

Time-weighted web authoritative ranking

Bundit Manaskasemsak · Arnon Rungasawang · Hayato Yamana

Received: 14 August 2009 / Accepted: 24 May 2010 / Published online: 15 June 2010
© Springer Science+Business Media, LLC 2010

Abstract We investigate temporal factors in assessing the authoritativeness of web pages. We present three different metrics related to time: age, event, and trend. These metrics measure recentness, special event occurrence, and trend in revisions, respectively. An experimental dataset is created by crawling selected web pages for a period of several months. This data is used to compare page rankings by human users with rankings computed by the standard PageRank algorithm (which does not include temporal factors) and three algorithms that incorporate temporal factors, including the Time-Weighted PageRank (*TWPR*) algorithm introduced here. Analysis of the rankings shows that all three temporal-aware algorithms produce rankings more like those of human users than does the PageRank algorithm. Of these, the *TWPR* algorithm produces rankings most similar to human users', indicating that all three temporal factors are relevant in page ranking. In addition, analysis of parameter values used to weight the three temporal factors reveals that age factor has the most impact on page rankings, while trend and event factors have the second and the least impact. Proper weighting of the three factors in *TWPR* algorithm provides the best ranking results.

Keywords Web ranking algorithm · Time-weighted ranking · Web authoritativeness · PageRank · Link analysis · Search engine

1 Introduction

Web search engines are critical to enabling users to find the most useful resources for their specific interest, one of the goals being to bring the most relevant web pages to the top

B. Manaskasemsak (✉) · A. Rungasawang
Department of Computer Engineering, Kasetsart University, Bangkok, Thailand
e-mail: un@mikelab.net

A. Rungasawang
e-mail: arnon@mikelab.net

H. Yamana
Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo, Japan

ranked list for a given query. Most search engines include a ranking algorithm that computes a page's authoritativeness based on either the link structure of the Web, such as HITS (Kleinberg 1999) and PageRank (Page et al. 1999; Brin and Page 1998), or mining of users' browse histories, such as BrowseRank (Liu et al. 2008), Traffic-weighted Ranking (Meiss et al. 2008), and BookRank (Gonçalves et al. 2009). While all of these algorithms produce useful rankings, overall, they do not incorporate temporal information of pages which can be important to users' interest in a particular page. Indeed, these algorithms may be biased against more recent pages (Adamic and Huberman 2001; Cho and Roy 2004), as such pages have had less time to accumulate in-links that contribute to their link-based ranking, and less chance to be incorporated into a browse history.

In many situations a web user may be more interested in recent or actively updated resources rather than older ones, even though the latter may appear to be more authoritative. A web page for an upcoming conference is clearly more interesting than a page for a past conference; a link to the current version of an open-source software is probably more useful than a link to an old one. Indeed, one can argue that search engines are most needed for finding newer, more dynamic content; such items are less likely to be findable via web directories, the Wikipedia, or other resources that a human user can turn to for help.

In this paper, we investigate the value of incorporating temporal aspects into the authority ranking of web pages. Three temporal factors are examined: *age*, *event*, and *trend*; these three capture the intuitive notion that a page with recent updates, special time occurrence, or trend in revisions is potentially more interesting to users. These attributes are computed from page timestamps (directly or by inference) and a modification history that is stored during each web crawl; the event and trend factors require categorizing pages and comparing the temporal history of pages in the same category, as described in Sect. 4. We discuss other algorithms that use temporal information in Sect. 3, notably *T-Rank* (Berberich et al. 2006) and *TimedPR* (Yu et al. 2005), which are included in our experimental study.

We conducted experiments on selected-topic web data crawled daily for several months to study the effectiveness of the proposed temporal factors in page ranking. We compared PageRank, *T-Rank*, *TimedPR*, and an experimental algorithm by examining following points: the time evolution effect in changing authoritativeness of web pages and a user study in order to evaluate the quality of ranked results. For the user study, search results returned by a given query are separately re-ranked according to authoritative scores produced by those algorithms and agreement on order asserted by a number of users. We judged the quality by comparison of the rankings obtained from each algorithm to the users' rankings. Finally, we performed sensitivity analysis on parameter values used to weight the three temporal factors.

The rest of this paper is organized as follows. We briefly describe the PageRank algorithm in Sect. 2, and review some related work in Sect. 3. In Sect. 4, we define measures of temporal factors, and then describe how to incorporate them in a temporal-aware modification of PageRank. In Sect. 5, we report and discuss the results of experiments. In Sect. 6, we suggest some future research directions. Finally, conclusions are given in Sect. 7.

2 Review of PageRank

The basic definition of PageRank (Page et al. 1999) can be stated as follows. For each crawl of the Web, we can construct a directed graph $G = (V, E)$, with pages as the set of

nodes V and their hyperlinks as the edges $E \subset V^2$, without multiple edges. If page $u \in V$ has a link to page $v \in V$, it implies that the author of page u implicitly confers some importance to page v . Let $r(u)$ be the PageRank score of page u and $t(u, v)$ be the proportion of importance propagated from u to v . This value is normally set to $1/out(u)$, where $out(u)$ is the out-degree of page u in G . Therefore, a link $(u, v) \in E$ confers $r(u)/out(u)$ units of rank to page v . This yields the following equation to calculate the PageRank score for all web pages:

$$\begin{aligned} \forall v \in V, r^{(k+1)}(v) &= \sum_{u \in B(v)} t(u, v)r^{(k)}(u) \\ &= \sum_{u \in B(v)} \frac{r^{(k)}(u)}{out(u)}, \end{aligned} \tag{1}$$

where $B(v)$ represents the set of pages linking to v . The total amount of score conferred on v is the summation of the score of each source page u divided by its out-degree. This recursive definition requires an iterative computation, which repeats until the rank scores are stable. In other words, the computation will stop when the size of the residual vector is less than a pre-defined threshold δ :

$$\|\mathbf{r}^{(k+1)} - \mathbf{r}^{(k)}\|_1 < \delta. \tag{2}$$

This process is equivalent to a random walk on the directed graph G . A random surfer visits a page u at time k then randomly chooses and follows a link from among u 's out-links to page v —with the probability $1/out(u)$, also illustrated by $t(u, v)$ in Eq. 1. Consequently, the PageRank score of a page v can be defined as the probability of a random surfer landing on v at a sufficiently large time step. Let n be the total number of pages contained in G , and M be the $(n \times n)$ -square matrix describing the transition over G . The matrix entry m_{ij} is thus assigned as $1/out(i)$ if there is a link from page i to page j , and the remainder are set to 0. From the power iteration of the matrix-vector multiplication of an eigensystem (Golub and Loan 1996), the PageRank vector (\mathbf{r}) can be considered as the principal eigenvector of M^T corresponding to eigenvalue 1:

$$\mathbf{r} = M^T \times \mathbf{r}. \tag{3}$$

Consider the Markov chain induced by a random walk on the web graph G , the transition probability matrix M is said to be valid if it is a row-stochastic matrix—its row totals equal 1. Hence, M must have no row consisting of all zeros. A general approach for dealing with dangling pages¹ is to add a set of artificial links from those pages to all pages. Let M' be the $(n \times n)$ -square matrix with all entries in each row corresponding to a dangling page are given by $1/n$; otherwise, they are 0. Then, the modified PageRank is:

$$\mathbf{r} = (M^T + M'^T) \times \mathbf{r}. \tag{4}$$

Moreover, the Ergodic Theorem for Markov chains (Grimmett and Stirzaker 2001; Ross 2002) guarantees that the PageRank formula will converge to a stationary probability distribution if M is also aperiodic and irreducible. These requirements hold for a proper Markov model; the former is in regard to the surfer possibly returns to a visited page in any

¹ The pages with zero out-degree cause the PageRank leak problem (Page et al. 1999).

time step, while the latter is in regard to every page directly links to every other page. In fact, the web graph often contains sinks or isolated clusters.² This problem can be solved by adding a complete teleportation term to every page. This solution is reasonable for the random surfer analogy: the random surfer, who follows links on the Web, at some time step abandons link surfing and instead enters a new destination in the browser's URL line. Let E be the $(n \times n)$ -square matrix representing the teleportation over the web graph G in which all entries have value $1/n$. Then, the modification of PageRank can be expressed as:

$$\begin{aligned} \mathbf{r} &= (\alpha(M^T + M'^T) + (1 - \alpha)E^T) \times \mathbf{r} \\ &= \alpha(M^T + M'^T) \times \mathbf{r} + (1 - \alpha)E^T \times \mathbf{r} \\ &= \alpha(M^T + M'^T) \times \mathbf{r} + (1 - \alpha) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1}. \end{aligned} \quad (5)$$

Here, the coefficient α is used to determine the weights of rank propagation and teleportation. Since the total probability $\|\mathbf{r}\|_1 = 1$, the multiplication $E^T \times \mathbf{r}$ then returns a uniform n -dimensional column vector, i.e., the total probability of random jumps to each page is equal to $1/n$.

3 Related work

Since the PageRank algorithm was published, it has been applied in many commercial Web search engines and applications. In reality, however, the Web is dynamic; considering only the number of page referrers for assessing authoritativeness is incomplete and inaccurate. PageRank tends to prefer older to newer pages since the older ones tend to have accumulated many in-links over time. Cho and Roy (2004) describe this “rich-get-richer” phenomenon where search engines usually report higher-PageRank pages at the top of results. As a result, these pages receive more and more references from other Web's authors over time, while new-born pages are continually left behind. However, the old pages may be outdated; the new pages may be more valuable. Cho et al. (2005) propose a model to estimate quality of pages by analyzing the growth of their popularity—how many references they receive over time. A page with fast growing trend should be considered as potentially valuable, even though it is still unpopular.

There are also some studies that concentrate on integrating temporal aspects into the PageRank algorithm. Baeza-Yates et al. (2002) assume that a page should be considered good if it has a recent modification time, and already has other pages linked to it. They thus propose a modified PageRank by weighting pages with an exponential decay function according to their age, to counteract bias against new pages.

Yu et al. (2005) propose another temporal modification of PageRank in the context of scientific publications, called TimedPageRank (*TimedPR*). The key concept is to weight each citation by an exponential decay function according to publication time of the papers. However, their technique cannot be directly applied to the Web. Although the citation concept of scientific papers is quite similar to web hyperlinks, other aspects differ from general web content. For instance, scientific papers have static information fixed at publication time. Since articles cannot be deleted, their citation counts are monotonically increasing. In contrast, web pages can be modified and deleted both of content and links

² The small self-loop cluster, having no connection back to main community, causes the PageRank sink problem (Page et al. 1999).

over time. Scientific papers can only refer to others further in the past; there are obviously no citation loops. On the Web, the link structure contains many loops and references to pages that are subsequently modified—becoming newer than their referrers and possibly providing higher quality information. This factor affects the authoritativeness and should be considered.

Berberich et al. (2006) specify a temporal window of interest to analyze freshness and activity of both pages and links. They incorporate these two factors into PageRank, called Time-Aware Authority Ranking (*T-Rank*), by weighting the page transition and random jump probabilities. However, their proposal utilized a dataset on one specific web domain; it is unclear how one can determine the temporal window of interest for the Web which contains several domains. Their results may not be representative since each domain usually has its own change behavior. For example, news pages having daily update should not dominate other pages. Therefore, in dealing with the Web, the Web's topics also should be considered.

Recently, several researchers, such as Burges et al. (2005), Richardson et al. (2006), Xu and Li (2007), Geng et al. (2008), and Liu et al. (2010), have applied machine learning techniques to train the ranking model using queries previously submitted to the search engine and relevance information on retrieved results derived from user's browse behavior, to improve ranking quality. The success of this "learning to rank" approaches depends on both query and document information, including their relevancy derived from users, as well as parameter tuning of the machine learning techniques, in contrast to a simple link-based ranking process investigated in this study.

4 Temporal aspects

In this section, we first describe quantitative measures of three temporal factors. To evaluate the usefulness of these factors, we then present an experimental algorithm that incorporates all three factors into the standard PageRank algorithm. Finally, we briefly provide the detail of the other two temporal-based page ranking algorithms.

4.1 Temporal factors

The Web is continually changing; web pages are created, modified, or deleted over time. Based on the assumptions: (1) authors modify their web pages to include higher quality and more current information, (2) they update their web pages for special occurrences as well as evolutionary trends, and (3) users often need or are interested in up-to-date information, we can infer some authoritativeness of pages by their change behavior. We investigate the first two factors (age and event) by examining the last modification time of pages, and the last one (trend) by analyzing the pattern of modification times.

4.1.1 Age and inverse-age factors

We define the age factor to describe how old a web page is. This factor does not mean an actual age of a web page since it was created, but rather the duration since last modification. A recently updated page is assumed to be of higher quality than an older one.

Let TS_{Start} and TS_{End} be global constants that define the start- and end-time for which to observe changes in web pages, with $TS_{Start} < TS_{End}$. Let $TS_{Mod}(u)$ denote a set of modification times of page u , and $ts_{Last}(u) \in TS_{Mod}(u)$ be the last modification time. The age factor of a page u , $AF(u)$, is simply defined in a linear time-scale as:

$$AF(u) = \begin{cases} 1 & \text{if } ts_{Last}(u) \leq TS_{Start}, \\ \frac{TS_{End} - ts_{Last}(u)}{TS_{End} - TS_{Start}} & \text{if } TS_{Start} < ts_{Last}(u) \leq TS_{End}, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

This age function is bound by $[0, 1]$. Given the observation period TS_{Start} through TS_{End} , the age factor of page u has the value 1 if it has never changed since TS_{Start} , linearly decreasing to 0 for pages with $ts_{Last}(u) \in (TS_{Start}, TS_{End}]$.

In order to assign more importance to recently modified pages, we complement $AF(u)$ to create the inverse-age factor $IAF(u)$ defined as:

$$IAF(u) = 1 - AF(u). \quad (7)$$

4.1.2 Event factor

When an event occurs, some web pages will be modified to provide up-to-date information about the event. The more important the event is, the more web pages will be changed. Observing the amount of change, we can infer not only that an event occurs but also how important it is. However, we would expect to observe event-related updates primarily in web pages related to the same topic as the event. Pages related to “sports” would likely be updated for the Olympic Games; pages related to “entertainment” may be updated for a new movie, for instance. In this study, we consider the last event that causes the change of each web page, i.e., the event factor is analyzed by considering the last-modified time of a page based on its topic.

Let $C = \{C_1, C_2, \dots, C_p\}$ be a set of categories or topics containing non-overlapping pages, i.e., if a page $u \in C_i$ then $u \notin C_j$ for all $j \neq i$. Let $I = \{I_1, I_2, \dots, I_q\}$ be a set of time intervals such that $I_i = (TS_{i-1}, TS_i]$ for $1 \leq i \leq q$, where TS_i is a pre-defined timestamp with $TS_{Start} = TS_0 < TS_1 < \dots < TS_{q-1} < TS_q = TS_{End}$. Let $page(C_i, I_j)$ be the number of web pages that are categorized in C_i and have last-modified time in I_j . Then, the event factor $EF(u)$ of a page u is defined as:

$$EF(u) = \begin{cases} \frac{page(C_i, I_j)}{\sum_{k=1}^q page(C_i, I_k)} & \text{if } u \in C_i \text{ and } ts_{Last}(u) \in I_j, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

That is, the event factor of page u is the proportion of pages in the same category as u that last changed in the same time interval. A large event value, i.e., a large fraction of web pages changed, can infer that an important event occurred. Note that some web pages may not be classified into any categories, or they may not be updated over the observation period. In this case, we reasonably assign a zero value to them.

4.1.3 Trend factor

The age and event factors only consider the last modification time of web pages. We hypothesize that the pattern of page updates in the past may also have significance in estimating a page’s importance. A page that is modified in accordance with a particular or general direction (e.g., business trend, fashion trend, etc.) reflects activeness in itself. However, considering only trend in revisions of such a page without examining its kind of topic may be biased and unfair to those pages which have already presented either correct or reliable information. For instance, a news page will likely be updated daily, which differs extremely from one containing historical information.

A trend factor describes how similar the change behavior of a web page is to others in the same category, based on the hypothesis that authoritative pages would behave in a similar manner. We first introduce an “update profile” of each page based on the changes over the observation period. Let $\mathbf{x}(u)$ be the m -dimensional vector representing the update profile of page u , where m is the number of time units that u has been observed. Each entry $x_i(u)$ corresponding to the i -th timestamp, $ts_i(u)$, is defined as:

$$x_i(u) = \begin{cases} 1 & \text{if } ts_i(u) \in TS_{Mod}(u), \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

However, directly comparing page profile of u with different pages will produce great differences since there may be less possibility that web pages will be updated at the same specific timestamps during an observation period. We thus reduce the dimension of the profile from m to r by defining another set of time intervals $I' = \{I'_1, I'_2, \dots, I'_r\}$ such that $I'_i = (TS'_{i-1}, TS'_i]$ for $1 \leq i \leq r$, where TS'_i is a pre-defined timestamp, $TS_{Start} = TS'_0 < TS'_1 < \dots < TS'_{r-1} < TS'_r = TS_{End}$, and an r -dimensional vector $\mathbf{x}'(u)$ to represent a new update profile of u , where

$$x'_i(u) = \sum_{\forall j: ts_j(u) \in I'_i} x_j(u). \tag{10}$$

Similarly, we let $\mathbf{x}'(C_j)$ be the r -dimensional vector representing the update profile of a category, where $C_j \in C$ and

$$x'_i(C_j) = \frac{\sum_{u \in C_j} x'_i(u)}{|C_j|}. \tag{11}$$

That is, $\mathbf{x}'(C_j)$ refers to the average update profile of all pages contained in the same category. Finally, the trend factor $TF(u)$ of a page u is defined as:

$$TF(u) = \begin{cases} Sim(\mathbf{x}'(u), \mathbf{x}'(C_j)) & \text{if } u \in C_j, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Here, $Sim(\cdot, \cdot)$ can simply be the cosine similarity function used in the classical IR (Baeza-Yates and Ribeiro-Neto 1998). The trend factor of u is thus considered as the similarity of its update profile to the profile of its category. For those web pages that may not be classified into any categories, we assign a zero value to them.

4.2 Time-weighted ranking

In this section, we describe a new ranking function that incorporates the inverse-age, event, and trend factors into PageRank. Let D be the set of dangling pages in the web graph G . Let $t(u, v)$ be the weight for a transition from page u to page v , and $s(v)$ be the weight for a random jump from any page to v . Then, the PageRank formula in Eq. 5 can be rewritten in a general form as follows:

$$\forall v \in V, r(v) = \alpha \sum_{u \in (B(v) \cup D)} t(u, v)r(u) + (1 - \alpha)s(v). \tag{13}$$

Note that in the original PageRank, $t(u, v)$ is set to either $1/out(u)$ if $u \in B(v)$ or $1/n$ if $u \in D$, and $s(v)$ is set to $1/n$. Some studies have suggested biasing the PageRank transition matrix (Richardson and Domingos 2002) and jump vector (Haveliwalla 2003). In this study, our purpose is primarily to investigate the contribution of temporal information (i.e.,

inverse-age, event, and trend) to page ranking by weighting in page transitions, and subsequently to reduce the effect of bias against new-born pages by weighting page jumps with only the inverse-age information.

4.2.1 Time-weighted transition

We introduce a transition probability based on the three temporal values of the target page. Each value affects its authoritativeness. For example, a target page that is recently modified but not updated for events or trend may still have importance based on its inverse-age value.

For each page $u \in V$, let $F(u)$ be the set of web pages that a page u connects to by either its actual links or artificial links (in case u is a dangling page). Then, the time-weighted transition is

$$t(u, v) = \frac{\omega_1 IAF(v) + \omega_2 EF(v) + \omega_3 TF(v)}{\sum_{w \in F(u)} (\omega_1 IAF(w) + \omega_2 EF(w) + \omega_3 TF(w))} \quad (14)$$

with coefficients $\omega_1, \omega_2, \omega_3 \in [0, 1]$. Equation 14 defines the rank propagation from page u to target page v based on inverse-age, event, and trend values of v . In other words, authoritativeness of the target page is determined by its up-to-dateness, event importance, as well as trend in revisions, in comparison to all targets.

4.2.2 Age-biased vector

In terms of the random-walk model, a complete set of transition edges is artificially added to cover all web pages with respect to the Ergodic Theorem (Grimmett and Stirzaker 2001; Ross 2002). In Eq. 5 these artificial transitions were assigned a uniform probability distribution $\left[\frac{1}{n}\right]_{n \times 1}$. To reduce the PageRank bias against new pages, we replace the uniform vector by a vector s where the entry for each page v is its normalized inverse-age:

$$s(v) = \frac{IAF(v) + \epsilon}{\sum_{w \in V} (IAF(w) + \epsilon)}. \quad (15)$$

Equation 15 determines the probability in jumping to v based on a bias of page age. Note that a positive minimal value ϵ is also assigned to the entire pages to follow the ergodicity for the Markov chains (Grimmett and Stirzaker 2001; Ross 2002).

4.3 Temporal-based page ranking algorithms

Three temporal-aware algorithms, i.e., our proposed *TWPR*, *T-Rank* (Berberich et al. 2006), and *TimedPR* (Yu et al. 2005), considered in this paper are the variants of the temporal based PageRank algorithms. *TWPR* incorporates three temporal factors: age, event, and trend, in page ranking. It also needs category information to compute event and trend factors. The age factor considers how much a page is up-to-date; the event factor considers the group of pages that is updated according to the event; and finally the trend factor considers the change of a page following those in the same group.

T-Rank incorporates two factors: freshness and activity, in its ranking calculation. The freshness of both page and link is assigned as the highest freshness among the values returned by a pre-defined freshness function corresponding to their modification times. The activity of both page and link, indicating the frequency of change, is defined as the sum of

their freshness values. For *TimedPR*, it incorporates only the age factor, which is defined by a function of decay rate corresponding to the timestamp of a page, in its ranking calculation.

At the methodological level, *TWPR* needs to detect page changes over a time period. The modification time of a page is used in the age function; the category information of that page is also needed in event and trend functions. All the three factors are then combined, normalized, and employed in weighting the page transitions of the standard PageRank algorithm. Moreover, only the normalized inverse-age value is employed in weighting the random jumps. Similarly, *T-Rank* needs to detect both page and its links' changes over a time period. It then weights both page transitions and random jumps using both normalized freshness and activity values of that page and its links. Finally, *TimedPR* observes page changes and uses the last modification time to calculate the age value. This age value is then employed in weighting the page transitions.

5 Experimental results

We implemented four ranking algorithms: *PR* (Page et al. 1999), *TWPR* (new), *T-Rank* (Berberich et al. 2006), and *TimedPR* (Yu et al. 2005), using the Java programming language, and integrated them into our prototype Lucene-based searching system (Lucene 2009) for comparison of the quality of ranking results.

5.1 Dataset

As the source of web data, we selected a portion of the set of web pages from the e-Society Project (Muraoka and Yamana 2005). To study the evolution of web pages, we daily downloaded them from August 15th, 2008 until January 31st, 2009. Each snapshot contains roughly 410,000 pages, and there are 390,717 pages (i.e., URLs) appearing in common for every snapshot. The modification time of pages was detected by comparing consecutive versions, but ignoring some special cases (e.g., a visit counter, date, time, etc.) that change automatically when users visit a page. As described in Sects. 4.1.2 and 4.1.3, *TWPR* needs category information, we then simply used the category data obtained from the Open Directory project (ODP) (DMOZ 2009). There are around 96% of pages in every snapshot having their categories labeled by the ODP. Since the ODP provides hierarchical topics indexed from most general to most specific, a page can then be categorized into more than one topic according to a defined level. For example, a page in “Computers/Internet/Searching/Directories/...” can be labeled as “Computers”, “Computers/Internet”, and “Computers/Internet/Searching” corresponding to the 1st, 2nd, and 3rd topic levels, respectively.

For the experiments, each snapshot was translated into a graph structure for computing authoritative scores; however, only the last one was indexed and retrieved by our searching system. We conducted the experiments using a set of 30 sample queries, in which some of them come from Dwork et al. (2001) and Haveliwala (2003); the remainder were randomly chosen from the set of words appearing in the title field of web pages in the dataset. To study the effect of the three temporal factors in page ranking, we asked five web researchers in the Computer Engineering Department, Kasetsart University, Thailand, to participate in the experiments, and fill out the questionnaire—in which class among “event”, “trend”, “history/fact”, and “others” you would like to classify each query after seeing the corresponding search results? Each researcher was then free to examine the

Table 1 Queries used in the experiments

Event	Trend	History/Fact	Others
christmas	blues	affirmative action	amusement park
disaster	business	architecture	music
earthquake	cruise	astronomy	rock climbing
film festival	fashion	lyme disease	soccer
hurricane	game	parallel computing	sushi
olympic	graphic design	risk management	
presidential election	innovation	socialism	
science conference	marketing		
tournament	tattoo		

resulting pages ranked by our *tf-idf* (Baeza-Yates and Ribeiro-Neto 1998) weight-based search prototype. The queries were finally grouped in Table 1 by majority vote.

5.2 Evaluation measures

In this section, we detail the three measures used in this paper. The *OSim* and *KSim* metrics, proposed by Haveliwala (2003), measure the similarity of any two ranked lists. $OSim_k(\tau_1, \tau_2)$ determines the degree of overlap between the top- k documents of two rankings τ_1 and τ_2 .

$$OSim_k(\tau_1, \tau_2) = \frac{|R_1 \cap R_2|}{k}, \quad (16)$$

where R_1 and R_2 are the lists of top- k documents contained in τ_1 and τ_2 , respectively.

$KSim_k(\tau_1, \tau_2)$, a variant of Kendall's τ distance measure, determines the degree of agreement in which pairwise distinct documents u and v within top- k rank have the same relative order in both rankings τ_1 and τ_2 . Consider two lists R_1 and R_2 of top- k document rankings. Let U be the union of documents contained in both lists and define R'_1 as the extension of R_1 to add the elements $U - R_1$ after all the documents in R_1 . Similarly, R'_2 is also defined as the extension of R_2 . *KSim* is then given as:

$$KSim_k(\tau_1, \tau_2) = \frac{|\{(u, v) : R'_1, R'_2 \text{ agree on order of } (u, v) \text{ and } u \neq v\}|}{|U| \times (|U| - 1)}. \quad (17)$$

As in most Kendall-based measures, *KSim* also suffers from disagreements that occur at any part of the long ranked list (Melucci 2007). However, the main goal of the experiments is to use *KSim* to simply measure the similarity of two short-list rankings, i.e., the first 5, 10, or 20 results that most web searchers look for. Further, a higher *KSim* value implies that the former pair of rankings are more similar than the latter.

The normalized discounted cumulative gain (*NDCG*), proposed by Järvelin and Kekäläinen (2000, 2002), measures ranking accuracies when there are multiple levels of relevance judgement. Given a query and a ranking τ , *NDCG* computed at top- k documents is given as:

$$NDCG_k(\tau) = Z_k \sum_{i=1}^k \frac{2^{r(i)} - 1}{\log_2(i + 1)}, \quad (18)$$

where $r(i)$ is the gain value associated with the document at the i -th position of the ranking and Z_k is a normalization factor to guarantee that a perfect of this top- k ranking will receive the score of one. Note that the gain value of a document is usually defined as a relevant score; however, in our experimental evaluation we define it as a users' preference score when seeing that document corresponding to a given query.

5.3 Results and discussions

We conducted experiments on our crawled web data to study the page authority assessment of the *PR*, *TWPR*, *T-Rank*, and *TimedPR* methods. In the following, we first discuss the time evolution of authoritativeness. Then, we describe results of a user study to evaluate the quality of ranked results obtained by a pre-defined set of queries listed in Table 1.

To assess the authoritativeness of web pages in each crawling snapshot, we first set parameters used in the computation of *TWPR* as follows: TS_{Start} = the date of the first crawl (August 15th, 2008), TS_{End} = the date of the last crawled version under consideration, C = a set of the 3rd level categories of ODP, I = a set of all 5-day intervals starting from the first crawl, the coefficients $\omega_1 = \omega_2 = \omega_3 = 1/3$, and a small constant $\epsilon = 10^{-10}$. We simply set the interval I' to be the same as I in all experiments. For the *T-Rank* method (Berberich et al. 2006), the temporal window of interest was set to cover the period from TS_{Start} to TS_{End} , which are the same values defined for *TWPR*. For all of other parameters, including for the *PR* (Page et al. 1999) and *TimedPR* (Yu et al. 2005) methods, we used the default values as defined by the authors. In addition, we fix the value of the parameter $\alpha = 0.85$ (Page et al. 1999; Haveliwala 1999) and $\delta = 10^{-10}$.

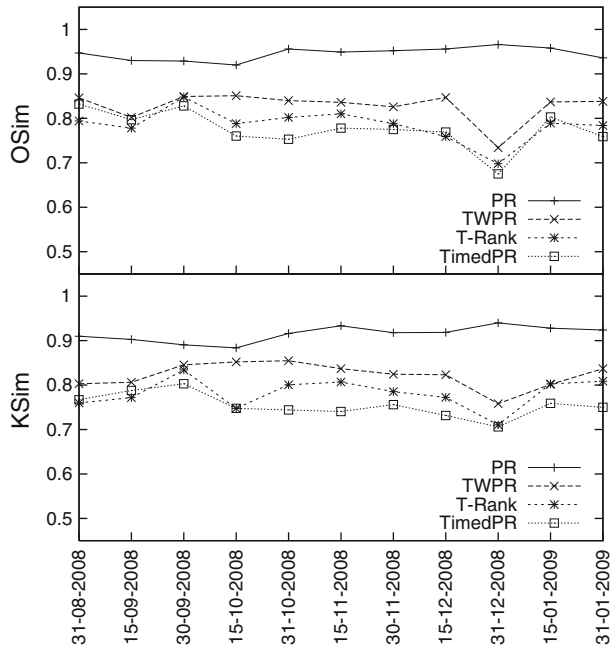
For a crawling snapshot, we first computed an authoritative score for each web page according to a ranking method, and then sorted them from highest to lowest to produce a ranked list. To investigate the time evolution effect (i.e., how the authoritativeness changes over time), we compared the lists of top authoritative web pages obtained from those four ranking methods for several consecutive fortnights. Figure 1 shows the evolution of top-1000 authorities in terms of *OSim* and *KSim*. Note that the observation period starts from August 15th, 2008 until January 31st, 2009. Hence, each point on the graph is the similarity between the ranking at that time and that of the previous fortnight.

As shown in the figure, *PR* yields *OSim* values of 0.92–0.97 (average 0.95) and *KSim* values of 0.88–0.94 (average 0.91). This implies that *PR* produces nearly the same rankings over the entire observation period on the experimental dataset. In contrast, the consecutive rankings of *TWPR*, *T-Rank*, and *TimedPR* have lower similarities because of their temporal approach. We also see that *TWPR* produces more similar rankings than *T-Rank* and *TimedPR*, respectively. This phenomenon may come from the effect of both event and trend factors in *TWPR* that allow some authoritative web pages in the past to be ranked in the top authorities. In addition, all three temporal-aware ranking methods show the lowest consecutive-ranking similarities at 31-12-2008, due to several factors; one of which may be the New Year event.

To evaluate the quality of page rankings of those methods with respect to human users' notion of pages' preference, we conducted a user study involving twenty-seven web search researchers in the Computer Science and Engineering Department of Waseda University, Japan. In the study, all ranking methods were integrated into the search prototype. We employed the last snapshot (i.e., January 31st, 2009) and the 30 sample queries in Table 1 in the experiments.

To simulate a normal searching situation, all users had been briefly told the objective of the research, but not known in advance the algorithm used to rank the results. Since we

Fig. 1 Similarities between lists of top-1000 authoritative web pages produced by the same ranking methods for consecutive fortnights



interest in studying the effect of temporal factors on the page ranking, we simply assume that a resulting page is relevant to a given query if it contains the query term(s). For a given query, we then separated the top-30 web pages from the *tf-idf* weight (Baeza-Yates and Ribeiro-Neto 1998) as the result, and asked those twenty-seven users to grade them. Each user can freely assign a score to each resulting page from 0 (unsatisfactory) to 3 (most satisfactory) based on his/her own preference or perspective, although some queries seem ambiguous.

We finally aggregated the users' agreement assessments and sorted them from highest to lowest to create a baseline ranking for comparisons. In any case, when some web results have equal aggregated scores, we simply ordered them by their original *tf-idf* scores. We chose the queries "christmas", "fashion", and "astronomy", to extensively study the effect of the event, trend, and age factors, respectively. Table 2 concludes the users' ranking agreement of those three queries, and the average ones for all queries listed in the Table 1, in term of the Fleiss' kappa (Fleiss 1971). According to the guideline proposed by Landis and Koch (1977), all these values are quite significant, indicating that the baseline rankings by this set of human users are in substantial agreement.

We now discuss the results of the comparisons between the baseline ranking and the ranking of the methods for the selected queries relating to event ("christmas"), trend ("fashion"), and fact ("astronomy"), respectively. We illustrate the rank similarities and average in-degrees per web page of the top-10 results in Fig. 2 and Table 3. In Fig. 2, it is clearly seen that for the "christmas" and "fashion" queries, all temporal-aware methods present much higher *OSim* similarity to the users' ranking than does the *PR* method (the *KSim* is higher as well). This probably comes from the fact that the temporal-aware methods rank web pages which contain the recent Christmas event as well as new fashion trend. Supported by Table 3, the *PR* method more highly ranks web pages with high in-degrees, while the others boost the more recent or active ones that have lower

Table 2 Users’ ranking agreement in term of the Fleiss’ kappa

Query	Top-5	Top-10	Top-20
christmas	0.943	0.775	0.743
fashion	0.857	0.743	0.735
astronomy	0.786	0.733	0.681
Event	0.881	0.786	0.696
Trend	0.814	0.740	0.660
History/Fact	0.861	0.754	0.687
Others	0.834	0.748	0.664

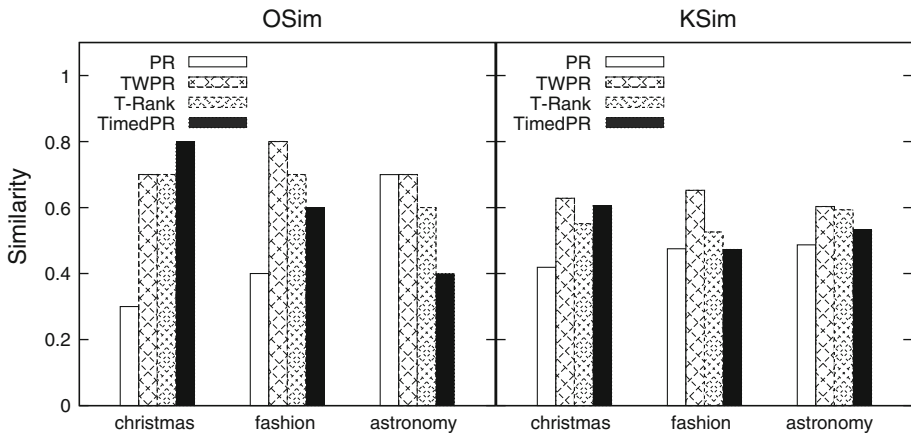


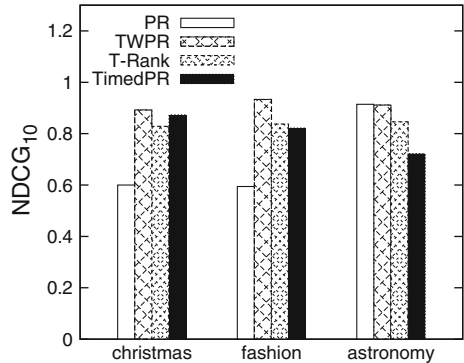
Fig. 2 Similarities at top-10 results between the users’ rankings and the rankings produced by the four methods for the queries “christmas”, “fashion”, and “astronomy”, respectively

Table 3 Average in-degrees per web page for top-10 results produced by the users and the ranking methods

Query	Users	PR	TWPR	T-Rank	TimedPR
christmas	24	58	23	26	20
fashion	31	46	29	27	17
astronomy	49	63	42	30	24

in-degrees in the top of rankings. We also observe that the average in-degrees resulted from TWPR are closest to the users’ rankings. In addition, for the “christmas” query, TimedPR shows the highest OSim since it prefers most recent pages, while TWPR agrees on some important pages that occurred in the past and were also affected by the event and trend factors. However, for the “fashion” query, TWPR shows the highest OSim since the users expect the new fashion trend continually acting from the past. The NDCG scores for the top-10 results corresponding to these three queries, produced by the four methods are given in Fig. 3. Note that the function $r(i)$ in Eq. 18 refers to the average users’ assessment scores associated with the result page at the i -th position in the ranking. Similarly, all three temporal-aware methods can produce higher NDCG score than does the PR method for the “christmas” and “fashion” queries.

Fig. 3 *NDCG* scores at top-10 of rankings produced by the four methods for the queries “christmas”, “fashion”, and “astronomy”, respectively



We also investigated how well the temporal-aware methods can induce web pages relating to fact. The results of the “astronomy” query, shown in Fig. 2, show that *TWPR* still results the highest *OSim* (and *KSim* as well), equal to that of the *PR* method. It is not surprising that *PR* yields much higher *OSim* than those of the previous two queries, and also higher than *T-Rank* and *TimedPR*, respectively. Likewise, for the *NDCG* scores shown in Fig. 3, *PR* gives the highest value (that of *TWPR* is almost equal to that of *PR* as well). It is possible that the users expect web pages to contain some truthful information. Also shown in Table 3, the users, the *PR*, and *TWPR* methods prefer web pages with high in-degrees. To further investigate this preference, the top-10 results obtained from the users and the ranking methods are shown in Tables 4, 5, and 6, respectively.

As shown in the tables, the *PR* method produces seven of ten web pages overlapping with the users’ ranking, while the temporal-aware methods, i.e., *TWPR*, *T-Rank*, and *TimedPR*, produce overlaps of seven, six, and four, respectively, that yield the *OSim* values given in Fig. 2. We manually examined each of these results and observed that there are four unchanged (static) web pages contained in the users’ ranking (rank numbers 4, 7, 8, and 10). The first one is a web portal while the latter three are online articles. As expected, the *PR* method produces the highest number of static pages (rank numbers 1, 4, 5, 7, and 8); however, four of them overlap with the users’ ranking. *TWPR* agrees on two static pages at rank numbers 9 and 10 that have very high in-degrees. These two pages overlap with the users’ ranking, too. Moreover, we observe that the page at rank number 2 contains several sub-topics of science articles; some part mentions about astronomy. Although this page appears to have low relevancy for the “astronomy” query, it is recently updated and also active. So it may still be interesting to users. Finally, the other two temporal-aware methods produce all of the new web pages; especially, *TimedPR* captures most recently updated ones. In addition, their results contain the same two low relevance pages at rank numbers 1 and 7 as obtained from *T-Rank*, and rank numbers 1 and 8 as obtained from *TimedPR*, respectively.

To present the overall quality of rankings, for every query in the same list, we summed up and normalized the scores obtained from the comparisons between the baseline ranking (i.e., users’ ranking) and the ranking of that method. We illustrate here the average *OSim*, *KSim*, and *NDCG* scores for the queries corresponding to pages related to “Event”, “Trend”, “History/Fact”, and “Others” for the top-5, 10, and 20 results in Fig. 4. As expected, for “Event” and “Trend” related web pages all three temporal-aware methods provide much higher *OSim*, *KSim*, and *NDCG* to users’ rankings than does PageRank. The similarities *OSim* and *KSim* increase as more results are included (i.e., from top-5 to top-20);

Table 4 Top-10 results of the “astronomy” query obtained from *Users* and produced by *PR*Top-10 results obtained from *Users*

- 1 NASA - Home
<http://www.nasa.gov>
- 2 Astronomy Now Online
<http://www.astronomynow.com>
- 3 NASA - NASA News and Features
<http://www.nasa.gov/news>
- 4 Astronomy Picture of the Day’s Educational Links
<http://antwrp.gsfc.nasa.gov/apod/lib/edlinks.html>
- 5 Astronomy Picture of the Day Archive
<http://apod.nasa.gov/apod/archivepix.html>
- 6 Physics and Astronomy Classification Scheme
<http://publish.aps.org/PACS>
- 7 Public Education and Outreach - Center for Astrophysics & Space Sciences
<http://cass.ucsd.edu/public/astroed.html>
- 8 Astronomy: Stars and Planets
<http://jumk.de/astronomie/astronomy.shtml>
- 9 Astronomy articles, forums, listings, and other resources about Astronomy
<http://www.astronomy.net>
- 10 astronomy 18th century
<http://www.cosmicelk.net/astronomy18c.htm>

Top-10 results produced by *PR*

- 1 Public Education and Outreach - Center for Astrophysics & Space Sciences
<http://cass.ucsd.edu/public/astroed.html>
- 2 NASA - Home
<http://www.nasa.gov>
- 3 Physics and Astronomy Classification Scheme
<http://publish.aps.org/PACS>
- 4 Institute for Astronomy - Department of Physics - ETH
<http://www.astro.phys.ethz.ch>
- 5 Astronomy: Stars and Planets
<http://jumk.de/astronomie/astronomy.shtml>
- 6 Physics and Astronomy - University of Canterbury - New Zealand
<http://www.phys.canterbury.ac.nz>
- 7 astronomy 18th century
<http://www.cosmicelk.net/astronomy18c.htm>
- 8 Astronomy Picture of the Day’s Educational Links
<http://antwrp.gsfc.nasa.gov/apod/lib/edlinks.html>
- 9 Department of Astronomy - University of Florida
<http://www.astro.ufl.edu>
- 10 Astronomy Picture of the Day Archive
<http://apod.nasa.gov/apod/archivepix.html>

Table 5 Top-10 results of the “astronomy” query produced by *TWPR* and *T-Rank*Top-10 results produced by *TWPR*

- 1 NASA – NASA News and Features
<http://www.nasa.gov/news>
- 2 Science Daily: News & Articles in Science, Health, Environment & Technology
<http://www.sciencedaily.com>
- 3 NASA – Home
<http://www.nasa.gov>
- 4 Astronomy Now Online
<http://www.astronomynow.com>
- 5 Astronomy Picture of the Day Archive
<http://apod.nasa.gov/apod/archivepix.html>
- 6 SkyandTelescope.com – News from Sky & Telescope
<http://www.skyandtelescope.com/news>
- 7 Space news and articles from Scientific American
<http://www.scientificamerican.com/space>
- 8 Astronomy articles, forums, listings, and other resources about Astronomy
<http://www.astronomy.net>
- 9 Astronomy: Stars and Planets
<http://jumk.de/astronomie/astronomy.shtml>
- 10 Astronomy Picture of the Day’s Educational Links
<http://antwrp.gsfc.nasa.gov/apod/lib/edlinks.html>

Top-10 results produced by *T-Rank*

- 1 Science Daily: News & Articles in Science, Health, Environment & Technology
<http://www.sciencedaily.com>
- 2 NASA – NASA News and Features
<http://www.nasa.gov/news>
- 3 Astronomy Now Online
<http://www.astronomynow.com>
- 4 NASA – Home
<http://www.nasa.gov>
- 5 SkyandTelescope.com – News from Sky & Telescope
<http://www.skyandtelescope.com/news>
- 6 Astronomy Picture of the Day Archive
<http://apod.nasa.gov/apod/archivepix.html>
- 7 EarthSky Science Podcasts: A clear Voice for Science
<http://www.earthsky.org>
- 8 Astronomy articles, forums, listings, and other resources about Astronomy
<http://www.astronomy.net>
- 9 Solarark high accurate Astronomy Softwares
<http://www.solarark.fi>
- 10 Physics and Astronomy Classification Scheme
<http://publish.aps.org/PACS>

Table 6 Top-10 results of the “astronomy” query produced by *TimedPR*

 Top-10 results produced by *TimedPR*

- 1 Science Daily: News & Articles in Science, Health, Environment & Technology
<http://www.sciencedaily.com>
 - 2 NASA - NASA News and Features
<http://www.nasa.gov/news>
 - 3 Astronomy News & Links
<http://www.astrophys.org>
 - 4 Astronomy Now Online
<http://www.astronomynow.com>
 - 5 SkyandTelescope.com - News from Sky & Telescope
<http://www.skyandtelescope.com/news>
 - 6 NASA - Home
<http://www.nasa.gov>
 - 7 Universe Today
<http://www.universetoday.com>
 - 8 EarthSky Science Podcasts: A clear Voice for Science
<http://www.earthsky.org>
 - 9 Solarark high accurate Astronomy Softwares
<http://www.solarark.fi>
 - 10 Astronomy articles, forums, listings, and other resources about Astronomy
<http://www.astronomy.net>
-

these values increase rapidly until the top-10 for all temporal-aware methods, indicating that they can identify the pages most relevant to human users' preference within the top few results. This is an advantage since, as reported by Silverstein et al. (1999), most users usually browse only the top-ten results. Moreover, the *NDCG* scores of all temporal-aware methods are quite high and unvaried, indicating that each result appearing earlier in the ranked list is indeed preferred by users.

For the results concerning “History/Fact” documents shown in Fig. 4c, as expected the *PR* method can produce better rankings to users than the temporal-aware methods; it shows all the highest scores in *OSim* as well as some in both *KSim* and *NDCG*, proving that users in general prefer web pages contained reliable factual information although they are not recently updated. However, for all of those results, *TWPR* produces the second highest scores (in some cases, the highest), while the scores of *T-Rank* and *TimedPR* are the third highest and the lowest, respectively. Finally, for the “Others” results as shown in Fig. 4d, all temporal-aware methods still give higher *OSim*, *KSim*, and *NDCG* to users' ranking than does the *PR* method. In conclusion, we see that the *TWPR* method outperforms the others in most cases in the experiments.

5.4 Sensitivity analysis

We now investigate the effect of parameters specified in the *TWPR* method. The parameters can be divided into two groups: external and internal parameters. The former is in regard to the levels of hierarchical topics in ODP (DMOZ 2009) and the width of the time

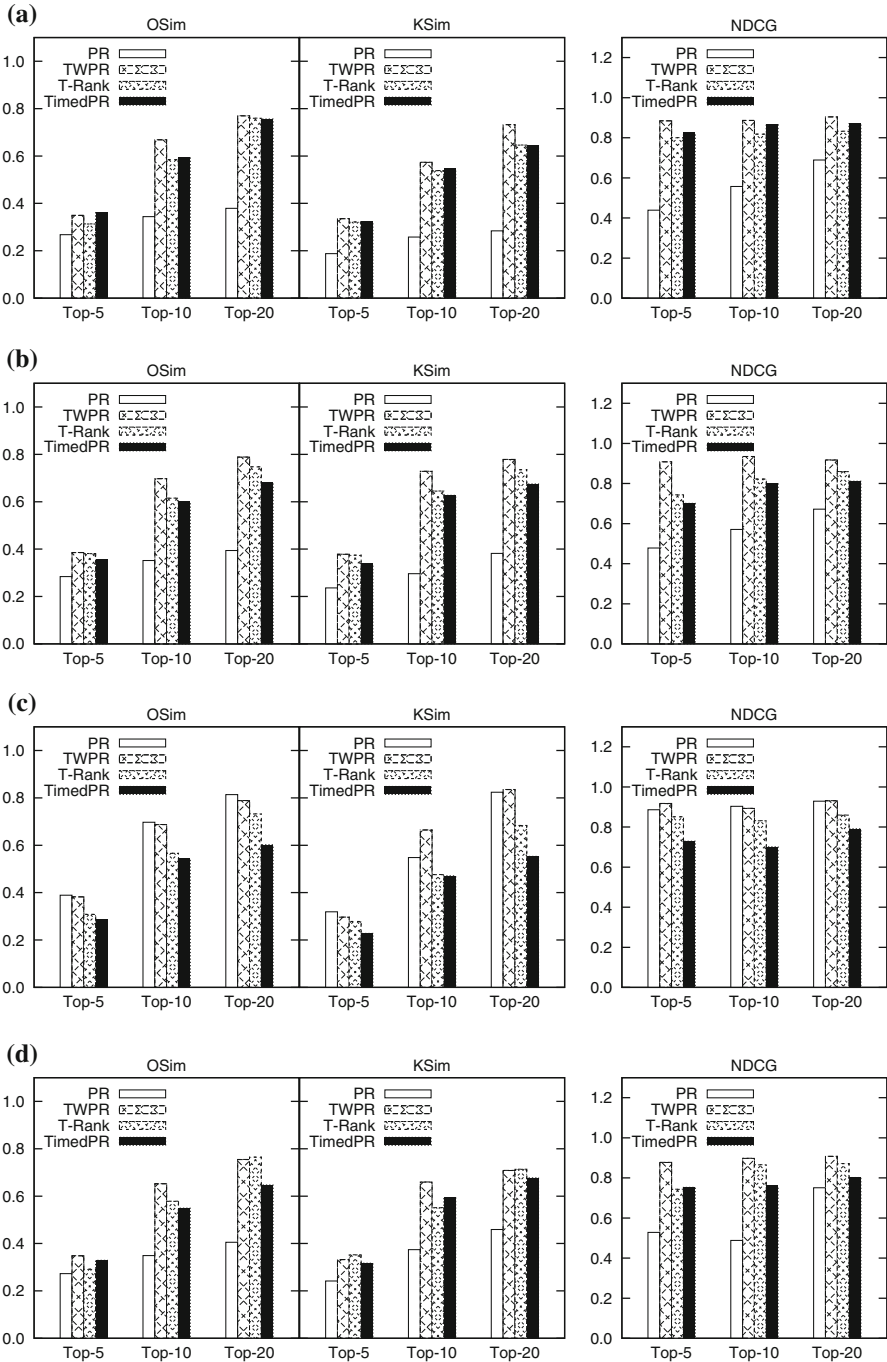


Fig. 4 Average similarities and *NDCG* scores at top-5, 10, and 20 results. (a) Results for queries in “Event” list. (b) Results for queries in “Trend” list. (c) Results for queries in “History/Fact” list. (d) Results for queries in “Others” list

intervals, i.e., C and I , respectively. The latter is in regard to the coefficients ω_1 , ω_2 , and ω_3 , which are the weights of the inverse-age, event, and trend factors. For the remaining parameters, the coefficient α in Eq. 13 is assigned as 0.85, a small constant ϵ in Eq. 15 is set to 10^{-10} , and the residual error δ is 10^{-10} . We conducted the experiments by employing the last crawled web pages on January 31st, 2009 as well as the 30 queries. We also used the baseline rankings obtained from the users for comparisons. In the following, we examine only the top-10 results.

To study the effects of the external parameters, we set the weights ω_i to the same constant value (i.e., $1/3$) but vary C and I . We let $C@x$ be the set of categories corresponding to the x -th level in ODP, and $I@y$ be the set of observation time intervals for every y consecutive days. For the experiments, we compared the rankings produced from twelve combinations of C and I with the baseline rankings as well as calculated the $NDCG$ scores for the queries corresponding to “Event”, “Trend”, “History/Fact”, and “Others” lists, as depicted in Table 7. We observe that the results for queries in each list behave in the similar way as follows: at $C@1$, varying I does not affect the rankings; the $OSim$, $KSim$, and $NDCG$ scores are quite unvaried. The reason is that categories are too general at the first level; there are many web pages contained in each category. Since a proportion of modified web pages within a specific time interval to the whole pages in the same category is very small, when varying I (i.e., from $I@1$ through $I@7$) the event effect for a web page is almost the same. Similarly, an update profile of a category would not be affected when varying I since it is created from averaging the large amount of pages’ profiles contained in that category. The trend effect corresponding to each defined observation time interval I for a web page is then quite similar. However, the time interval I has much more effect when changing C to $C@2$ and $C@3$, respectively. Setting $C@2$ or $C@3$, and varying I from $I@1$ through $I@7$ can produce better rankings. The reason may be that $I@1$ represents too short period. Modifying certain web pages within the same short time interval has less probability, while $I@3$ through $I@7$ gaps are more flexible. From the table, we can observe that the combination of $C@3$ and $I@5$ is an optimal setting for this experimental dataset.

Subsequently, we examine the effect of varying the remaining internal parameters. For this, we fixed the external parameter values at $C@3$ and $I@5$. The variations of three coefficients employed for the following experiments are concluded in Table 8. We compared the rankings produced from these seven combinations of $TWPR$ with the baseline rankings and calculated the $NDCG$ scores for the queries corresponding to “Event”, “Trend”, “History/Fact”, and “Others” lists, respectively. The results are shown in Table 9. Note that we also repeat the $OSim$, $KSim$, and $NDCG$ scores for the top-10 results of the PR method, given in Fig. 4, here for comparisons.

From the overall results, we can conclude that the inverse-age factor is the most influential, while the trend and event factors have the second and the least impact, respectively. This means that users prefer newer as well as up-to-date web pages rather than those which were updated in the past. However, the combinations with more than one factor can produce better rankings; especially, $TWPR_{AET}$ provides the best results in this dataset. Moreover, we found that the event factor is more significant in page ranking for queries in the “Event” list. The combination $TWPR_{AE}$ produces the second highest value. This indicates that the event factor can influence web pages containing recent event information closed to users’ preferences. Similarly, the results for queries in the “Trend” list, the trend factor can influence web pages containing information continually changed by trend, closed to users’ preferences, as well. $TWPR_{AT}$ can produce the second highest $OSim$ and $KSim$, and $TWPR_{ET}$ produce the second highest $NDCG$. However, the results for

Table 7 Average similarities and *NDCG* scores for the queries in “Event”, “Trend”, “History/Fact”, and “Others” lists at top-10 results of *TWPR* setting $\omega_1 = \omega_2 = \omega_3 = 1/3$ but varying the values of *C* and *I*

		<i>I</i> @1	<i>I</i> @3	<i>I</i> @5	<i>I</i> @7
(a) Results for queries in “Event” list					
<i>C</i> @1	<i>OSim</i>	0.548	0.548	0.553	0.553
	<i>KSim</i>	0.531	0.539	0.551	0.547
	<i>NDCG</i>	0.748	0.752	0.755	0.753
<i>C</i> @2	<i>OSim</i>	0.582	0.589	0.652	0.649
	<i>KSim</i>	0.541	0.564	0.569	0.565
	<i>NDCG</i>	0.855	0.878	0.881	0.879
<i>C</i> @3	<i>OSim</i>	0.614	0.649	0.668	0.665
	<i>KSim</i>	0.552	0.558	0.573	0.575
	<i>NDCG</i>	0.874	0.885	0.886	0.879
(b) Results for queries in “Trend” list					
<i>C</i> @1	<i>OSim</i>	0.553	0.553	0.557	0.561
	<i>KSim</i>	0.529	0.543	0.549	0.551
	<i>NDCG</i>	0.776	0.784	0.789	0.787
<i>C</i> @2	<i>OSim</i>	0.585	0.591	0.655	0.652
	<i>KSim</i>	0.532	0.607	0.684	0.682
	<i>NDCG</i>	0.834	0.855	0.869	0.871
<i>C</i> @3	<i>OSim</i>	0.609	0.654	0.697	0.693
	<i>KSim</i>	0.561	0.624	0.728	0.679
	<i>NDCG</i>	0.863	0.924	0.935	0.919
(c) Results for queries in “History/Fact” list					
<i>C</i> @1	<i>OSim</i>	0.551	0.554	0.559	0.559
	<i>KSim</i>	0.521	0.525	0.531	0.533
	<i>NDCG</i>	0.803	0.797	0.826	0.813
<i>C</i> @2	<i>OSim</i>	0.577	0.579	0.593	0.594
	<i>KSim</i>	0.529	0.532	0.535	0.535
	<i>NDCG</i>	0.835	0.841	0.856	0.839
<i>C</i> @3	<i>OSim</i>	0.662	0.684	0.687	0.682
	<i>KSim</i>	0.658	0.663	0.664	0.661
	<i>NDCG</i>	0.884	0.882	0.893	0.887
(d) Results for queries in “Others” list					
<i>C</i> @1	<i>OSim</i>	0.549	0.551	0.554	0.561
	<i>KSim</i>	0.506	0.533	0.552	0.544
	<i>NDCG</i>	0.785	0.804	0.816	0.796
<i>C</i> @2	<i>OSim</i>	0.579	0.588	0.636	0.628
	<i>KSim</i>	0.516	0.566	0.641	0.639
	<i>NDCG</i>	0.813	0.824	0.856	0.849
<i>C</i> @3	<i>OSim</i>	0.603	0.642	0.652	0.658
	<i>KSim</i>	0.597	0.635	0.659	0.648
	<i>NDCG</i>	0.859	0.884	0.897	0.892

The bold text highlights the highest value

Table 8 Internal parameter settings correspond to the notation of *TWPR*

Notation	ω_1	ω_2	ω_3
$TWPR_A$	1	0	0
$TWPR_E$	0	1	0
$TWPR_T$	0	0	1
$TWPR_{AE}$	1/2	1/2	0
$TWPR_{AT}$	1/2	0	1/2
$TWPR_{ET}$	0	1/2	1/2
$TWPR_{AET}$	1/3	1/3	1/3

queries in the “History/Fact” list, although the *PR* method shows better rankings, *TWPR* which incorporates all three temporal factors can result the second highest value. Finally, we observe that inverse-age factor alone (i.e., $TWPR_A$) can produce better results for all queries, except those in “History/Fact” list, than does the standard PageRank.

6 Future research directions

There are still several research issues to pursue. We will mention some of them as follows. For the model itself, many questions are still interesting to answer. For example, we simply use a linear time-scale to define the decay rate of the age factor. It will thus be interesting to study the effect of the age factor using other decay functions. Second, we set both observed time interval equally in studying the event and trend effects. Therefore, it will also be interesting to further explore those effects with different pairs of the observation periods. Third, as most of the experimental web pages have been selected in common with those in the ODP, we are then simply able to employ the category information from the ODP, on leaving the effect of non-categorized pages to zero. Hence, it will be interesting to experiment with larger datasets in which many pages have no categorical data. Last but not least, it will be interesting to study the effect from the size of the category, i.e., the number of documents in a category, as well.

To scale up to the Web, it may be interesting to explore existing categorization techniques, such as those proposed by Fan et al. (2008), Tang et al. (2009), and Liu et al. (2005). However, in many situations it may be unnecessary to apply categorization technique to the whole web collection. Sometimes search engine administrators just would like to boost some groups of pages for some special purposes. For example, only some groups of web pages that are modified within a gap of 3 months before the up-coming new year, with the expect for some special occurrences, will just be categorized.

To deal with multiple topics on a web page, it can be planned to exploit a page segmentation algorithm, e.g., VIPS (Cai et al. 2003a, b), to extract semantic blocks and treat them as topic units (Cai et al. 2004). For such units, they have the same modification time if they belong to the same page. *TWPR* can then be used to compute an authoritative score for a unit. All units’ scores of a page can be later aggregated and normalized to produce the final ranked score.

Since update information of a web page used in the experiments is extracted by daily crawling and simply differing the consecutive versions, one possible optimization would adapt a model for predicting change (Cho and Garcia-Molina 2003a, b) in order to save time and resource. In addition, many pages are automatically changed in corresponding with a user visit (e.g., a visit counter, the time of day, or even parameters in the URL, etc.)

Table 9 Average similarities and *NDCG* scores for the queries in “Event”, “Trend”, “History/Fact”, and “Others” lists at top-10 results of *TWPR* (setting *C@3*, *I@5*, and weights as in Table 8), and *PR*

Notation	<i>OSim</i>	<i>KSim</i>	<i>NDCG</i>
(a) Results for queries in “Event” list			
<i>TWPR_A</i>	0.586	0.553	0.887
<i>TWPR_E</i>	0.574	0.502	0.883
<i>TWPR_T</i>	0.563	0.523	0.849
<i>TWPR_{AE}</i>	0.652	0.561	0.894
<i>TWPR_{AT}</i>	0.649	0.558	0.877
<i>TWPR_{ET}</i>	0.627	0.553	0.881
<i>TWPR_{AET}</i>	0.668	0.573	0.886
<i>PR</i>	0.344	0.258	0.557
(b) Results for queries in “Trend” list			
<i>TWPR_A</i>	0.612	0.564	0.901
<i>TWPR_E</i>	0.524	0.463	0.876
<i>TWPR_T</i>	0.548	0.516	0.903
<i>TWPR_{AE}</i>	0.635	0.624	0.896
<i>TWPR_{AT}</i>	0.688	0.717	0.898
<i>TWPR_{ET}</i>	0.631	0.597	0.905
<i>TWPR_{AET}</i>	0.697	0.728	0.935
<i>PR</i>	0.352	0.296	0.571
(c) Results for queries in “History/Fact” list			
<i>TWPR_A</i>	0.584	0.531	0.877
<i>TWPR_E</i>	0.483	0.352	0.864
<i>TWPR_T</i>	0.537	0.404	0.869
<i>TWPR_{AE}</i>	0.628	0.552	0.883
<i>TWPR_{AT}</i>	0.634	0.608	0.891
<i>TWPR_{ET}</i>	0.597	0.487	0.884
<i>TWPR_{AET}</i>	0.687	0.664	0.893
<i>PR</i>	0.697	0.548	0.903
(d) Results for queries in “Others” list			
<i>TWPR_A</i>	0.576	0.513	0.893
<i>TWPR_E</i>	0.496	0.434	0.874
<i>TWPR_T</i>	0.533	0.471	0.879
<i>TWPR_{AE}</i>	0.602	0.566	0.895
<i>TWPR_{AT}</i>	0.647	0.615	0.886
<i>TWPR_{ET}</i>	0.586	0.526	0.874
<i>TWPR_{AET}</i>	0.652	0.659	0.897
<i>PR</i>	0.349	0.374	0.488

The bold and italic texts highlight the first and the second highest values, respectively

In this study, we simply employed a modification time in HTTP header, and size of a page in bytes, to detect changes. However, in the future, it will be interesting to approximate a static version of each dynamic page by exploring the techniques proposed by Minamide (2005) to detect such changes.

Finally, as in most ranking algorithms, combating spam is currently an important issue for web search designers, our approach here has also no exception. As long as a novel ranking technique is continually published, spammers can always follow, try, and finally

succeed to spam it. However, the key factor as claimed in many studies (e.g., Richardson et al. 2006; Liu et al. 2010, etc.) is that the technique should be difficult for spamming. However, we believe that web pages with up-to-dateness, event-related changes, and trend in revisions, are authoritative. In addition, the ranking effect from event and trend also needs the neighbor pages to change within the same observation period. It is then expected that spammers could succeed to spam such pages using much more effort, as those pages must often be modified, and changed in corresponding to event and trend of the other pages during an observation period.

7 Conclusions

For many web queries, users seek current or actively maintained information; hence, one would expect that time-related factors of a web page can contribute to establishing its usefulness to human web searchers. In this paper, we present a novel link-based ranking algorithm that incorporates the temporal factors into the authority ranking of web pages. We also study the effectiveness of the other two temporal-aware ranking algorithms. The experiments with the user study, using a sample set of web pages and a set of selected queries, show that the temporal-aware algorithms outperform the standard PageRank almost all cases. However, in overall quality of rankings, our *TWPR* algorithm, which incorporates age, event, and trend factors, provides rankings most similar to human users' preference.

Acknowledgments This research is supported by the Thailand Research Fund through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0122/2548). We thank the members of Yamana Laboratory for their help in data preparation and query evaluations. We also thank James Brucker for his valuable comments in the paper. In addition, we thank the anonymous reviewers for their comments and suggestions that have been incorporated into this article.

References

- Adamic, L. A., & Huberman, B. A. (2001). The web's hidden order. *Communications of the ACM*, 44(9), 55–59.
- Baeza-Yates, R. A., & Ribeiro-Neto, B. A. (1999). *Modern information retrieval*. New York: ACM Press & Addison Wesley.
- Baeza-Yates, R. A., Saint-Jean, F., & Castillo, C. (2002). Web structure, dynamics and page quality. In *SPIRE '02: Proceedings of the 9th international symposium on string processing and information retrieval* (pp. 117–130).
- Berberich, K., Vazirgiannis, M., & Weikum, G. (2006). Time-aware authority ranking. *Internet Mathematics*, 2(3), 301–332.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117.
- Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton N., & Hullender, G. (2005). Learning to rank using gradient descent. In *ICML '05: Proceedings of the 22nd international conference on machine learning* (pp. 89–96).
- Cai, D., Yu, S., Wen, J. R., & Ma, W. Y. (2003a). Extracting content structure for web pages based on visual representation. In *APWeb '03: Proceedings of the 5th Asia Pacific web conference* (pp. 406–417).
- Cai, D., Yu, S., Wen, J. R., & Ma, W. Y. (2003b). Vips: A vision-based page segmentation algorithm. Technical report, Microsoft Research.
- Cai, D., He, X., Wen, J. R., & Ma, W. Y. (2004). Block-level link analysis. In *SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 440–447).

- Cho, J., & Garcia-Molina, H. (2003a). Effective page refresh policies for web crawlers. *ACM Transactions on Database Systems*, 28(4), 390–426.
- Cho, J., & Garcia-Molina, H. (2003b). Estimating frequency of change. *ACM Transactions on Internet Technology*, 3(3), 256–290.
- Cho, J., & Roy, S. (2004). Impact of search engines on page popularity. In *WWW '04: Proceedings of the 13th international world wide web conference* (pp. 20–29).
- Cho, J., Roy, S., & Adams, R. E. (2005). Page quality: In search of an unbiased web ranking. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on management of data* (pp. 551–562).
- DMOZ. (2009). The open directory project. <http://www.dmoz.org>.
- Dwork, C., Kumar, R., Naor, M., & Sivakumar, D. (2001). Rank aggregation methods for the web. In *WWW '01: Proceedings of the 10th international world wide web conference* (pp. 613–622).
- Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J. (2008). Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, 9, 1871–1874.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382.
- Geng, X., Liu, T. Y., Qin, T., Arnold, A., Li, H., & Shum, H. Y. (2008). Query dependent ranking using k-nearest neighbor. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval* (pp. 115–122).
- Golub, G. H., & Loan, C. F. V. (1996). *Matrix computations*. Baltimore and London: Johns Hopkins University Press.
- Gonçalves, B., Meiss, M. R., Ramasco, J. J., Flammini, A., & Menczer, F. (2009). Remembering what we like: Toward an agent-based model of web traffic. In *WSDM '09: Proceedings of the 2nd ACM international conference on web search and data mining*.
- Grimmett, G. R., & Stirzaker, D. R. (2001). *Probability and random processes*. New York: Oxford University Press.
- Haveliwala, T. H. (1999). Efficient computation of pagerank. Technical report, Stanford InfoLab.
- Haveliwala, T. H. (2003). Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4), 784–796.
- Järvelin, K., & Kekäläinen, J. (2000). Ir evaluation methods for retrieving highly relevant documents. In *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrieval* (pp. 41–48).
- Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, 20(4), 422–446.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604–632.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Liu, T. Y., Yang, Y., Wan, H., Zhou, Q., Gao, B., Zeng, H. J., Chen, Z., & Ma, W. Y. (2005). An experimental study on large-scale web categorization. In *WWW '05: Special interest tracks and posters of the 14th international world wide web conference* (pp. 1106–1107).
- Liu, Y., Gao, B., Liu, T. Y., Zhang, Y., Ma, Z., He, S., & Li, H. (2008). Browserank: Letting web users vote for page importance. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval* (pp. 451–458).
- Liu, Y., Liu, T. Y., Gao, B., Ma, Z., & Li, H. (2010). A framework to compute page importance based on user behaviors. *Information Retrieval*, 13(1), 22–45.
- Lucene. (2009). The apache software foundation. <http://lucene.apache.org>.
- Meiss, M. R., Menczer, F., Fortunato, S., Flammini, A., & Vespignani, A. (2008). Ranking web sites with real user traffic. In *WSDM '08: Proceedings of the 1st ACM international conference on web search and data mining* (pp. 65–76).
- Melucci, M. (2007). On rank correlation in information retrieval evaluation. *ACM SIGIR Forum*, 41(1), 18–33.
- Minamide, Y. (2005). Static approximation of dynamically generated web pages. In *WWW '05: Proceedings of the 14th international world wide web conference* (pp. 432–441).
- Muraoka, Y., & Yamana, H. (2005). e-society project. http://www.yama.info.waseda.ac.jp/~yamana/e-society/index_eng.htm.
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.
- Richardson, M., & Domingos, P. (2002). The intelligent surfer: Probabilistic combination of link and content information in pagerank. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems 14* (pp. 1441–1448). Cambridge, MA: MIT Press.

- Richardson, M., Prakash, A., & Brill, E. (2006). Beyond pagerank: Machine learning for static ranking. In *WWW '06: Proceedings of the 15th international world wide web conference* (pp. 707–715).
- Ross, S. M. (2002). *Introduction to probability models*. San Diego: Academic Press.
- Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum*, 33(1), 6–12.
- Tang, L., Rajan, S., & Narayanan, V. K. (2009). Large scale multi-label classification via metalabeler. In *WWW '09: Proceedings of the 18th international world wide web conference* (pp. 211–220).
- Xu, J., & Li, H. (2007). Adarank: A boosting algorithm for information retrieval. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 391–398).
- Yu, P. S., Li, X., & Liu, B. (2005). Adding the temporal dimension to search—A case study in publication search. In *WI '05: Proceedings of the 2005 IEEE/WIC/ACM international conference on web intelligence* (pp. 543–549).