Freshness Matters
- In Flowers, Food, and Web Authority -

Andreas Sander

*Hot Topics in Information Retrieval, WS 10/11*

15th November 2010
Overview

1. Motivation
2. Related Work
   - PageRank
   - T-Rank
3. New Approach
   - Idea
   - Web Freshness
   - Authority Propagation
4. Experimental Results
5. Future Work
6. Strengths and Weakness of the Paper
Why do we need freshness in web authority rankings?
Freshness is all about links & pages which were created or changed recently.

Contrary to that: stale pages and links.
Freshness Matters

Motivation
Welcome to SIGIR | Home
... General for Iraq Reconstruction (SIGIR), a temporary federal agency serving ... Awards Leadership Organizational Chart SIGIR FAQ Law ...
www.sigir.mil - Im Cache

SIGIR: Information Retrieval - Übersetzen
Addresses issues ranging from theory to user demands in the application of computers to the acquisition, SIGIR 2011 (Beijing, China) - July 24-28, 2011 ...
www.sigir.org - Im Cache

27th Annual International ACM SIGIR Conference 2004 - Übersetzen
The 27th Annual International ACM SIGIR Conference will be held at The University of Sheffield, ... SIGIR is the major international forum for the presentation ...
www.sigir.org/sigir2004/index.html - Im Cache

SIGIR - Übersetzen
www.sigir.com

SIGIR - Übersetzen
Contents - SIGIR 2008 Home Page ... Proceedings of the SIGIR 2007 International Workshop on Plagiarism Analysis, ... 29. SIGIR 2006: Seattle, WA, USA ...
www.informatik.uni-kl.de/~wyd/conf/sigir/index.html - 61k - Im Cache

Обслуживание компьютеров. ИТ Аутсорсинг IT сервис Сервисное ... ИТ аутсорсинг. Обслуживание компьютеров. Абонентское обслуживание компьютеров по ... Квалифицированные менеджеры ответят на вопросы, помогут подобрать эффективные ...
sigir.ru - Im Cache

IT outsourcing — ИТ Аутсорсинг Сервисное обслуживание организаций фирм ... İT outsourcing İT Aoutosrinsing CİGİR Обслуживание компьютеров в Санкт-Петербурге ... Надежную работу информационной системы в целом ...
www.sigir.ru/ - Im Cache

SIGIR - Wikipedia, the free encyclopedia - Übersetzen
SIGIR. From Wikipedia, the free encyclopedia. Jump to: navigation, ... Retrieved from http://en.wikipedia.org/wiki/SIGIR Categories: Disambiguation pages ...
en.wikipedia.org/wiki/SIGIR - Im Cache

SIGIR '08 - Singapre - Übersetzen
SIGIR 2008 ... SIGIR is the major international forum for the presentation of new research ...
... Powered by Check Pagenace © Copyright 2007-2008 ACM SIGIR ...
www.sigir2008.org - Im Cache

home ACM SIGIR 2010 - Übersetzen
www.sigir2010.org/doku.php - Im Cache

1 2 3 4 5 6 7 8 9 10 11 Weitere >
**Freshness Matters**

**Motivation**
Freshness Matters

Motivation
It is a common strategy to estimate the page authority by accumulation of in-link (links pointing to the specific site) contributions (e.g. PageRank).
It is a common strategy to estimate the page authority by accumulation of in-link (links pointing to the specific site) contributions (e.g., PageRank).
Millions of people (interested in web activities) change web link-structure and content day by day.

**Figure:** Various Web 2.0 Tools
(Source: Flickr User pipeapple)
Millions of people (interested in web activities) change web link-structure and content day by day.

**BUT**

Web authority of a webpage is often computed based on a single website snapshot.

**Figure:** Various Web 2.0 Tools (Source: Flickr User pipeapple)
Freshness Matters

Motivation

Figure: Wikipedia Statistic: Average Changes per Day (Source: Wikipedia)
### Figure: Wikipedia Statistic: Top 20 Wikipedia Sites (Germany) with the most changes from 01.06.2009 to 30.06.2009 (Source: Wikipedia)

<table>
<thead>
<tr>
<th>#</th>
<th>Seite</th>
<th>Namensräume</th>
<th>Bearb. (30 T.)</th>
<th>Bearb.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wikipedia:Vandalismusmeldung</td>
<td>4</td>
<td>3891</td>
<td>3891</td>
</tr>
<tr>
<td>2</td>
<td>Wikipedia:Auskunft</td>
<td>4</td>
<td>3079</td>
<td>3079</td>
</tr>
<tr>
<td>3</td>
<td>Wikipedia:Auskunft/alt6</td>
<td>4</td>
<td>2647</td>
<td>5143</td>
</tr>
<tr>
<td>4</td>
<td>Wikipedia:Fragen zur Wikipedia</td>
<td>4</td>
<td>2582</td>
<td>3680</td>
</tr>
<tr>
<td>5</td>
<td>Wikipedia:Kandidaten für exzellente Bilder/Archiv2</td>
<td>4</td>
<td>1737</td>
<td>4843</td>
</tr>
<tr>
<td>6</td>
<td>Wikipedia:Löschrückung</td>
<td>4</td>
<td>1525</td>
<td>2053</td>
</tr>
<tr>
<td>7</td>
<td>Wikipedia:Vandalismusmeldung/Alt11</td>
<td>4</td>
<td>1137</td>
<td>4823</td>
</tr>
<tr>
<td>8</td>
<td>Wikipedia:Löschkandidaten/I. Juni 2009</td>
<td>4</td>
<td>972</td>
<td>973</td>
</tr>
<tr>
<td>9</td>
<td>Wikipedia:Kandidaten für lesenswerte Artikel/Alt2</td>
<td>4</td>
<td>838</td>
<td>4775</td>
</tr>
<tr>
<td>10</td>
<td>Wikipedia:Importwünsche</td>
<td>4</td>
<td>880</td>
<td>4524</td>
</tr>
<tr>
<td>14</td>
<td>Diskussion:All-France-Flug 447</td>
<td>1</td>
<td>814</td>
<td>814</td>
</tr>
<tr>
<td>15</td>
<td>Wikipedia:Spielewiese</td>
<td>4</td>
<td>790</td>
<td>872</td>
</tr>
<tr>
<td>18</td>
<td>Wikipedia:Administratoren/Anfragen</td>
<td>4</td>
<td>747</td>
<td>3710</td>
</tr>
<tr>
<td>20</td>
<td>Wikipedia:Persönliche Bekanntschaften/newe Einträge</td>
<td>4</td>
<td>742</td>
<td>9619</td>
</tr>
</tbody>
</table>
Freshness Matters

Motivation

Figure: Picture from Flickr
Overview

1. Motivation
2. Related Work
   - PageRank
   - T-Rank
3. New Approach
   - Idea
   - Web Freshness
   - Authority Propagation
4. Experimental Results
5. Future Work
6. Strengths and Weakness of the Paper
PageRank [Lawrence Page & Sergey Brin, 1998]

- is the most popular ranking algorithm.

- concept is based on a site’s in-links (links pointing to the specific site).

- represents the probability that a random web-surfer model enters a page.
Moves of the Random Web Surfer Model

In every step the random web surfer model either **follows** an out-linked page **OR** does a random **jump** to any page.
The PageRank Algorithm

Assuming n pages $T_1, T_2, \ldots, T_n$ link to a page A.
The PageRank Algorithm

Assuming \(n\) pages \(T_1, T_2, \ldots, T_n\) link to a page \(A\).

Then

\[
PR(A) = \frac{1-d}{N} + d \left( \frac{PR(T_1)}{C(T_1)} + \cdots + \frac{PR(T_n)}{C(T_n)} \right)
\]

with

\(N\): \# of overall pages

\(d\): damping factor \((0 \leq d \leq 1)\) and

\(C(T_i)\): \# of out-links of \(T_i\), \(1 \leq i \leq n\).
No freshness of links and pages is considered here!
T-Rank [K. Berberich et al, 2005]

- is build upon the PageRank idea & algorithm.
- extends PageRank algorithm with a bias for following an out-link and jumping to a new page to introduce freshness of pages and links.
The probability that a random surfer is clicking at page $x$ on an out-link to page $y$ is a weighted combination of

- freshness of page $y$,
- freshness of the link from $x$ to $y$ and
- average freshness of all incoming links of $y$. 
The probability that a random surfer takes a random jump to page $y$ is a weighted combination of

- freshness of page $y$,
- activity of page $y$,
- average freshness of all incoming links of $y$ and
- average activity of the pages that link to $y$. 
The T-Rank Algorithm

\[ r(y) = (1 - \epsilon) \cdot s(y) + \sum_{(x,y) \in E} \epsilon \cdot t(x,y) \cdot r(x) \]

with

\( \epsilon \): probability that a random surfer is clicking on an out-link

\( 1 - \epsilon \): probability that a random surfer makes a random jump to a page.
The T-Rank Algorithm

\[ r(y) = (1 - \varepsilon) \cdot s(y) + \sum_{(x,y) \in E} \varepsilon \cdot t(x,y) \cdot r(x) \]

with
\[ \varepsilon: \text{ probability that a random surfer is clicking on an out-link} \]
\[ (1 - \varepsilon): \text{ probability that a random surfer makes a random jump to a page.} \]
\[ s(y): \text{ the random jump probability} \]
\[ t(x,y): \text{ the transition probability} \]
The T-Rank Algorithm

- T-Rank is based on the fact that the users have a specific temporal focus of attention where the user is interested in.

- Freshness and activity computation is based on that temporal focus.

Figure: Temporal proximity window for a temporal interval
The T-Rank Algorithm

T-Rank is based on a good approach, **BUT**: Activities at different time points are not distinguished.

Instead, influence of activities on web freshness should decay over time.

**Figure:** Temporal proximity window for a temporal interval
Overview

1. Motivation

2. Related Work
   - PageRank
   - T-Rank

3. New Approach
   - Idea
   - Web Freshness
   - Authority Propagation

4. Experimental Results

5. Future Work

6. Strengths and Weakness of the Paper
Probabilistic algorithm to estimate web page authority by considering two temporal aspects:

1. Web freshness

2. Multiple snapshots collected at different timepoints $t_0, t_1, \cdots, t_n$. 
Web Freshness

- is regarded from two points of view: the information recommenders and the information providers.
Web Freshness

- is regarded from two points of view: the information recommenders and the information providers.

- hence is divided into in-link freshness \((InF)\) and page freshness \((PF)\).
Web Freshness

- is regarded from two points of view: the information recommenders and the information providers.

- hence is divided into in-link freshness ($InF$) and page freshness ($PF$).

- computation is based on temporal page profiles and temporal link profiles.
The Web Freshness Computation

Computation of the in-link and page freshness at a certain time point $t_i$: 
The Web Freshness Computation

Computation of the in-link and page freshness at a certain time point $t_i$:

$$\ln F(p)_{t_i} = \beta_1 e^{-\beta_2 \Delta t} \ln F(p)_{t_{i-1}} + \Delta \ln F(p)|_{t_{i-1}}^{t_i}$$
The Web Freshness Computation

Computation of the in-link and page freshness at a certain time point $t_i$:

$$\text{InF}(p)_{t_i} = \beta_1 e^{-\beta_2 \Delta t} \text{InF}(p)_{t_i-1} + \Delta \text{InF}(p)_{t_i-1}$$

$$\text{PF}(p)_{t_i} = \beta_3 e^{-\beta_4 \Delta t} \text{PF}(p)_{t_i-1} + \Delta \text{PF}(p)_{t_i-1}$$
Computation of the in-link and page freshness at a certain time point $t_i$:

$$\text{InF}(p)_{t_i} = \beta_1 e^{-\beta_2 \Delta t} \text{InF}(p)_{t_{i-1}} + \Delta \text{InF}(p)|_{t_{i-1}}^{t_i}$$

$$\text{PF}(p)_{t_i} = \beta_3 e^{-\beta_4 \Delta t} \text{PF}(p)_{t_{i-1}} + \Delta \text{PF}(p)|_{t_{i-1}}^{t_i}$$

How can we compute $\Delta \text{InF}(p)|_{t_{i-1}}^{t_i}$ and $\Delta \text{PF}(p)|_{t_{i-1}}^{t_i}$?
## In-Link Activities

<table>
<thead>
<tr>
<th>Link activity</th>
<th>Infl. on p’s InF</th>
<th>Gain of p’s InF</th>
</tr>
</thead>
<tbody>
<tr>
<td>creation of link $l : q \to p$</td>
<td>↑↑↑↑</td>
<td>3</td>
</tr>
<tr>
<td>update on link $l : q \to p$ (changed anchor)</td>
<td>↑↑</td>
<td>2</td>
</tr>
<tr>
<td>update on link $l : q \to p$ (unchanged anchor)</td>
<td>↑</td>
<td>1.5</td>
</tr>
<tr>
<td>removal of link $l : q \to p$</td>
<td>↓↓</td>
<td>-0.5</td>
</tr>
</tbody>
</table>
Computation of $\Delta lnF(p)|_{t_i}^{t_i-1}$

$$\Delta lnF_0(p)|_{t_i}^{t_i-1} = \sum_{l:q\to p} \sum_{j\in LA} \omega_j C_j(l)$$

with

$C_j(l) : \# \text{ of the } j^{th} \text{ type of link activity on link } l \text{ in } [t_{i-1}, t_i]$. 
Computation of $\Delta \ln F(p)|^L_{t_i}$

\[
\Delta \ln F_0(p)|^L_{t_i-1} = \sum_{l:q \rightarrow p} \sum_{j \in LA} \omega_j C_j(l)
\]

\[
\Delta \ln F(p)|^L_{t_i-1} = \lambda_{\ln F} \Delta \ln F_0(p)|^L_{t_i-1} + (1 - \lambda_{\ln F}) \sum_{l:q \rightarrow p} m_{qp} \Delta \ln F(q)|^L_{t_i-1}
\]

with

$C_j(l)$: # of the $j^{th}$ type of link activity on link l in $[t_{i-1}, t_i]$.

$m_{qp}$: one-step transition probability from q to p with $\sum m_{q*} = 1$
### Page Activities

<table>
<thead>
<tr>
<th>Page activity</th>
<th>Infl. on $q$’s PF</th>
<th>Gain of $q$’s PF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 creation of page $q$</td>
<td>↑↑↑↑</td>
<td>3</td>
</tr>
<tr>
<td>2 update on page $q$</td>
<td>↑</td>
<td>1.5</td>
</tr>
<tr>
<td>3 removal of page $q$</td>
<td>↓↓</td>
<td>-0.5</td>
</tr>
</tbody>
</table>
Computation of $\Delta PF(q)|_{t_i}^{t_{i-1}}$

$$\Delta PF_0(q)|_{t_i}^{t_{i-1}} = \sum_{j \in PA} \omega'_j C'_j(q)$$

with

$C'_j(q) : \# \text{ of the } j^{th} \text{ type of page activity on page } q \text{ in } [t_{i-1}, t_i].$
Computation of $\Delta PF(q)|_{t_i-1}^{t_i}$

$$\Delta PF_0(q)|_{t_i-1}^{t_i} = \sum_{j \in PA} \omega_j C_j'(q)$$

$$\Delta PF(q)|_{t_i-1}^{t_i} = \lambda_{PF} \Delta PF_0(q)|_{t_i-1}^{t_i} + (1 - \lambda_{PF}) \sum_{l: q \rightarrow p} m'_{qp} \Delta PF(p)|_{t_i-1}^{t_i}$$

with

$C_j'(q)$: # of the $j^{th}$ type of page activity on page q in $[t_{i-1}, t_i]$.  
$m'_{qp}$: Inverted one-step transition probability from q to p with  
$\sum m'_{*p} = 1$
How to use web freshness to control authority propagation in an archival link graph?
The difference between PageRank’s surfer model and this model is:

- The extended model has specific temporal intent
- The extended model can choose a snapshot for targeted page (based on the temporal intent).
- The extended model prefers fresh web resources.
A move of the extended surfer model takes the following steps:
A move of the extended surfer model takes the following steps:

1. Follow an out-linked page OR do a random jump to any page at the same timepoint. (see PageRank surfer model)
Moves of the Surfer

A move of the extended surfer model takes the following steps:

1. **Follow** an out-linked page **OR** do a random **jump** to any page at the same timepoint. (see PageRank surfer model)

2. After reaching target page, choose the snapshot based on the temporal intent.
Use a semi-Markov process for page authority estimation.

Long-run proportion of time that the Markov process is in state $i$ is given by:

$$A(i) = \frac{\pi_i \mu_i}{N \sum_{j=1}^{N} \pi_j \mu_j}, \quad i = 1, 2, \cdots, N$$
Goal

Use a semi-Markov process for page authority estimation.

Long-run proportion of time that the Markov process is in state $i$ is given by:

$$A(i) = \frac{\pi_i \mu_i}{\sum_{j=1}^{N} \pi_j \mu_j}, \quad i = 1, 2, \cdots, N$$

with

- $\pi_i$: probability that markov process reaches state $i$
- $\mu_i$: mean time process is remaining at state $i$. 

Computation of $\pi_{p,i}$

How to compute the probability $\pi_{p,i}$ that a web surfer reaches page $p$ at snapshot $t_i$?
Computation of $\pi_{p,i}$

We need some probability values:
We need some probability values:

\[ P_{t_j}(\text{Follow}|q) = (1 - d) \quad P_{t_j}(\text{Jump}|q) = d \]
Computation of $\pi_{p,i}$

We need some probability values:

\[ P_{t_j}(\text{Follow}|q) = (1 - d) \quad P_{t_j}(\text{Jump}|q) = d \]

\[ P_{t_j}(p|q, \text{Follow}) = F_{t_j}(p, q) \quad P_{t_j}(p|q, \text{Jump}) = \frac{1}{N_{t_j}} \]

with

$F_{t_j}(p, q)$: web surfer’s preference for following out-linked pages

$N_{t_j}$: total # of pages at $t_j$
Due to the fact, that fresh web resources are likely to attract surfer’s attention, we define:

\[ F_{tj}(p, q) = \frac{PF_{tj}(p)}{\sum_{p': q \rightarrow p'} PF_{tj}(p')} \]
Computation of $\pi_{p,i}$

Probability that a surfer reaches page $p$ at $t_i$ from page $p$ at $t_j$ is

$$P_{t_i|t_j}(p) = \frac{\omega(t_i, t_j)}{\sum_{V_k : p \in V_k, t_k \in T_i} \omega(t_i, t_k)}$$
Computation of $\pi_{p,i}$

Probability that a surfer reaches page $p$ at $t_i$ from page $p$ at $t_j$ is

$$P_{t_i|t_j}(p) = \frac{\omega(t_i,t_j)}{\sum_{V_k : p \in V_k, t_k \in T_i} \omega(t_i,t_k)}$$

with

$\omega(t_i,t_k)$ : weight that represents the influence between the snapshots at $t_i$ and $t_k$ (modeled with kernels)

$V_j$: set of pages at time point $t_j$

$T_i$: set of snapshots which can directly distribute authority to $t_i$ within one step (depending on window size $|T|$).
Kernels for $|T|=4$

- Gaussian Kernel
- Triangle Kernel
- Cosine Kernel
- Circle Kernel
Computation of $\pi_{p,i}$

Probability that a web surfer reaches page $p$ at snapshot $t_i$: 

\[
\pi_{p,i} = \sum_{t_j \in T_i} P_{t_i | t_j}(p) \sum_{q: q \to p | t_j} P_{t_j | q}(p | q, \text{Follow}) \pi_{q,j} + \sum_{t_j \in T_i} P_{t_i | t_j}(p) \sum_{q: q \to t_j} P_{t_j | q}(p | q, \text{Jump}) \pi_{q,j}
\]
Computation of $\pi_{p,i}$

Probability that a web surfer reaches page $p$ at snapshot $t_i$:

$$\pi_{p,i} = \sum_{t_j \in T_i} P_{t_i \mid t_j}(p) \sum_{q: q \rightarrow p \mid t_j} P_{t_j}(Follow \mid q)P_{t_j}(p \mid q, Follow)\pi_{q,j}$$

with

$T_i$ : set of snapshots which can directly distribute authority to $t_i$ within one step (depending on window size $|T|$).
Computation of $\pi_{p,i}$

Probability that a web surfer reaches page $p$ at snapshot $t_i$:

$$\pi_{p,i} = \sum_{t_j \in T_i} P_{t_i|t_j}(p) \sum_{q: q \rightarrow p|t_j} P_{t_j}(\text{Follow}|q) P_{t_j}(p|q, \text{Follow}) \pi_{q,j}$$

$$+ \sum_{t_j \in T_i} P_{t_i|t_j}(p) \sum_{q|t_j} P_{t_j}(\text{Jump}|q) P_{t_j}(p|q, \text{Jump}) \pi_{q,j}$$

with

$T_i$ : set of snapshots which can directly distribute authority to $t_i$ within one step (depending on window size $|T|$).
Computation of $\mu_{p,i}$

How to compute the average staying time $\mu_{p,i}$ that a web surfer surfs on page $p$ at snapshot $t_i$?
Computation of $\mu_{p,i}$

- Pages with more in-link activity are likely to attract a surfer to spend time browsing it.

- Using in-link freshness to model the time of a surfer staying on a web page.
Computation of $\mu_{p,i}$

$$
\mu_{p,i} = \sum_{t_j \in T_{t_i}'} \omega'(t_i, t_j) \ln F(p)_{t_j}
$$

with

$T_{t_i}':$ Set of snapshots included in the sliding window centered on $t_i$

$$
\omega'(t_i, t_j) = \frac{1}{|T_{t_i}'|} \text{ for any } t_j \in T_{t_i}'
$$

$$
\sum_{t_j \in T_{t_i}'} \omega'(t_i, t_j) = 1
$$
Final Computation

\[ A(p, i) = \frac{\pi_{p,i} \mu_{p,i}}{\sum_{j=1}^{n} \sum_{q \in V_j} \pi_{q,j} \mu_{q,j}} \]

with

n: # of snapshots

\( V_j \): set of pages at time point \( t_j \)
T-Fresh & Attributes

T-Fresh(kernel, window, snapshot)

kernel: The kernel controlling authority propagation among different web snapshots

window: The window size used in calculating average in-link freshness for estimating staying time

snapshot: The number of month spanned over the temporal graph.
Overview

1. Motivation

2. Related Work
   - PageRank
   - T-Rank

3. New Approach
   - Idea
   - Web Freshness
   - Authority Propagation

4. Experimental Results

5. Future Work

6. Strengths and Weakness of the Paper
Experimental Setup
Used archival web pages in the .ie domain (collected from January 200 to December 2007)
- Used archival web pages in the .ie domain (collected from January 200 to December 2007)
- Corpus contained 158 million web pages and 12 billion temporal links
- Used archival web pages in the .ie domain (collected from January 200 to December 2007)

- Corpus contained 158 million web pages and 12 billion temporal links

- Each month one snapshot
Used archival web pages in the .ie domain (collected from January 200 to December 2007)

Corpus contained 158 million web pages and 12 billion temporal links

Each month one snapshot

Time period of interest was April 2007.
Ninety queries.
Ninety queries.

For each query 84.6 URLs on average were judged according to freshness and relevance by human editors.
Ninety queries.

For each query 84.6 URLs on average were judged according to freshness and relevance by human editors.

For judging about ranking qualities two metrics were used: the Normalized Discounted Cumulative Gain (NDCG) and Precision@k.
Experimental Results
(a) Relevance performance: NDCG@10
Freshness Matters

Experimental Results

(b) Freshness performance: NDCG@10
<table>
<thead>
<tr>
<th>Method</th>
<th>P@10</th>
<th>NDCG@3</th>
<th>NDCG@5</th>
<th>NDCG@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM25</td>
<td>0.4695</td>
<td>0.2478</td>
<td>0.2740</td>
<td>0.3344</td>
</tr>
<tr>
<td>PageRank</td>
<td>0.4894</td>
<td>0.2589</td>
<td>0.2840</td>
<td>0.3457</td>
</tr>
<tr>
<td>BuzzRank</td>
<td>0.4770</td>
<td>0.2770</td>
<td>0.2980</td>
<td>0.3460</td>
</tr>
<tr>
<td>TemporalRank</td>
<td>0.4841</td>
<td>0.2706</td>
<td>0.2875</td>
<td>0.3524</td>
</tr>
<tr>
<td>TimedPageRank</td>
<td>0.5031</td>
<td>0.2830</td>
<td>0.3063</td>
<td>0.3587</td>
</tr>
<tr>
<td>T-Random</td>
<td>0.4904</td>
<td>0.2690</td>
<td>0.2877</td>
<td>0.3495</td>
</tr>
<tr>
<td>T-rank</td>
<td>0.4875</td>
<td>0.2669</td>
<td>0.2870</td>
<td>0.3496</td>
</tr>
<tr>
<td>T-Fresh(1,1,30)</td>
<td><strong>0.5051</strong></td>
<td><strong>0.3229</strong></td>
<td><strong>0.3347</strong></td>
<td><strong>0.3729</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>P@10</th>
<th>NDCG@3</th>
<th>NDCG@5</th>
<th>NDCG@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM25</td>
<td>0.3138</td>
<td>0.2137</td>
<td>0.2379</td>
<td>0.2805</td>
</tr>
<tr>
<td>PageRank</td>
<td>0.3325</td>
<td>0.1946</td>
<td>0.2345</td>
<td>0.2838</td>
</tr>
<tr>
<td>BuzzRank</td>
<td>0.3327</td>
<td>0.2043</td>
<td>0.2234</td>
<td>0.2797</td>
</tr>
<tr>
<td>TemporalRank</td>
<td><strong>0.3473</strong></td>
<td>0.2312</td>
<td>0.2510</td>
<td>0.2992</td>
</tr>
<tr>
<td>TimedPageRank</td>
<td>0.3398</td>
<td><strong>0.2443</strong></td>
<td>0.2514</td>
<td>0.2972</td>
</tr>
<tr>
<td>T-Random</td>
<td>0.3316</td>
<td>0.2054</td>
<td>0.2403</td>
<td>0.2879</td>
</tr>
<tr>
<td>T-rank</td>
<td>0.3356</td>
<td>0.2269</td>
<td>0.2498</td>
<td>0.2950</td>
</tr>
<tr>
<td>T-Fresh(1,1,30)</td>
<td>0.3412</td>
<td>0.2411</td>
<td><strong>0.2662</strong></td>
<td><strong>0.3076</strong></td>
</tr>
</tbody>
</table>
Future Work

Sensitivity of web activity detection accuracy.

Try as best as possible to find a way to close the gap caused by some pages having archival copies and some without in the searching process.
Strengths and Weakness of the Paper

<table>
<thead>
<tr>
<th>Good</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nice background informations</td>
<td>Wrong formulas</td>
</tr>
<tr>
<td>Good explanations of the ideas</td>
<td>Improvable detailed explanations</td>
</tr>
</tbody>
</table>
Thank you for your attention!