Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 447–456

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

Scientometric re-ranking approach to improve search results

Nedra Ibrahim*, Anja Habacha Chaibi, Henda Ben Ghézala

RIADI Laboratory, National School of Computer Science (ENSI)
University of Manouba, 2010, Tunisia

Abstract

Common personalization approaches involve re-ranking search results. In such way, documents likely to be preferred by the user are presented higher. In this paper, we focus on research-paper retrieval. We propose a scientometric re-ranking approach based on the scientometric preferences of a particular researcher. The researcher creates its own definition of document quality by the mean of scientometric indicators. These indicators are the base of the scientometric score calculation, which serves to results re-ranking. The originality of our approach was the incorporation of different scientometric indicators into researcher's preferences which have significantly improved ranking performance.

Peer-review under responsibility of KES International

Keywords: Personalization, scientometrics, retrieval quality, user profile, profile modeling, re-ranking ;

1. Introduction

Personalizing search results for individual users is increasingly being recognized as an important future direction for retrieval systems. Providing results specific to individual users is particularly important because different users expect different information even given the same query[1]. User profiles have played a significant role in adapting search results. If a user profile represents faithfully the interests and preferences of the user; therefore it improves the search process[2]. User profile exploitation can be at different levels:

- At query level: by query substitution or query enrichment[3].
- At search level: by matching query, document and profile.

* Corresponding author. Tel.: +966-502-075-690.
E-mail address: nedra.ibrahim@ensi-uma.tn

Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.020

- At result presentation: by filtering, recommendation or re-ranking results.

In this paper, we focus our study on the improvement of re-ranking approach. One main problem is how to rank the results returned by a search engine or a combination of search engines? For individual search engine, there are many techniques for ranking results, such as ranging from counting the similarity score between profile, query and document. Ranking search results is a fundamental problem in information retrieval (IR). Most common approaches primarily focus on similarity of query and a page, as well as the overall page quality[4]. Typically, users expect to find such information in the top-ranked results. They not only look at the document snippets in the first few result pages[5] and then they give up or reformulate the query. This can introduce a significant bias to their information finding process and calls for ranking schemes. That takes into account the overall page quality and its relevance to the query. Besides, it considers the match with the users' real search intent when they formulate the query.

Particularly, in scientific research, the quality of search results is different from the classical definition of results quality. In fact, the researcher is not only interested on results' accuracy but also on their impact and popularity which reflect their scientific quality. In this paper, we focus on improving the scientific quality of search results by re-ranking them according to their qualitative score. The quality score is based on a combination of scientometric indicators which measure the scientific quality of documents. One potential method for assisting the user in searching for required information is to use a personal definition of quality to focus information search results. Thus, we discuss the feasibility of our approach by showing how a user can create his definition of quality. This definition will be used to drive an IR by ranking search results based on these quality preferences. We discuss the problems associated with defining quality. Then, we present an overview of our previous work on the creation of hierarchical, flexible and extensible multidimensional user profile model. This work comes as part of the proposal of a personalized IR system dedicated to scientific research. Our system is based on the scientific quality of the information processed at different stages.

This paper is organized as follows: Section 2 summarizes the state of the art on personalizing search results. In Section 3, we present an overview of our proposed approach. In Section 4, we discuss how we assess the research paper quality. In Section 5, we detail our scientometric re-ranking method. In Section 6, we present the experimentation and evaluation. We finish with a conclusion and future works in Section 7.

2. Overview of related work

We are interested on the personalization of search results in the information retrieval field generally and the research paper retrieval field particularly. The most commonly used results personalization approaches are: filtering results, re-ranking of the returned elements and the recommendation of elements.

Re-ranking search results is a post treatment process aimed to adapt the content of the result to a given user. Challam et al.[6] proposed a conceptual re-ranking approach based on contextual user profile. The results are re-ranked using a combination of the original rank and the conceptual one. To determine the conceptual rank, the conceptual similarity score should be calculated. Vallet et al.[7] proposed a conceptual score. For each document found, the system identifies the first N most similar concepts to this document. Then, it calculates for each a personalized score based on measurement of cosine similarity with the document. Daoud et al.[8] proposed a contextual score. This one is calculated according to a measure of similarity. The similarity measure is deducted from the representative vector of the document and the context vector representative of the adequate center of interest. The final score of the document will be calculated by combining the original score and the contextual score.

In recent years, new similarity measures include learning to rank and personalization-based ranking. Plansangket and Gan[9] proposed an effective web document ranking method. They used Latent Dirichlet Allocation (LDA) classification scores to re-rank Google search returned webpages. Du and Hai[10] proposed a method for measuring webpage similarity based on Formal Concepts Analysis (FCA). Xiang et al.[11] developed different ranking principles for different types of contexts. They encoded the context information as features of the model using a learning-to-rank approach. Context information includes previous queries and the search results clicked on or skipped by users. Lu et al.[12] proposed a user model based on ranking method. Their proposed model is mainly used to capture and record the user's interests. Wang et al.[13] proposed a general ranking model adaptation framework for personalized search. For that purpose, they used a user-independent ranking model and the number of adaptation queries from individual users.

In the field of personalizing the retrieval of research-paper, much research articles were published. We reviewed some ranking and recommendation approaches. Singh et al.[14] proposed to rank the research-papers based on citation network using a modified version of the PageRank algorithm[15]. Tang et al.[16] ranked the authors on h-index and conferences' impact.

In research-paper recommendation, the Content-Based Filtering (CBF) was the predominant recommendation class. The majority utilized plain terms contained in the documents, others used n-grams, or topics based on LDA[17]. A few approaches also utilized non-textual features, such as citations or authors. Yang et al. intended to develop collaborative filtering system, but he has not successfully used explicit ratings[18]. Other recommendation approaches were co-occurrence based which analyzed how often papers were co-viewed during a browsing session. Another approach used co-citations to calculate document relatedness[19]. Some recommendation approaches built graphs to generate recommendations. Such graphs typically included papers that were connected via citations. Some graphs included authors, users/customers and publishing years of the papers[20].

3. Overview of the proposed approach

In this section, we present the different parts of our proposed approach. We motivate our approach from the critical of existing works and the drawbacks which they present. On one hand, the firstly mentioned re-ranking approaches did not serve to scientific researchers. Therefore, they didn't take into account the scientific quality of returned results. Some research-paper ranking approaches did not take into account the user preferences. Other approaches focused on ranking authors or conferences according to one of the impact criteria which cannot match all users' preferences. On the other hand, the reviewed research-paper recommendation approaches present some drawbacks. The majority was a content based approach. In which, the authors focused on extracting text from the title, abstract, introduction, keywords, bibliography, body text and social tags. Some other approaches used different information such as citation or authors. Moreover, we can observe that they did not allow users to define their preferences. In fact, they did not take into account that researcher satisfaction might depend not only on accuracy or on citations. Researchers are also interested on other factors such as paper impact, journal impact or authors popularity.

To overcome the drawbacks of existing personalization approaches, our work is oriented to propose a new personalization approach. Our approach is dedicated to researchers and applicable to the field of research papers retrieval. We propose a new scientometric-based personalization approach which consists of two parts:

- User modeling: a scientometric user profile. We proposed a multidimensional user profile model enriched by a new scientometric dimension: "scientometric preferences". This dimension represents the expectations of researchers by incorporating different scientometric indicators to the user profile.
- Personalization search results: a scientometric approach for re-ranking research papers. This part is the focus of this paper. Our contribution is the proposal of a scientometric-based re-ranking approach based on users' quality preferences. We allow researchers to choose their own definition of quality from their scientometric preferences. We consider both known and unknown users. We define a scientometric score based on scientometric indicators deriving from user profile. This score is used to re-rank search results and to deliver qualitative information relevant and appropriate to the researcher needs at the top ranks.

In this paper, we present our scientometric re-ranking approach based on our proposed user profile model[21]. In Section 4, we present how the user can create his own definition of research paper quality based on his profile. In Section 5, we discuss the method of re-ranking search results based of the user preferences.

4. Research paper quality

The definition of research paper quality is different according to the researcher point of view. A researcher may not be interested to papers having a high impact considering the diversity of information needs. In one research process, a researcher wants, for example, a comprehensive historical survey, a paper that introduced a particular technique or a negative citation to criticize. For this, a general definition of research paper quality cannot suit all the researchers' needs. Therefore, we offer to the researcher the possibility of creating his own definition of quality.

4.1. Assessing tools

The assessment of research papers can be performed by a set of quantitative and qualitative measures. In this context, scientometrics is defined as all quantitative aspects of the science of science, communication science and science policy[22]. In our paper[23], we studied all elements affecting the research paper quality. Amongst the large set of scientometric indicators existing in the literature, we selected the most relevant ones which reflect the real paper impact. We showed that we could assess paper quality by combining a set of scientometric indicators. Then, we performed a user study[1] to ensure that the selected indicators are really the ones that interest the researchers and are the ones that the researchers use. These selected indicators were incorporated into the user profile. They include: publications number, citations number, h-index, journal impact factor and conference ranking. Other incorporated indicators will be detailed in Section 4.2. The scientometric indicators have been used by bibliographic databases, such as Science Citation Index (SCI[24]), Google Scholar[25], CiteSeer[26] and Microsoft Academic Search[2]. On the other hand, we note the existing of several ranking systems providing scientific journal ranking and conference ranking according to their impact. Thomson ISI annually publishes the Journal Citation Report (JCR[3]) which includes a number of indicators among which the journal impact factor (JIF). The portal of the Association Core[4] provides access to the logs of journal and conference classification. The SCImago Journal & Country Ranking portal (SJR[5]) provides a set of journal classification metrics and quality evaluation.

4.2. User's scientometric preferences

The user's scientometric preferences can be defined as the set of scientometric indicators which interest him. Scientometric preferences corresponding to each user vary according to his needs which depend on his institution's research strategy. From the user's preferences, the user profile is constructed according to the profile model described in our paper[21]. Our contribution was the integration of a new dimension "scientometric preferences" based on scientometric indicators. We aimed to construct a user profile concerned with the quality of research.

Table 1. Scientometric criteria corresponding to SubDimensions/ExtSubDimensions

		Criteria (C_j)
SubDimensions (Q_{SUB})	Author quality (Q_{AUT})	-author position-h-index-publication number--citation number
	Content quality (Q_{CNT})	-citation number-co-authors number
	Journal quality (Q_{JOU})	-impact factor-ranking-citation number-self-citation number-publication number-response time
	Conference quality (Q_{CNF})	-publication number-citation number-self-citation number-ranking
	Affiliation quality (Q_{AFF})	-publication number-citation number-self-citation number-group stability-group h-index
ExtSub-Dimensions (Q_{EXT})	Career quality (Q_{CAR})	-Status/grade-Number of research years
	Source quality (Q_{SRC})	-field coverage-time coverage-publications coverage
	publisher quality (Q_{PBL})	-specialty coverage-number of books-journals number
	Association quality (Q_{ASS})	-specialty coverage-number of conferences
	Organization quality (Q_{ORG})	-publication number-citation number-Shanghai ranking

Scientometric indicators are presented in the user profile as attributes classified under SubDimensions or ExtSubDimensions as shown in Table1. To each attribute, which we call criteria, we associate the user preferences as a vector (p_j, op_j, w_i):

---

[1] The study was performed on the researchers of our research laboratory RIADI (www.riadi.rnu.tn/)
[2] www.academic.research.microsoft.com/
[3] Thomson, R. (2014), Journal Citation Reports® Science Edition.
[4] www.portal.core.edu.au/conf-ranks/
[5] www.scimagojr.com/index.php

- p_j: a threshold value of preference,
- op_j: a comparison operator, such as op_j in{<,<=,>,>=,=}
- w_j: an importance weight describing the importance of the associated attribute compared to the other attributes of the same SubDimension or ExtSubDimension, such as w_j in [0,1].

p_j, op_j and w_j are determined implicitly and/or explicitly from the preferences of the user. We can cite an example of user preferences: CitationNumber>25 (p_j = 25, op_j = ">").

To each SubDimension and ExtSubDimension the user associates a weight W_{SUB} in [0,1] and W_{EXT} in [0,1]. These weights describe the importance of each SubDimension or ExtSubDimension compared to other SubDimensions or ExtSubDimensions relatively to the user.

Based on user preferences, we construct the user profile. Users can be known or unknown. A known user has an account and he is individually identified by the system while an unknown user does not possess proper profile. For unknown users, we construct a real-time user profile. For known users, we activate the concerned user model and update his real-time profile.

Our user profile is based on implicit and explicit interactions with the user to collect his preferences. The interactions are implicit because the user isn't directly asked to give opinion. The implicit interactions are based on the user navigation to measure his interest to a given entity. We collect user preferences from the number of pages the user reads, user's interaction with the papers (downloads, edits, views) and citations. Then, we measure the similarities between attributes and entities of interest. Otherwise, the interactions are explicit because we ask the unknown user to define his quality preferences according to a set of scientometric preferences which are:

- Author and his career quality preferences,
- Content and source quality preferences,
- Affiliation and organization quality preferences,
- Journal and publisher quality preferences,
- Conference and association quality preferences.

5. Scientometric re-ranking

Our personalization approach is based on the exploitation of the user profile to re-rank documents according to the user preferences. For each of the returned results we compare its similarity to the user profile. Then, we re-rank the search results according to the similarity score.

To implement this idea we propose a scientometric score, which is the base of the calculation of the scientometric rank. Then, we determine the final rank based on the initial rank and the scientometric rank. We define the formula of the final rank based on the works of Challam et al.[6] and Daoud et al.[8] such as:

$$FinalRank = \alpha * InitialRank + (1-\alpha) * ScientometricRank, 0 < \alpha < 1 \quad (1)$$

The initial rank is the original rank returned by a retrieval system and the Scientometric rank is calculated according to the scientometric score. Our approach is based on the scientometric preferences of the user, which represent the definition of document quality.

5.1. Calculation method

The scientometric score (Q) is calculated according to the steps in Fig. 1. The calculation method is applied to the different hierarchical levels of the user profile. The first step consists of collecting data from the appropriate data resource (user profile and bibliographic database). The online bibliographic database must be chosen according to the scientometric indicators it uses. It should cover the set of scientometric criteria incorporated into our user profile. We extract the results returned by the bibliographic database, the preferences of the user and the scientometric data corresponding to the set of returned results. These data will be used in the following steps. The second step consists of calculating the quality scores that correspond to each SubDimension (Q_{SUB}) and ExtSubDimension (Q_{EXT}). The calculation of the quality score is made by adapting the weighted sums method according to the user preferences. The quality scores obtained will be the basis for the scientometric score calculation. The third step consists of applying the proposed user profile model to define the formula of the scientometric score Q. In the following sections, we detail each of the presented steps.

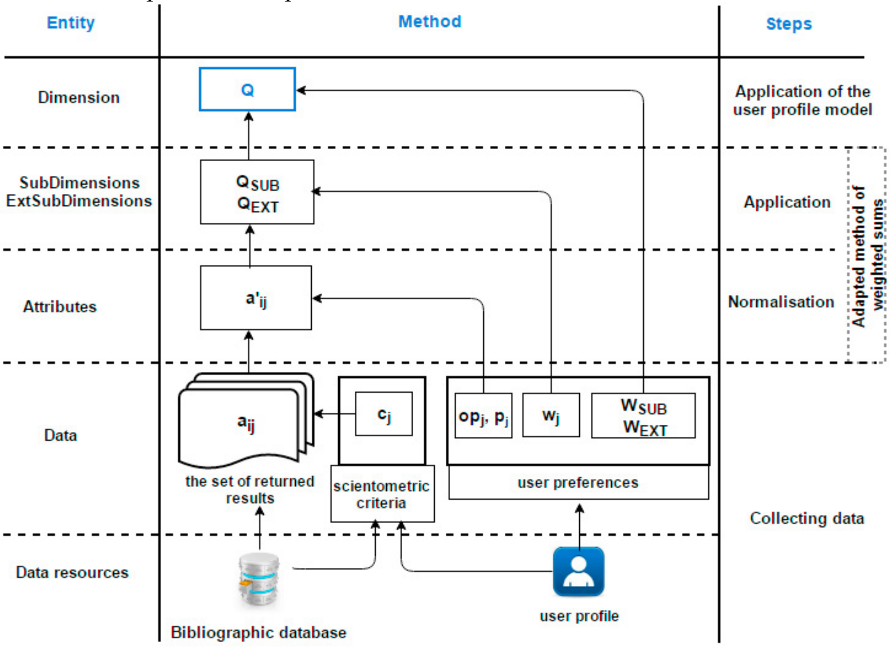

Fig. 1 Calculation method

5.2. Extracting data

The data used in this study were collected from the bibliographic database and the user profile as following:

- m articles A_1, A_2, A_3...A_m : returned from the bibliographic database
- n scientometric criteria C_1, C_2, C_3...C_n corresponding to the attributes of each SubDimension or ExtSubDimension. (Table 1).
- Importance weights vector for each scientometric criteria (w_1, w_2,...w_n) and w_j >=0. These weights are the user preferences extracted from his profile.
- $a_{ij} = U_j(A_i)$, This data represent the scientometric data of each article (A_i) corresponding to each criteria (C_j). Such as, $i \in [1,m]$ and $j \in [1,n]$.
- (p_j,op_j) = (threshold preference value, comparison operator). This couple represents the user preferences for each scientometric criteria C_j extracted from the user profile.

5.3. Adapting the mathematical model of weighted sums

To each SubDimension and ExtSubDimension, we associate a quality score (Q_{SUB} and Q_{EXT}). Each score is calculated from its associated attributes (C_j) and according to the user preferences (p_j, op_j, w_j). We note, for example, Q_{AUT}: the score of author quality and Q_{CAR}: the score of the author's career quality. The other notations are shown in Table 1. Since all the quality scores are calculated by the same way, we note Q_S: the quality score. We consider that $S \in$ {SUB, EXT}. Q_{SUB} is the quality score corresponding to all SubDimentions (Q_{AUT}, Q_{JOU}, etc.). Q_{EXT} is the quality score corresponding to all ExtSubDimensions (Q_{CAR}, Q_{PBL}, etc.). To determine Q_{SUB} and Q_{EXT}, we adapted the mathematical model of weighted sums according to our data and our needs as following:

1. Normalization of all a_{ij} given the preferences (p_j,op_j): we compare scientometric data of each document (a_{ij}) to the threshold value (p_j) preferred by the user. According to the operator (op_j), the normalization will be the difference between the scientometric attributes (a_{ij}) and the threshold value (p_j).
2. Normalization of weights: the sum of weights = 1.

3. Implementation of the weighted sum method presented in equation 2:

$$Q_S(A_i) = \sum_{j=1}^{n} w_j * a_{ij} \qquad such\ as\ i \in [1,m]\ and\ S \in \{SUB, EXT\} \tag{2}$$

We present an example of the calculation of the quality score Q_{CNT} (the score of content quality). We note C_1 and C_2 as the scientometric criteria associated to Q_{CNT}. C_1 is the citations number and C_2 is the co-authors number. For each article A_i we extract the scientometric data a_{i1} and a_{i2} from the bibliographic database. a_{i1} and a_{i2} correspond respectively to the scientometric criteria C_1 and C_2. From the user profile, we extract the user preferences corresponding to C_1: "citations number>50" and corresponding to C_2: "co-authors number<3". The user associated an importance of 60% to C_1 and 40% to C_2. From the extracted user preferences, we determine (p_1=50, op_1=">", w_1=0.6) and (p_2=3, op_2="<", w_2=0.5). Then, we apply the adapted mathematical model of weighted sums to calculate Q_{CNT} according to equation 2.

5.4. Scientomtric score of document-profile similarity

We remind that the scientometric rank is the base of the FinalRank determination (equation 1). Following the user model, we calculate the scientometric score, which we note as Q, corresponding to the scientometric dimension. We define the Scientometric score Q of an article A_i in equation 3:

$$Q(A_i) = \sum_{j=1}^{n} W_{SUB} * Q_{SUB}(A_i) + \sum_{j=1}^{n} W_{EXT} * Q_{EXT}(A_i) + \sum_{j=1}^{n} W_{SUB} * Q_{EXT}(A_i), \forall i \in [1,m] \tag{3}$$

Q_{SUB} and Q_{EXT} were defined in equation 2. The scientometric score $Q(A_i)$ is defined as a sum of three terms. The first term corresponds to the weighted sum of the quality score of all profile's SubDimensions. The second term corresponds to the weighted sum of the quality score of all profile's ExtSubDimensions. The third term corresponds to the extend relation between each SubDimension and its ExtSubDimension. In fact, the quality of the SubDimension depends on the quality of the ExtSubDimension but the reciprocal is not true.

6. Experimentation and evaluation

Our objective is to evaluate our proposed scientometric re-ranking algorithm among an initial ranking. Then, we produce our personalized results and compare it to initial ones. We used the $nDCG_p$[27] as a measure of ranking performance.

6.1. Data set

Our data set contains the following elements:

- Users: we performed our evaluation based on users' database containing 171 researchers working in our research laboratory RIADI. Users are divided into two categories: 20 known users and 151 unknown users. All the users are identified by their research card, which contain their general information, specialty and their research interests. We constructed a complete user profile for the known researchers from their research cards and collected information. We collected the user's scientometric preferences by launching a survey.
- Data source: we opted for the bibliographic database "MS Academic Search" to extract the initial ranking and the corresponding scientometric data. Our choice is justified by the broad set of scientometric indicators covered by MS Academic Search. It includes the scientometric criteria of our user profile.
- Queries and subjects: we used keywords based queries to perform our experimentations. All the known users executed 30 queries on the MS Academic Search. Their queries were related to their domains of research.
- Relevance judgement: only the first hundred results were presented to the users to judge their relevance to the query. Then, after considering the user profile, we attributed to each returned document a scientometric relevance judgment. 0: irrelevant document, 1: relevant having Q<threshold and 3: relevant having Q>threshold. The threshold is the minimal scientometric score according to the user preferences.

6.2. Evaluation procedure

In Fig. 2, we present our evaluation procedure aiming to measure the researcher's satisfaction. One evaluation session consists on the following steps:

- The user executes his query on the MS Academic Search platform.
- Receiving then recording the initial rank corresponding to the top hundred results returned by MS Academic Search.
- Receiving then recording the relevance judgment of the user.
- Establishing the scientometric relevance judgment, referring to user expectations.
- Matching the initial results of the system to the scientometric relevance values.
- Applying the scientometric re-ranking approach to calculate scientometric score for each article in the list of initial results based on the user profile.
- Re-ranking the initial results according to the scientometric score. Then, we record the scientometric rank of the top hundred results.
- Calculating then recording the final ranks for different α values.
- Matching the final results to the scientometric relevance values (for each α value).
- Calculating $nDCG_p$ for the initial ranked list and the final ranked list (for each α value).
- Comparing between the $nDCG_p$ results of the initial ranking and the final ranking (for each α value).

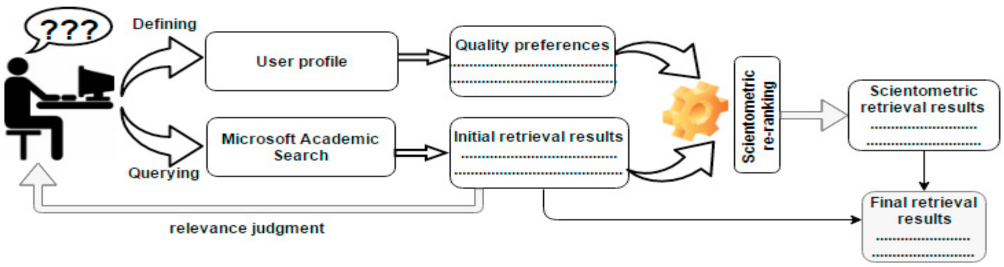

Fig. 2 Evaluation procedure.

6.3. Experimental results and discussion

It should be noted that this evaluation is not intended to assess system performance but to show the contribution of the scientomtrics in personalization search results. For each evaluation session we calculate the $nDCG_p$[27] for different values of α . We computed $nDCG_p$ at rank 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100. We report the results in Fig. 3(a) and Fig. 3(b). In Fig. 3(a), we present the different $nDCG_p$ curves for different α values. By decreasing α, the performance of $nDCG_p$ curves increase. This shows that the integration of the scientometric rank improved the ranking performance. Fig. 3(b) presents the $nDCG_p$ curves at the two extreme α values. When $\alpha = 0$, initial rank is not given any weight, and it is equivalent to pure scientometric based ranking. If $\alpha = 1$, scientometric based ranking is ignored and pure initial rank is considered. Both the scientometric and initial based rankings can be blended by varying the values of α. The results in Fig. 3(b) indicate that the best rank is the scientometric one. It corresponds more to the user expectations comparing to the initial rank given by MS Academic Search.

6.4. Synthesis and comparison

We proposed a scientometric approach of re-ranking search results according to the user's qualitative preferences. Our method was the application of our user profile model and the adaptation of the mathematical model of weighted sums. We adopted our method to calculate the scientometric score used for re-ranking the top hundred results returned by MS Academic Search. Based on our experimentation, we evaluated our re-ranking approach. The findings of this study clearly show that the integration of the scientometric ranking has enhanced the relevance of results from the user point of view.

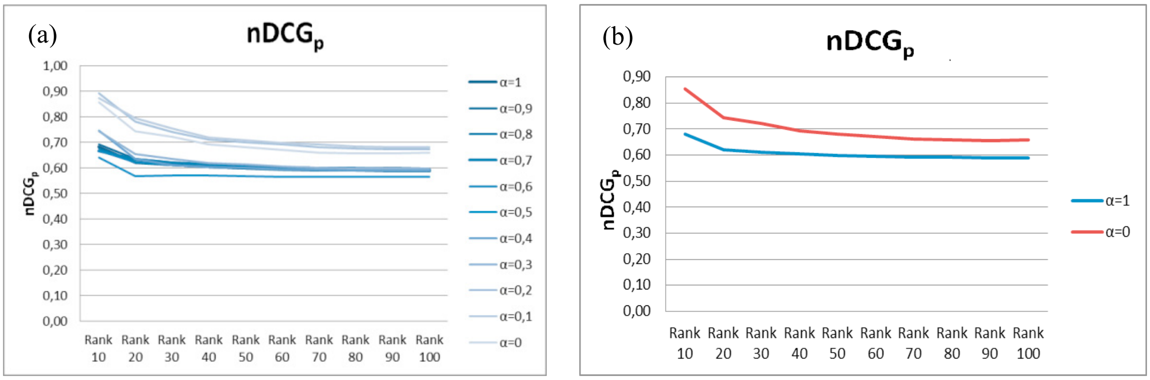

Fig. 3 (a) $nDCG_p$ variation curves for different α ; (b) $nDCG_p$ variation curves for the extreme α values.

In Table 2, we present the $nDCG_p$ results for all ranks corresponding to the initial rank and the final rank. The scientometric ranking realized an improvement at all ranks. By considering the mean of $nDCG_p$, the improvement rate was rated for 14.75% compared to the MS Academic Search ranking.

Table 2. Improvement rate

	nDCG10	nDCG20	nDCG30	nDCG40	nDCG50	nDCG60	nDCG70	nDCG80	nDCG90	nDCG100	**Mean nDCGp**	**Improvement rate**
Initial ranking (α =1)	0,68	0,62	0,61	0,6	0,6	0,6	0,59	0,59	0,59	0,59	**0,61**	
Scientometric ranking (α =0)	0,86	0,74	0,72	0,69	0,68	0,67	0,66	0,66	0,66	0,66	**0,7**	**14.75%**

We notice that several online bibliographic databases, such as Google Scholar[25] and MS Academic Search, used scientometrics. However, they enriched search results with scientometric information when displaying without having to involve the user in the process of results presentation. They initially rank search results according to key words similarity. Yet, CiteSeer[26] has a user profiling system which tracks the interests of users and recommends new citations and documents. It used only citations to find similar scientific articles. The novelty of our approach is the integration of multiple scientometric indicators to re-rank search results based on the user own quality definition.

7. Conclusion and future work

Given the large number of publications, it is impossible to read everything. Therefore, researchers are unable to select relevant qualitative publications. The purpose of our study was helping researchers retrieve effective results corresponding to their preferences. For this reason, we proposed a re-ranking scientometric approach based on scientometric indicators. The preferences of the user served to the construction of the user profile. It includes all elements affecting the scientific quality (author and career quality, content and source quality, journal and publisher quality, conference and association quality, affiliation and organization quality).

Our proposed re-ranking approach was based on a scientometric score calculated for each scientific article returned by the bibliographic database. This scientometric score was the result of the application of an adapted mathematical model of weighted sums considering the scientometric preferences of the user. A strong point of the proposed method is the incorporation of different scientometric indicators, the option to allow users define their preferences and the explicit consideration of known and unknown users.

Based on the proposed scientometric score we performed our experimentations on the first hundred results of the MS Academic Search. Our evaluation method showed a best performance given by the integration of the scientometric rank.

Practically, our contribution can improve the research quality and relevance. Indirectly, it can positively influence research attitudes and affect the quality of research by limiting unscientific practices. Our ranking approach can have various applications: looking up the quality of a given research paper, comparing the research papers, comparing conferences or journals and comparing authors and affiliations.

Our work can be considerably enhanced by taking into account the year of publication of the papers. Considering the time factor can reduce the bias against the recent papers. In fact, they get less time for being studied compared to the older papers.

References

1. Teevan J, Dumais S T, Horvitz E. Beyond the commons: Investigating the value of personalizing web search. In Proceedings of the Workshop on New Technologies for Personalized Information Access (PIA'05), Edinburgh, UK, 24-30 July 2005. p. 84–92.
2. Ramana T V, Rao K V. User search personalization in semantic web mining. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 2012; 1(34):34-40.
3. Zingla M A, Chiraz L, Slimani Y. Short Query Expansion for Microblog Retrieval. In Proceedings of the 20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES'16), Procedia Computer Science, York, UK; 5-7 September 2016. 96:225-234.
4. Chowdhury G. *Introduction to modern information retrieval*. 3rd ed. Facet publishing; 2010.
5. Jansen J, Spink A. An Analysis of Web Documents Retrieved and Viewed. In Proceedings of the International Conference on Internet Computing, Las Vegas, Nevada; 23-26 June 2003. p. 65-69.
6. Challam V, Gauch S, Chandramouli A. Contextual search using ontology-based user profiles. In Proceedings of Large Scale Semantic Access to Content (Text, Image, Video, and Sound) Conference (RIAO'07), Pittsburgh, Pennsylvania; 30 May-01 June 2007. P. 612-617.
7. Vallet D, Fernandez M, Castells P, Mylonas P, Avrithis Y. Personalized information retrieval in context. In Proceedings of the 21st National Conference on Artificial Intelligence, Boston, USA ; July 2006. p. 16-17.
8. Daoud M, Tamine L, Boughanem M, Chebaro B. Construction des profils utilisateurs à base d'une ontologie pour une recherche d'information personnalisée. In francophone en Recherche d'Information et Applications (CORIA'08) ; Mars 2008, Trégastel, France. p. 225-240.
9. S. Plansangket, and J Q. Gan. 2017. Re-ranking Google search returned web documents using document classification scores. Artificial Intelligence Research 2017; 6(1):59-68.
10. Du Y, Hai Y. Semantic ranking of web pages based on formal concept analysis. Journal of Systems and Software 2013; 86(1):187-97.
11. Xiang B, Jiang D, Pei J, et al. Contex-aware ranking in web search. In Proceedings of the International Conference on Research and Development in Information Retrieval (SIGIR'10), Geneva, Switzerland; July 19-23 2010. p. 451-458.
12. Lu Y, Li Y, Xu M, et al. A user model based ranking method of query results of meta-search engines. In Proceedings of the International Conference on Network and Information Systems for Computers (ICNISC'15), Wuhan, China; 23-25 January 2015. p. 426-430.
13. Wang H, He X, Chang, MW et al. Personalized ranking model adaptation for web search. In Proceedings of the International Conference on Research and Development in Information Retrieval (SIGIR'13), Dublin, Ireland; 28 July-01 August 2013. p. 323-332.
14. Singh A P, Shubhankar K., Pudi V. An efficient algorithm for ranking research papers based on citation network. In Proceedings of the 3rd IEEE Conference on Data Mining and Optimization (DMO'11), Putrajaya, Malaysia; 28-29 June 2011. p. 88-95.
15. Page L, Brin S, Motwani R, Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report, Stanford InfoLab, 1999.
16. Tang J, Zhang J, Yao L, Li J, Zhang L, Su Z. Arnetminer: Extraction and Mining of Academic Social Networks. In Proceedings of 14th ACM International Conference on Knowledge Discovery and Data Mining (KDD'08), Las Vegas, Nevada, USA; 24-27 August 2008. p. 990-998.
17. Beel J, Gipp B, Langer S, Breitinger C. Research-paper recommender systems: a literature survey. International Journal on Digital Libraries 2016; 17(4):305-338.
18. 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 (JCDL'09), Austin, TX, USA; 15-19 June 2009. p. 203–212.
19. 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 (JCDL'07), Vancouver, BC, Canada; 18-23 June 2007. p. 417–418.
20. 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 (CIKM'12), Maui, Hawaii, USA; 29 October-02 November 2012. p. 1910–1914.
21. Ibrahim N, Habacha Chaibi A, Ben Ghézala H. A new Scientometric Dimension for User Profile. In The 9th International Conference on Advances in Computer-Human Interactions (ACHI'16), Venice, Italy, 24-28 April 2016. p. 261-267.
22. Hood W, Wilson C. The literature of bibliometrics, scientometrics, and informetrics. Scientometrics 2001; 52(2):291-314.
23. Ibrahim N, Habacha Chaibi N, Ben Ahmed M. New scientometric indicator for the qualitative evaluation of scientific production. New Library World Journal 2015; 116(11/12):661-676.
24. Alireza N. Google Scholar: The New Generation of Citation Indexes. Libri 2005; 55(4):170–180.
25. Lawrence S, Lee C G, Bollacker K. Digital libraries and autonomous citation indexing. Computer 1999; 32(6):67-71.
26. Harzing A W. *The publish or perish book: your guide to effective and responsible citation analysis*. Australia: Tarma software research; 2012.
27. Jurafsky D, Martin J H. *Speech and language processing: an introduction to natural language processing*. New York: Prentice Hall; 2008.