Mobile Netw Appl (2015) 20:508–520
DOI 10.1007/s11036-014-0539-2

A Novel Ranking Model for a Large-Scale Scientific Publication

Bong-Soo Sohn · Jai E. Jung

Published online: 22 November 2014

Abstract With a large number of scientific literature, it has been difficult to search for a set of relevant articles and to rank them. In this work, we propose a generalized network analysis approach (called N-star ranking model) for sorting them based on . The ranking of the result is considered in the mutual relationships between another classes: keyword, publication, citation. From the model, we propose two ranks for this problem: *the Universal-Publication rank - (UP rank)* and *Topic-Publication rank (TP rank)*. We also study two simple *ranks based on citation counting (RCC rank) and content matching (RCM rank)*. We propose the metrics for ranking comparison and analysis on two criteria value and order. We have conducted the experimentations for confirming the predictions and studying the features of the ranks. The results show that the proposed ranks are very impressive for the given problem since they consider the query/topic, the content of publication and the citations in the ranking model.

Keywords Scientific search engine · Scientific recommendation system · Keyword-based query · Scientific topic ranking · PageRank · N-star ranking system

B. S. Sohn
Department of Computer Engineering, Chung-Ang University, Seoul, South Korea

J. E. Jung (✉)
Department of Computer Science and Engineering, Chung-Ang University, Seoul, South Korea
e-mail: ontology.society@gmail.com

1 Introduction

Nowadays, search engines are the foundation of the Internet. Most users choose search engines as the best solution for finding the information they want. In the case the database of the documents is very huge and increases quickly, the number of the matched documents becomes very big. Search engine has to apply different techniques to rank search results for the user query. Most of search engines suggest the users type the *keywords* to identify the content relevant to their needs. Hence, ranking the result for keyword based query/topic is very important and interesting problem.

In this work, we study ranking results on keyword based scientific search engines which are designed for retrieving scientific publications[1]. Keywords represent for queries/topics and the content of publications. There are two main factors effecting to the ranking results: *citations* and *contents*. Users prefer to see the quality publications which have many valuable citations. Moreover, the content of *chosen* publications should be hot topics and must match to the query/topic strongly. We propose a new concept *state* representing for the relationship of keyword and publication. Finally, each keyword, citation, publication and state are proposed to be ranked in our model.

The proposed model are built around the N-star ranking models, which are studied in our prior works [11, 43, 44]. A N-star ranking model consists the rank scores of N classes, which depend to each other. The dependent relationships

[1] Several famous keyword based scientific search engine are Google Scholar (http://scholar.google.com), Microsoft Academic Search(http://academic.research.microsoft.com/), ArnetMiner (http://arnetminer.org/), etc

are described by a system of linear equations called constraint system. Moreover, there is a core class which affects and reflects directly to other classes. The existence of the rank scores is unique and can be estimated by a loop of computing a linear function.

Figure 1 represents an example of 4 publications p, q, r, s with 4 citations c, d, e, f. The principle of PageRank's 2-star model is: (i) a *quality publication* has many *valuable citations*; (ii) a *valuable citation* comes from a *quality publication* and (iii) User may choose a publication randomly. Thus, the 2-star model is described by the equations showed in Fig. 1, in which (i) the rank score of a citation is determined from the rank score of the its citing publication (Fig. 1- Eqs. 3, 4) and (ii) The rank score of a publication is determined by d % from the rank scores of cited-citations and $(100 - d)$ % from the random event that user choose some publication (Fig. 1- Eqs. 5, 6 and 7 with $d = 50$ %).

PageRank is not suitable for the problem since the content factor is ignored. As an example in Fig. 1, suppose three keywords A, B, C belong to the publications. Given query C which publication should be in the first position, q or s? Also, suppose C is a hot topic. How is it effecting to the rank scores of publications q and s? Score s has lower PageRank score, but it is only about topic C. Score q has higher PageRank score, but it has three topics inside. Thus, in this case it is not quite convinced by PageRank in which q is proposed in the first position. Finally, PageRank does not answer question (ii) since it does not reflect information linking between the keywords and publications.

To overcome the above questions, we propose a novel computational model which can measure the probability of the event that user reads some topic (keyword) and its publication. Here are four main hypotheses of the new model.

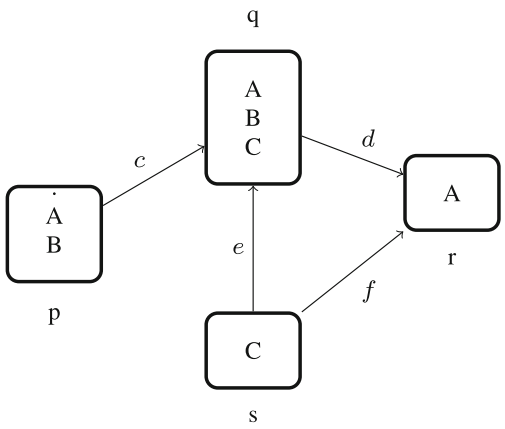

Fig. 1 Example for N-star ranking models of PageRank and the proposed model

1. Publication and its states have mutual affects: The quality of publications is determined by interesting states and vice versa.
2. Keyword and its states have mutual affects: The hot keyword is built on its interesting states and vice versa.
3. Citation factor: It is inherited from the PageRank model. A valuable citation comes from a quality publication and quality publication receives many valuable citations.
4. Randomness: It is inherited from the PageRank model. A user can choose a paper and them choose a state randomly.

The proposed N-star ranking model consists the rank scores of 4 classes: State, Keyword, Publication and Citation. State is the core class of the model. All mutual dependent relationship between classes are expressed by the help of the core class. For example, the relationship between Keyword and Citation is expressed by two mutual direct relationships: Keyword-State and State-Citation. The main improvement of the proposed ranking system from PageRank's ranking system is that we bring the content (keywords) and the relationship keyword-publication (states) to the model. Now, the keyword C has mutual affects with two states (C, q) and (C, s) whose have the mutual affect to the publications q and s (See Fig. 1). Our proposed ranking will consider the ranking scores of two states (C, q) and (C, s) for deciding if q or s is the first position for C. The rank score of a publication depends on its citations and contents. The proposed ranking systems are more adaptive to keywords based queries than PageRank.

We classify the ranks for the given problem into two groups: (i) *universal ranks* which are independent from topic/query and (ii) *local ranks* which are dependent on the topic/query. Straightforward, universal ranks are more suitable for general publication ranking and local ranks are more suitable for scientific topic ranking. In another way, we classify the ranks into three groups: *citation based*, *content based* and *hybrid*. The classification is based on the rank's behavior that if it considers only citation, only content or both. PageRank is universal and citation based. We introduce two simple ranks: (i) RCC (*Ranking based on Citation Counting*) which is citation based and universal; (ii) RCM (*Ranking based on Content Matching*) which is content based and local. From the proposed 4-star model, we propose universal rank, UP (*Universal-Publication rank*), and local rank, TP (*Topic-Publication rank*). Both of UP and TP are hybrid since they consider citation and content both in their formula. The experimental results point out that: (i) UP is the good solution for general ranking publication based on the citation and the content and (ii) TP is the best solution for scientific ranking problem among considered ranks.

2 Mathematical background

2.1 N-star ranking models

Given a set $\mathcal{A}$ of finite objects $\{a_1, \ldots, a_n\}$, R is called a *score* if it is a non-negative real function on $\mathcal{A}$ and $\sum_{i=1}^{n} R(a_i) = 1$. In such a case, $(\mathcal{A}, R)$ is called a *ranking model*, $\mathcal{A}$ is called a *class*, and $R(a_i)$ is called a *score of the object* a_i. Given N ranking models $\{(\mathcal{A}_i, R_i)\}_{i=1}^{N}$, they are *mutually linear* if each score R_i is a linear transformation of others, i.e. there exists a sequence $\{\alpha_{ij}, W_{ij}\}_{i,j=1}^{N}$, where α_{ij} is non-negative, $W_{ij} = (\omega_{uv}^{(ij)})_{|\mathcal{A}_i|\times|\mathcal{A}_j|}$ is a non-negative and normalized columns.

In such a case, $\{(\mathcal{A}_i, R_i)\}_{i=1}^{N}$ is called a *N-linear mutual ranking model*. Note that, a linear mutual ranking model is not a Markov chain except for the case $\sum_i \alpha_{ij} = 1$ for all $j = 1, \ldots, N$. A linear mutual ranking model allows us to describe a lot of complicate realistic multi-ranking systems in which each their sub-ranking system affects mutually other sub-ranking systems. However, the model generally does not guarantee for the existing significant ranking scores (*a unique property*), and from that we can not achieve the main purpose. Recently, we have proposed a *N-star ranking model*, a special case of N-linear mutual ranking model, and proved that this model exists significant ranking score [19, 44]. A N-star ranking model is simply defined as: Given a N-linear mutual ranking model $\{(\mathcal{A}_i, R_i)\}_{i=1}^{N}$, firstly we emphasize some following concepts: An object a_{jv} *affects* to an object a_{iu} (denote $a_{jv} \to a_{iu}$) if $\alpha_{ij}\omega_{uv}^{(ij)} > 0$; A class $\mathcal{A}_i$ *affects and reflects directly* to a class $\mathcal{A}_j$ (denote $\mathcal{A}_i \to \mathcal{A}_j$) if $\forall a_{jv} \in \mathcal{A}_j$: $\exists a_{iu_1}, a_{iu_2} \in \mathcal{A}_i$: $a_{jv} \to a_{iu_2}$ and $a_{iu_1} \to a_{jv}$, and *affects and reflects* to a class $\mathcal{A}_j$ (denote $\mathcal{A}_i \rightsquigarrow \mathcal{A}_j$) if $\mathcal{A}_i \to \mathcal{A}_j$ or $\exists \mathcal{A}_k : \mathcal{A}_i \to \mathcal{A}_k$ and $\mathcal{A}_k \rightsquigarrow \mathcal{A}_j$.

Definition 1 N-star ranking model $\{(\mathcal{A}_i, R_i)\}_{i=1}^{N}$ is represented when there exists a class $\mathcal{A}_i$ such that for all $j \neq i$, (i) $\alpha_{ji} = 1$ and $\mathcal{A}_i$ affects and reflects to $\mathcal{A}_j$ ($\mathcal{A}_i \rightsquigarrow \mathcal{A}_j$), and (ii) $\alpha_{ii} > 0$ and W_{ii} has positive entries. $\mathcal{A}_i$ is called a core of the model.

2.2 PageRank as 2-star ranking model

PageRank model [20, 33] is famous and effective on the *Webpage-ranking field*. This model can be applied on the *Scientific ranking field* when we consider publications as Webpages and citations as hyperlinks. Let $\mathcal{P}$ and $\mathcal{L}$ be the sets of all *publications* and *citations* respectively. For each citation $l \in \mathcal{L}$ from publication p to publication q, $(p, q \in \mathcal{P})$, we denote $p = in(l)$ and $q = out(l)$. For each publication $p \in \mathcal{P}$, we denote $\mathcal{I}(p) = \{l \in \mathcal{L} : p = out(l)\}$ and $\mathcal{O}(p) = \{l \in \mathcal{L} : p = in(l)\}$. In the case p does not cite anywhere, we assume that p cites everywhere, or $\mathcal{O}(p) = \mathcal{P}$. PageRank model constructs the ranking score for publications based on the following formula:

$$R_{\mathcal{P}}^{pr}(p) = d \sum_{l\in\mathcal{I}(p), q=in(l)} \frac{R_{\mathcal{P}}^{pr}(q)}{|\mathcal{O}(q)|} + \frac{1-d}{|\mathcal{P}|} \tag{1}$$

where $d \in (0, 1)$ is a constant. Suppose the rank score of a citation is define as follows:

$$R_{\mathcal{L}}^{pr}(l) = \frac{R_{\mathcal{P}}^{pr}(p)}{|\mathcal{O}(p)|}, \qquad \forall l \in \mathcal{I}, p = in(l) \tag{2}$$

Equation 1 implies that:

$$R_{\mathcal{P}}^{pr}(p) = d \sum_{l\in\mathcal{I}(p)} R_{\mathcal{L}}^{pr}(l) + \frac{1-d}{|\mathcal{P}|}. \tag{3}$$

Equations 2 and 3 imply that PageRank can be presented as a 2-star ranking model.

Equation 3 is called the *master equation* of system, which represents all mutual dependent relationships among classes. It has two factors: (i) *Citation factor*, $\sum_{l\in\mathcal{I}(p)} R_{\mathcal{L}}^{pr}(l)$ and (ii) *Randomness*, $\frac{1}{|\mathcal{P}|}$. d is the *citation parameter* of the model. There is no factor reflecting contents of publications in the master equation. It will be appended in the proposed model (Section 4).

2.3 Scientific topic ranking

In this subsection, we propose an idea to evaluate the scientific publications related with a *given topic* based on their *keywords*. It is simply explained as follows: Let $\mathcal{K}$ be a set of all *keywords* given by publications in $\mathcal{P}$. We denote $K(p) \subseteq \mathcal{K}$ is a set of keywords given in the publication p, and $P(A) = \{q \in \mathcal{P} : A \in K(q)\}$ is a set of all publications containing the keyword $A \in \mathcal{K}$.

Definition 2 T is called a topic if T is a set of finite keywords in $\mathcal{K}$. $P(T) = \bigcap_{A\in T} P(A)$, a set of all publications containing all keywords in the topic T, is called a T-scientific publications.

The aim of our work is firstly to give a significant ranking score on a class T-scientific publications for any given topic T, secondly to evaluate how topics have an effect on the ranking score of publications. The N-star ranking model is used to study these problems. A new class named *state*, $\mathcal{S} = \mathcal{K} \times \mathcal{P}$, a set of all couple (A, p) where A is a keyword and p is a publication, is considered. Of course, a ranking score on $\mathcal{S}$ is the best answer for the question "*How do keywords and topics have an effect on publications?*". Moreover, we also consider two following concepts to study this question. Given a topic T and a ranking score R_T on the class T-scientific publications $P(T)$.

Definition 3 R_T is called universal if it does not depend on T, and called local if it is not universal.

The concept *universal* is understood as follows: R_T is universal if there exists a *universal ranking score* R_U on $\mathcal{P}$ which is determined based on all topics such that for all $p \in P(T) \subseteq \mathcal{P}$, $R_T(p) = R_U(p)$.

3 Citation and content matching for ranking

We propose two simple ranks for the scientific topic ranking problem. The first one is based only the citation counting. The second one is based on the matching content between the topic and the publication.

3.1 Rank based on citation counting

Since there is no loop in a graph-structure of Scientific ranking field in which nodes are publications and links are citations, the PageRank model can be estimated by the following simple model. It is a *citations counting model* which is introduced shortly as follows:

Definition 4 Given any publications $p \in \mathcal{P}$, the rank based citations counting (RCC) of p is given:

$$R^{cc}_{\mathcal{P}}(p) = \frac{|\mathcal{I}(p)|}{|\mathcal{L}|}.$$

We can predict that the rank order of PageRank is very closed to RCC's and is not changed for any citation parameter, d.

3.2 Rank based content matching

We propose the important occurrence function, $O : \mathcal{K} \times \mathcal{P} \mapsto \mathbb{R}^+$, for any couple keyword and publication as follows: For all $A \in \mathcal{K}$, $p \in \mathcal{P}$:

$$O(A, p) = \begin{cases} 3 \text{ if in the title and the abstract of } p \\ 2 \text{ if only in the title of } p \\ 1 \text{ if only in the abstract of } p \\ 0 \text{ otherwise} \end{cases}$$

The important occurrence function, $O : 2^{\mathcal{K}} \times \mathcal{P} \mapsto \mathbb{R}^+$, is generalized for any couple topic and publication as follows:

$$O(\mathbf{T}, p) = \sum_{A \in \mathbf{T}} O(A, p) \tag{4}$$

Matching content function, $M : 2^{\mathcal{K}} \times \mathcal{P} \mapsto \mathbb{R}^+$ is for any couple topic and publication. $M(\mathbf{T}, p)$ is proportional with $O(\mathbf{T}, p)$ and inversely proportional with the number of keywords with p.

$$M(\mathbf{T}, p) = \frac{O(\mathbf{T}, p)}{|K(p)|} \tag{5}$$

Definition 5 Given a topic T and publications $p \in P(T)$, the rank based content matching (RCM) of p is given by

$$R^{cm}_{\mathbf{T}}(p) = M(\mathbf{T}, p).$$

4 Four-star ranking model for scientific topics

Given an object $s = (A, p)$ of class states $\mathcal{S}$, s is called a *related state* of the keyword A and the publication p. The model is constructed on four classes, i.e., $\mathcal{S}$*-states*, $\mathcal{P}$*-publications*, $\mathcal{K}$*-keywords*, and $\mathcal{L}$*-citations* with the following hypotheses.

1. A rank score of a publication equals to a sum of rank scores of its related states:
$$R_{\mathcal{P}}(p) = \sum_{A \in K(p), s=(A,p)} R_{\mathcal{S}}(s), \forall p \in \mathcal{P}. \tag{6}$$
2. A rank score of a keyword equals to a sum of the rank scores of its related states:
$$R_{\mathcal{K}}(A) = \sum_{p \in P(A), s=(A,p)} R_{\mathcal{S}}(s), \forall A \in \mathcal{K}. \tag{7}$$
3. A rank score of a citation equals to a sum of the rank scores of the states where the citation is from:
$$R_{\mathcal{L}}(l) = \frac{R_{\mathcal{P}}(p)}{|\mathcal{O}(p)|}, \forall l \in \mathcal{L}, p = in(l) \tag{8}$$
$$= \sum_{A \in K(p), s=(A,p)} \frac{R_{\mathcal{S}}(s)}{|\mathcal{O}(p)|}. \tag{9}$$
4. A rank score of a state is given:
$$R_{\mathcal{S}}(s) = \alpha_1 \frac{R_{\mathcal{P}}(p)}{|K(p)|} \tag{10}$$
$$+\alpha_2 \frac{R_{\mathcal{K}}(A)}{|P(A)|} \tag{11}$$
$$+\alpha_3 \sum_{l \in \mathcal{I}(p)} \frac{R_{\mathcal{L}}(l)}{|K(p)|} \tag{12}$$
$$+\alpha_4 \frac{1}{|\mathcal{P}||K(p)|}, \forall s = (A, p) \in \mathcal{S} \tag{13}$$
where $\alpha_i > 0$ and $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$.

Let us explain the above model is an improvement of the PageRank model by considering extra information of keywords of publications. It can be verified that Eq. 8 is similar to Eq. 2 where the rank score of a citation is determined by a rank score of its publication. By considering a new class states $\mathcal{S} = \mathcal{K} \times \mathcal{P}$, the improvement here is a difference

Table 1 Properties of the data set

#Pub	#Cit.	#Key.	#State	Avg($\frac{Key.}{Pub.}$)	Avg($\frac{Pub.}{Key.}$)	Avg($\frac{Cit.}{Pub.}$)	Avg(Pub. Year)
121166	604672	18527	530750	4.3	28.6	5.0	2001.3

between Eqs. 2 and 10. While Eq. 3 has two factors, *citation factor*, $\sum_{l\in\mathcal{I}(p)} R_{\mathcal{L}}(l)$; and *randomness*, $\frac{1}{|\mathcal{P}|}$, Eq. 10 has four factors, *quality publication*, $\frac{R_{\mathcal{P}}(p)}{|K(p)|}$; *hot topic*, $\frac{R_{\mathcal{K}}(A)}{|P(A)|}$; *citation factor*, $\sum_{l\in\mathcal{I}(p)} \frac{R_{\mathcal{L}}(l)}{|K(p)|}$; and *randomness*, $\frac{1}{|\mathcal{P}||K(p)|}$. α_3 is the *citation parameter* of the model. Clearly, the citation factor and the randomness in Eqs. 3 and 10 are similar. The factors *quality publication* and *hot topic* have since there always exist a relationship between any publication's keywords and its content, and a relationship between any keywords and any topics respectively. The proposed model utilizes these relationships to increase information of publications for ranking them. The improvement is sufficient to confirm that the new model has a significant ranking score.

5 Experiments

5.1 Data set and query set

Data set We do experiments on real data collected from Microsoft Academic Search (MAS)[2]. There are 121166 publications belonging to database sub-domain which includes publications from 643 conferences and 337 journals. Citations are internal, in which each publication refers to only other publications in the current data set. Number of internal citations (*#Cit.*) is 604072. Keywords are extracted from MAS. Number of keywords (*#Key.*) is 18527. A size of the space of all states (keyword, publication) (*#State*) is 530750. The average keywords contained in each publication (*Avg Key./Pub.*) is 4.3, the average publications having a given keywords (*Avg Pub./Key.*) is 28.6, the average citations on each publication (*Avg Cit./Pub.*) is 5, and the average published year (*Avg Pub. Year*) is 2001.3. Table 1 gives these information of the data set.

5.2 Metrics

We propose different metrics for evaluating and comparing the ranks in two criteria: (i) ranking value; (ii)ranking order. They are as follows:

Ranking value comparison Given two rank scores R_1, R_2 on the class $\mathcal{A}$ of finite objects, we use following formulas to compare the difference of the ranking values of R_1 and R_2: $\forall\omega \in \mathcal{A}$:

$$\Delta^{R_1,R_2}(\omega) = |R_1(\omega) - R_2(\omega)| \tag{14}$$

$$\%\Delta^{R_1,R_2}(\omega) = \frac{\Delta^{R_1,R_2}(\omega)}{R_1(\omega)} (R_1(\omega) > 0) \tag{15}$$

w is a *increasing (decreasing) point* of R_2 compare to R_1, if $R_2(\omega) > R_1(\omega)(R_2(\omega) < R_1(\omega))$.

$TopN^{\uparrow}_{\Delta}(R_1, R_2)$ and $TopN^{\downarrow}_{\Delta}(R_1, R_2)$ are the top N increasing and decreasing points of $\mathcal{A}$ sorting by Δ^{R_1,R_2}. By studying these top-N points, we can find the reasons for explaining the different values of R_2 compare to R_1.

Ranking order comparison The measure of the ranking order difference between R_1 and R_2 is proposed based on the *concordant* and *discordant* concepts see [21, 26, 29]. R_1 and R_2 are called "concordant" when large values of R_1 go with large values of R_2 and "discordant" when large values of R_1 go with small values of R_2. More precisely, two objects (ω_i, ω_j) are *concordant* if

$$[R_1(\omega_i) - R_1(\omega_j)][R_2(\omega_i) - R_2(\omega_j)] > 0$$

or

$$R_1(\omega_i) = R_1(\omega_j) \wedge R_2(\omega_i) = R_2(\omega_j),$$

and *discordant* if

$$[R_1(\omega_i) - R_1(\omega_j)][R_2(\omega_i) - R_2(\omega_j)] < 0.$$

And, R_1 and R_2 are *concordant* if the probability of (ω_i, ω_j) being concordant is very high, and R_1 and R_2 are *discordant* if vice versa. We propose that R_1 and R_2 is *similar* if they are concordant and *different* if they are discordant, and use the *Kendall measure* [22, 25] to measure these quantities. It can be verified that $-1 \leqslant \Delta_K(R_1, R_2) \leqslant 1$, and it receives value -1 if R_1 and R_2 are totally different and 1 if R_1 and R_2 are totally similar. If $\Delta_K(R_1, R_2) = 1-\alpha$, we guarantee that the probability of the objects whose ranking order by R_1 and R_2 are different is around $\alpha/2$. Finally, given R_1, R_2 and R_3 ranks over the same class of objects. R_1 is *closer* to R_2 than R_3, if we have:

$$\Delta_K(R_1, R_2) > \Delta_K(R_2, R_3)$$

And R_2 is *in middle* of R_1 and R_3, if R_2 is closer to R_1 than R_3 and R_2 is closer to R_3 than R_1.

[2] http://academic.research.microsoft.com/ - Accessed on December 2013

Table 2 Kendall (Not Δ_K) value for all pairs of PageRank's results

d	0.1	0.2	0.3	0.4	0.5
0.1					
0.2	0.993				
0.3	0.987	0.994			
0.4	0.981	0.988	0.994		
0.5	0.976	0.982	0.988	0.994	
0.6	0.970	0.977	0.983	0.989	0.995
0.7	0.965	0.971	0.977	0.983	0.989
0.8	0.960	0.966	0.972	0.978	0.984
0.9	0.955	0.961	0.967	0.973	0.979

5.3 Experiment results

We design 9 parameter sets for PageRank (PR) in which citation weighting changes from 0.1 to 0.9 and the randomness weighting changes from 0.9 to 0.1, respectively. For each parameter set for PageRank, we generate a ranked list of publications. Then, we compare those parameter sets with each other by computing the Δ_K value between each pair of their results. We also rank publications by RCC. Then, we compare PageRank and RCC by computing the Δ_K value between each result of PageRank and RCC.

- The average of the Δ_K value for all pairs of PageRank's results is 0.9811. The minimum and maximum value are 0.9547 and 0.9949, respectively. Table 2 shows the detailed results. Please note that some results are obvious, so they are omitted, e.g., Δ_K between a ranked list and itself equals 1, and the Δ_K is symmetric.
- The average of the Δ_K value between each PageRank's results and RCC's result is 0.9259. The minimum and maximum value are 0.9242 and 0.9266, respectively. Table 3 shows the detailed results.

Result RES1 implies that the probability of a publication having different ranking order by two arbitrary PageRank in regard to parameter set is around 0.0094. Similarly, result RES2 implies that the probability of a publication having different ranking order by PageRank and RCC is around 0.0371.

Based on the parameter sets for PageRank, we define 9 parameter sets for UP rank (UP) using two rules: (i) when the citation weighting changes from 0.1 to 0.9, in the first 8 cases, the randomness weighting is kept stable at 0.1 except the last case it is 0.05; (ii) we keep keyword weighting equals publication weighting. Table 4 shows the detailed information on parameter sets. For each parameter set for UP, we generate a ranked list of publications. Then, for each parameter set, we compare UP with PR and RCC by computing the Δ_K value between each pair of their results, respectively.

The average of the Δ_K value between each UP's results and RCC's result is 0.4415. The minimum and maximum value are 0.1980 and 0.5983, respectively. The average of the Δ_K value between each UP's results and its respective PageRank's result is 0.4695. The minimum and maximum value are 0.2049 and 0.6602, respectively. Table 4 shows the detailed results.

To further examine UP, we compare UP ranking value with PR (using parameter set 5) and RCC on many aspects. We list Top-3 most increasing and decreasing publications by UP vs. PR with detailed information on title, authors, and keywords. We also analyze the average statistic of Top-20 most increasing and decreasing publications. Finally, we compare UP, PR, and RCC average ranking value grouping by published year, the number of keywords, and the number of citations.

Table 5 shows the detailed information of Top-3 increasing and decreasing publications.

Table 6 shows the average values of some features of Top-20 most increasing/ decreasing publications by UP rank vs. PageRank. The average published year are 1977.75 and 2001.6 for Top-20 increasing and decreasing publications, respectively. Similarly, the average number of keywords in each publication are 5.2 and 1.3, the average number of citations are 600.2 vs. 7.45.

UP, PR, and RCC average ranking value grouping by published year, the number of keywords, and the number of citations are presented in Figs. 2, 3, and 4, respectively.

The two first increasing publications were written by E. F. Codd, who is the Father of Relational Database. They have big numbers of citations (1329 and 257 respectively). They are the novel works in Relational Database. The third increasing publication was about R-tree index, which is one of the most important topic in database field. Those topics of Top-3 increasing ones are the foundations of Relational Database. E.g., *Relational Data* are presented in 712 publications and cited by 5617 publications, *Data Integrity* are presented in 1611 publications and cited by 8934 publications, *Normal Form* are presented in 267 publications and

Table 3 Δ_K value between each PageRank's results and RCC's result

d	0.1	0.2	0.3	0.4	0.5
Δ_K	0.924	0.925	0.926	0.926	0.927

Table 4 Parameter sets for UP rank and the corresponding Δ_K value between UP, PR, and RCC

Param. set	α_1	α_2	α_3	α_4	Δ_K **UP/PR**	Δ_K **UP/RCC**
1	0.4	0.4	0.1	0.1	0.2049	0.1980
2	0.35	0.35	0.2	0.1	0.3131	0.3043
3	0.3	0.3	0.3	0.1	0.3852	0.3732
4	0.25	0.25	0.4	0.1	0.4395	0.4230
5	0.2	0.2	0.5	0.1	0.4858	0.4633
6	0.15	0.15	0.6	0.1	0.5299	0.4997
7	0.1	0.1	0.7	0.1	0.5765	0.5365
8	0.05	0.05	0.8	0.1	0.6307	0.5774
9	0.025	0.025	0.9	0.05	0.6602	0.5983

Table 5 Top-3 increasing and decreasing publications in details

Pub.	Title	Author	Keyword
1 (Inc.)	A Relational Model for Large Shared Data Banks	Edgar Frank Codd	Data Structure, Normal Form, Data Representation, Network Model, Relational Model, Data Integrity, Tree Structure, Large Data, and External Representation
2 (Inc.)	Further Normalization of the Data Base Relational Model	Edgar Frank Codd	Relational Data and Relational Model
3 (Inc.)	R-trees: a dynamic index structure for spatial searching	Antonin Guttman	Search Space, Computer Aided Design, Multi Dimensional, k-d tree, Index Structure, Database System, Spatial Data, Data Grid, and Scientific Research
1 (Dec.)	How to develop a differentiated supply chain strategy	Per Hilletofth	Supply Chain
2 (Dec.)	Measuring overlap in logistic regression	Peter J. Rousseeuw and Andreas Christmann	Logistic Regression
3 (Dec.)	Logistics information systems: An analysis of software solutions for supply chain co-ordination	Petri T. Helo and Bulcsu Szekely	Information System and Supply Chain

Table 6 Average values of Top-20 increasing/decreasing publications by UP over PR

Publications.	**PR**$\times 10^6$	**UP**$\times 10^6$	$\Delta \times 10^6$	**Year**	**#Key.**	**#Cit.**
Top-20 Increasing	448.48	753.15	304.67	1977.75	5.2	600.2
Top-20 Decreasing	14.53	10.46	-4.07	2001.6	1.3	7.45

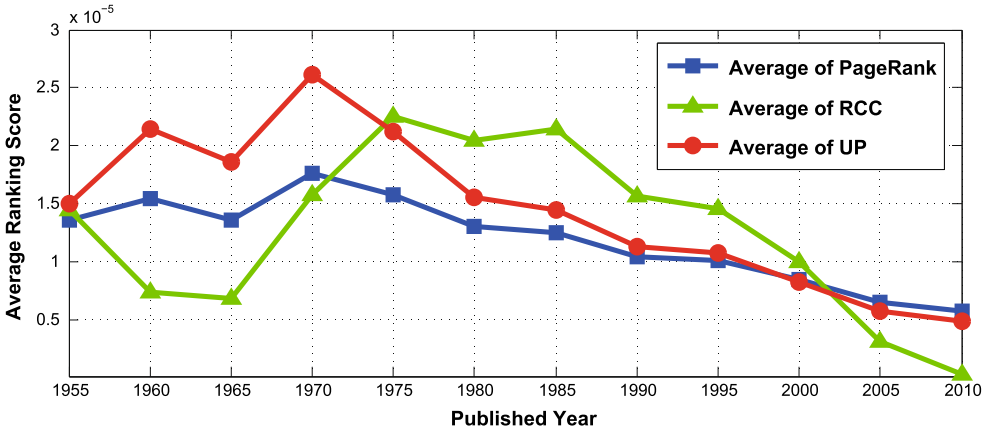

Fig. 2 Average ranking value by published time

cited by 3112 publications. Top-3 decreasing publications are about: *Supply Chain, Logistic Regression, Information System*. Those topics are not very important in database field, or too general. E.g., *Supply Chain* are presented in 681 publications and cited by 906 publications, *Logistic Regression* are presented in 114 publications and cited by 103 publications.

The hotness of a keyword can be reflected by the ratio C/P (citation/publication). The C/P of *Relational Data*, *Data Integrity* and *Normal Form* keywords are 7.89, 5.54 and 11.65 respectively. The C/P of *Supply Chain* and *Logistic Regression* are 1.05 and 0.9 respectively. The average number of C/P is 5 (See Table 1). Thus the keywords of the first group are *hot keywords*, the second's one are not hot keywords.

We examine 3 basic aspects: published year, the number of keywords, and the number of citations. We group publications by those aspects, then, we compute the average ranking value for RCC, PR, and UP.

We divide published time into group of 5 consecutive years. The results are shown on Fig. 2. We could see that all ranking values are generally decreasing by time. This happens because all three ranks are based on the number citations, which is generally larger for older publications. Based on the results, we have some interesting comments: In the history of database field, the period between 1970 and 1975 is the most active one with the introduction of relational database. This is the most important event in the modern database development. We could list some dominating topics such as *Relational Data*, *Data Integrity*, *Normal Form*. This is the reason why this period does not have the highest RCC ranking value but it have the highest UP and PR ranking value.

We group publications by the number of keywords. The results are shown on Fig. 3. Based on the results, we have some interesting comments: The explanation for this phenomenon is that UP takes into account content information. When the number of keywords increases, the content information increases. We also observe a very interesting phenomenon. That is, when the number of keywords equal to 23, RCC, PR, and UP are all increasing drastically. Taking a closer look into the dataset, we see that there are just a little of publications having that number of keywords. They generally have many citations, but the citations are not very

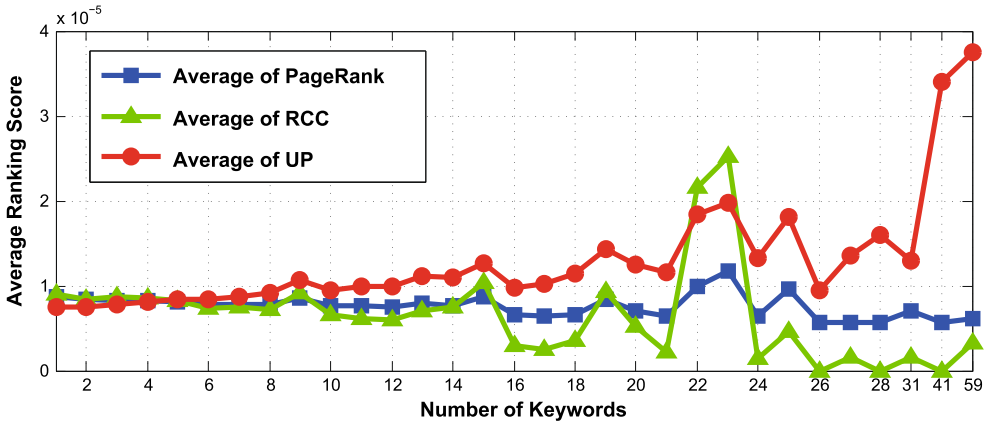

Fig. 3 Average ranking value by keyword count

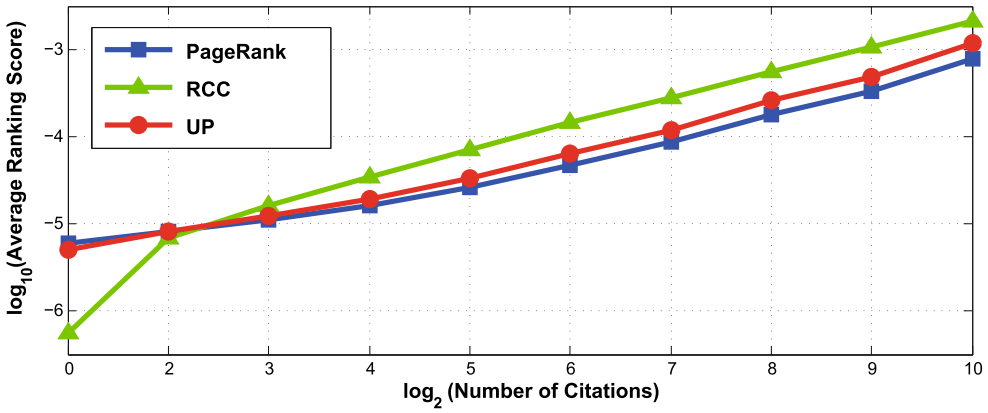

Fig. 4 Average ranking value by citation count

quality, as reflected in smaller PR and UP ranking value. However, the increases in UP and PR ranking value indicate that these publications are good publications.

We divide the publications by the integer value of the binary logarithm of the number of citations. The values are from 0 to 10. Then, we compute the average ranking value for each group of publications with 0 to 1 citation, 2 to 3 citations,..., with 1024 to 2048 citations. The results are shown on Fig. 4. Please note that the number of publications which have no citations is 53568, which is 44.21 % of the dataset. Those publications which have a little of citations are not preferred by UP.

We valuate the efficient of the ranks over the Scientific Topic Ranking problem with following parameters: (i) The citation of parameter of PageRank and N-star ranking model is 0.5; (ii) Other parameters of N-star ranking model is chosen as described in the parameter set 5 (See Table 4). We compute the rank orders and the value of Δ_K of the ranks for each topic in the query set. We compute the average value of Δ_K of each pair of ranks over each type of topics (1 keyword, 2-keywords, 3 -keywords and 4-keywords) and overall.

The explanation for the above remark is that: (i) TP is a local rank which is quite different to universal ranks, UP and PageRank; (ii) TP considers the citation factor which is not considered in RCM. Hence TP is quite different to UP, PageRank or RCM. Therefore, TP is a good candidate for Scientific Topic Ranking problem, since it is local and based on both of citation and content.

We choose 4 topics representing for 4 types of topics in the query set: 1-keyword, 2-keywords, 3-keywords and 4-keywords. The result orders of these topics have the least value of $\Delta_K(TP, PR)$ different among their type, since we want to analyze the typical case studies. We also focus on analyzing three rank: TP (representing for the approach based on citations, content) , PageRank (representing for the approach based on citations) and RCM (representing for the approach based on content). The detail information of 4 topics are shown in Table 7. For each topic, we list all publications of the results and their properties.

There are many publications in the same topic whose PagerRank score are the same with the others. All T1, T2, T3 and T4 have 6 cases of the same score. We have examine overall the query set and the proportion of such cases are about 60 % for 1-key., 42 % for 2-key, 32.6 % for 3-key. and 30 % for 4-key. See Table 8 for detail. Hence, we conclude: The explanation of this remark is that there are many publications which have no citations and the citation graph is acyclic. It is another reason that PageRank does not work well in this problem.

RCM meets this problem too. The proportion of publications which have the same rank value of RCM in some topic is 44.2 %. T1, T3 and T4 have 4, 2, 4 cases of the same RCM score respectively. TP does not have this problem. The

Table 7 Four specific topics with the most difference of Δ_K(TP,PR)

ID	Keywords	#Res	UP/TP	UP/PR	TP/PR	TP/RMC	PR/RMC
T1	Life Insurance	10	-0.73	0.53	-0.31	0.56	-0.49
T2	Natural Language, Semantic Analysis	10	-0.60	0.58	-0.27	0.33	-0.09
T3	Decision Making, Decision Support System, Knowledge Management	10	-0.78	0.44	-0.53	0.58	-0.16
T4	Query Language, Database System, Query Processing, Data Model	13	0.24	0.58	0.33	0.33	-0.22

Table 8 The average percent of the equal rank score cases for the query set

Topics	1-key	2-key	3-key	4-key	All
PR	0.600	0.420	0.326	0.301	0.464
RCM	0.614	0.383	0.276	0.210	0.442
TP	0.018	0.006	0.005	0.003	0.009

proportion of publications which have the same rank value of TP in some topic is just very small 0.9 %.

The rank order analysis over four case studies shows again that PageRank is not suitable for Scientific Topic Ranking problem, since it is confused by the publications which have the same score and it is not considered the content factor. RCM is not suitable either, since the citation is the most important factor for evaluating the quality of publication. The analysis confirms again that both of content and citation factor have the effect to the rank order of TP in Scientific Topic Ranking. Hence, we consider TP is the best solution among the ranks for the given problem.

6 Related works

In this section we have a brief review of related work belonging to following topics: (i) Keyword based querying and ranking; (ii) Web-pages Ranking; (iii) Link-based object ranking; (iv) Bibliometric ranking and (v) N-tier ranking system.

Keyword-based querying and ranking play crucial roles in information retrieval mechanism for data on the Internet, Database, Information Systems, and so on, because of its user-friendly query interface [7, 30, 47]. Most of current works focus on exploiting graph-based structures of keywords and data, since graph is a strong mathematical model for representing for the relationship between keywords and data. According to [45], the data graph can be XML (semi-structured data), relational databases (structured data), and all kinds of schema-free graph data. In the topic of keyword search over database, database is viewed as a labeled graph where records in different tables are treated as nodes connected via foreign-key relationships ([1, 15] and [48]). In the topic of keyword search over XML, the basic structure of the data graph is a tree and the results of keyword-based queries are defined as the meaningful smallest sub-trees which often refer to the lowest common ancestors (LCA) [4, 10]. In the topic of keyword search over schema-free graphs, many algorithms like BLINKS [12], BANKS [16] or R-cliques [23] consider the answer as sub-graphs in LCA semantics. Meanwhile, other approaches rank the result nodes with different criteria, such as high prestige authority value in ObjectRank [2] or high relevance combined with user preference value in PerK [40]. Finally, [3] claimed that ranking the results is the most important component for keyword-based system and it strongly depends on the data model. N-linear ranking model has two advances for ranking results of above these systems. First, it is based on graph-based structure. Second, its linear constraint system is a good tool for express the semantic meaning of the mutual relationships between objects.

The topic inspires us and our works with many foundation works, since it is the symbol of the era of the Internet and big data. The brief overviews of Web-pages rankings can be found at [39] and [38]. Two most famous and important algorithms are PageRank ([33]) utilized by Google search engine and HITS (Hyperlink-Induced Topic Search) introduced by [27]. We have shown that PageRank can be considered a special case of N-star ranking model (Section 2.2). The idea behind HITS algorithm classify the webs into two classes: (i) hubs, served as large directories point to (ii) authoritative pages. A good hub represented a page that pointed to many other pages, and a good authority represented a page that was linked by many different hubs. The model can be rewritten into 3-star ranking model of hub, link and page. Finally, both PageRank and HITS consider the network of web pages as a directed homogeneous network with the weight of the edge is binary.

With the emerge of semantic web (Web of data), many strategies have been proposed to deal with the ranking problems, with different approaches, exploiting different aspects and characteristics of designed systems like queries (user/non-user dependency), ranking granularity (items, documents, entities), features of data (domain, authority, locality, predictability) [9, 35]. [31] proposed PopRank for ranking the popularity of objects based on their web popularity and the object relationship graph. It uses the Popularity Propagation Factor to express the relationship between classes. PopRank is based on Markov Chain model, which is different to the mathematical model of N-linear mutual ranking systems. In N-star ranking model, each class has its own rank score, rather than consider them in a unified framework. [42] proposed NetClus algorithm that utilizes links between objects of multi-typed to rank cluster of multi-typed heterogeneous networks. This work is the successor of their previous work, RankClus [41], which can rank and cluster one-typed objects mutually. Their model is

the same idea with ours when limiting only within a star network schema and giving rank distribution for each type of objects. The different points between ours and theirs are as follows. In NetClus model: (i) The center class does not permit to have the relationship with itself and (ii) The convergence of the rank scores is not proven completely under mathematics view point. In N-star ranking model: (i) The center class can has relationship with itself (The randomness for example) (ii) The convergence of the rank scores is guaranteed by the condition that the center class must *affect and reflect* to all classes. Other complex ranking systems have been already explored using a different formalism for ranking or classification in heterogeneous networks. For example, the quantum ranking [36] is based on quantum navigation. Their formula is come from the quantum theory and quite different to ours.

In a very recent scientometric study, [28] compare expert assessments to bibliometric measures for determining a tiered structure of information systems (IS) journals. One of their noticeable conclusion is that bibliometrics can be a complete, less expensive and more efficient substitute for expert assessment. For bibliometric approaching, the most used citation metrics are ISI Impact Factor see [8, 32], H-index [13, 14] and it variants like G-index [6], E-index [49], and so on. Another study of bibliometric graph-based algorithms focus on ranking researchers was conducted by [18]. They compare sophisticated citation analysis algorithms like PageRank [33], SARA [34], CoRank [50], FutureRank [37], P-Rank [46], BiRank [17] with some simpler methods like citation count and sum of paper ranks, similar to the way we evaluate the experiment results. Further information about bibliometrics and web-based citation analysis can be seen on [5].

7 Discussion

The proposed model is a new application of N-star ranking model which is based on a quite new mathematical model. Thus there are many interesting issues, which are related to N-star ranking model and the proposed model, need to be considered in the future. Some of them are follows.

The choice of the parameter values in the master equation of a N-star ranking model is very important step since it reflect the importance of the mutual dependent relationships between classes. Literally, it decides the quality of the ranking systems. The parameter's optimization is not considered in this work, and it is the first priority for our future work. We have two main ideas for the optimization of parameter values of the proposed model. First, we use the ranking list of venues which is published by experts for adjusting our ranking scores of venues fitting their ranking list best (Example: The ranking list of venues published by CORE - The Computing Research and Education Association of Australasia[3]). Second, we apply the techniques in Artificial Intelligence to reduce the complexity of the optimization since the data set is very big. We will study also the difference and the precise of the optimized parameters over difference domains such as Database, Data mining, Software, and so on to have a deep understanding about the optimization.

Most of scientific indexes (such as H-index, and Impact factor) consider the citations equally. However, the values of the citations from *famous* publications should be considered higher than the values of the ones from *junk* publications. We have planned to study NS-indexes[4] in which each citation, publication, and author are ranked by the N-star ranking model. There are two approaches for the proposed indexes. First approach is the improvement from the classical indexes. For each classical index, we develop a new NS-index which inherits all principle features of the original classical index but considers the difference of citations by the their ranking scores. In the second approach, we design the NS-indexes for ranking authors, venues according two criteria *popularity* and *quality*. The desired properties of NS-indexes will be examined the metrics of ranking comparison over the real data sets.

The experiments have shown that there are many *hot* topics were introduced and studied in 1970, when the foundations of relational database were established. It is a reason that the time series should be integrated in the N-star ranking model. The time point of publication and citation are assigned by the published day. We can mine interesting knowledge based on the history of citations, publications for each topic, author, venue, research institute or even nation. There are two special mining techniques will be emphasized in our time series N-star ranking model. First, based on the formula we can detect the effect of some class to another in the time line. Second, by comparing the difference of the rank scores we can detect the events or predict the trends.

8 Conclusion

We have introduced and studied a new approach for sorting publications matching a keyword based query. The ranks are based on 4-star ranking model of 4 classes: Keyword, Publication, Citation, State. State represents for the correlation of a keyword and a publication. It is the core class of the system, in which all mutual relationships between

[3] http://core.edu.au/

[4] NS is the short of N-star ranking model

classes are represented by a system of linear equations. The proposed model is improved from the PageRank, which is also represented by N-star ranking model. The random and citation factor of the PageRank's are inherited in the proposed model. By adding the content factor, the proposed model supports the given problem better than the PageRank's.

We have proposed two ranks: UP rank and TP rank. We have introduced two simple ranks: (i) RCC for ranking publication based on the number of citations and (ii) RCM for ranking scientific topic based on the matching content. UP, PageRank and RCC are universal rankings in which the ranking of the result of a query is independent from any query. We give 2 predictions about the universal ranks: (i) The rank order of PageRank is not change when the citation parameter is changed and it closes to RCC's; (ii) UP is quite different from RCC, PageRank and the difference increases when the citation parameter decreases. TP and RCM are local ranks since they consider the keywords of the topic in ranking. We give a prediction that TP is the best solution among considered ranks since it is local and combining the citation and content factors.

We have proposed the metrics for ranking comparison on two criteria value and order. We have discovered the features of a given ranking based on studying the Top-N difference elements determined from the rank value comparison. The metric of rank order comparison is improved from the Kendall measure. It is based on the concordant concept. It is a very useful tool for the analysis of the difference of rank orders. The metric of related content is based on two functions: (i) Matching function for evaluating the matching of the content of a publication and the query. (ii) Traffic function for evaluating the importance of a given i-th position in the ranked result. It is the tool for measuring the related content support for a query of a rank function.

We have done the experiments to confirm our predictions and study the main features of the ranks. We found that the rank order of PageRank is not changed when citation parameter is changed. Moreover, PageRank is too closed to RCC, thus we should use RCC instead of PageRank for ranking publication. The experiments have showed that UP is quite different from RCC and PageRank. The difference will increase when the citation factor decreases. The study on Top-N difference has shown some special features of ranks. We have examined the different ranks over the results of queries. We found that TP is the best solution for ranking keyword based query from the related content viewpoint. The experiments have confirm that the proposed model and ranks are quite different from the classical ones. Their results are impressive since the content factor is considered in the model.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2014R1A2A2A05007154). Also, this work was supported by the MSIP (Ministry of Science, ICT&Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2014-H0301-14-1044) supervised by the NIPA (National ICT Industry Promotion Agency).

References

1. Agrawal S, Chaudhuri S, Das G (2002) DBXplorer: a system for keyword-based search over relational databases. In: Proceedings of the 18th International Conference on Data Engineering, February 2002, San Jose, California, pp 5–16
2. Balmin A, Hristidis V, Papakonstantinou Y (2004) Objectrank: Authority-based keyword search in databases. In: Proceedings of the Thirtieth International Conference on Very Large Data Bases, June 2004, Toronto, Canada, pp 564–575
3. Bergamaschi S, Guerra F, Simonini G (2014) Keyword search over relational databases: Issues, approaches and open challenges. In: Bridging Between Information Retrieval and Databases, LNCS, vol 8173, pp 54–73
4. Cohen S, Mamou J, Kanza Y, Sagiv Y (2003) XSEarch: A semantic search engine for XML. In: Proceedings of the 29th International Conference on Very Large Data Bases, September 2003, Berlin, Germany, pp 45–56
5. Cronin B (2001) Bibliometrics and beyond: some thoughts on web-based citation analysis. J Inf Sci 27(1):1–7
6. Egghe L (2006) Theory and practise of the G-index. Scientometrics 69:131–152
7. Fuhr N (2014) Bridging information retrieval and databases. In: Bridging Between Information Retrieval and Databases, LNCS 8173. Springer, Berlin Heidelberg, pp 97–115
8. Garfield E (1999) Journal impact factor: A brief review. Can Med Assoc J 161(8):979–980
9. Getoor L, Diehl CP (2005) Link mining: A survey. ACM SIGKDD Explor Newsl 7(2):3–12
10. Guo L, Shao F, Botev C, Shanmugasundaram J (2003) XRANK: Ranked keyword search over XML documents. In: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, June 2003, San Diego, California, pp 16–27
11. Hoang HH, Jung JJ, Tran CP (2014) Ontology-based approaches for cross-enterprise collaboration: a literature review on semantic business process management. Enterp Inf Syst 8(6):648–664
12. He H, Wang H, Yang J, Yu PS (2007) BLINKS: Ranked keyword searches on graphs. In: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, June 2007, Beijing, China, pp 305–316
13. Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad of Sci USA 572(46):16,569–16
14. Hirsch JE (2007) Does the H index have predictive power? Proc Natl Acad Sci USA 198(49):19,193–19
15. Hristidis V, Papakonstantinou Y (2002) Discover: Keyword search in relational databases. In: Proceedings of the 28th International Conference on Very Large Data Bases, August 2002, Hong Kong, China, pp 670–681
16. Hulgeri A, Nakhe C (2002) Keyword searching and browsing in databases using BANKS. In: Proceedings of the 18th International Conference on Data Engineering, pp 431–440
17. Jiang X, Sun X, Zhuge H (2012) Towards an effective and unbiased ranking of scientific literature through mutual

reinforcements. In: Proceedings of the 21st ACM Conference on Information and Knowledge Management, November 2012, Hawaii, USA, pp 714–723
18. Jiang X, Sun X, Zhuge H (2013) Graph-based algorithms for ranking researchers: not all swans are white! Scientometrics 96(3):743–759
19. Jung JJ (2011) Ubiquitous Conference Management System for Mobile Recommendation Services Based on Mobilizing Social Networks: a Case Study of u-Conference. Expert Syst Appl 38(10):12786–12790
20. Jung JJ (2012) ContextGrid: A Contextual Mashup-based Collaborative Browsing System. Inf Syst Front 14(4):953–961
21. Jung JJ (2013) Contextual Synchronization for Efficient Social Collaborations in Enterprise Computing: a Case Study on TweetPulse. Concurr Eng-Res Appl 21(3):209–216
22. Jung JJ (2014) Understanding information propagation on online social tagging systems: a case study on Flickr. Qual Quant 48(2):745–754
23. Kargar M, An A (2011) Keyword search in graphs: Finding r-cliques. Proc VLDB Endowment 4(10):681–692
24. Keener JP (1993) The Perron-Frobenius theorem and the ranking of football teams. SIAM Rev 35(1):80–93
25. Kendall M (1938) A new measure of rank correlation. Biometrika pp 81–93
26. Kien LT (2012) Information Dependency and Its Applications. Doctoral Dissertation, Faculty of Mathematics and Natural Sciences. University of Greifswald, Germany
27. Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. J ACM 46(5):604–632
28. Lowry PB, Moody GD, Gaskin J, Galletta DF, Humpherys SL, Barlow JB, Wilson DW (2013) Evaluating journal quality and the association for information systems senior scholars' journal basket via bibliometric measures: Do expert journal assessments add value? MIS Q 37(4):993–1012
29. Nelsen R (2006) An Introduction to Copulas, 2nd ed. Springer Series in Statistics
30. Nguyen DT, Jung JJ (2014) Privacy-preserving Discovery of Topic-based Events from Social Sensor Signals: An Experimental Study on Twitter. Scientific World Journal 2014:Article ID 204785
31. Nie Z, Zhang Y, Wen JR, Ma WY (2005) Object-level ranking: Bringing order to web objects. In: Proceedings of the 14th International Conference on World Wide Web, May 2005, Chiba, Japan, pp 567–574
32. Opthof T (1997) Sense and nonsense about the impact factor. Cardiovasc Res 33(1):1–7
33. Page L, Brin S, Motwani R (1999) The pagerank citation ranking: Bringing order to the web
34. Radicchi F, Fortunato S, Markines B, Vespignani A (2009) Diffusion of scientific credits and the ranking of scientists. Phys Rev E 112(5):056,103–056
35. Roa-Valverde AJ, Sicilia MA (2014) A survey of approaches for ranking on the web of data. Information Retrieval pp 1–31
36. Sánchez-Burillo E, Duch J, Gómez-Gardeñes J, Zueco D (2012) Quantum navigation and ranking in complex networks. Sci Rep 2:605–612
37. Sayyadi H, Getoor L (2009) FutureRank: ranking scientific articles by predicting their future pagerank. In: Proceedings of 2009 SIAM Conference on Data Mining, pp 533–544
38. Selvan MP, chandra Sekar A, Dharshini AP (2012) Article: Survey on web page ranking algorithms. Int J Computer Appl 41(19):1–7
39. Sharma DK, Sharma A (2010) A comparative analysis of web page ranking algorithms. Int J Comput Sci Eng 02(8):2670–2676
40. Stefanidis K, Drosou M, Pitoura E (2010) PerK: Personalized keyword search in relational databases through preferences. In: Proceedings of the 13th International Conference on Extending Database Technology, March 2010, Lausanne, Switzerland, pp 585–596
41. Sun Y, Han J, Zhao P, Yin Z, Cheng H, Wu T (2009a) RankClus: Integrating clustering with ranking for heterogeneous information network analysis. In: Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, March 2009, Saint Petersburg, Russia, pp 565–576
42. Sun Y, Yu Y, Han J (2009b) Ranking-based clustering of heterogeneous information networks with star network schema. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, June 2009. Paris, France, pp 797–806
43. Vu LA, Hoang HV, Kien LT, Hieu LT, Jung JJ (2014a) Evaluating scientific publications by n-linear ranking model. Annales University Science Budapest, Sect Comp to appear
44. Vu LA, Hoang HV, Kien LT, Hieu LT, Jung JJ (2014b) A general model for mutual ranking systems. In: Intelligent Information and Database Systems. Springer, pp 211–220
45. Wang H, Aggarwal CC (2010) A survey of algorithms for keyword search on graph data. In: Managing and Mining Graph Data, Advances in Database Systems, vol 40. Springer, pp 249–273
46. Yan E, Ding Y, Sugimoto C (2011) P-Rank: an indicator measuring prestige in heterogeneous scholarly networks. J Am Soc Inf Sci Technol 62(3):467–477
47. Yu JX, Qin L, Chang L (2009) Keyword Search in Databases. Morgan and Claypool Publishers
48. Zeng Z, Bao Z, Lee M, Ling T (2013) A semantic approach to keyword search over relational databases. In: Conceptual Modeling, LNCS, vol 8217, pp 241–254
49. Zhang CT (2009) The e-index, complementing the h-index for excess citations. PLoS ONE 4(5):1–4
50. Zhou D, Orshanskiy S, Zha H, Giles C (2007) Co-ranking authors and documents in a heterogeneous network. In: Proceedings of the Seventh IEEE International Conference on Data Mining, pp 739–744