. For instance, 32 % of the approaches were evaluated with data from CiteSeer but no single CiteSeer dataset exists. Authors collected CiteSeer data at different times and pruned datasets differently. Some authors removed documents with fewer than two citations from the CiteSeer corpus [24], others with fewer than three citations [117], and others with fewer than four citations [137]. Other datasets were pruned even more heavily. Caragea et al. removed papers with fewer than ten and more than 100 citations, as well as papers citing fewer than 15 and more than 50 papers [20]. From 1.3 million papers in the corpus, around 16,000 remained (1.2 %). Pennock et al. removed documents from the corpus with fewer than 15 implicit ratings [106]: from originally 270,000 papers, 1575 remained (0.58 %). It is, therefore, safe to say that no two studies, performed by different authors, used the same dataset. This raises the question to what extent results based on different datasets are comparable.

Naturally, recommendation approaches perform differently on different datasets [224,292,293]. This is particularly true for the absolute effectiveness of recommendation approaches. For instance, an algorithm that achieved a recall of 4 % on an IEEE dataset achieved a recall of 12 % on an ACM dataset [101]. The *relative* effectiveness of two approaches is also not necessarily the same with different datasets. For instance, because approach A is more effective than approach B on dataset I, it does not mean that A is also more effective than B on dataset II. However, among the few reviewed approaches that were evaluated on different datasets, the effectiveness was surprisingly consistent.

Of the evaluated approaches, seven were evaluated on multiple offline datasets. Dataset combinations included CiteSeer and some blogs [100], CiteSeer and Web-kd [65], CiteSeer and CiteULike [56], CiteSeer and Eachmovie [106], and IEEE, ACM and ScienceDirect [101]. Results differed notably among the different datasets only in one study. However, the absolute ranking of the approaches remained the same [56] (Table 7). In that article, the proposed approach (CTM) performed best on two datasets (CiteULike and CiteSeer), with a MRR of 0.529 and 0.467, respectively. Three of the four baselines performed similarly on the CiteSeer dataset (all with a MRR between 0.238 and 0.288). However, for the CiteULike dataset, the TM approach performed four times as well as CRM. Consequently, if TM had been compared with CRM, rankings would have been similar on the CiteSeer dataset but different on the CiteULike dataset.

Table 7 MRR on different datasets used for offline evaluations

Rank	Approach	Dataset	
		CiteSeer	CiteUlike
1	CTM	0.529	0.467
2	TM	0.288	0.285
3	cite-LDA	0.285	0.143
4	CRM	0.238	0.072
5	link-LDA	0.028	0.013

Overall, a sample size of seven is small, but it gives at least some indication that the impact of the chosen dataset is rather low in the domain of research-paper recommender systems. This finding is interesting because in other fields it has been observed that different datasets lead to different results [224,292]. Nevertheless, we doubt that pruning datasets drastically should be considered good practice, especially if just a fraction of the original data remains.

4.7 Reproducibility and the butterfly effect

The reproducibility of experimental results is the “fundamental assumption” in science [294], and the “cornerstone” that allows drawing meaningful conclusions about the generalizability of ideas [295]. Reproducibility describes the situation when (slightly) different ideas, scenarios, and evaluations lead to similar experimental results [294], where we define “similar results” as results that allow the same conclusions to be drawn. *Reproducibility* should not be confused with *replicability*. Replicability describes an exact copy of an experiment that uses the same tools, follows the same steps, and produces the same results [296]. Therefore, replicability is important when analyzing whether the original experiment was conducted thoroughly and whether the results can be trusted.

Conversely, if changes in the ideas, scenarios, or evaluations cause dissimilar results, i.e., results that do not allow the same conclusions to be drawn, we speak of

[25] http://core.kmi.open.ac.uk/intro/data_dumps.

non-reproducibility. Non-reproducibility is expected when significant changes are made to the ideas, scenarios, or evaluations. However, if minor changes are made but results are unexpectedly dissimilar, then we speak of what we term the "butterfly effect".

During the review, we found several examples of this butterfly effect, i.e., variations in experimental results that we considered unexpected and non-reproducible. For instance, the developers of the recommender system *bx* report that the effectiveness of their recommender system varied by a factor of three at different institutions, although the same recommendation approach was used [116]. Lu et al. reported that the *translation* model had twice the accuracy of the *language* model [86], but in another evaluation, accuracy was only 18 % higher [49]. Huang et al. report that the *Context-aware Relevance Model* (CRM) and *cite-LDA* performed similarly, but in another evaluation by the same authors, CRM performed significantly worse than cite-LDA [56]. Lu et al. found that sometimes terms from the abstract performed better than terms from the body-text, while sometimes the opposite was true [86]. Zarrinkalam and Kahani found that sometimes terms from the title *and* abstract were most effective, while sometimes terms from the title, abstract, *and* citation context were most effective [125]. Bethard and Jurafsky reported that citation counts *strongly* increased the effectiveness of their recommendation approach [13], while He et al. reported that citation counts *slightly* increased the effectiveness of their approach [51].

Most interesting with respect to the butterfly effect, there were some evaluations by the TechLens team (Table 8). The TechLens team evaluated several content-based (CBF) and collaborative filtering (CF) approaches for research-paper recommendations. In 2002, McNee et al. conducted an offline evaluation in which CF and CBF performed similarly [93]. However, their additional user study led to a different result—CBF outperformed CF. A user study by Torres et al. in 2004 reports results similar to the user study by McNee et al. (CBF outperformed CF) [117]. However, the offline evaluation from Torres et al. contradicted the previous results—this time, CF outperformed CBF. In 2006, another user study by McNee et al. indicated that CF (slightly) outperforms CBF [87], which showed the opposite of the previous user studies. In 2009, Dong et al., who are not affiliated with TechLens, evaluated the approaches of Torres et al. with an offline evaluation [24]. In this evaluation, CBF outperformed CF, contradicting the previous offline results from Torres et al. In 2010, Ekstrand et al. found that CBF performed worse than CF in both an offline evaluation and a user study, which again did not align with the previous findings [26].

The authors of the studies provide some potential reasons for the variations, such as different datasets (as discussed in Sect. 4.6) differences in user populations, and variations in the implementations. However, these reasons can only explain *some* of the variations. Overall, we consider most of the different outcomes to be unexpected. We view this as a problem, since we see the primary purpose of evaluations in aiding developers and researchers to identify the most effective recommendation approaches (for a given scenario). Consequently, a developer looking for an effective recommendation approach, or a researcher needing an appropriate baseline to compare a novel approach against, would not find much guidance in the existing evaluations. Similarly, the currently existing evaluations do not help to identify whether CF or CBF is more promising for research-paper recommender systems.

Interestingly, reproducibility is widely ignored by the (research-paper) recommender-system community, even by researchers focusing on recommender-systems evaluation. For instance, Al-Maskari et al. analyzed how well classic IR evaluation metrics correlated with user satisfaction in recommender systems [297]. Gunawardana and Shani published a survey about accuracy metrics [224]. Herlocker et al. wrote an article on how to evaluate collaborative filtering approaches [225]. Various authors showed that offline and online evaluations often provide contradictory results [93,269,280]. Many papers about various aspects of recommender-system evaluation have been published [226,268,275,280,283,298–300]. However, while many of the findings in these papers are important with respect to reproducibility, none of the authors mentioned or discussed their findings in the context of reproducibility.

The neglect of reproducibility in recommender-systems evaluation is also observed by Ekstrand et al. and Konstan and Adomavicius. They state that "it is currently difficult to reproduce and extend recommender systems research results," evaluations are "not handled consistently" [301], and many research papers "contribute little to collective knowledge," primarily due to non-reproducibility of the results [302]. They concluded:

Table 8 Results of different CBF and CF evaluations

	McNee et al. [93]		Torres et al. [117]		McNee et al. [87]		Dong et al. [24]		Ekstrand et al. [26]	
	Offline	User std.	Offline	User std.	Offline	User std.	Offline	User std.	Offline	User std.
CBF	Draw	Win	Lose	Win	–	Lose	Win	–	Lose	Lose
CF	Draw	Lose	Win	Lose	–	Win	Lose	–	Win	Win

> "[T]he Recommender Systems research community is facing a crisis where a significant number of papers present results that contribute little to collective knowledge […] often because the research lacks the […] evaluation to be properly judged and, hence, to provide meaningful contributions".

Not all researchers agree that the primary purpose of evaluations is to aid developers and researchers identify the most effective recommendation approaches. When we submitted a paper on reproducibility to the ACM RecSys conference observing that the evaluations were largely non-reproducible, one reviewer commented:

> "I think that it is widely agreed in the community that this [non-reproducibility] is just the way things are – if you want a recsys for a specific application, there is no better way than just test and optimize a number of alternatives. This probably cannot be avoided – there will never be a sufficient set of experiments that would allow "practitioners" to make decisions without running through this optimization process for their specific app and dataset".

When it comes to the lack of reproducibility, we hope that the research community can move beyond the observation that "*this is just the way things are*", given that reproducibility is a "bedrock principle in the conduct and validation of experimental science" [294]. The view of the reviewer above also leads to the question: How should the "number of [promising] alternatives" be determined? At least for research-paper recommender systems, there is no small number of promising alternatives; there are only *all* alternatives, because nearly all the evaluated approaches were most effective in at least one evaluation. Practitioners could hardly implement all approaches to find the most effective approach for their scenario. Even if a few promising approaches were identified, how should they be optimized? There is no list of parameters that might be worth optimizing, and even if there were, there would probably be dozens of parameters, each with dozens or even hundreds of possible values, that would require testing. Again, this would hardly be feasible for someone who wanted to implement a recommender system. In addition, datasets and "specific features" of recommender systems change over time. What does this mean for the operation of a recommender system? Would the operator have to reevaluate the "number of alternatives" every time documents are added to the recommendation corpus, or whenever minor features of the recommender system were changed?

5 Survey of the recommendation classes

Aside from collaborative and content-based filtering, which were briefly introduced in Sect. 2, there are feature-based, knowledge-based, behavior-based, citation-based, context-based, ruse-based, and many more recommendation classes [133,187,300,303–306]. We consider the following seven classes to be most appropriate for distinguishing the approaches in the field of research-paper recommender systems:

1. Stereotyping
2. Content-based Filtering
3. Collaborative Filtering
4. Co-Occurrence
5. Graph-based
6. Global Relevance
7. Hybrid

Originally, we planned to review the most promising approach or approaches of each recommendation class. However, as the review of the evaluations showed, most approaches were evaluated in ways making them nearly impossible to compare. Therefore, the most promising approaches could not be determined. Instead, we provide an overview of the most important aspects and techniques that have been used in the field. The analysis is based on the "in-depth" dataset, i.e., the 127 articles on 62 recommendation approaches that we classified as significant.

5.1 Stereotyping

Stereotyping is one of the earliest user modeling and recommendation classes. It was introduced by *Rich* in the recommender system *Grundy*, which recommended novels to its users [307]. *Rich* was inspired by stereotypes from psychology that allowed psychologists to quickly judge people based on a few characteristics. *Rich* defined stereotypes—which she called "facets"—as collections of characteristics. For instance, Grundy assumed that male users have "a fairly high tolerance for violence and suffering, as well as a preference for thrill, suspense, fast plots, and a negative interest in romance" [307]. Consequently, Grundy recommended books that had been manually classified to match the facets.

One major problem with stereotypes is that they can pigeonhole users. While many men may have a negative interest in romance, this is not true for *all* men. In addition, building stereotypes is often labor intensive, since the items typically need to be manually classified for each facet. This limits the number of items, for example, books that can reasonably be personalized [308].

Advocates of stereotype approaches argue that once the stereotypes are created the recommender system needs little computing power and may perform quite well in practice. For instance, Weber and Castillo observed that female users were usually searching for the composer Richard Wagner when they entered the search query 'Wagner' on *Yahoo!* [309]. In contrast, male users entering the same query were usually looking for the Wagner paint sprayer. Weber and Castillo modified the search algorithm to show the Wikipedia page for Richard Wagner to female users, and the homepage of the Wagner paint sprayer company to male users searching for ‘Wagner.’ As a result, user satisfaction increased. Similarly, the travel agency *Orbitz* observed that Macintosh users were “40 % more likely to book a four- or five-star hotel than PC users” and when booking the same hotel, Macintosh users booked the more expensive rooms [310]. Consequently, Orbitz assigned their website visitors to either the “Mac User” or “PC user” stereotype, and Mac users received recommendations for pricier hotels than PC users. All parties benefited—users received more relevant search results, and Orbitz received higher commissions.

In the domain of research-paper recommender systems, only Beel et al. applied stereotypes [311,312]. The authors assume that all users of their reference-management software *Docear* are researchers or students. Hence, papers and books are recommended that are potentially interesting for researchers and students (for example, papers about optimizing scholarly literature for Google Scholar [313]). Beel et al. used stereotypes as a fallback model when other recommendation approaches could not deliver recommendations. They report mediocre performance of the stereotype approach with click-through rates (CTR) around 4 %, while their content-based filtering approaches achieved CTRs over 6 %.

5.2 Content-based filtering

Content-based filtering (CBF) is one of the most widely used and researched recommendation class [262]. One central component of CBF is the user modeling process, in which the interests of users are inferred from the items that users interacted with. “Items” are usually textual, for instance emails [314] or webpages [315]. “Interaction” is typically established through actions, such as downloading, buying, authoring, or tagging an item. Items are represented by a content model containing the items’ features. Features are typically word-based, i.e., single words, phrases, or n-grams. Some recommender systems also use non-textual features, such as writing style [316,317], layout information [318,319], and XML tags [320]. Typically, only the most descriptive features are used to model an item and users and these features are commonly weighted. Once the most discriminative features are identified, they are stored, often as a vector that contains the features and their weights. The user model typically consists of the features of a user’s items. To generate recommendations, the user model and recommendation candidates are compared, for example using the vector space model and the cosine similarity coefficient.

In the research-paper recommender-system community, CBF is the predominant recommendation class: of the 62 reviewed approaches, 34 (55 %) applied the idea of CBF [7,51,91]. “Interaction” between users and items was typically established through authorship [2,115,118], having papers in one’s personal collection [7,19,60], adding social tags [29], or downloading [106], reading [124], and browsing papers [15,48,92].

Most of the reviewed approaches use plain words as features, although some use *n-grams* [29,101], *topics* (words and word combinations that occurred as social tags on CiteULike) [61], and *concepts* that were inferred from the Anthology Reference Corpus (ACL ARC) via Latent Dirichlet Allocation [13], and assigned to papers through machine learning. A few approaches utilize non-textual features, and if they did then these non-textual features were typically utilized *in addition* to words. Giles et al. used citations in the same way as words were used and weighted the citations with the standard TF-IDF measure (they called this method *CC-IDF*) [43]. Others adopted the idea of CC-IDF or used it as a baseline [1,7,27]. However, Beel recently provided some initial evidence that CC-IDF might not be an ideal weighting scheme [311]. Zarrinkalam and Kahani considered *authors* as features and determined similarities by the number of authors two items share [125].

The approaches extracted words from the title [79,85,112], abstract [26,51,61], header [43], introduction [57], foreword [57], author-provided keywords [26,57,58], and bibliography [27], as well as from the papers’ body text [65,101,112]. The approaches further extracted words from external sources, such as social tags [58,62], ACM classification tree and DMOZ categories [91,94], and citation context [51,55,65]. Using citation context is similar to the way search engines use anchor analysis for webpage indexing since the 1990s [320,322]. Citation context analysis was also used in academic search before it was used by research-paper recommender systems [323].

Words from different document fields have different discriminative powers [324]. For instance, a word occurring in the title is usually more meaningful than a word occurring in the body text. Nascimento et al. accounted for this and weighted terms from the title three times stronger than terms from the text body, and text from the abstract twice as strong [101]. This weighting scheme was arbitrarily selected and not based on empirical evidence. Huang et al. experimented with different weights for papers’ content and citation context [55]. They found that an equal weight for both fields achieved the highest precision. The other reviewed approaches that used text from different fields did not report on any field weighting.

The most popular model to store item representations and user models was the Vector Space Model (VSM), which was used by 9 (64 %) of those 14 approaches that reported how they stored the user and item models. Other approaches modeled their users as a graph [102,103,122], as a list of topics that were assigned through machine learning [91], or as an ACM hierarchy [66]. Of those who used the VSM, all but one used the cosine measure to calculate similarities between user models and recommendation candidates. In 1998, Giles et al. compared headers of documents with a string distance measure [1], but neither they nor others mentioned this technique again, which might imply that string edit distance was not effective.

TF-IDF was the most popular weighting scheme (70 %) among those approaches for which the scheme was specified. Other weighting schemes included plain term frequency (TF) [29,101,115], and techniques that the authors called “phrase depth” and “life span” [29].

CBF has a number of advantages compared to stereotypes. CBF allows a user-based personalization so that the recommender system can determine the best recommendations for each user individually, rather than being limited by stereotypes. CBF also requires less up-front classification work, since user models can be created automatically.

On the downside, content-based filtering requires more computing power than stereotyping. Each item must be analyzed for its features, user models must be built, and similarity calculations must be performed. If there are many users and many items, these calculations require significant resources. The weakness of content-based filtering is its low serendipity and overspecialization leading it to recommend items as similar as possible to the ones a user already knows [262]. Content-based filtering also ignores quality and popularity of items [24]. For instance, two research papers may be considered equally relevant by a CBF recommender system if the papers share the same terms with the user model. This relevance might not always be justified, for example if one paper was written by an authority in the field and presents original results, while another paper was written by a student who paraphrases the results of other research papers. Ideally, a recommender system should recommend only the first paper but a CBF system would fail to do so. Another criticism of content-based filtering is that it is dependent on access to the item’s features [24]. For research-paper recommendations, usually PDFs must be processed and converted to text, document fields must be identified, and features, such as terms must be extracted. None of these tasks are trivial and they may introduce errors into the recommendations [256,325,326].

5.3 Collaborative filtering

The term “collaborative filtering” (CF) was coined in 1992 by Goldberg et al., who proposed that “information filtering can be more effective when humans are involved in the filtering process” [327]. The concept of collaborative filtering as it is understood today was introduced 2 years later by Resnick et al. [221]. Their theory was that users like what like-minded users like, where two users were considered like-minded when they rated items alike. When like-minded users were identified, items that one user rated positively were recommended to the other user, and vice versa. Compared to CBF, CF offers three advantages. First, CF is content independent, i.e., no error-prone item processing is required [117,264,286]. Second, because humans do the ratings, CF takes into account real quality assessments [24]. Finally, CF is supposed to provide serendipitous recommendations because recommendations are not based on *item* similarity but on *user* similarity [87,286].

From the reviewed approaches, only 11 (18 %) applied collaborative filtering [93,106,119]. Yang et al. intended to let users rate research papers, but users were “too lazy to provide ratings” [124]. Naak et al. faced the same problem and created artificial ratings for their evaluation [98]. This illustrates one of the main problems of CF: CF requires user participation, but often the motivation to participate is low. This problem is referred to as the “cold-start” problem, which may occur in three situations [264]: new users, new items, and new communities or disciplines. If a new user rates few or no items, the system cannot find like-minded users and, therefore, cannot provide recommendations. If an item is new in the system and has not been rated yet by at least one user, it cannot be recommended. In a new community, no users have rated items, so no recommendations can be made and as a result, the incentive for users to rate items is low.

To overcome the cold-start problem, implicit ratings may be inferred from the interactions between users and items. Yang et al. inferred implicit ratings from the number of pages the users read: the more pages users read, the more the users were assumed to like the documents [124]. Pennock et al. interpreted interactions, such as downloading a paper, adding it to ones’ profile, editing paper details, and viewing its bibliography as positive votes [106]. McNee et al. assumed that an author’s citations indicate a positive vote for a paper [93]. They postulated that when two authors cite the same papers, they are like-minded. Similar, if a user reads or cites a paper the citations of the cited paper are supposed to be liked by the user.

Using *inferred* ratings voids CF’s advantage of being based on real user quality assessments. This criticism applies to citations as well as to other types of implicit ratings [328–330]. For example, we cite papers in this survey that had inadequate evaluations, or were written in barely understandable English. Thus, interpreting citations always as a positive vote can be misguiding. Similarly, when a user spends a lot of time reading a paper this *could* mean that the paper contains interesting information that the user would rate positively, but it could also mean that the paper is difficult to understand and requires a lot of effort to read. Consequently, CF’s advantage of explicit human quality assessments mostly vanishes when implicit ratings are used.

Using citations as inferred ratings might also void CF's second advantage of being content-independent. Typically, reliable citation data are not widely available. Therefore, access to the paper's content is required to build a citation network, but this process is even more fault-prone than word extraction in CBF. In CBF, the text of the papers must be extracted, and maybe fields such as the title or abstract must be identified. For citation-based CF, the text must also be extracted but in the text, the bibliography and its individual references must be identified, including their various fields including title and author. This is an error-prone task [325].

A general problem of collaborative filtering in the domain of research-paper recommender systems is sparsity. Vellino compared the implicit ratings on Mendeley (research papers) and Netflix (movies), and found that sparsity on Netflix was three orders of magnitude lower than on Mendeley [211]. This is caused by the different ratio of users and items. In domains like movie recommendations, there are typically few items and many users. For instance, the movie recommender MovieLens has 65,000 users and 5,000 movies [225]. Typically, many users watched the same movies. Therefore, like-minded users can be found for most users and recommendations can be given effectively. Similarly, most movies have been watched by at least some users and hence most movies can be recommended. The situation is different in the domain of research papers. There are typically few users but millions of papers, and very few users have rated the same papers. Finding like-minded users is often not possible. In addition, many papers are not rated by any users and, therefore, cannot be recommended.

There are further critiques of CF. Computing time for CF tends to be higher than for content-based filtering [264]. Collaborative filtering is generally less scalable and requires more offline data processing than CBF [331]. Torres et al. note that collaborative filtering creates similar users [117] and Sundar et al. observe that collaborative filtering dictates opinions [332]. Lops makes the criticism that collaborative filtering systems are black boxes that cannot explain why an item is recommended except that other users liked it [262]. Manipulation is also a problem: since collaborative filtering is based on user opinions, blackguards might try to manipulate ratings to promote their products so they are recommended more often [333–335].

5.4 Co-occurrence recommendations

To give *co-occurrence* recommendations, those items are recommended that frequently co-occur with some source items. One of the first applications of co-occurrence was co-citation analysis introduced by Small [233]. *Small* proposed that two papers are more related to each other, the more often they are co-cited. Many others adopted this concept, the most popular example being Amazon's "*Customers Who Bought This Item Also Bought…*." Amazon analyzes which items are frequently bought together, and when a customer browses a product, items frequently bought with that item are recommended.

One advantage of co-occurrence recommendations is the focus on relatedness instead of similarity. Similarity expresses how many features two items have in common. Recommending similar items, as CBF is doing, is often not ideal because similar items are not serendipitous [336]. In contrast, *relatedness* expresses how closely coupled two items are, not necessarily dependent on their features. For instance, two papers sharing the same features (words) are similar. In contrast, paper and pen are not similar but related, because both are required for writing letters. Hence, co-occurrence recommendations provide more serendipitous recommendations and, in this way, are comparable to collaborative filtering. In addition, no access to content is needed and complexity is rather low. It is also rather easy to generate anonymous recommendations, and hence to assure users' privacy. On the downside, recommendations are not highly personalized and items can only be recommended if they co-occur at least once with another item.

Six of the reviewed approaches are based on co-occurrences (10 %). Three of those approaches analyze how often papers are *co-viewed* during a browsing session [15,48,92]. Whenever a user views a paper, those papers that were frequently co-viewed with the browsed paper are recommended. Another approach uses proximity of co-citations to calculate document relatedness [44]: the closer the proximity of two references within a paper, the more related the cited papers are assumed to be. Pohl et al. compared the effectiveness of co-citations and co-downloads and found that co-downloads are more effective than co-citations only in the first 2 years after a paper is published [110].

Calculating co-occurrence recommendations is not always feasible. For instance, on arXiv.org, two-thirds of all papers have no co-citations, and those that do usually have no more than one or two [110]. Despite its limitations, co-occurrence recommendations seem to perform quite well. Two popular research-paper recommender systems, bX and BibTip, both rely on co-occurrence recommendations and deliver millions of recommendations every month [15,92].

5.5 Graph based

Ten of the reviewed approaches utilize the inherent connections that exist in academia (16 %). Based on these connections, the approaches build graph networks that typically show how papers are connected through citations [10,72,84]. Sometimes, graphs include authors [4,79,127], users/customers [57], venues [10,79,127], genes and proteins [4,79], and the years the papers were published [79]. Lao et al. even included terms from the papers' titles in the graph, which makes their approach a mixture of the graph and content based approach [79]. Depending on the entities in the graph, connections can be citations [10,79,84], purchases [57], "published in" relations, [10,79,127], authorship [4,10,127], relatedness between genes[26] [4], or occurrences of genes in papers [79]. Some authors connected entities based on non-inherent relations. For instance, Huang et al. and Woodruff et al. calculated text similarities between items and used the text similarity to connect papers in the graph [57,123]. Other connections were based on attribute similarity,[27] bibliographic coupling, co-citation strength [57,123,127], or demographic similarity [57]. Once a graph was built, graph metrics were used to find recommendation candidates. Typically, one or several input papers were given and from this input random walks with restarts were conducted to find the most popular items in the graph [46,72,79].

5.6 Global relevance

In its simplest form, a recommender system adopts a one-fits-all approach and recommends items that have the highest global relevance. In this case, the relevance is not calculated specific to a user. Instead, some global measures are used, such as overall popularity. For instance, a movie-rental system could recommend those movies that were most often rented or that had the highest average rating over all users. In this case, the basic assumption would be that users like what most other users like.

From the reviewed approaches, none used global relevance exclusively but many used it as an additional ranking factor. For instance, five CBF approaches used global popularity metrics in their rankings [13,51,125]. They first determined a list of recommendation candidates with a user-specific CBF approach. Then, the recommendation candidates were re-ranked based on the global relevance metrics. Popular metrics were PageRank [13], HITS [51], Katz metric [51], citation counts [13,51,112], venues' citation counts [13,112], citation counts of the authors' affiliations [112], authors' citation count [13,112], h-index [13], recency of articles [13], title length [112], number of co-authors [112], number of affiliations [112], and venue type [112].

Strohman et al. report that the Katz metric, which quantifies relevance as a function of the paths between two nodes (the shorter the paths the higher the relevance), strongly improved precision [114]. All variations that included Katz were about twice as good as those variations without. Bethard and Jurafsky report that a simple citation count was the most important factor, and age (recency) and *h*-index were even counterproductive [13]. They also report that considering these simple metrics doubled mean average precision compared to a standard content-based filtering approach.

5.7 Hybrid recommendation approaches

Approaches of the previously introduced recommendation classes may be combined in hybrid approaches. Many of the reviewed approaches have some hybrid characteristics. For instance, several of the CBF approaches use global relevance attributes to rank the candidates, or graph methods are used to extend or restrict potential recommendation candidates. This type of hybrid recommendation technique is called "feature augmentation" [303]. It is a weak form of hybrid recommendation technique, since the primary technique is still dominant. In true hybrids, the combined concepts are similarly important [303,337]. From the reviewed approaches, only some of the TechLens approaches may be considered true hybrid approaches.

TechLens [26,63,64,87,93,117] is one of the most influential research-paper recommender systems, although it was not the first like some have claimed (e.g., [109]). TechLens was developed by the GroupLens[28] team. Currently TechLens is not publicly available, although the GroupLens team is still active in the development and research of recommender systems in other fields. Between 2002 and 2010, Konstan, Riedel, McNee, Torres, and several others published six articles related to research-paper recommender systems. Often, McNee et al.'s article from 2002 is considered to be the original TechLens article [93]. However, the 2002 article 'only' introduced some algorithms for recommending citations, which severed as a foundation for TechLens, which was introduced in 2004 by Torres et al. [117]. Two articles about TechLens followed in 2005 and 2007 but added nothing new with respect to recommendations [63,64]. In 2006, McNee et al. analyzed potential pitfalls of recommender systems [87]. In 2010, Ekstrand et al. published another article on the approaches of TechLens and suggested enhancements for the approaches [26].

TechLens' algorithms were adopted from Burke [303] and consisted of three CBF variations, two CF variations, and five hybrid approaches.

Content-Based Filtering: *Pure-CBF* served as a baseline in the form of standard CBF in which a term-based user model was compared with the recommendation candidates. In the case of TechLens, terms from a single input paper were used. In *CBF-Separated*, for each paper being cited by the input paper, similar papers are determined separately and at the end the different recommendation lists are merged and presented to the user.

[26] Relatedness between genes was retrieved from an external data source that maintained information about gene relatedness.

[27] Attribute similarity was calculated, e.g., based on the number of pages.

[28] http://grouplens.org/.

In *CBF-Combined*, terms of the input paper and terms of all papers being cited by the input paper are combined in the user model. Then, the papers most similar to this user model are recommended.

Collaborative Filtering: *Pure-CF* served as another baseline and represented the collaborative filtering approach from *McNee et al*., in which papers were interpreted as users and citations were interpreted as votes [93]. In *Denser-CF*, citations of the input paper were additionally included in the user model.

Hybrid: With *Pure-CF->CBF Separated*, recommendations were first created with Pure-CF. These recommendations were then used as input documents for CBF-Separated. Similarly, *Pure-CF->CBF Combined*, *CBF Separated-> Pure-CF*, and *CBF-Combined->Pure-CF* were used to generate recommendations. *Fusion* created recommendations with both CBF and CF independently and then merged the recommendation lists.

Despite various evaluations of the approaches, it remains unclear which are most promising (refer to the explanations in Sect. 4.7).

6 Survey of the research field and shortcomings

Our survey already revealed that there are some shortcomings and challenges in the field of research-paper recommender systems. This is especially the case when it comes to evaluations, which are often non-reproducible and incomparable. However, during the review, we identified more challenges. In the next sections, we introduce these challenges, and hope to stimulate a discussion about future research directions to enhance research-paper recommender systems.

6.1 Neglect of user modeling

A fundamental part of generating recommendations is the user modeling process that identifies a user's information needs [263]. Ideally, a recommender system identifies the needs automatically by inferring the needs from the user's item interactions. Alternatively, the recommender system asks users to specify their needs by providing a list of keywords or through some other method. However, in this case a recommender system becomes very much like a search engine and loses one of its main features, namely the capability to recommend items even if users do not know exactly what they need.

Of the 62 reviewed approaches, 50 (81 %) required users to either explicitly provide keywords [52,98,119], or to provide text snippets (e.g., an abstract) [13,111,125], or to provide a single paper as input [26,101,114], or the approaches ignored the user modeling process entirely. These approaches neglect one of the most important parts of a recommender system, which makes the approaches very similar to classic search, or related document search [338–340], where users provide search terms or an input paper, and receive a list of search results or similar papers. Of course, neither classic search nor related-document search are trivial tasks in themselves, but they neglect the user modeling process and we see little reason to label such systems as recommender systems.

Of the reviewed approaches, 12 (19 %) inferred information from the items the users interacted with. However, most approaches that inferred information automatically used *all* papers that a user authored, downloaded, etc. [62,94,122]. This is not ideal. When inferring information automatically, a recommender system should determine the items that are currently relevant for the user-modeling process [341]. For example, 10-year-old user-authored papers are probably not suitable to describe a user's current information needs. This aspect is called "concept drift" and it is important for creating meaningful user models. In the research-paper recommender systems community, concept drift is widely ignored: a mere three approaches considered concept drift in detail. Middleton et al. weight papers by the number of days since the user last accessed them [91]. Watanabe et al. use a similar approach [122]. Sugiyama and Kan, who use the user's authored papers, weight each paper based on the difference between a paper's publication year, and the year of the most recently authored paper [115]. In addition, they found that it makes sense to include only those papers that the user authored in the past 3 years [115].

Another important question in user modeling is the user-model size. While in search, user models (i.e., search queries) typically consist of a few words; user models in recommender systems may consist of hundreds or even thousands of words. Of the reviewed CBF approaches, 30 (88 %) did not report the user-model size. The few authors who reported the user-model size, usually stored fewer than 100 terms. For instance, Giles et al. made us in the top 20 words of the papers [43].

6.2 Focus on accuracy

The research-paper recommender-system community places a strong focus on accuracy, and seems to assume that an accurate recommender system will lead to high user satisfaction. However, outside the research-paper recommender-system community, it is agreed that many aspects beyond accuracy affect user satisfaction. For instance, users might become dissatisfied with accurate recommendations when they have no trust in the recommender system's operator [342], their privacy is not ensured [300], they need to wait too long for recommendations [300], or they find the user interfaces unappealing [343]. Other factors that affect user satisfaction are confidence in a recommender system [263], data security [344], diversity [345], user tasks [87], item's lifespan [346] and novelty [347], risk of accepting recommendations [348], robustness against spam and fraud [349], transparency and explanations [350], time to first recommendation [225], and interoperability [351].

Among the reviewed articles, a few authors considered aspects beyond accuracy, as shown in the following sections.

6.2.1 Users' tasks

Torres et al. from TechLens' considered a user's current task in the recommendation process. The authors distinguished between users who wanted to receive authoritative recommendations and novel recommendations [117]. They showed that different recommendation approaches were differently effective for these tasks. The developers of *TheAdvisor* let users specify whether they are interested in classic or recent papers [73]. Uchiyama et al. found that students are typically not interested in finding papers that are "similar" to their input paper [118]. This finding is interesting because content-based filtering is based on the assumption that user want similar papers. However, the study by Uchiyama et al. was based on 16 participants. As such, it remains uncertain how significant the results are.

6.2.2 Diversity

Diversity of recommendations was mentioned in a few articles, but considered in depth only by two author groups (Vellino et al. and Küçüktunç et al.). Vellino et al. measured diversity as the number of journals from which articles were recommended [119]. If recommendations were all from the same journals, diversity was zero. They compared diversity of a CF approach with the co-occurrence approach from bX and found that CF had a diversity of 60 % while diversity of bX was 34 %. Küçüktunç et al. from TheAdvisor published two articles about diversity in research-paper recommender systems [70,71]. They provided a survey on diversification techniques in graphs and proposed some new techniques to measure diversity.

6.2.3 Layout

Farooq et al. from CiteSeer analyzed which information users wanted to see when receiving recommendations in RSS feeds [138]. They found that the information to display varies for the type of recommendation. In one approach, Farooq et al. recommended papers that cited the user's papers. In this case, users preferred to see the citing paper's bibliographic data (e.g., title, author) and the context of the citation—the sentence in which the citation appeared. When papers were recommended that were co-cited with the users' papers, citation context was not that important. Rather, the users preferred to see the bibliographic data and abstract of the co-cited paper. When papers were recommended that had a similar content to the users' papers, users preferred to see bibliographic data and abstract. These findings are interesting because from the reviewed recommender systems the majority display only the title and not the abstract.

As part of our work, we researched the impact of labeling and found that papers labeled as 'sponsored recommendation' performed worse than recommendations with a label that indicated that the recommendations were 'organic,' though the recommended papers were identical [5]. It also made a difference if paper recommendations were labeled as 'Sponsored' or 'Advertisement' although both labels indicate that recommendations are displayed for commercial reasons.

6.2.4 User characteristics

For our own recommender system Docear, we found that researchers who registered tended to have higher click-through rates than unregistered users (6.95 vs. 4.97 %) [8]. In addition, older users seem to have higher average click-through rates (40–44 years: 8.46 %) than younger users (20–24 years: 2.73 %) [8]. Middleton et al. also report differences for different user groups. Click-through rates in their recommender system, Quickstep, was around 9 %, but only around 3.5 % for Foxtrot, although both systems applied very similar approaches. However, Quickstep users were recruited from a computer science laboratory, while Foxtrot was a real-world system offered to 260 faculty members and students (although only 14 % of users used Foxtrot at least three times).

Click-through rates from the bX recommender are also interesting [116]. They varied between 3 and 10 % depending on the university in which recommendations were shown (bX provides more than 1000 institutions with recommendations) [25]. This could have been caused by different layouts of the recommendations, but it might also have been caused by the different backgrounds of students.

6.2.5 Usage duration

Middleton et al. reported that the longer someone used the recommender system; the lower click-through rates became [94]. Jack reports the opposite, namely that precision increased over time ($p = 0.025$ in the beginning, $p = 0.4$ after 6 months) and depended on a user's library size ($p = 0.08$ for 20 articles and $p = 0.40$ for 140 articles) [58]. We showed that it might make sense to be "persistent" and show the same recommendations to the same users multiple times—even recommendations that users had clicked before were often clicked again [6].

6.2.6 Recommendation medium

User satisfaction also depends on the medium through which recommendations are made. Middleton et al. report that recommendations via email received half the click-through rate as the same recommendations delivered via a website [94]. Of the reviewed recommender systems, only Docear [7] and Mendeley [58] provide recommendations through a desktop software; CiteSeer provided recommendations in a news feed [138]; and all others deliver their recommendations through websites. If and how click rates differ when recommendations are delivered via desktop software compared to a website remains unknown.

6.2.7 Relevance and profile feedback

Relevance feedback is a common technique to improve recommendations [263] but it is widely ignored in the research-paper recommender-system community. Middleton et al. showed that profile feedback is better than relevance feedback: allowing users to edit their user models is more effective than just learning from relevance feedback [94]. Bollacker et al. from CiteSeer allowed their users to edit their profiles but conducted no research on the effectiveness of this feature [80].

6.3 Translating research into practice

Translating research into practice is a current challenge in the research paper recommender system community. Out of the large number of proposed approaches in the field, 24 research-paper recommender systems could be used by users in practice (Table 9).[29] Of these 24 recommender systems, eight (33 %) never left the prototyping stage—and today only one of the prototypes is still publicly available. Of the remaining recommender systems, four are offline (25 %), five are no longer actively maintained (31 %),[30] while seven are running and actively maintained (44 %). Of the seven active recommender systems, four operators are involved with the recommender-system research community (see footnote 29) and publish information about their systems.

Most of the real-world recommender systems apply simple recommendation approaches that are not based on recent research. For instance, as far as we could tell, PubMed was still using an approach introduced in 2007; ResearchGate is using a simple content-based filtering approach similar to classic search[31]; CiteULike apparently uses two approaches from 2008/2009; and BibTip and bX are using simple co-occurrence approaches. Whether the RefSeer system applies all the results from their research remains unclear. In other words, the reviewed research typically did not affect real-world recommender systems.

6.4 Persistence and authorities

One reason the research is not transferred directly into practice might be a lack of persistence and authorities in the field. Of the 276 authors who authored the 185 articles, 201 (73 %) published just one article (Fig. 5). 15 authors (5 %) published five or more articles, but of these authors, several were co-authors publishing the same articles. This means that there are only a few groups that consistently publish research in the field of research-paper recommender systems.

The most productive authors are C. Lee Giles and his co-author P. Mitra from CiteSeer/RefSeer (Tables 10, 11). No other authors have published as many articles (17) over as long a period of time (16 years) on as many different aspects of research-paper recommender systems. Other productive authors are A. Geyer-Schulz and his co-authors M. Hahsler, and M. Jahn from BibTip. They published 13 articles, but these were less often cited in the community than those of Giles et al. The articles are also narrower in scope than those of the CiteSeer authors. Our research group, i.e., J. Beel, S. Langer, M. Genzmehr, and B. Gipp from Docear, authored eight papers between 2009 and 2013, including posters or short papers. The research concentrated on aspects beyond accuracy, such as the impact of labeling recommendations and the impact of demographics on click-through rates. O. Küçüktunç and his co-authors E. Saule and K. Kaya from TheAdvisor published six articles focusing on diversity and graph-based recommendations. J. A. Konstan, S. M. McNee, R. Torres, and J.T. Riedel, who are highly recognized authors in the field of recommender systems, developed TechLens and authored six articles relating to research-paper recommender systems during 2002 and 2010. Two of their articles influenced the work of several others and are among the most cited articles we reviewed [93,117]. W. W. Cohen and his PhD student N. Lao are also two productive authors. They authored six articles from 2008 to 2012 (some of which are unpublished). SE. Middleton and his co-authors published five articles. It is interesting to note that five of the six most productive research groups have access to real-world recommender systems.

[29] The recommender systems of Mendeley, CiteULike, and CiteSeer are counted twice because they offer or offered two independent recommender systems.

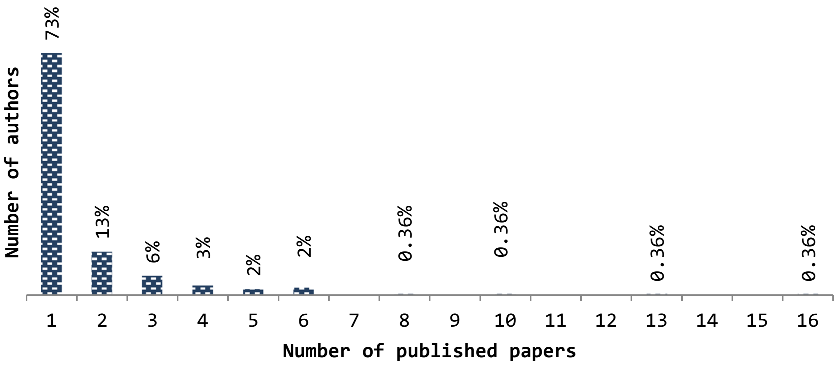
[30] We classified a recommender system as not actively maintained if no article was published or no changes were made to the system for a year.

[31] ResearchGate also applied other recommender systems, e.g., for people or news, and it seems that these approaches are more sophisticated.

Table 9 Overview of recommender systems surveyed

Status	Name	Maturity	Research oriented	Type	Presentation
Active	BibTip	Real system	No[1]	Stand-alone	Webpage
Active	bx	Real system	No	Stand-alone	Webpage
Active	Docear	Real system	Yes	On-top	Software
Active	Mendeley	–	–	–	–
	Related Papers	Real system	Yes	On-top	Software
	Suggest	Real system	Yes	On-top	Software
Active	RefSeer	Real system	Yes	Stand-alone	Webpage
Active	Scholar Update	Real system	No	On-top	Webpage
Idle	CiteULike	–	–	–	–
	CF	Real system	No	On-top	Webpage
	Item–Centric	Real system	No	On-top	Webpage
Idle	PubMed PRMA	Real system	No	On-top	Webpage
Idle	ResearchGate	Real system	No	On-top	Webpage
Idle	TheAdvisor	Real system	Yes	Stand-alone	Webpage
Idle	Who shoud I Cite?	Prototype	Yes	Stand-alone	Webpage
Offline	CiteSeer	–	–	–	–
	Alert	Real system	Yes	On-top	Feed
	Related Documents	Real system	Yes	On-top	Webpage
Offline	Foxtrot	Real system	Yes	Stand-alone	Webpage, Email
Offline	TechLens	Real system	Yes	Stand-alone	Webpage
Offline	NSYSU-ETD	Prototype	Yes	On-top	Webpage
Offline	OSUSUME	Prototype	Yes	On-top	?
Offline	Papits	Prototype	Yes	On-top	Webpage
Offline	Papyres	Prototype	Yes	On-top	?
Offline	Pirates	Prototype	Yes	Stand-alone	?
Offline	Quickstep	Prototype	Yes	Stand-alone	Webpage
Offline	Sarkanto & Synthese	Prototype	Yes	Stand-alone	Webpage

Fig. 5 Papers published per author in the research paper recommender field



6.5 Cooperation

Most articles were authored by multiple authors: the majority of articles had two (26 %), three (26 %) or four authors (18 %) (Fig. 6).[32] 17 % of articles were authored by a single researcher. On first glance, these numbers indicate a high degree of collaboration. However, we noticed that between different co-author groups little cooperation took place. The closest cooperation we could identify was that Giles was part of a committee for a thesis that Cohen supervised [74]. Leading authors of different research groups did not co-author articles together.

Co-author groups frequently seemed to work alone and did not always build on the results of the work done by their peers.

[32] Median author count was three, maximum count eleven.

Table 10 Most productive authors in the research paper recommender field

Authors	Papers
C. Lee Giles	16
A. Geyer-Schulz	13
M. Hahsler	10
J. Beel	8
O. Küçüktunç	6
E. Saule	6
K. Kaye	6
P. Mitra	6
J. A. Konstan	6
W. W. Cohen	6
S. Langer	5
M. Genzmehr	5
M. Jahn	5
S. E. Middleton	5
D. C. De Raure	5

Table 11 Most productive author groups

Groups	Papers
C. Lee Giles; P. Mitra (CiteSeer/RefSeer)	17
A. Geyer-Schulz; M. Hahsler; M. Jahn (BibTip)	13
J. Beel; S. Langer, M. Genzmehr, B. Gipp (Docear)	8
J. A. Konstan; S. M. McNee; R. Torres, J. T. Riedl (TechLens)	6
O. Küçüktunç; E. Saule; K. Kaya (TheAdvisor)	6
W. W. Cohen; N. Lao	6
S. E. Middleton; D. C. De Raure, N. R. Shadbolt (Quickstep & Foxtrot)	5

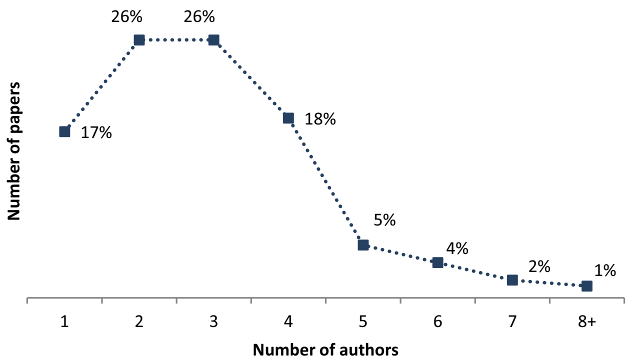


Fig. 6 Number of authors of the reviewed papers

Among the reviewed articles, it rarely happened that authors reported to have built their novel approach based upon an existing approach. This lack of cooperation also becomes apparent when looking at the citations. Although some of the reviewed articles gained many citations, these citations usually resulted from articles outside the research-paper recommender domain. For instance, the paper "Learning multiple graphs for document recommendations" attracted 63 citations since 2008 [127]. From these citations, only three were made by the reviewed articles. Another article, from the BibTiP developers, gained 24 citations since 2002 [32]. From the 24 citations, ten were self-citations and none was from the reviewed articles. Both examples are typical for most of the reviewed articles. One of the few articles that is constantly cited in the research-paper recommender community is an article about TechLens, which accumulated more than 100 citations [117]. However, many authors cited the article for authoritative reasons. In the citing papers, TechLens is mentioned but, with few exceptions, its approaches are neither adopted nor used as a baseline.

6.6 Information scarcity

Most authors provided little detail about their approaches, which makes a re-implementation difficult, if not impossible. For instance, for 24 of the 34 content-based filtering approaches (71 %), the authors did not report the weighting scheme they used (e.g., TF-IDF). The feature representation model (e.g., vector space model) was not reported for 20 approaches (59 %). Whether stop words were removed was not reported for 23 approaches (68 %). For 16 approaches (47 %), no information was given on the fields the terms were extracted from (e.g., title or abstract). This information scarcity means, when an evaluation reports promising results for an approach, that other researchers would not know how to re-implement the approach in detail. If they tried, and guessed the specifics of an approach, the outcome would probably differ significantly from the original. This might cause problems in replicating evaluations, and reproducing research results and hinders the re-implementation and application of promising approaches in real-word recommender systems.

7 Summary and outlook

In the 16 years from 1998 to 2013, more than 200 research articles were published in the field of research-paper recommender systems. The articles consisted primarily of peer-reviewed conference papers (59 %), journal articles (16 %), pre-prints (5 %), and other documents such as presentations and web pages (15 %). The few existing literature surveys in this field cover only a fraction of these articles, which is why we conducted a comprehensive survey of research-paper recommender systems. The review revealed the following information.

Content-based filtering (CBF) is the predominant recommendation class. From 62 reviewed approaches, 34 used CBF (55 %). From these CBF approaches, the majority utilized plain terms contained in the documents. Some used n-grams, or topics based on LDA. A few approaches also utilized non-textual features, such as citations or authors. The most popular model to store item representations was the Vector Space Model. Other approaches modeled their users as graphs, as lists with topics that were assigned through machine learning, or as ACM classification hierarchies. The reviewed approaches extracted text from the title, abstract, header, introduction, foreword, author-provided keywords, bibliography, body text, social tags, and citation context.

Only eleven approaches applied collaborative filtering, and none of them successfully used explicit ratings. Yang et al. intended to develop such a system, but their users were "too lazy to provide ratings" [124]. Hence, implicit instead of explicit ratings were used. Implicit ratings were inferred from the number of pages the users read, users' interaction with the papers (e.g., downloads, edits, views) and citations. The main problem of collaborative filtering for research papers seems to be scarcity. Vellino compared implicit ratings on Mendeley (research papers) and Netflix (movies), and found that scarcity on Mendeley differs from Netflix by a magnitude of three.

Although stereotyping is one of the oldest user modeling approaches, and is successfully applied by services, such as Yahoo!, only one of the reviewed approaches applied stereotypes as a fallback model when other recommendation approaches could not deliver recommendations. The authors reported reasonable performance of the stereotype approach with click-through rates (CTR) around 4 % while the CBF approach achieves CTRs around 6 %. Given these results, we believe that a more thorough analysis of stereotype recommendations could be interesting.

Six of the reviewed approaches were co-occurrence based. Three approaches analyzed how often papers were *co-viewed* during a browsing session: whenever users browsed a paper, the system recommended those papers that had frequently been co-viewed with the browsed paper in the past. Another approach used co-citations to calculate document relatedness. The higher the proximity of two references within a paper, the more related they are assumed to be. Pohl et al. compared the effectiveness of co-citations and co-downloads and found that co-downloads are more effective than co-citations only in the first 2 years after a paper was published [110].

Ten recommendation approaches built graphs to generate recommendations. Such graphs typically included papers that were connected via citations. Some graphs included authors, users/customers, venues, genes and proteins, and publishing years of the papers. Lao et al. even included terms from the papers' titles in the graph. Depending on the entities in the graph, connections were citations, purchases, "published in" relations, authorship, relatedness between genes, and occurrences of genes in papers. Some authors connected the entities based on non-inherent relations. For instance, Huang et al. and Woodruff et al. calculated text similarities between items and used the text similarity to connect papers in the graph. Other connections were based on attribute similarity, bibliographic coupling, co-citation strength, and demographic similarity.

Despite a lot of research about research-paper recommender systems, we identified two main problems in the research field.

First, it is currently not possible to determine the most effective recommendation approaches. If we were asked which recommendation approach to apply in practice or to use as baseline, there is no definite answer. We do not even have a clue as to which of the approaches might be most promising. This problem mainly relates to poor experimental design and lack of information, which includes inadequate evaluations and too little information given by the authors: 22 % of the approaches were not evaluated. Of the remaining approaches, 20 % were not evaluated against a baseline; the majority of the remaining approaches were compared to simple baselines but not to approaches of other researchers in the field. The majority (71 %) of approaches were evaluated using offline evaluations, which are subject to various shortcomings. Some claim that offline evaluations should not be used for evaluating research-paper recommender systems [191]. If that is true, most of the reviewed evaluations would be of little significance. Even if this criticism is unjustified, some problems remain. Many authors pruned their offline datasets in drastic ways. For instance, Pennock et al. removed all documents with fewer than 15 implicit ratings from the corpus. Therefore, 1575 papers remained from the original 270,000 (0.58 %). Results based on such datasets do not allow drawing reliable conclusions about how the approaches might perform in real-world recommender systems. In addition, the majority of the user studies (58 %) had 15 or less participants, which questions the significance of these evaluations. Only 8 % of the approaches were evaluated with online evaluations in real-world recommender systems with real users. Additionally, many authors provided little information about their approaches, which hinders re-implementation. For instance, most authors did not report on the text fields they utilized, or which weighting schemes were used.

To solve this problem, we believe it is crucial that the community discusses and develops frameworks and best-practice guidelines for the evaluation of research-paper recommender-systems. This should include an analysis and discussion of how suitable offline evaluations are; to what extent datasets should be pruned; the creation of datasets comparable to existing TREC datasets; the minimum number of participants in user studies; and which factors influence the results of evaluations (e.g., user demographics). Ideally, a set of reference approaches would be implemented that could be used as baselines. In addition, more details on implementations are needed, based on a discussion of the information needed in research articles.

It is crucial to find out why seemingly minor differences in algorithms or evaluations lead to major variations in the evaluation results. As long as the reasons for these variations are not found, scholars cannot rely on existing research results because it is not clear whether the results can be reproduced in a new recommendation scenario.

Second, we identified unused potential in recommender systems research. This problem has two root causes.

(A) Research results are often not transferred into practice, or considered by peers. Despite the large number of research articles, just a handful of active recommender systems exist, and most of them apply simple recommendation approaches that are not based on recent research results. As such, the extensive research conducted from 1998 to 2013 apparently had a rather minor impact on research-paper recommender systems in the real world. Additionally, developers of several of the active recommender systems do not engage in the research community or publish information about their systems. Some researchers also seem to be unaware of developments in related research domains, such as user modeling, scientometrics, and the reviewer-assignment problem. In addition, the major co-author groups in the domain of research-paper recommender systems do not cooperate much with each other. One reason for some of these problems might be a relatively short-lived interest in the research field. Most authors (73 %) published only a single paper on research-paper recommender systems.

(B) The majority of authors did not take into account that user satisfaction might depend not only on accuracy but also on factors such as privacy, data security, diversity, serendipity, labeling, and presentation. The operator perspective was widely neglected. Information about runtime was provided for 10 % of the approaches. Complexity was covered by very few authors and the costs of running a recommender system were reported by a single article. We also observed that many authors neglect the user-modeling process: 81 % of the approaches made their users provide some keyword, text snippets, or a single input paper to represent their information need. Few approaches automatically inferred information needs from the users' authored, tagged, or otherwise connected papers.

One step towards using the full potential of research-paper recommender systems could be to establish a platform for researchers to collaborate, work on joint publications, communicate ideas, or to establish conferences or workshops focusing solely on research-paper recommender systems. An open-source recommender framework containing the most promising approaches could help transfer the research results into practice. Such a framework would also help new researchers in the field access a number of baselines they could compare their own approaches with. A framework could either be built from scratch, or be based on existing frameworks such as MyMediaLite,[33] LensKit,[34] Mahout,[35] Duine,[36] RecLab Core,[37] easyrec,[38] or Recommender101.[39] Finally, the community could benefit from considering research results from related disciplines. In particular, research in the area of user modeling and scientometrics appears promising, as well as research from the general recommender-systems community about aspects beyond accuracy.

References

1. Bollacker, K.D., Lawrence, S., Giles, C.L.: CiteSeer: an autonomous web agent for automatic retrieval and identification of interesting publications. In: Proceedings of the 2nd international conference on Autonomous agents, pp. 116–123 (1998)
2. Google Scholar, Scholar Update: Making New Connections, Google Scholar Blog. http://googlescholar.blogspot.de/2012/08/scholar-updates-making-new-connections.html
3. Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P., Jaakkola, T.: Mixed membership stochastic block models for relational data with application to protein–protein interactions. In: Proceedings of the International Biometrics Society Annual Meeting, pp. 1–34 (2006)
4. Arnold, A., Cohen, W.W.: Information extraction as link prediction: using curated citation networks to improve gene detection. In: Proceedings of the 4th International Conference on Wireless Algorithms, Systems, and Applications, pp. 541–550 (2009)
5. Beel, J., Langer, S., Genzmehr, M.: Sponsored vs. Organic (Research Paper) Recommendations and the Impact of Labeling. In: Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), pp. 395–399 (2013)
6. Beel, J., Langer, S., Genzmehr, M., Nürnberger, A.: Persistence in Recommender Systems: Giving the Same Recommendations to the Same Users Multiple Times. In: Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), vol. 8092, pp. 390–394 (2013)

[33] http://www.mymedialite.net/.

[34] http://lenskit.grouplens.org/.

[35] http://mahout.apache.org/.

[36] http://www.duineframework.org/.

[37] http://code.richrelevance.com/reclab-core/.

[38] http://easyrec.org/.

[39] http://ls13-www.cs.uni-dortmund.de/homepage/recommender101/index.shtml.

7. Beel, J., Langer, S., Genzmehr, M., Nürnberger, A.: Introducing Docear's Research Paper Recommender System. In: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'13), pp. 459–460 (2013)
8. Beel, J., Langer, S., Nürnberger, A., Genzmehr, M.: The Impact of Demographics (Age and Gender) and Other User Characteristics on Evaluating Recommender Systems. In: Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), pp. 400–404 (2013)
9. Böhm, W., Geyer-schulz, A., Hahsler, M., Jahn, M.: Repeat-Buying Theory and Its Application for Recommender Services. In: Proceedings of the 25th Annual Conference of the Gesellschaft für Klassifikation e.V., pp. 229–239 (2003)
10. Baez, M., Mirylenka, D., Parra, C.: Understanding and supporting search for scholarly knowledge. In: Proceeding of the 7th European Computer Science Summit, pp. 1–8 (2011)
11. Beel, J., Gipp, B., Langer, S., Genzmehr, M.: Docear: an academic literature suite for searching, organizing and creating academic literature. In: Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries (JCDL), pp. 465–466 (2011)
12. Beel, J., Gipp, B., Mueller, C.: SciPlore MindMapping'—a tool for creating mind maps combined with PDF and reference management. D-Lib Mag. **15**(11) (2009)
13. Bethard, S., Jurafsky, D.: Who should I cite: learning literature search models from citation behavior. In: Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 609–618 (2010)
14. Bogers, T., van den Bosch, A.: Recommending scientific articles using citeulike. In: Proceedings of the 2008 ACM conference on Recommender systems, pp. 287–290 (2008)
15. Bollen, J., Van de Sompel, H.: An architecture for the aggregation and analysis of scholarly usage data. In: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries, pp. 298–307 (2006)
16. CiteSeerX, T.: About RefSeer. http://refseer.ist.psu.edu/about (2012)
17. CiteULike: My Top Recommendations. Website http://www.citeulike.org/profile/username/recommendations (2011)
18. CiteULike: Science papers that interest you. Blog. http://blog.citeulike.org/?p=11 (2009)
19. CiteULike: Data from CiteULike's new article recommender. Blog, http://blog.citeulike.org/?p=136 (2009)
20. Caragea, C., Silvescu, A., Mitra, P., Giles, C.L.: Can't See the Forest for the Trees? A Citation Recommendation System. In: iConference 2013 Proceedings, pp. 849–851 (2013)
21. Chandrasekaran, K., Gauch, S., Lakkaraju, P., Luong, H.: Concept-based document recommendations for citeseer authors. In: Proceedings of the 5th international conference on Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 83–92 (2008)
22. Choochaiwattana, W.: Usage of tagging for research paper recommendation. In: Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), vol. 2, pp. 439–442 (2010)
23. Councill, I., Giles, C., Di Iorio, E., Gori, M., Maggini, M., Pucci, A.: Towards next generation CiteSeer: a flexible architecture for digital library deployment. In: Research and Advanced Technology for Digital Libraries, pp. 111–122 (2006)
24. Dong, R., Tokarchuk, L., Ma, A.: Digging Friendship: Paper Recommendation in Social Network. In: Proceedings of Networking and Electronic Commerce Research Conference (NAEC 2009), pp. 21–28 (2009)
25. ExLibris: bX Usage-Based Services transform your discovery experience!, Web page, http://www.exlibrisgroup.com/category/bXUsageBasedServices (2013)
26. Ekstrand, M.D., Kannan, P., Stemper, J.A., Butler, J.T., Konstan, J.A., Riedl, J.T.: Automatically building research reading lists. In: Proceedings of the 4th ACM conference on Recommender systems, pp. 159–166 (2010)
27. Erosheva, E., Fienberg, S., Lafferty, J.: Mixed-membership models of scientific publications. Proc. Natl. Acad. Sci. U. S. Am. **101**(Suppl 1), 5220–5227 (2004)
28. Franke, M., Geyer-Schulz, A.: Using restricted random walks for library recommendations and knowledge space exploration. Int. J. Pattern Recognit. Artif. Intell. **21**(02), 355–373 (2007)
29. Ferrara, F., Pudota, N., Tasso, C.: A Keyphrase-Based Paper Recommender System. In: Proceedings of the IRCDL'11, pp. 14–25 (2011)
30. Geyer-Schulz, A., Hahsler, M.: Comparing two recommender algorithms with the help of recommendations by peers. In: Proceedings of the WEBKDD 2002—Mining Web Data for Discovering Usage Patterns and Profiles, pp. 137–158 (2003)
31. Geyer-Schulz, A., Hahsler, M.: Evaluation of recommender algorithms for an internet information broker based on simple association rules and on the repeat-buying theory. In: Proceedings of the 4th WebKDD Workshop: Web Mining for Usage Patterns and User Profiles, pp. 100–114 (2002)
32. Geyer-Schulz, A., Hahsler, M., Jahn, M.: A customer purchase incidence model applied to recommender services. In: Proceedings of the 3rd International Workshop on Mining Web Log Data Across All Customers Touch Points, pp. 25–47 (2002)
33. Geyer-Schulz, A., Hahsler, M., Jahn, M.: Recommendations for virtual universities from observed user behavior. In: Proceedings of the 24th Annual Conference of the Gesellschaft für Klassifikation e.V., pp. 273–280 (2002)
34. Geyer-Schulz, A., Hahsler, M., Jahn, M., Geyer, A.: Wissenschaftliche Recommendersysteme in Virtuellen Universitäten. In: Proceedings of the Symposiom of Unternehmen Hochschule, pp. 101–114 (2001)
35. Geyer-Schulz, A., Hahsler, M., Neumann, A., Thede, A.: An integration strategy for distributed recommender services in legacy library systems. In: Between Data Science and Applied Data Analysis. Springer, pp. 412–420 (2003)
36. Geyer-Schulz, A., Hahsler, M., Neumann, A., Thede, A.: Behavior-based recommender systems as value-added services for scientific libraries. Statistical Data Mining and Knowledge Discovery, pp. 433–454 (2003)
37. Geyer-Schulz, A., Hahsler, M., Thede, A.: Comparing Simple Association-Rules and Repeat-Buying Based Recommender Systems in a B2B Environment. In: Proceedings of the 26th Annual Conference of the Gesellschaft für Klassifikation e.V., pp. 421–429 (2003)
38. Geyer-Schulz, A., Neumann, A., Thede, A.: An architecture for behavior-based library recommender systems. Inf. Technol. Libr. **22**(4), 165–174 (2003)
39. Geyer-Schulz, A., Neumann, A., Thede, A.: Others also use: a robust recommender system for scientific libraries. In: Proceedings of the 7th European Conference on Digital Libraries, pp. 113–125 (2003)
40. Gillitzer, B.: Der Empfehlungsdienst BibTip - Ein flächendeckendes Angebot im Bibliotheksverbund Bayern. http://www.b-i-t-online.de/heft/2010-01/nachrichtenbeitrag3. pp. 1–4 (2010)
41. Gottwald, S.: Recommender Systeme fuer den Einsatz in Bibliotheken/Survey on recommender systems. Konrad-Zuse-Zentrum für Informationstechnik Berlin, ZIB-Report **11–30** (2011)
42. Geyer-Schulz, A., Hahsler, M., Jahn, M.: Educational and scientific recommender systems: designing the information channels of the virtual university. Int. J. Eng. Educ. **17**(2), 153–163 (2001)

43. Giles, C.L., Bollacker, K.D., Lawrence, S.: CiteSeer: an automatic citation indexing system. In: Proceedings of the 3rd ACM conference on Digital libraries, pp. 89–98 (1998)
44. Gipp, B., Beel, J.: Citation proximity analysis (CPA)—a new approach for identifying related work based on co-citation analysis. In: Proceedings of the 12th international conference on Scientometrics and informetrics (ISSI'09), vol. 2, pp. 571–575 (2009)
45. Gipp, B., Beel, J., Hentschel, C.: Scienstein: a research paper recommender system. In: Proceedings of the international conference on Emerging trends in computing (ICETiC'09), pp. 309–315 (2009)
46. Gori, M., Pucci, A.: Research paper recommender systems: a random-walk based approach. In: Proceedings of the 2006 IEEE/WIC/ACM international conference on Web intelligence, pp. 778–781 (2006)
47. Henning, V., Reichelt, J.: Mendeley-a last. fm for research? In: Proceedings of the IEEE 4th international conference on eScience, pp. 327–328 (2008)
48. Hwang, S.-Y., Hsiung, W.-C., Yang, W.-S.: A prototype WWW literature recommendation system for digital libraries. Online Inf. Rev. **27**(3), 169–182 (2003)
49. He, J., Nie, J.-Y., Lu, Y., Zhao, W.X.: Position-aligned translation model for citation recommendation. In: Proceedings of the 19th international conference on String processing and information retrieval, pp. 251–263 (2012)
50. He, Q., Kifer, D., Pei, J., Mitra, P., Giles, C.L.: Citation recommendation without author supervision. In: Proceedings of the 4th ACM international conference on Web search and data mining, pp. 755–764 (2011)
51. He, Q., Pei, J., Kifer, D., Mitra, P., Giles, L.: Context-aware citation recommendation. In: Proceedings of the 19th international conference on World wide web, pp. 421–430 (2010)
52. Hess, C.: Trust-Based Recommendations in Multi-Layer Networks. IOS Press, Amsterdam (2008)
53. Hess, C.: Trust-based recommendations for publications: a multilayer network approach. TCDL Bull. **2**(2), 190–201 (2006)
54. Hess, C., Stein, K., Schlieder, C.: Trust-enhanced visibility for personalized document recommendations. In: Proceedings of the 2006 ACM symposium on Applied computing, pp. 1865–1869 (2006)
55. Huang, S., Xue, G.R., Zhang, B.Y., Chen, Z., Yu, Y., Ma, W.Y.: Tssp: a reinforcement algorithm to find related papers. In: Proceedings of the IEEE/WIC/ACM international conference on Web intelligence (WI), pp. 117–123 (2004)
56. Huang, W., Kataria, S., Caragea, C., Mitra, P., Giles, C.L., Rokach, L.: Recommending citations: translating papers into references. In: Proceedings of the 21st ACM international conference on Information and knowledge management, pp. 1910–1914 (2012)
57. Huang, Z., Chung, W., Ong, T.H., Chen, H.: A graph-based recommender system for digital library. In: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, pp. 65–73 (2002)
58. Jack, K.: Mendeley: recommendation systems for academic literature. Presentation at Technical University of Graz (TUG) (2012)
59. Jack, K.: Mendeley suggest: engineering a personalised article recommender system. Presentation at RecSysChallenge workshop 2012 (2012)
60. Jack, K.: Mahout becomes a researcher: large scale recommendations at Mendeley. Presentation at big data week conferences (2012)
61. Jiang, Y., Jia, A., Feng, Y., Zhao, D.: Recommending academic papers via users' reading purposes. In: Proceedings of the 6th ACM conference on Recommender systems, pp. 241–244 (2012)
62. Jomsri, P., Sanguansintukul, S., Choochaiwattana, W.: A framework for tag-based research paper recommender system: an IR approach. In: Proceedings of the 24th international conference on Advanced information networking and applications (WAINA), pp. 103–108 (2010)
63. Kapoor, N., Chen, J., Butler, J.T., Fouty, G.C., Stemper, J.A., Riedl, J., Konstan, J.A.: Techlens: a researcher's desktop. In: Proceedings of the 2007 ACM conference on Recommender systems, pp. 183–184 (2007)
64. Konstan, J.A., Kapoor, N., McNee, S.M., Butler, J.T.: Techlens: exploring the use of recommenders to support users of digital libraries. In: Proceedings of the coalition for networked information fall 2005 task force meeting, pp. 111–112 (2005)
65. Kataria, S., Mitra, P., Bhatia, S.: Utilizing context in generative bayesian models for linked corpus. In: Proceedings of the 24th AAAI conference on Artificial intelligence, pp. 1340–1345 (2010)
66. Kodakateri Pudhiyaveetil, A., Gauch, S., Luong, H., Eno, J.: Conceptual recommender system for CiteSeerX. In: Proceedings of the 3rd ACM conference on Recommender systems, pp. 241–244 (2009)
67. Kuberek, M., Mönnich, M.: Einsatz von Recommendersystemen in Bibliotheken Recommender systems in libraries. Presentation (2012)
68. Küçüktunç, O., Kaya, K., Saule, E., Catalyürek, U.V.: Fast recommendation on bibliographic networks. In: Proceedings of the IEEE/ACM international conference on Advances in social networks analysis and mining (ASONAM), pp. 480–487 (2012)
69. Küçüktunç, O., Kaya, K., Saule, E., Catalyürek, U.V.: Fast recommendation on bibliographic networks with sparse-matrix ordering and partitioning. Soc. Netw. Anal. Min. **3**(4), 1097–1111 (2013)
70. Küçüktunç, O., Saule, E., Kaya, K., Çatalyürek, Ü.V.: Result Diversification in automatic citation recommendation. In: Proceedings of the iConference workshop on Computational scientometrics: theory and applications, pp. 1–4 (2013)
71. Küçüktunç, O., Saule, E., Kaya, K., Çatalyürek, Ü.V.: Diversifying citation recommendations. arXiv preprint. arXiv:1209.5809. pp. 1–19 (2012)
72. Küçüktunç, O., Saule, E., Kaya, K., Çatalyürek, Ü.V.: Recommendation on academic networks using direction aware citation analysis. arXiv preprint. arXiv:1205.1143. pp. 1–10 (2012)
73. Küçüktunç, O., Saule, E., Kaya, K., Çatalyürek, Ü.V.: Direction awareness in citation recommendation. In: Proceedings of DBRank workshop in conjunction with VLDB'12. pp. 161–166 (2012)
74. Lao, N.: Efficient random walk inference with knowledge bases. PhD Thesis. The Carnegie Mellon University (2012)
75. Lao, N., Cohen, W.W.: Personalized reading recommendations for Saccharomyces genome database. Unpublished Paper. http://www.cs.cmu.edu/nlao/publication/2012/2012.dils.pdf. pp. 1–15 (2012)
76. Lao, N., Cohen, W.W.: Personalized reading recommendations for Saccharomyces genome database. Unpublished Poster. http://www.cs.cmu.edu/nlao/publication/2012/2012.dils.poster.portrat.pdf (2012)
77. Lao, N., Cohen, W. W.: Contextual recommendation with path constrained random walks. Unpublished. http://www.cs.cmu.edu/nlao/doc/2011.cikm.pdf. pp. 1–9 (2011)
78. Lakkaraju, P., Gauch, S., Speretta, M.: Document similarity based on concept tree distance. In: Proceedings of the 19th ACM conference on Hypertext and hypermedia, pp. 127–132 (2008)
79. Lao, N., Cohen, W.W.: Relational retrieval using a combination of path-constrained random walks. Mach. Learn. **81**(1), 53–67 (2010)
80. Lawrence, K.D.B.S.: A system for automatic personalized tracking of scientific literature on the web. In: Proceedings of the 4th ACM conference on Digital libraries, pp. 105–113 (1999)

81. Lawrence, S.R., Bollacker, K.D., Giles, C.L.: Autonomous citation indexing and literature browsing using citation context. U.S. Patent US 6,738,780 B2Summer-2004
82. Lawrence, S.R., Giles, C. L., Bollacker, K.D.: Autonomous citation indexing and literature browsing using citation context. U.S. Patent US 6,289,342 B1Nov-2001
83. Li, H., Councill, I., Lee, W.-C., Giles, C. L.: CiteSeerx: an architecture and web service design for an academic document search engine. In: Proceedings of the 15th international conference on World wide web, pp. 883–884 (2006)
84. Liang, Y., Li, Q., Qian, T.: Finding relevant papers based on citation relations. In: Proceedings of the 12th international conference on Web-age information management, pp. 403–414 (2011)
85. Lin, J., Wilbur, W.J.: PubMed related articles: a probabilistic topic-based model for content similarity. BMC Bioinform. **8**(1), 423–436 (2007)
86. Lu, Y., He, J., Shan, D., Yan, H.: Recommending citations with translation model. In: Proceedings of the 20th ACM international conference on Information and knowledge management, pp. 2017–2020 (2011)
87. McNee, S. M., Kapoor, N., Konstan, J.A.: Don't look stupid: avoiding pitfalls when recommending research papers. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, pp. 171–180 (2006)
88. Middleton, S.E., Alani, H., De Roure, D.C.: Exploiting synergy between ontologies and recommender systems. In: Proceedings of the semantic web workshop, pp. 1–10 (2002)
89. Middleton, S.E., De Roure, D., Shadbolt, N.R.: Ontology-based recommender systems. In: Handbook on Ontologies, pp. 779–796, Springer, Berlin (2009)
90. Middleton, S.E., De Roure, D.C., Shadbolt, N.R.: Foxtrot recommender system: user profiling, ontologies and the World Wide Web. In: Proceedings of the WWW conference, pp. 1–3 (2002)
91. Middleton, S.E., De Roure, D.C., Shadbolt, N.R.: Capturing knowledge of user preferences: ontologies in recommender systems. In: Proceedings of the 1st international conference on Knowledge capture, pp. 100–107 (2001)
92. Mönnich, M., Spiering, M.: Adding value to the library catalog by implementing a recommendation system. D-Lib Mag. **14**(5), 4–11 (2008)
93. McNee, S.M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S.K., Rashid, A.M., Konstan, J.A., Riedl, J.: On the recommending of citations for research papers. In: Proceedings of the ACM conference on Computer supported cooperative work, pp. 116–125 (2002)
94. Middleton, S.E., Shadbolt, N.R., De Roure, D.C.: Ontological user profiling in recommender systems. ACM Trans. Inf. Syst. (TOIS) **22**(1), 54–88 (2004)
95. Monnich, M., Spiering, M.: Einsatz von BibTip als Recommendersystem im Bibliothekskatalog. Bibliotheksdienst **42**(1), 54 (2008)
96. Naak, A.: Papyres: un système de gestion et de recommandation d'articles de recherche. Master Thesis. Université de Montréal (2009)
97. Neumann, A.W.: Recommender Systems for Information Providers. Springer, Berlin (2009)
98. Naak, A., Hage, H., Aimeur, E.: A multi-criteria collaborative filtering approach for research paper recommendation in papyres. In: Proceedings of the 4th international conference MCETECH, pp. 25–39 (2009)
99. Naak, A., Hage, H., Aimeur, E.: Papyres: a research paper management system. In: Proceedings of the 10th E-Commerce Technology Conference on Enterprise Computing, E-Commerce and E-Services, pp. 201–208 (2008)
100. Nallapati, R.M., Ahmed, A., Xing, E.P., Cohen, W.W.: Joint latent topic models for text and citations. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 542–550 (2008)
101. Nascimento, C., Laender, A.H., da Silva, A.S., Gonçalves, M.A.: A source independent framework for research paper recommendation. In: Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries, pp. 297–306 (2011)
102. Ozono, T., Goto, S., Fujimaki, N., Shintani, T.: P2p based knowledge source discovery on research support system papits. In: Proceedings of the 1st international joint conference on Autonomous agents and multiagent systems: part 1, pp. 49–50 (2002)
103. Ozono, T., Shintani, T.: P2P based information retrieval on research support system Papits. In: Proceedngs of the IASTED international conference on Artificial and computational intelligence, pp. 136–141 (2002)
104. Ozono, T., Shintani, T.: Paper classification for recommendation on research support system papits. IJCSNS Int. J. Comput. Sci. Netw. Secur. **6**, 17–23 (2006)
105. Ozono, T., Shintani, T., Ito, T., Hasegawa, T.: A feature selection for text categorization on research support system Papits. In: Proceedings of the 8th Pacific Rim international conference on Artificial intelligence, pp. 524–533 (2004)
106. Pennock, D.M., Horvitz, E., Lawrence, S., Giles, C.L.: Collaborative filtering by personality diagnosis: a hybrid memory-and model-based approach. In: Proceedings of the 16th conference on Uncertainty in artificial intelligence, pp. 473–480 (2000)
107. Petinot, Y., Giles, C.L., Bhatnagar, V., Teregowda, P.B., Han, H.: Enabling interoperability for autonomous digital libraries: an API to citeseer services. In: Digital Libraries, 2004. Proceedings of the 2004 joint ACM/IEEE conference on, pp. 372–373 (2004)
108. Petinot, Y., Giles, C.L., Bhatnagar, V., Teregowda, P.B., Han, H., Councill, I.: A service-oriented architecture for digital libraries. In: Proceedings of the 2nd international conference on Service oriented computing, pp. 263–268 (2004)
109. Pohl, S.: Using access data for paper recommendations on ArXiv.org. Master Thesis. Technical University of Darmstadt (2007)
110. Pohl, S., Radlinski, F., Joachims, T.: Recommending related papers based on digital library access records. In: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, pp. 417–418 (2007)
111. Researchgate, T.: Researchgate recommender. http://www.researchgate.net/directory/publications/ (2011)
112. Rokach, L., Mitra, P., Kataria, S., Huang, W., Giles, L.: A supervised learning method for context-aware citation recommendation in a large corpus. In: Proceedings of the large-scale and distributed systems for information retrieval workshop (LSDS-IR), pp. 17–22 (2013)
113. Sarkanto: About the Sarkanto Recommender Demo. http://lab.cisti-icist.nrc-cnrc.gc.ca/Sarkanto/about.jsp (2013)
114. Strohman, T., Croft, W.B., Jensen, D.: Recommending citations for academic papers. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 705–706 (2007)
115. Sugiyama, K., Kan, M.-Y.: Scholarly paper recommendation via user's recent research interests. In: Proceedings of the 10th ACM/IEEE annual joint conference on Digital libraries (JCDL), pp. 29–38 (2010)
116. Thomas, D., Greenberg, A., Calarco, P.: Scholarly usage based recommendations: evaluating bX for a Consortium, Presentation. http://igelu.org/wp-content/uploads/2011/09/bx_igelu_presentation_updated_september-13.pdf (2011)
117. Torres, R., McNee, S.M., Abel, M., Konstan, J.A., Riedl, J.: Enhancing digital libraries with TechLens+. In: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries, 2004, pp. 228–236

118. Uchiyama, K., Nanba, H., Aizawa, A., Sagara, T.: OSUSUME: cross-lingual recommender system for research papers. In: Proceedings of the 2011 workshop on context-awareness in retrieval and recommendation, pp. 39–42 (2011)
119. Vellino, A.: A comparison between usage-based and citation-based methods for recommending scholarly research articles. Proc. Am. Soc. Inf. Sci. Technol. **47**(1), 1–2 (2010)
120. Vellino, A., Zeber, D.: A hybrid, multi-dimensional recommender for journal articles in a scientific digital library. In: Proceedings of the 2007 IEEE/WIC/ACM international conference on Web intelligence, pp. 111–114 (2007)
121. Wang, Y., Zhai, E., Hu, J., Chen, Z.: Claper: recommend classical papers to beginners. Seventh international conference on Fuzzy systems and knowledge discovery **6**, 2777–2781 (2010)
122. Watanabe, S., Ito, T., Ozono, T., Shintani, T.: A paper recommendation mechanism for the research support system papits. In: Proceedings of the international workshop on Data engineering issues in E-Commerce, pp. 71–80
123. Woodruff, A., Gossweiler, R., Pitkow, J., Chi, E.H., Card, S.K.: Enhancing a digital book with a reading recommender. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 153–160 (2000)
124. Yang, C., Wei, B., Wu, J., Zhang, Y., Zhang, L.: CARES: a ranking-oriented CADAL recommender system. In: Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries, pp. 203–212 (2009)
125. Zarrinkalam, F., Kahani, M.: SemCiR—a citation recommendation system based on a novel semantic distance measure. Program: Electron. Libr. Inf. Syst. **47**(1), 92–112 (2013)
126. Zarrinkalam, F., Kahani, M.: A new metric for measuring relatedness of scientific papers based on non-textual features. Intell. Inf. Manag. **4**(4), 99–107 (2012)
127. Zhou, D., Zhu, S., Yu, K., Song, X., Tseng, B.L., Zha, H., Giles, C.L.: Learning multiple graphs for document recommendations. In: Proceedings of the 17th international conference on World Wide Web, pp. 141–150 (2008)
128. Avancini, H., Candela, L., Straccia, U.: Recommenders in a personalized, collaborative digital library environment. J. Intell. Inf. Syst. **28**(3), 253–283 (2007)
129. Agarwal, N., Haque, E., Liu, H., Parsons, L.: A subspace clustering framework for research group collaboration. Int. J. Inf. Technol. Web Eng. **1**(1), 35–58 (2006)
130. Agarwal, N., Haque, E., Liu, H., Parsons, L.: Research paper recommender systems: a subspace clustering approach. In: Proceedings of the 6th international conference on Advances in Web-Age Information Management (WAIM'05), pp. 475–491 (2005)
131. Bollen, J., Rocha, L.M.: An adaptive systems approach to the implementation and evaluation of digital library recommendation systems. In: Proceedings of the 4th European conference on Digital libraries, Springer, pp. 356–359 (2000)
132. Bancu, C., Dagadita, M., Dascalu, M., Dobre, C., Trausan-Matu, S., Florea, A.M.: ARSYS-article recommender system. In: Proceedings of the 14th international symposium on Symbolic and numeric algorithms for scientific computing, pp. 349–355 (2012)
133. Cazella, S.C., Alvares, L.O.C.: Combining data mining technique and users' relevance opinion to build an efficient recommender system. Revista Tecnologia da Informação, UCB, 4(2) (2005)
134. Cazella, S., Alvares, L.: Modeling user's opinion relevance to recommending research papers. In: Proceedings of the UMAP Conference, pp. 150–150 (2005)
135. Chirawatkul, P.: Structured Peer-to-Peer Search to Build a Bibliographic Paper Recommendation System. Saarland University, Saarland (2006)
136. Dattolo, A., Ferrara, F., Tasso, C.: Supporting personalized user concept spaces and recommendations for a publication sharing system. In: Proceedings of the 17th international conference on User modeling, adaptation, and personalization, pp. 325–330 (2009)
137. Daud, A.: Muhammad Akramand Rajpar Shaikh, A.H.: Scientific reference mining using semantic information through topic modeling. Res. J. Eng. Technol. **28**(2), 253–262 (2009)
138. Farooq, U., Ganoe, C.H., Carroll, J.M., Councill, I.G.: Lee Giles, C.: Design and evaluation of awareness mechanisms in CiteSeer. Inf. Process. Manag. **44**(2), 596–612 (2008)
139. Fernández, L., Sánchez, J.A., García, A.: Mibiblio: personal spaces in a digital library universe. In: Proceedings of the 5th ACM conference on Digital libraries, pp. 232–233 (2000)
140. Gross, T.: CYCLADES: a distributed system for virtual community support based on open archives. In: Proceedings of the 11th Euromicro Conference on Parallel, distributed and network-based orocessing, pp. 484–491 (2003)
141. Geisler, G., McArthur, D., Giersch, S.: Developing recommendation services for a digital library with uncertain and changing data. In: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries, pp. 199–200 (2001)
142. Hong, K., Jeon, H., Jeon, C.: UserProfile-based personalized research paper recommendation system. In: Proceedings of the 8th international conference on Computing and networking technology, pp. 134–138 (2012)
143. Huang, Y.: Combining Social Networks and Content for Recommendation in a Literature Digital Library. National Sun Yat-Sen University, Taiwan (2007)
144. Kang, S., Cho, Y.: A novel personalized paper search system. In: Proceedings of the international conference on Intelligent computing, pp. 1257–1262 (2006)
145. Martin, G.H., Schockaert, S., Cornelis, C., Naessens, H.: Metadata impact on research paper similarity. In: 14th European Conference on Digital libraries, pp. 457–460 (2010)
146. Morales-del-Castillo, J.M., Peis, E., Herrera-Viedma, E.: A filtering and recommender system prototype for scholarly users of digital libraries. In: Proceedings of the Second World Summit on the Knowledge Society, Springer, pp. 108–117 (2009)
147. Mao, Y., Vassileva, J., Grassmann, W.: A system dynamics approach to study virtual communities. In: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, pp. 178–197 (2007)
148. Matsatsinis, N.F., Lakiotaki, K., Delia, P.: A system based on multiple criteria analysis for scientific paper recommendation. In: Proceedings of the 11th Panhellenic Conference on Informatics, pp. 135–149 (2007)
149. Mishra, G.: Optimised research paper recommender system using social tagging. Int. J. Eng. Res. Appl. **2**(2), 1503–1507 (2012)
150. Nakagawa, A., Ito, T.: An implementation of a knowledge recommendation system based on similarity among users' profiles. In: Proceedings of the 41st SICE annual conference, vol. 1, pp. 326–327 (2002)
151. Pan, C., Li, W.: Research paper recommendation with topic analysis. In: Proceedings of the international conference on Computer design and applications (ICCDA), pp. 264–268 (2010)
152. Popa, H.-E., Negru, V., Pop, D., Muscalagiu, I.: DL-AgentRecom-A multi-agent based recommendation system for scientific documents. In: Proceedings of the 10th international symposium on Symbolic and numeric algorithms for scientific computing, pp. 320–324 (2008)
153. Ratprasartporn, N., Ozsoyoglu, G.: Finding related papers in literature digital libraries. In: Proceedings of the 11th European Conference on Digital libraries, pp. 271–284 (2007)
154. Rocha, L.M.: TalkMine: a soft computing approach to adaptive knowledge recommendation. Stud. Fuzziness Soft Comput. **75**, 89–116 (2001)

155. Rocha, L.M.: Talkmine and the adaptive recommendation project. In: Proceedings of the fourth ACM conference on Digital libraries, pp. 242–243 (1999)
156. Stock, K., Robertson, A., Reitsma, F., Stojanovic, T., Bishr, M., Medyckyj-Scott, D., Ortmann, J.: eScience for Sea Science: a semantic scientific knowledge infrastructure for marine scientists. In: Proceedings of the 5th IEEE international conference on e-Science, pp. 110–117 (2009)
157. Straccia, U.: Cyclades: an open collaborative virtual archive environment. Poster (http://www.ercim.eu/cyclades/cyclades-fs.pdf) (2003)
158. Shaoping, Z.: ActiveCite: an interactive system for automatic citation suggestion. Master Thesis. National University of Singapore (2010)
159. Stock, K., Karasova, V., Robertson, A., Roger, G., Small, M., Bishr, M., Ortmann, J., Stojanovic, T., Reitsma, F., Korczynski, L., Brodaric, B., Gardner, Z.: Finding science with science: evaluating a domain and scientific ontology user interface for the discovery of scientific resources. Trans. GIS **1**, 1–28 (2013)
160. Tang, T.Y., McCalla, G.: Towards pedagogy-oriented paper recommendations and adaptive annotations for a web-based learning system. In: Knowledge representation and automated reasoning for E-Learning systems, pp. 72–80 (2003)
161. Tang, J., Zhang, J.: A discriminative approach to topic-based citation recommendation. Advances in Knowledge Discovery and Data Mining, pp. 572–579 (2009)
162. Tang, T., McCalla, G.: Utilizing artificial learners to help overcome the cold-start problem in a pedagogically-oriented paper recommendation system. In: Adaptive hypermedia and adaptive web-based systems, pp. 245–254 (2004)
163. Tang, T., McCalla, G.: Beyond learners' interest: personalized paper recommendation based on their pedagogical features for an e-learning system. In: Proceedings of the 8th Pacific Rim international conference on Artificial intelligence, Springer, pp. 301–310 (2004)
164. Tang, T.Y., McCalla, G.: Mining implicit ratings for focused collaborative filtering for paper recommendations. In: Proceedings of the workshop on User and group models for web-based adaptive collaborative environments (2003)
165. Tang, T.Y., McCalla, G.: Smart recommendation for an evolving e-learning system. In: Proceedings of the workshop on Technologies for electronic documents for supporting learning, at the international conference on Artificial intelligence in education, pp. 699–710 (2003)
166. Tang, T.Y.: The design and study of pedagogical paper recommendation. PhD Thesis. University of Saskatchewan (2008)
167. Tang, T.Y., McCalla, G.: A multidimensional paper recommender: experiments and evaluations. Internet Comput. IEEE **13**(4), 34–41 (2009)
168. Tang, T.Y., McCalla, G.: The pedagogical value of papers: a collaborative-filtering based paper recommender. J. Digit. Inf. **10**(2), 1–12 (2009)
169. Tang, T.Y., McCalla, G.: On the pedagogically guided paper recommendation for an evolving web-based learning system. In: Proceedings of the FLAIRS Conference, pp. 86–91 (2004)
170. Tang, T.Y., McCalla, G.: The social affordance of a paper. In: Proceedings of the workshop of assessment of group and individual learning through intelligent visualization on the 13th international conference on Artificial intelligence in education, pp. 34–42 (2007)
171. Tang, X., Zeng, Q.: Keyword clustering for user interest profiling refinement within paper recommender systems. J. Syst. Softw. **85**(1), 87–101 (2012)
172. Vassileva, J.: Harnessing p2p power in the classroom. In: Proceedings of the conference on Intelligent tutoring systems, pp. 305–314 (2004)
173. Vassileva, J.: Supporting peer-to-peer user communities. In: Proceedings of the conference on the move to meaningful internet systems, pp. 230–247 (2002)
174. Vassileva, J., Detters, R., Geer, J., Maccalla, G., Bull, S., Kettel, L.: Lessons from deploying I-Help. In: Workshop on Multi-agent architectures for distributed learning environments. In: Proceedings of international conference on AI and Education, San Antonio, TX, pp. 3–11 (2001)
175. Vivacqua, A.S., Oliveira, J., de Souza, J.M.: i-ProSE: inferring user profiles in a scientific context. Comput. J. **52**(7), 789–798 (2009)
176. Weng, S.-S., Chang, H.-L.: Using ontology network analysis for research document recommendation. Expert Syst. Appl. **34**(3), 1857–1869 (2008)
177. Winoto, P., Tang, T.Y., McCalla, G.I.: Contexts in a paper recommendation system with collaborative filtering. Int. Rev. Res. Open Distance Learn. **13**(5), 56–75 (2012)
178. Wu, H., Hua, Y., Li, B., Pei, Y.: Enhancing citation recommendation with various evidences. In: Proceedings of the 9th international conference on Fuzzy systems and knowledge discovery (FSKD), pp. 1160–1165 (2012)
179. Xia, H., Li, J., Tang, J., Moens, M.-F.: Plink-LDA: using link as prior information in topic modeling. In: Proceedings of the conference on Database systems for advanced applications (DASFAA), pp. 213–227 (2012)
180. Yang, Q., Zhang, S., Feng, B.: Research on personalized recommendation system of scientific and technological periodical based on automatic summarization. In: Proceedings of the 1st international symposium on Information technologies and applications in education, pp. 34–39 (2007)
181. Yang, S.-Y., Hsu, C.-L.: A new ontology-supported and hybrid recommending information system for scholars. In: Proceedings of the 13th international conference on Network-based information systems (NBiS), pp. 379–384 (2010)
182. Yin, P., Zhang, M., Li, X.: Recommending scientific literatures in a collaborative tagging environment. In: Proceedings of the 10th international conference on Asian digital libraries, Springer, pp. 478–481 (2007)
183. Zarrinkalam, F., Kahani, M.: A multi-criteria hybrid citation recommendation system based on linked data. In: Proceedings of the 2nd international eConference on Computer and knowledge engineering, pp. 283–288 (2012)
184. Zhang, M., Wang, W., Li, X.: A paper recommender for scientific literatures based on semantic concept similarity. In: Proceedings of the international conference on Asian Digital Libraries, pp. 359–362 (2008)
185. Zhang, Z., Li, L.: A research paper recommender system based on spreading activation model. In: Proceedings of the 2nd international conference on Information Science and Engineering (ICISE), pp. 928–931 (2010)
186. Gottwald, S., Koch, T.: Recommender systems for libraries. In: Proceedings of the ACM international conference on Recommender systems, pp. 1–5 (2011)
187. Leong, S.: A survey of recommender systems for scientific papers. Presentation. http://www.liquidpub.org/mediawiki/upload/f/ff/RecommenderSystems.pdf (2012)
188. Smeaton, A.F., Callan, J.: Personalisation and recommender systems in digital libraries. Int. J. Digit. Libr. **5**(4), 299–308 (2005)
189. Alotaibi, S., Vassileva, J.: Trust-based recommendations for scientific papers based on the researcher's current interest. In: Artificial Intelligence in Education, pp. 717–720 (2013)
190. Beel, J., Langer, S., Genzmehr, M., Gipp, B., Breitinger, C., Nürnberger, A.: Research paper recommender system evaluation: a quantitative literature survey. In: Proceedings of the Workshop on Reproducibility and Replication in Recommender Systems Eval-

uation (RepSys) at the ACM Recommender System Conference (RecSys), pp. 15–22 (2013)

191. Beel, J., Langer, S., Genzmehr, M., Gipp, B., Nürnberger, A.: A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation. In: Proceedings of the Workshop on Reproducibility and Replication in Recommender Systems Evaluation (RepSys) at the ACM Recommender System Conference (RecSys), pp. 7–14 (2013)
192. Chen, C., Mao, C., Tang, Y., Chen, G., Zheng, J.: Personalized recommendation based on implicit social network of researchers. In: Joint international conference, ICPCA/SWS, pp. 97–107 (2013)
193. De Nart, D., Ferrara, F., Tasso, C.: Personalized access to scientific publications: from recommendation to explanation. In: Proceedings of the international conference on User modeling, adaptation, and personalization, pp. 296–301 (2013)
194. De Nart, D., Ferrara, F., Tasso, C.: RES: a personalized filtering tool for CiteSeerX queries based on keyphrase extraction. In: Proceedings of the international conference on User modeling, adaptation, and personalization (UMAP), pp. 341–343 (2013)
195. Franke, M., Geyer-Schulz, A., Neumann, A.: Building recommendations from random walks on library opac usage data. In: Data Analysis, Classification and the Forward Search, Springer, pp. 235–246 (2006)
196. Kim, S.: iScholar: a mobile research support system. PhD Thesis. University of Regina (2013)
197. Küçüktunç, O.: Result Diversication on Spatial, Multidimensional, Opinion, and Bibliographic Data. Ohio State University, Columbus (2013)
198. Küçüktunç, O., Saule, E., Kaya, K., Çatalyürek, Ü. V.: TheAdvisor: a webservice for academic recommendation. In: Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries, pp. 433–434 (2013)
199. Küçüktunç, O., Saule, E., Kaya, K., Çatalyürek, Ü. V.: Towards a personalized, scalable, and exploratory academic recommendation service. In: Proceedings of the 2013 IEEE/ACM international conference on Advances in social networks analysis and mining, pp. 636–641 (2013)
200. Lai, Y., Zeng, J.: A cross-language personalized recommendation model in digital libraries. Electron. Libr. **31**(3), 164–277 (2013)
201. Li, Y., Yang, M., Zhang, Z.M.: Scientific articles recommendation. In: Proceedings of the 22nd ACM International conference on information and knowledge management, pp. 1147–1156 (2013)
202. Lee, J., Lee, K., Kim, J.G.: Personalized academic research paper recommendation system. ArXiv Preprint, vol. arXiv:1304.5457. pp. 1–8 (2013)
203. Manouselis, N., Verbert, K.: Layered evaluation of multi-criteria collaborative filtering for scientific paper recommendation. Procedia Comput. Sci. **18**, 1189–1197 (2013)
204. Meng, F., Gao, D., Li, W., Sun, X., Hou, Y.: A unified graph model for personalized query-oriented reference paper recommendation. In: Proceedings of the 22nd ACM international conference on Conference on information and knowledge management, pp. 1509–1512 (2013)
205. Pera, M.S., Ng, Y.-K.: Exploiting the wisdom of social connections to make personalized recommendations on scholarly articles. J. Intell. Inf. Syst. **42**(3), 371–391 (2014)
206. Pera, M.S., Ng, Y.-K.: Exploiting the wisdom of social connections to make personalized recommendations on scholarly articles. J. Intell. Inf. Syst. **42**(3), 371–391 (2014)
207. Sugiyama, K., Kan, M.-Y.: Exploiting potential citation papers in scholarly paper recommendation. In: Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries, pp. 153–162 (2013)
208. Sun, J., Ma, J., Liu, X., Liu, Z., Wang, G., Jiang, H., Silva, T.: A novel approach for personalized article recommendation in online scientific communities. In: Proceedings of the 46th Hawaii international conference on System sciences (HICSS) (2013)
209. Sun, J., Ma, J., Liu, Z., Miao, Y.: Leveraging content and connections for scientific article recommendation. Comput. J. 60–71 (2013)
210. Tian, G., Jing, L.: Recommending scientific articles using birelational graph-based iterative RWR. In: Proceedings of the 7th ACM conference on Recommender systems, pp. 399–402 (2013)
211. Vellino, A.: Usage-based vs. citation-based methods for recommending scholarly research articles. Arxiv, vol. arXiv:1303.7149 (2013)
212. Yan, R., Yan, H. et al.: Guess what you will cite: personalized citation recommendation based on users's preference. In: Proceedings of the annual I&R training and education conference, pp. 428–439 (2013)
213. Yang, W.-S., Lin, Y.-R.: A task-focused literature recommender system for digital libraries. Online Inf. Rev. **37**(4), 581–601 (2013)
214. Yao, W., He, J., Huang, G., Cao, J., Zhang, Y.: Personalized recommendation on multi-layer context graph. In: Web Information Systems Engineering (WISE 2013), pp. 135–148 (2013)
215. Yu, L., Yang, J., Yang, D., Yang, X.: A decision support system for finding research topic based on paper recommendation. In: Proceedings of the Pacific Asia conference on Information systems (2013)
216. Zarrinkalam, F., Kahani, M.: Using semantic relations to improve quality of a citation recommendation system. Soft Comput. J. **1**(2), 36–45 (2013)
217. Zhang, Z.P., Li, L.N., Yu, H.Y.: A hybrid document recommender algorithm based on random walk. Appl. Mech. Mater. **2270**, 336–338 (2013)
218. Beel, J., Gipp, B.: Academic search engine spam and Google Scholar's resilience against it. J. Electron. Publ. **13**(3) (2010)
219. Bar-Ilan, J.: Which h-index?—A comparison of WoS. Scopus Google Scholar Scientometr. **74**(2), 257–271 (2007)
220. Noruzi, A.: Google Scholar: the new generation of citation indexes. Libri **55**(4), 170–180 (2005)
221. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: GroupLens: an open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM conference on Computer supported cooperative work, pp. 175–186 (1994)
222. Melville, P., Mooney, R.J., Nagarajan, R.: Content-boosted collaborative filtering for improved recommendations. In: Proceedings of the National Conference on Artificial Intelligence, pp. 187–192 (2002)
223. Shi, Y., Larson, M., Hanjalic, A.: Collaborative filtering beyond the user-item matrix: a survey of the state of the art and future challenges. ACM Comput. Surv. 47(1), 3:1–3:45 (2014)
224. Gunawardana, A., Shani, G.: A survey of accuracy evaluation metrics of recommendation tasks. J. Mach. Learn. Res. **10**, 2935–2962 (2009)
225. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. (TOIS) **22**(1), 5–53 (2004)
226. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: Proceedings of the 4th ACM conference on Recommender systems, pp. 257–260 (2010)
227. Ritchie, A., Teufel, S., Robertson, S.: Using terms from citations for IR: some first results. In: Macdonald, C., Ounis, I., Plachouras, V., Ruthven, I., White, R.W. (eds.) Advances in Information Retrieval, pp. 211–221. Springer (2008)
228. Ritchie, A., Teufel, S., Robertson, S.: Using terms from citations for IR: some first results. In: Macdonald, C., Ounis, I., Plachouras, V., Ruthven, I., White, R.W. (eds.) Advances in Information Retrieval, pp. 211–221. Springer (2008)

229. Ritchie, A.: Citation context analysis for information retrieval. PhD Thesis. University of Cambridge (2008)
230. Dumais, S.T., Nielsen, J.: Automating the assignment of submitted manuscripts to reviewers. In: Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 233–244 (1992)
231. Wang, F., Shi, N., Chen, B.: A comprehensive survey of the reviewer assignment problem. Int. J. Inf. Technol. Decis. Mak. **9**(04), 645–668 (2010)
232. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. U. S. Am. **102**(46), 16569 (2005)
233. Small, H.: Co-citation in the scientific literature: a new measure of the relationship between two documents. J. Am. Soc. Inf. Sci. **24**, 265–269 (1973)
234. Kessler, M.M.: Bibliographic coupling between scientific papers. Am. Documentation **14**, 10–25 (1963)
235. Zyczkowski, K.: Citation graph, weighted impact factors and performance indices. Scientometrics **85**(1), 301–315 (2010)
236. Fischer, G.: User modeling in human–computer interaction. User Model. User-Adapt. Interact. **11**(1), 65–86 (2001)
237. Eirinaki, M., Vazirgiannis, M.: Web mining for web personalization. ACM Trans. Internet Technol. (TOIT) **3**(1), 1–27 (2003)
238. Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C.D.: Web usage mining as a tool for personalization: a survey. User Model. User-Adapt. Interact. **13**(4), 311–372 (2003)
239. Mooney, R.J., Roy, L.: Content-based book recommending using learning for text categorization. In: Proceedings of the 5th ACM conference on Digital libraries, pp. 195–204 (2000)
240. Brusilovsky, P., Farzan, R., Ahn, J.: Comprehensive personalized information access in an educational digital library. In: Digital Libraries, 2005. JCDL'05. In: Proceedings of the 5th ACM/IEEE-CS joint conference on, pp. 9–18 (2005)
241. Faensen, D., Faultstich, L., Schweppe, H., Hinze, A., Steidinger, A.: Hermes: a notification service for digital libraries. In: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries, pp. 373–380 (2001)
242. Das, S., Mitra, P., Giles, C.L.: Similar researcher search'. In: Academic Environments. In: Proceedings of the JCDL'12, pp. 167–170 (2012)
243. Abu-Jbara, A., Radev, D.: Coherent citation-based summarization of scientific papers. In: Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 500–509 (2011)
244. Mohammad, S., Dorr, B., Egan, M., Hassan, A., Muthukrishan, P., Qazvinian, V., Radev, D., Zajic, D.: Using citations to generate surveys of scientific paradigms. In: Proceedings of human language technologies: the 2009 annual conference of the North American Chapter of the Association for Computational Linguistics, 2009, pp. 584–592
245. Teufel, S., Moens, M.: Summarizing scientific articles: experiments with relevance and rhetorical status. Comput. Linguist. **28**(4), 409–445 (2002)
246. Collins, L.M., Mane, K.K., Martinez, M.L., Hussell, J.A., Luce, R.E.: ScienceSifter: facilitating activity awareness in collaborative research groups through focused information feeds. In: 1st international conference on e-Science and grid computing, pp. 40–47 (2005)
247. Klamma, R., Cuong, P.M., Cao, Y.: You never walk alone: recommending academic events based on social network analysis. In: Zhou, J. (ed.) Complex Sciences, pp. 657–670. Springer (2009)
248. Klamma, R., Cuong, P.M., Cao, Y.: You never walk alone: recommending academic events based on social network analysis. In: Zhou, J. (ed.) Complex Sciences, pp. 657–670. Springer (2009)
249. Yang, Z., Davison, B. D.: Venue recommendation: submitting your paper with style. In: Machine learning and applications (ICMLA), 2012 11th international conference on, vol. 1, pp. 681–686 (2012)
250. Oh, S., Lei, Z., Lee, W.-C., Mitra, P., Yen, J.: CV-PCR: a context-guided value-driven framework for patent citation recommendation. In: Proceedings of the 22nd ACM international conference on Conference on information and knowledge management, pp. 2291–2296 (2013)
251. Singhal, A., Kasturi, R., Sivakumar, V., Srivastava, J.: Leveraging web intelligence for finding interesting research datasets. In: Web intelligence (WI) and intelligent agent technologies (IAT), 2013 IEEE/WIC/ACM international joint conferences on, vol. 1, pp. 321–328 (2013)
252. Gipp, B., Beel, J.: Citation based plagiarism detection–a new approach to identify plagiarized work language independently. In: Proceedings of the 21st ACM conference on Hypertext and hypermedia, pp. 273–274 (2010)
253. Zhan, S., Byung-Ryul, A., Ki-Yol, E., Min-Koo, K., Jin-Pyung, K., Moon-Kyun, K. (2008) Plagiarism detection using the Levenshtein distance and Smith-Waterman algorithm. In: Proceedings of the 3rd international conference on Innovative computing information and control, pp. 569–569
254. Zini, M., Fabbri, M., Moneglia, M., Panunzi, A.: Plagiarism detection through multilevel text comparison. In: Proceedings of the 2nd conference on Automated production of cross media content for multi-channel distribution, pp. 181–185 (2006)
255. Ley, M., Reuther, P.: Maintaining an online bibliographical database: the problem of data quality, EGC'2006, Actes des sixièmes journées Extraction et Gestion des Connaissances, pp. 17–20 (2006)
256. Beel, J., Langer, S., Genzmehr, M., Müller, C.: Docears PDF inspector: title extraction from PDF files. In: Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries (JCDL'13), pp. 443–444 (2013)
257. Beel, J., Gipp, B., Shaker, A., Friedrich, N.: SciPlore Xtract: extracting titles from scientific PDF documents by analyzing style information (Font Size). In: Research and Advanced Technology for Digital Libraries. Proceedings of the 14th European conference on Digital libraries (ECDL'10), vol. 6273, pp. 413–416 (2010)
258. Han, H., Giles, C.L., Manavoglu, E., Zha, H., Zhang, Z., Fox, E.A.: Automatic document metadata extraction using support vector machines. In: Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries, pp. 37–48 (2003)
259. Hu, Y., Li, H., Cao, Y., Teng, L., Meyerzon, D., Zheng, Q.: Automatic extraction of titles from general documents using machine learning. Inf. Process. Manag. **42**(5), 1276–1293 (2006)
260. Peng, F., McCallum, A.: Information extraction from research papers using conditional random fields. Inf. Process. Manag. **42**(4), 963–979 (2006)
261. Lawrence, S., Giles, C.L., Bollacker, K.D.: Autonomous citation matching. In: Proceedings of the 3rd annual conference on Autonomous agents, pp. 392–393 (1999)
262. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds): Recommender Systems Handbook, pp. 1–35. Springer, Berlin (2011)
263. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds): Recommender Systems Handbook, pp. 1–35. Springer, Berlin (2011)
264. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. Lect. Notes Comput. Sci. **4321**, 291 (2007)
265. Rossi, P.H., Lipsey, M.W., Freeman, H.E.: Evaluation: A Aystematic Approach, 7th edn. Sage publications, Thousand Oaks (2004)
266. Gorrell, G., Ford, N., Madden, A., Holdridge, P., Eaglestone, B.: Countering method bias in questionnaire-based user studies. J. Documentation **67**(3), 507–524 (2011)
267. Leroy, G.: Designing User Studies in Informatics. Springer, Berlin (2011)

268. Said, A., Tikk, D., Shi, Y., Larson, M., Stumpf, K., Cremonesi, P.: Recommender systems evaluation: a 3d benchmark. In: ACM RecSys 2012 workshop on Recommendation utility evaluation: beyond RMSE, Dublin, Ireland, pp. 21–23 (2012)
269. Cremonesi, P., Garzotto, F., Turrin, R.: Investigating the persuasion potential of recommender systems from a quality perspective: an empirical study. ACM Trans. Interact. Intell. Syst. (TiiS) **2**(2), 11 (2012)
270. Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A.V., Turrin, R.: Looking for 'good' recommendations: a comparative evaluation of recommender systems. In: Human–computer interaction-INTERACT 2011, Springer, pp. 152–168 (2011)
271. Burns, C.A., Bush, F.R.: Marketing Research, 7th edn. Prentice Hall, Upper Saddle River (2013)
272. Loeppky, J.L., Sacks, J., Welch, W.J.: Choosing the sample size of a computer experiment: a practical guide. Technometrics **51**(4), 366–376 (2009)
273. Zheng, H., Wang, D., Zhang, Q., Li, H., Yang, T.: Do clicks measure recommendation relevancy?: an empirical user study. In: Proceedings of the 4th ACM conference on Recommender systems, pp. 249–252 (2010)
274. Konstan, J.A., Riedl, J.: Recommender systems: from algorithms to user experience. User Model. User-Adapt. Interact. **22**(1–2), 101–123 (2012)
275. Konstan, J.A., Riedl, J.: Recommender systems: from algorithms to user experience. User Model. User-Adapt. Interact. 22(1–2), 101–123 (2012)
276. Matejka, J., Li, W., Grossman, T., Fitzmaurice, G.: CommunityCommands: command recommendations for software applications. In: Proceedings of the 22nd annual ACM symposium on User interface software and technology, pp. 193–202 (2009)
277. Rashid, A.M., Albert, I., Cosley, D., Lam, S.K., McNee, S.M., Konstan, J.A., Riedl, J.: Getting to know you: learning new user preferences in recommender systems. In: Proceedings of the 7th international conference on Intelligent user interfaces, pp. 127–134 (2002)
278. Hersh, W., Turpin, A., Price, S., Chan, B., Kramer, D., Sacherek, L., Olson, D.: Do batch and user evaluations give the same results? In: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 17–24 (2000)
279. Hersh, W.R., Turpin, A., Sacherek, L., Olson, D., Price, S., Chan, B., Kraemer, D.: Further Analysis of whether batch and user evaluations give the same results with a question-answering task. In: Proceedings of the 9th Text REtrieval Conference (TREC 9) (2000)
280. Said, A.: Evaluating the accuracy and utility of recommender systems. PhD Thesis. Technische Universität Berlin (2013)
281. Turpin, A.H., Hersh, W.: Why batch and user evaluations do not give the same results. In: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 225–231 (2001)
282. Jannach, D., Lerche, L., Gedikli, F., Bonnin, G.: What recommenders recommend—an analysis of accuracy, popularity, and sales diversity effects. In: User Modeling, Adaptation, and Personalization, Springer, pp. 25–37 (2013)
283. Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. User Model. User-Adapt. Interact. **22**(4–5), 441–504 (2012)
284. Jannach, D., Zanker, M., Ge, M., Gröning, M.: Recommender systems in computer science and information systems–a landscape of research. In: Proceedings of the 13th international conference, EC-Web, pp. 76–87 (2012)
285. Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwar, B., Herlocker, J., Riedl, J.: Combining collaborative filtering with personal agents for better recommendations. In: Proceedings of the National Conference on Artificial Intelligence, pp. 439–446 (1999)
286. Palopoli, L., Rosaci, D., Sarné, G.M.: A multi-tiered recommender system architecture for supporting E-Commerce. In: Fortino, G., Badica, C., Malgeri, M., Unland, R. (eds.) Intelligent Distributed Computing VI, pp. 71–81. Springer (2013)
287. Palopoli, L., Rosaci, D., Sarné, G.M.: A multi-tiered recommender system architecture for supporting E-Commerce. In: Fortino, G., Badica, C., Malgeri, M., Unland, R. (eds.) Intelligent Distributed Computing VI, pp. 71–81. Springer (2013)
288. Lee, Y.-L., Huang, F.-H.: Recommender system architecture for adaptive green marketing. Expert Syst. Appl. **38**(8), 9696–9703 (2011)
289. Prieto, M.E., Menéndez, V.H., Segura, A.A., Vidal, C.L.: A recommender system architecture for instructional engineering. In: Emerging Technologies and Information Systems for the Knowledge Society, Springer, pp. 314–321 (2008)
290. Bhatia, S., Caragea, C., Chen, H.-H., Wu, J., Treeratpituk, P., Wu, Z., Khabsa, M., Mitra, P., Giles, C.L.: Specialized research datasets in the CiteSeerx digital library. D-Lib Mag. **18**(7/8) (2012)
291. Jack, K., Hristakeva, M., de Zuniga, R.G., Granitzer, M.: Mendeley's open data for science and learning: a reply to the dataTEL challenge. Int. J. Technol. Enhanc. Learn. **4**(1/2), 31–46 (2012)
292. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. Microsoft Research, MSR-TR-98-12 (1998)
293. Karypis, G.: Evaluation of item-based top-n recommendation algorithms. In: Proceedings of the 10th international conference on Information and knowledge management, pp. 247–254 (2001)
294. Casadevall, A., Fang, F.C.: Reproducible science. Infect. Immun. **78**(12), 4972–4975 (2010)
295. Rehman, J.: Cancer research in crisis: are the drugs we count on based on bad science? http://www.salon.com/2013/09/01/is_cancer_research_facing_a_crisis/ (2013)
296. Drummond, C.: Replicability is not reproducibility: nor is it good science. In: Proceedings of the evaluation methods for Machine-Learning Workshop at the 26th ICML (2009)
297. Al-Maskari, A., Sanderson, M., Clough, P.: The relationship between IR effectiveness measures and user satisfaction. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 773–774 (2007)
298. Knijnenburg, B.P., Willemsen, M.C., Kobsa, A.: A pragmatic procedure to support the user-centric evaluation of recommender systems. In: Proceedings of the 5th ACM conference on Recommender systems, pp. 321–324 (2011)
299. Pu, P., Chen, L., Hu, R.: Evaluating recommender systems from the user's perspective: survey of the state of the art. User Model. User-Adapt. Interact. **22**(4–5), 317–355 (2012)
300. Pu, P., Chen, L., Hu, R.: Evaluating recommender systems from the user's perspective: survey of the state of the art. User Model. User-Adapt. Interact. **22**(4–5), 317–355 (2012)
301. Ekstrand, M.D., Ludwig, M., Konstan, J.A., Riedl, J.T.: Rethinking the recommender research ecosystem: reproducibility, openness, and LensKit. In: Proceedings of the 5th ACM conference on Recommender systems, pp. 133–140 (2011)
302. Konstan, J.A., Adomavicius, G.: Toward identification and adoption of best practices in algorithmic recommender systems research. In: Proceedings of the international workshop on Reproducibility and replication in recommender systems evaluation, pp. 23–28 (2013)
303. Burke, R.: Hybrid recommender systems: survey and experiments. User Model. User-Adapt. Interact. **12**(4), 331–370 (2002)

304. Perugini, S., Gonçalves, M.A., Fox, E.A.: Recommender systems research: a connection-centric survey. J. Intell. Inf. Syst. **23**(2), 107–143 (2004)
305. Torre, I.: Adaptive systems in the era of the semantic and social web, a survey. User Model. User-Adapt. Interact. **19**(5), 433–486 (2009)
306. Zanker, M., Jessenitschnig, M., Jannach, D., Gordea, S.: Comparing recommendation strategies in a commercial context. IEEE Intell. Syst. **22**(3), 69–73 (2007)
307. Rich, E.: User modeling via stereotypes. Cogn. Sci. **3**(4), 329–354 (1979)
308. Barla, M.: Towards social-based user modeling and personalization. Inf. Sci. Technol. Bull. ACM Slovakia **3**, 52–60 (2011)
309. Weber, I., Castillo, C.: The demographics of web search. In: Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, pp. 523–530 (2010)
310. Mattioli, D.: On Orbitz, Mac users steered to pricier hotels. Wall Str. J. vol. http://online.wsj.com/news/articles/SB10001424052702304458604577488822667325882 (2012)
311. Beel, J.: Towards effective research-paper recommender systems and user modeling based on mind maps. PhD Thesis. Otto-von-Guericke Universität Magdeburg (2015)
312. Beel, J., Langer, S., Kapitsaki, G.M., Breitinger, C., Gipp, B.: Exploring the potential of user modeling based on mind maps. In: Proceedings of the 23rd conference on User modelling, adaptation and personalization (UMAP) (to appear) (2015)
313. Beel, J., Gipp, B., Wilde, E.: Academic search engine optimization (ASEO): optimizing scholarly literature for Google Scholar and Co. J. Sch. Publ. **41**(2), 176–190 (2010)
314. Paik, W., Yilmazel, S., Brown, E., Poulin, M., Dubon, S., Amice, C.: Applying natural language processing (nlp) based metadata extraction to automatically acquire user preferences. In: Proceedings of the 1st international conference on Knowledge capture, pp. 116–122 (2001)
315. Seroussi, Y.: Utilising user texts to improve recommendations. In: De Bra, P., Kobsa, A., Chin, D. (eds.) User Modeling, Adaptation, and Personalization, pp. 403–406. Springer, Berlin (2010)
316. Seroussi, Y., Zukerman, I., Bohnert, F.: Collaborative inference of sentiments from texts. In: De Bra, P., Kobsa, A., Chin, D. (eds.) User Modeling, Adaptation, and Personalization, pp. 195–206. Springer, Berlin (2010)
317. Seroussi, Y., Zukerman, I., Bohnert, F.: Collaborative inference of sentiments from texts. In: De Bra, P., Kobsa, A., Chin, D. (eds.) User Modeling, Adaptation, and Personalization, pp. 195–206. Springer, Berlin (2010)
318. Esposito, F., Ferilli, S., Basile, T.M.A., Mauro, N.D.: Machine learning for digital document processing: from layout analysis to metadata extraction. Stud. Comput. Intell. (SCI) **90**, 105–138 (2008)
319. Shin, C.K., Doermann, D.: Classification of document page images based on visual similarity of layout structures. In: Proceedings of the SPIE document recognition and retrieval VII, pp. 182–190 (2000)
320. Buttler, D.: A short survey of document structure similarity algorithms. In: Proceedings of the 5th international conference on Internet computing (2004)
321. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Comput. Netw. ISDN Syst. **30**(1–7), 107–117 (1998)
322. McBryan, O.A.: GENVL and WWWW: tools for taming the Web. In: Proceedings of the 1st international World Wide Web conference, vol. 341 (1994)
323. Shi, S., Xing, F., Zhu, M., Nie, Z., Wen, J.-R.: Anchor text extraction for academic search. In: Proceedings of the 2009 workshop on Text and citation analysis for scholarly digital libraries (ACL-IJCNLP 2009), pp. 10–18 (2009)
324. Manning, C.D., Raghavan, P., Schütze, H.: An Introduction to Information Retrieval, Online edn. Cambridge University Press, Cambridge (2009)
325. Councill, I.G., Giles, C.L., Kan, M.Y.: ParsCit: an open-source CRF reference string parsing package. Proc. LREC **2008**, 661–667 (2008)
326. Marinai, S.: Metadata extraction from PDF papers for digital library ingest. 10th international conference on Document analysis and recognition (2009)
327. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information Tapestry. Commun. ACM **35**(12), 61–70 (1992)
328. Brooks, T.A.: Private acts and public objects: an investigation of citer motivations. J. Am. Soc. Inf. Sci. **36**(4), 223–229 (1985)
329. Liu, M.: Progress in documentation the complexities of citation practice: a review of citation studies. J. Documentation **49**(4), 370–408 (1993)
330. MacRoberts, M.H., MacRoberts, B.: Problems of citation analysis. Scientometrics **36**, 435–444 (1996)
331. Sosnovsky, S., Dicheva, D.: Ontological technologies for user modeling. Int. J. Metadata Semant. Ontol. **5**(1), 32–71 (2010)
332. Sundar, S.S., Oeldorf-Hirsch, A., Xu, Q.: The bandwagon effect of collaborative filtering technology. In: CHI'08 extended abstracts on Human factors in computing systems, pp. 3453–3458 (2008)
333. Mehta, B., Hofmann, T., Fankhauser, P.: Lies and propaganda: detecting spam users in collaborative filtering. In: Proceedings of the 12th international conference on Intelligent user interfaces, pp. 14–21 (2007)
334. Mehta, B., Hofmann, T., Nejdl, W.: Robust collaborative filtering. In: Proceedings of the 2007 ACM conference on Recommender systems, pp. 49–56 (2007)
335. Mehta, B., Nejdl, W.: Attack resistant collaborative filtering. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, pp. 75–82 (2008)
336. Sugiyama, K., Kan, M.Y.: Serendipitous recommendation for scholarly papers considering relations among researchers. In: Proceeding of the 11th annual international ACM/IEEE joint conference on Digital libraries, pp. 307–310 (2011)
337. Burke, R.: Hybrid web recommender systems. The adaptive web, pp. 377–408 (2007)
338. Ahlgren, P., Colliander, C.: Document-document similarity approaches and science mapping: experimental comparison of five approaches. J. Informetr. **3**(1), 49–63 (2009)
339. Hammouda, K.M., Kamel, M.S.: Phrase-based document similarity based on an index graph model. In: Data mining, 2002. ICDM 2003. Proceedings. 2002 IEEE international conference on, pp. 203–210 (2002)
340. Lee, M.D., Pincombe, B., Welsh, M.: An empirical evaluation of models of text document similarity. In: Proceedings of the 27th annual conference of the Cognitive Science Society, pp. 1254–1259 (2005)
341. Tsymbal, A.: The Problem of Concept Drift: Definitions and Related Work. Computer Science Department, Trinity College, Dublin (2004)
342. Victor, P., De Cock, M., Cornelis, C.: Trust and recommendations. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P. B. (eds.) Recommender Systems Handbook, pp. 645–675. Springer (2011)
343. Verbert, K., Parra, D., Brusilovsky, P., Duval, E.: Visualizing recommendations to support exploration, transparency and controllability. In: Proceedings of the 2013 international conference on Intelligent user interfaces, pp. 351–362 (2013)

344. Lam, S., Frankowski, D., Riedl, J.: Do you trust your recommendations? An exploration of security and privacy issues in recommender systems. Emerging Trends in Information and Communication Security, pp. 14–29 (2006)
345. Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: Proceedings of the 14th international conference on World Wide Web, pp. 22–32 (2005)
346. Burke, R., Ramezani, M.: Matching recommendation technologies and domains. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 367–386. Springer (2011)
347. Zhang, Y., Callan, J., Minka, T.: Novelty and redundancy detection in adaptive filtering. In: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 81–88 (2002)
348. Pizzato, L., Rej, T., Yacef, K., Koprinska, I., Kay, J.: Finding someone you will like and who won't reject you In: A. Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) User Modeling, Adaption and Personalization, pp. 269–280. Springer, Berlin (2011)
349. Cosley, D., Lam, S.K., Albert, I., Konstan, J.A., Riedl, J.: Is seeing believing? How recommender system interfaces affect users' opinions. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 585–592 (2003)
350. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM conference on Computer supported cooperative work, pp. 241–250 (2000)
351. Carmagnola, F., Cena, F., Gena, C.: User model interoperability: a survey. User Model. User-Adapt. Interact. **21**(3), 285–331 (2011)