# Freshness Matters

## - In Flowers, Food, and Web Authority -

Andreas Sander

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

15th November 2010

# Overview

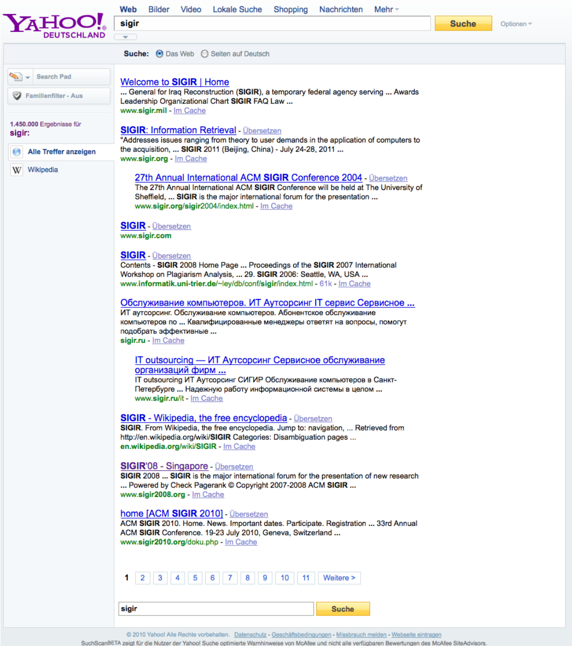
# Why do we need freshness in web authority rankings?

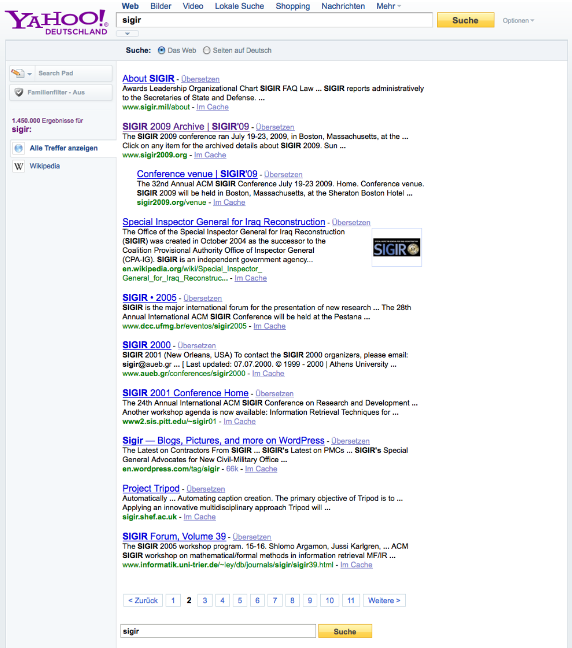
Freshness is all about links & pages which were created or changed recently.

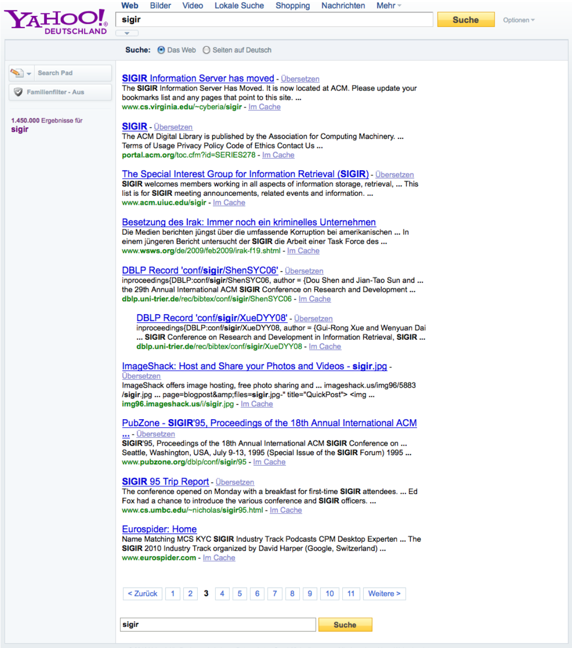
Contrary to that: stale pages and links.


Google
sigir
Suche
Google Instant ist aktiviert. ▼
Ungefähr 479.000 Ergebnisse (0,07 Sekunden)
Erweiterte Suche
Alles
Bilder
Videos
Blogs
Mehr
Riegelsberg
Standort ändern
Das Web
Seiten auf Deutsch
Seiten aus Deutschland
Alle
Letzte Woche
Alle Ergebnisse
Verwandte Suchanfragen
Wunderrad
Mehr Optionen
ACM SIGIR Special Interest Group on Information Retrieval Home Page
- [ Diese Seite übersetzen ]
"Addresses issues ranging from theory to user demands in the application of computers to the acquisition, organization, storage, retrieval, and distribution ...
www.sigir.org/ - Im Cache - Ähnliche Seiten
Publications SIG-IRList
Events The Forum
Resources General Information
Awards
Weitere Ergebnisse von sigir.org »
ACM SIGIR - Publications - [ Diese Seite übersetzen ]
The ACM SIGIR Conference focuses on research and development in information retrieval. It is the major international forum for the presentation of new ...
www.sigir.org/.../Proc-Browse.html - Im Cache - Ähnliche Seiten
Welcome to SIGIR | Home - [ Diese Seite übersetzen ]
The Office of the Special Inspector General for Iraq Reconstruction (SIGIR), a temporary federal agency serving the American public as a watchdog for fraud, ...
www.sigir.mil/ - Im Cache - Ähnliche Seiten
home [ACM SIGIR 2010] - 11. Nov. - [ Diese Seite übersetzen ]
26 Jul 2010 ... ACM-SIGIR 2010 was held at UniMail, Geneva, Switzerland between 19th and 23rd of July 2010. Thanks to all the participants! ...
www.sigir2010.org/ - Im Cache - Ähnliche Seiten
SIGIR - [ Diese Seite übersetzen ]
SIGIR 2003: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 28 - August 1, ...
www.informatik.uni-trier.de/.../sigir/index.html - Im Cache - Ähnliche Seiten
26. SIGIR 2003: Toronto, Canada - [ Diese Seite übersetzen ]
SIGIR 2003: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 28 - August 1, ...
www.informatik.uni-trier.de/.../sigir/sigir2003.html - Im Cache - Ähnliche Seiten
Weitere Ergebnisse anzeigen von informatik.uni-trier.de
SIGIR'08 - Singapore - [ Diese Seite übersetzen ]
24 Jul 2008 ... SIGIR is the major international forum for the presentation of new research results and for the demonstration of new systems and techniques ...
www.sigir2008.org/ - Im Cache - Ähnliche Seiten
SIGIR • 2005 - [ Diese Seite übersetzen ]
SIGIR is the major international forum for the presentation of new research results and the demonstration of new systems and techniques in the broad field ...
www.dcc.ufmg.br/eventos/sigir2005/ - Im Cache - Ähnliche Seiten
SIGIR - [ Diese Seite übersetzen ]
Browse SIGIR: Annual ACM Conference on Research and Development in ... 2010 : Proceeding of the 33rd international ACM SIGIR conference on Research and ...
portal.acm.org/browse_dl.cfm?linked=1&part=series&idx...coll...
David Isenberg: October 2010 SIGIR Quarterly Report - [ Diese Seite übersetzen ]
4 Nov 2010 ... The latest quarterly report from the Special Inspector General for Iraq Reconstruction (SIGIR), released October 30, has not received much ...
www.huffingtonpost.com/.../october-2010-sigir-quarte_b_779303.html - Im Cache
Goooooooooogle
1 2 3 4 5 6 7 8 9 10
Vorwärts
Anpassungen anzeigen Suchtipps Feedback geben
Google-Startseite Werben mit Google Unternehmensangebote Datenschutz Über Google


sigir
Suche
Google Instant ist aktiviert.
Seite 2 von ungefähr 479.000 Ergebnissen (0,10 Sekunden)
Erweiterte Suche
Alles
Bilder
Videos
Blogs
Mehr
Riegelsberg
Standort ändern
Das Web
Seiten auf Deutsch
Seiten aus Deutschland
Alle
Letzte Woche
Alle Ergebnisse
Verwandte Suchanfragen
Wunderrad
Mehr Optionen
sigir 2006 - [ Diese Seite übersetzen ]
30 Jul 2010 ... sigir 2006 .... Copyright © sigir 2006 | Dark Marble by CMS Developer Web. | Powered by Wordress. Return to pagetop.
www.sigir2006.org/ - Im Cache - Ähnliche Seiten
grants:google [ACM SIGIR 2010] - [ Diese Seite übersetzen ]
2 Beiträge - Letzter Eintrag: 26. Juli
To apply send, no later than 21 June 2010 , an e-mail to europe-events@google.com with the subject heading 'Google SIGIR 2010 Conference ...
www.sigir2010.org/doku.php?id=grants... - Im Cache - Ähnliche Seiten
[PDF] SIGIR 2010 - [ Diese Seite übersetzen ]
Dateiformat: PDF/Adobe Acrobat - Schnellansicht
7 Jul 2010 ... SIGIR Workshop on Feature Generation and Selection for Information Re- ...... SIGIR conference on Research and development in ...
alex.smola.org/workshops/sigir10/contrib/sigir.pdf
Federal News Radio 1500 AM: SIGIR: Defense can't account for $8.7 ... - [ Diese Seite übersetzen ]
27 Jul 2010 ... The Special Inspector General for Iraq Reconstruction (SIGIR) finds that only one Defense organization actually set up the accounts required ...
www.federalnewsradio.com/?nid=35&sid=2012362 - Im Cache
[PDF] SIGIR 2010 - [ Diese Seite übersetzen ]
Dateiformat: PDF/Adobe Acrobat - Schnellansicht
von V Carvalho - 2010 - Ähnliche Artikel
A Workshop at the 33rd Annual International ACM SIGIR Conference on ... Proceedings of the SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE ...
www.cs.cmu.edu/~vitor/papers/CSE2010-Proceedings.pdf
Welcome to The 34th Annual ACM SIGIR Conference - [ Diese Seite übersetzen ]
Welcome to the 34th Annual ACM SIGIR Conference Web Site ... Ricardo Baeza-Yates, Yahoo! Research ■ Tat-Seng Chua, National University of Singapore. SIGIR.
www.sigir2011.org/ - Im Cache
SIGIR'99 - SIMS - [ Diese Seite übersetzen ]
www2.sims.berkeley.edu/research/.../sigir99/ - Ähnliche Seiten
SIGIR - Wikipedia, the free encyclopedia - [ Diese Seite übersetzen ]
SIGIR may refer to: Special Inspector General for Iraq Reconstruction · Special Interest Group on Information Retrieval, a Special Interest Group (SIG) of ...
en.wikipedia.org/wiki/SIGIR - Im Cache - Ähnliche Seiten
SIGIR 2010: Query Representation and Understanding - [ Diese Seite übersetzen ]
SIGIR 2010: Query Representation and Understanding. Overview; Proceedings; Invited Speakers; Accepted Papers; Workshop Program; Call For Papers ...
ciir.cs.umass.edu/sigir2010/qru/ - Im Cache
[PDF] SIGIR 2010 - [ Diese Seite übersetzen ]
Dateiformat: PDF/Adobe Acrobat - Schnellansicht
SIGIR 2010. Geneva, Switzerland. July 19-23, 2010. Large-Scale Distributed Systems for Information Retrieval. Workshop of the 33 rd. Annual International ...
www.lsdsir.org/wp-content/uploads/2010/05/LSDSIR2010.pdf
Zurück 1 2 3 4 5 6 7 8 9 10 11 Vorwärts
Suchtipps Feedback geben
Google-Startseite Werben mit Google Unternehmensangebote Datenschutz Über Google


YAHOO! DEUTSCHLAND
Web Bilder Video Lokale Suche Shopping Nachrichten Mehr
sigir
Suche
Optionen
Suche: Das Web Seiten auf Deutsch
Search Pad
Familienfilter - Aus
1.450.000 Ergebnisse für sigir:
Alle Treffer anzeigen
Wikipedia
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-trier.de/~ley/db/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 ...
IT outsourcing ИТ Аутсорсинг СИГИР Обслуживание компьютеров в Санкт-Петербурге ... Надежную работу информационной системы в целом ...
www.sigir.ru/it - 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 - Singapore - Übersetzen
SIGIR 2008 ... SIGIR is the major international forum for the presentation of new research ... Powered by Check Pagerank © Copyright 2007-2008 ACM SIGIR ...
www.sigir2008.org - Im Cache
home [ACM SIGIR 2010] - Übersetzen
ACM SIGIR 2010. Home. News. Important dates. Participate. Registration ... 33rd Annual ACM SIGIR Conference. 19-23 July 2010, Geneva, Switzerland ...
www.sigir2010.org/doku.php - Im Cache
1 2 3 4 5 6 7 8 9 10 11 Weitere >
sigir
Suche
© 2010 Yahoo! Alle Rechte vorbehalten. Datenschutz - Geschäftsbedingungen - Missbrauch melden - Webseite eintragen
SuchScanBETA zeigt für die Nutzer der Yahoo! Suche optimierte Warnhinweise von McAfee und nicht alle verfügbaren Bewertungen des McAfee SiteAdvisors.


YAHOO! DEUTSCHLAND
Web Bilder Video Lokale Suche Shopping Nachrichten Mehr
sigir
Suche
Optionen
Suche: Das Web Seiten auf Deutsch
Search Pad
Familienfilter - Aus
1.450.000 Ergebnisse für sigir:
Alle Treffer anzeigen
Wikipedia
About SIGIR - Übersetzen
Awards Leadership Organizational Chart SIGIR FAQ Law ... SIGIR reports administratively to the Secretaries of State and Defense. ...
www.sigir.mil/about - Im Cache
SIGIR 2009 Archive | SIGIR'09 - Übersetzen
The SIGIR 2009 conference ran July 19-23, 2009, in Boston, Massachusetts, at the ... Click on any item for the archived details about SIGIR 2009. Sun ...
www.sigir2009.org - Im Cache
Conference venue | SIGIR'09 - Übersetzen
The 32nd Annual ACM SIGIR Conference July 19-23 2009. Home. Conference venue. SIGIR 2009 will be held in Boston, Massachusetts, at the Sheraton Boston Hotel ...
sigir2009.org/venue - Im Cache
Special Inspector General for Iraq Reconstruction - Übersetzen
The Office of the Special Inspector General for Iraq Reconstruction (SIGIR) was created in October 2004 as the successor to the Coalition Provisional Authority Office of Inspector General (CPA-IG). SIGIR is an independent government agency...
en.wikipedia.org/wiki/Special_Inspector_General_for_Iraq_Reconstruc... - Im Cache
SIGIR
SIGIR • 2005 - Übersetzen
SIGIR is the major international forum for the presentation of new research ... The 28th Annual International ACM SIGIR Conference will be held at the Pestana ...
www.dcc.ufmg.br/eventos/sigir2005 - Im Cache
SIGIR 2000 - Übersetzen
SIGIR 2001 (New Orleans, USA) To contact the SIGIR 2000 organizers, please email: sigir@aueb.gr ... [ Last updated: 07.07.2000. © 1999 - 2000 | Athens University ...
www.aueb.gr/conferences/sigir2000 - Im Cache
SIGIR 2001 Conference Home - Übersetzen
The 24th Annual International ACM SIGIR Conference on Research and Development ... Another workshop agenda is now available: Information Retrieval Techniques for ...
www2.sis.pitt.edu/~sigir01 - Im Cache
Sigir — Blogs, Pictures, and more on WordPress - Übersetzen
The Latest on Contractors From SIGIR ... SIGIR's Latest on PMCs ... SIGIR's Special General Advocates for New Civil-Military Office ...
en.wordpress.com/tag/sigir - 66k - Im Cache
Project Tripod - Übersetzen
Automatically ... Automating caption creation. The primary objective of Tripod is to ... Applying an innovative multidisciplinary approach Tripod will ...
sigir.shef.ac.uk - Im Cache
SIGIR Forum, Volume 39 - Übersetzen
The SIGIR 2005 workshop program. 15-16. Shlomo Argamon, Jussi Karlgren, ... ACM SIGIR workshop on mathematical/formal methods in information retrieval MF/IR ...
www.informatik.uni-trier.de/~ley/db/journals/sigir/sigir39.html - Im Cache
< Zurück 1 2 3 4 5 6 7 8 9 10 11 Weitere >
sigir
Suche



YAHOO! DEUTSCHLAND

Web Bilder Video Lokale Suche Shopping Nachrichten Mehr

sigir

Suche

Optionen

Suche: Das Web Seiten auf Deutsch

Search Pad

Familienfilter - Aus

1.450.000 Ergebnisse für **sigir**

**SIGIR** Information Server has moved - Übersetzen
The **SIGIR** Information Server Has Moved. It is now located at ACM. Please update your bookmarks list and any pages that point to this site. ...
www.cs.virginia.edu/~cyberia/sigir - Im Cache

**SIGIR** - Übersetzen
The ACM Digital Library is published by the Association for Computing Machinery. ... Terms of Usage Privacy Policy Code of Ethics Contact Us ...
portal.acm.org/toc.cfm?id=SERIES278 - Im Cache

The Special Interest Group for Information Retrieval (**SIGIR**) - Übersetzen
**SIGIR** welcomes members working in all aspects of information storage, retrieval, ... This list is for **SIGIR** meeting announcements, related events and information. ...
www.acm.uiuc.edu/sigir - Im Cache

Besetzung des Irak: Immer noch ein kriminelles Unternehmen
Die Medien berichten jüngst über die umfassende Korruption bei amerikanischen ... In einem jüngeren Bericht untersucht der **SIGIR** die Arbeit einer Task Force des ...
www.wsws.org/de/2009/feb2009/irak-f19.shtml - Im Cache

DBLP Record 'conf/**sigir**/ShenSYC06' - Übersetzen
inproceedings{DBLP:conf/**sigir**/ShenSYC06, author = {Dou Shen and Jian-Tao Sun and ... the 29th Annual International ACM **SIGIR** Conference on Research and Development ...
dblp.uni-trier.de/rec/bibtex/conf/sigir/ShenSYC06 - Im Cache

DBLP Record 'conf/**sigir**/XueDYY08' - Übersetzen
inproceedings{DBLP:conf/**sigir**/XueDYY08, author = {Gui-Rong Xue and Wenyuan Dai ... **SIGIR** Conference on Research and Development in Information Retrieval, **SIGIR** ...
dblp.uni-trier.de/rec/bibtex/conf/sigir/XueDYY08 - Im Cache

ImageShack: Host and Share your Photos and Videos - **sigir**.jpg - Übersetzen
ImageShack offers image hosting, free photo sharing and ... imageshack.us/img96/5883 /**sigir**.jpg ... page=blogpost&files=**sigir**.jpg-" title="QuickPost"> <img ...
img96.imageshack.us/i/sigir.jpg - Im Cache

PubZone - **SIGIR**'95, Proceedings of the 18th Annual International ACM ... - Übersetzen
**SIGIR**'95, Proceedings of the 18th Annual International ACM **SIGIR** Conference on ... Seattle, Washington, USA, July 9-13, 1995 (Special Issue of the **SIGIR** Forum) 1995 ...
www.pubzone.org/dblp/conf/sigir/95 - Im Cache

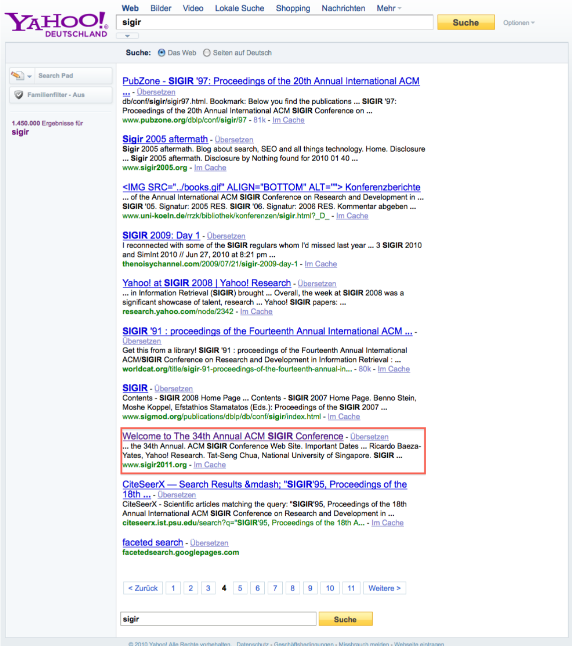
**SIGIR** 95 Trip Report - Übersetzen
The conference opened on Monday with a breakfast for first-time **SIGIR** attendees. ... Ed Fox had a chance to introduce the various conference and **SIGIR** officers. ...
www.cs.umbc.edu/~nicholas/sigir95.html - Im Cache

Eurospider: Home
Name Matching MCS KYC **SIGIR** Industry Track Podcasts CPM Desktop Experten ... The **SIGIR** 2010 Industry Track organized by David Harper (Google, Switzerland) ...
www.eurospider.com - Im Cache

< Zurück 1 2 3 4 5 6 7 8 9 10 11 Weitere >

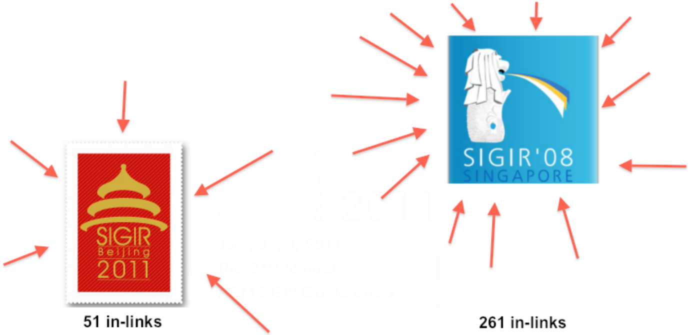
sigir

Suche


YAHOO! DEUTSCHLAND
Web Bilder Video Lokale Suche Shopping Nachrichten Mehr
sigir
Suche
Optionen
Suche: Das Web Seiten auf Deutsch
Search Pad
Familienfilter - Aus
1.450.000 Ergebnisse für sigir
PubZone - SIGIR '97: Proceedings of the 20th Annual International ACM ... - Übersetzen
db/conf/sigir/sigir97.html. Bookmark: Below you find the publications ... SIGIR '97: Proceedings of the 20th Annual International ACM SIGIR Conference on ...
www.pubzone.org/dblp/conf/sigir/97 - 81k - Im Cache
Sigir 2005 aftermath - Übersetzen
Sigir 2005 aftermath. Blog about search, SEO and all things technology. Home. Disclosure ... Sigir 2005 aftermath. Disclosure by Nothing found for 2010 01 40 ...
www.sigir2005.org - Im Cache
<IMG SRC="../books.gif" ALIGN="BOTTOM" ALT=""> Konferenzberichte
... of the Annual International ACM SIGIR Conference on Research and Development in ... SIGIR '05. Signatur: 2005 RES. SIGIR '06. Signatur: 2006 RES. Kommentar abgeben ...
www.uni-koeln.de/rrzk/bibliothek/konferenzen/sigir.html?_D_ - Im Cache
SIGIR 2009: Day 1 - Übersetzen
I reconnected with some of the SIGIR regulars whom I'd missed last year ... 3 SIGIR 2010 and SimInt 2010 // Jun 27, 2010 at 8:21 pm ...
thenoisychannel.com/2009/07/21/sigir-2009-day-1 - Im Cache
Yahoo! at SIGIR 2008 | Yahoo! Research - Übersetzen
... in Information Retrieval (SIGIR) brought ... Overall, the week at SIGIR 2008 was a significant showcase of talent, research ... Yahoo! SIGIR papers: ...
research.yahoo.com/node/2342 - Im Cache
SIGIR '91 : proceedings of the Fourteenth Annual International ACM ... - Übersetzen
Get this from a library! SIGIR '91 : proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval : ...
worldcat.org/title/sigir-91-proceedings-of-the-fourteenth-annual-in... - 80k - Im Cache
SIGIR - Übersetzen
Contents - SIGIR 2008 Home Page ... Contents - SIGIR 2007 Home Page. Benno Stein, Moshe Koppel, Efstathios Stamatatos (Eds.): Proceedings of the SIGIR 2007 ...
www.sigmod.org/publications/dblp/db/conf/sigir/index.html - Im Cache
Welcome to The 34th Annual ACM SIGIR Conference - Übersetzen
... the 34th Annual. ACM SIGIR Conference Web Site. Important Dates ... Ricardo Baeza-Yates, Yahoo! Research. Tat-Seng Chua, National University of Singapore. SIGIR ...
www.sigir2011.org - Im Cache
CiteSeerX — Search Results — "SIGIR'95, Proceedings of the 18th ... - Übersetzen
CiteSeerX - Scientific articles matching the query: "SIGIR'95, Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in ...
citeseerx.ist.psu.edu/search?q="SIGIR'95, Proceedings of the 18th A... - Im Cache
faceted search - Übersetzen
facetedsearch.googlepages.com
< Zurück 1 2 3 4 5 6 7 8 9 10 11 Weitere >
sigir
Suche

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)

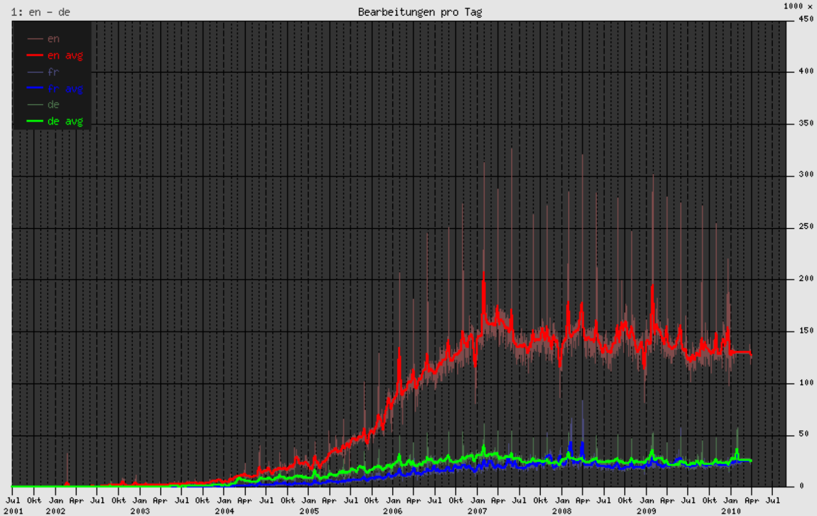



Figure: Wikipedia Statistic: Average Changes per Day (Source: Wikipedia)

| # | Seite | Namensräume | Bearb. (30 T.) | Bearb. |
|---|---|---|---|---|
| 1 | Wikipedia:Vandalismusmeldung | 4 | 3891 | 3891 |
| 2 | Wikipedia:Auskunft | 4 | 3079 | 3079 |
| 3 | Wikipedia:Auskunft/alt6 | 4 | 2647 | 5143 |
| 4 | Wikipedia:Fragen zur Wikipedia | 4 | 2582 | 3660 |
| 5 | Wikipedia:Kandidaten für exzellente Bilder/Archiv2 | 4 | 1737 | 4843 |
| 6 | Wikipedia:Löschprüfung | 4 | 1525 | 2053 |
| 7 | Wikipedia:Vandalismusmeldung/Alt11 | 4 | 1137 | 4823 |
| 8 | Wikipedia:Löschkandidaten/2. Juni 2009 | 4 | 972 | 973 |
| 9 | Wikipedia:Kandidaten für lesenswerte Artikel/alt2 | 4 | 938 | 4775 |
| 10 | Wikipedia:Importwünsche | 4 | 880 | 4524 |
| 11 | Wikipedia:Löschkandidaten/8. Juni 2009 | 4 | 849 | 850 |
| 12 | Wikipedia:Löschkandidaten/13. Juni 2009 | 4 | 842 | 842 |
| 13 | Wikipedia:Löschkandidaten/24. Juni 2009 | 4 | 830 | 830 |
| 14 | Diskussion:Air-France-Flug 447 | 1 | 814 | 814 |
| 15 | Wikipedia:Spielwiese | 4 | 790 | 872 |
| 16 | Wikipedia:Löschkandidaten/19. Juni 2009 | 4 | 772 | 772 |
| 17 | Wikipedia:Löschkandidaten/11. Juni 2009 | 4 | 764 | 765 |
| 18 | Wikipedia:Administratoren/Anfragen | 4 | 747 | 3710 |
| 19 | Wikipedia:Löschkandidaten/23. Juni 2009 | 4 | 744 | 744 |
| 20 | Wikipedia:Persönliche Bekanntschaften/neue Einträge | 4 | 742 | 9519 |

Figure: Wikipedia Statistic: Top 20 Wikipedia Sites (Germany) with the most changes from 01.06.2009 to 30.06.2009 (Source: Wikipedia)


Figure: Picture from Flickr

# Overview

PageRank [Lawrence Page & Sergey Brin, 1998]

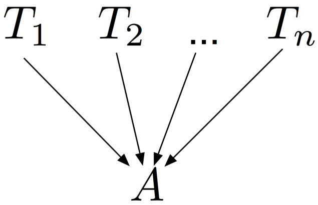
- is the most popular ranking algorithm.
- concept is based on a site's in-links (links pointing to the specific site).
- represents the probability that a random web-surfer model enters a page.

# Moves of the Random Web Surfer Model

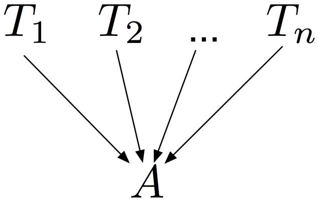
In every step the random web surfer model either follows an out-linked page **OR** does a random jump to any page.

# The PageRank Algorithm

Assuming n pages $T_1$, $T_2$, $\cdots$, $T_n$ link to a page A.



# The PageRank Algorithm




Assuming n pages $T_1$, $T_2$, $\cdots$, $T_n$ link to a page A.

Then

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

with

N: # of overall pages

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

$C(T_i)$: # of out-links of $T_i$, $1 \leq i \leq n$.

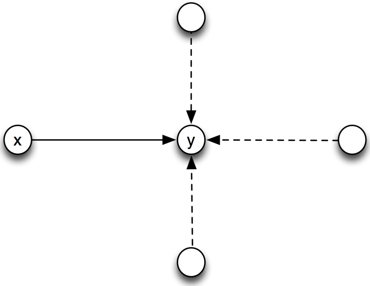
No freshness of links and pages is considered here!

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.

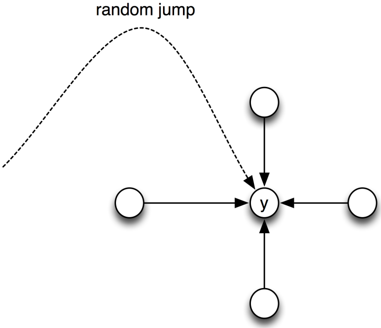
The probability that a random surfer is clicking at page x on an out-link to page y is a weighted combination of

- 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

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

with

$\varepsilon$: probability that a random surfer is clicking on an out-link

$(1-\varepsilon)$: 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

$\varepsilon$: probability that a random surfer is clicking on an out-link

$(1-\varepsilon)$: probability that a random surfer makes a random jump to a page.

$s(y)$: the random jump probability

$t(x,y)$: 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.

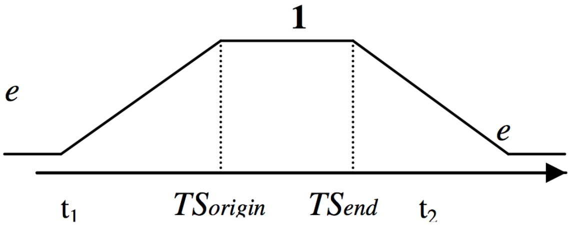



Figure: Temporal proximity window for a temporal interval

# The T-Rank Algorithm

T-Rank is based on a good approach, BUT:

Activities at different time points are not distinguished.

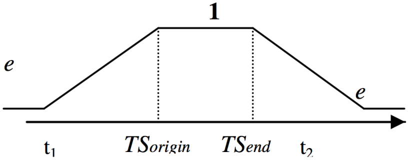



Figure: Temporal proximity window for a temporal interval

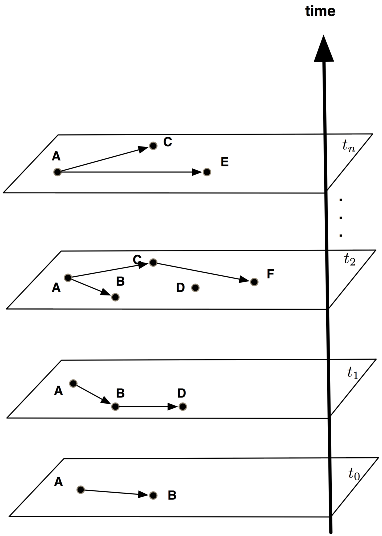
Instead, influence of activities on web freshness should decay over time.

# Overview

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 $t_i$:

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

# The Web Freshness Computation

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

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

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

# The Web Freshness Computation

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

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

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

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

# In-Link Activites

| Link activity | | Infl. on $p$'s InF | Gain of $p$'s InF ($\omega_j$) |
|---|---|---|---|
| 1 | creation of link $l : q \rightarrow p$ | $\uparrow\uparrow\uparrow$ | 3 |
| 2 | update on link $l : q \rightarrow p$ (changed anchor) | $\uparrow\uparrow$ | 2 |
| 3 | update on link $l : q \rightarrow p$ (unchanged anchor) | $\uparrow$ | 1.5 |
| 4 | removal of link $l : q \rightarrow p$ | $\downarrow\downarrow$ | -0.5 |

# Computation of $\Delta InF(p)|_{t_{i-1}}^{t_i}$

$$\Delta InF_0(p)|_{t_{i-1}}^{t_i} = \sum_{l:q \to p} \sum_{j \in LA} \omega_j C_j(l)$$

with

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

# Computation of $\Delta InF(p)|_{t_{i-1}}^{t_i}$

$$\Delta InF_0(p)|_{t_{i-1}}^{t_i} = \sum_{l:q\to p} \sum_{j\in LA} \omega_j C_j(l)$$

$$\Delta InF(p)|_{t_{i-1}}^{t_i} = \lambda_{InF}\, \Delta InF_0(p)|_{t_{i-1}}^{t_i} + (1-\lambda_{InF}) \sum_{l:q\to p} m_{qp}\, \Delta InF(q)|_{t_{i-1}}^{t_i}$$

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

$\omega_j^{'}$

| Page activity | | Infl. on $q$'s PF | Gain of $q$'s PF |
|---|---|---|---|
| 1 | creation of page $q$ | $\uparrow\uparrow\uparrow$ | 3 |
| 2 | update on page $q$ | $\uparrow$ | 1.5 |
| 3 | removal of page $q$ | $\downarrow\downarrow$ | -0.5 |

# Computation of $\Delta PF(q)|_{t_{i-1}}^{t_i}$

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

with

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

# Computation of $\Delta PF(q)|_{t_{i-1}}^{t_i}$

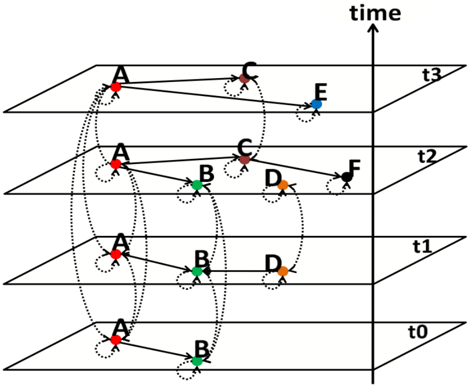
$$\Delta PF_0(q)|_{t_{i-1}}^{t_i} = \sum_{j \in PA} \omega_j^{'} C_j^{'}(q)$$

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

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?



# Extended Temporal Random Surfer Model

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.

# Moves of the Surfer

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

# 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)

# 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.

# 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:

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

# 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:

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

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:

# Computation of $\pi_{p,i}$

We need some probability values:

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

## Computation of $\pi_{p,i}$

We need some probability values:

$$P_{t_j}(Follow|q) = (1-d) \qquad P_{t_j}(Jump|q) = d$$

$$P_{t_j}(p|q, Follow) = F_{t_j}(p,q) \qquad P_{t_j}(p|q, Jump) = \frac{1}{N_{t_j}}$$

with

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

$N_{t_j}$ : total # of pages at $t_j$

# Computation of $\pi_{p,i}$

Due to the fact, that fresh web resources are likely to attract surfer's attention, we define:

$$F_{t_j}(p,q) = \frac{PF_{t_j}(p)}{\sum\limits_{p':q \rightarrow p'} PF_{t_j}(p')}$$

# Computation of $\pi_{p,i}$

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

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

# Computation of $\pi_{p,i}$

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

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

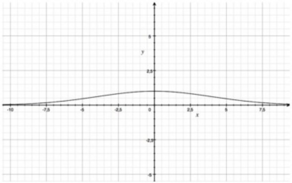
with

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

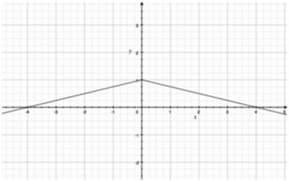
$V_j$: set of pages at time point $t_j$

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

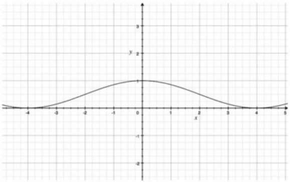
**Kernels for |T|=4**



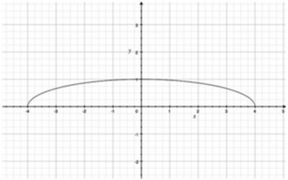
Gaussian Kernel



Triangle Kernel



Cosine Kernel



Circle Kernel

# Computation of $\pi_{p,i}$

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

# Computation of $\pi_{p,i}$

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

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

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 $t_i$:

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

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

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 $\mu_{p,i}$

$$\mu_{p,i} = \sum_{t_j \in T'_{t_i}} \omega'(t_i, t_j) InF(p)_{t_j}$$

with

$T'_{t_i}$: Set of snapshots included in the sliding window centered on $t_i$

$\omega'(t_i, t_j) = \frac{1}{|T'_{t_i}|}$ 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

# 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

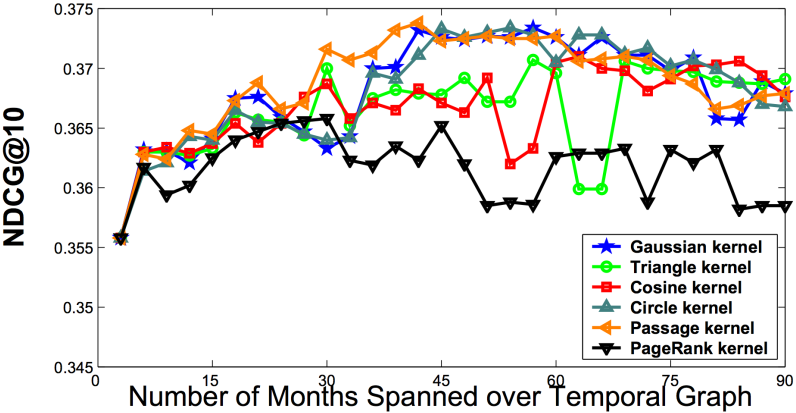
- Used archival web pages in the .ie domain (collected from January 200 to December 2007)
- Corpus contained 158 million web pages and 12 billion temporal links
- Each month one snapshot
- Time period of interest was April 2007.

- Ninety queries.

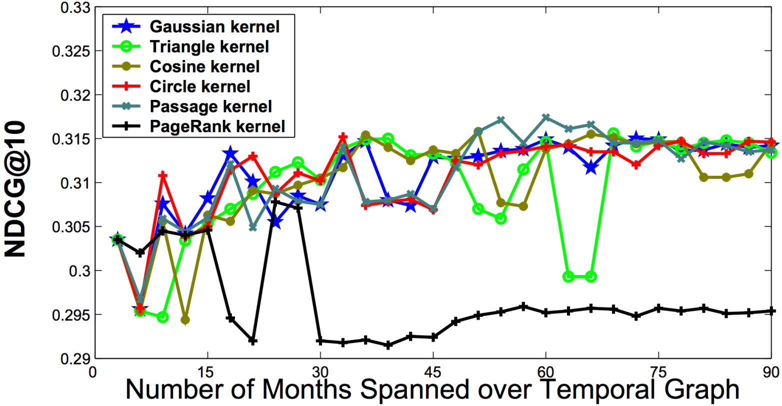
- Ninety queries.
- For each query 84.6 URLs on average were judged according to freshness and relevance by human editors.

- 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




(b) Freshness performance: NDCG@10

| Relevance | | | | |
|---|---|---|---|---|
| Method | P@10 | NDCG@3 | NDCG@5 | NDCG@10 |
| BM25 | 0.4695 | 0.2478 | 0.2740 | 0.3344 |
| PageRank | 0.4894 | 0.2589 | 0.2840 | 0.3457 |
| BuzzRank | 0.4770 | 0.2770 | 0.2980 | 0.3460 |
| TemporalRank | 0.4841 | 0.2706 | 0.2875 | 0.3524 |
| TimedPageRank | 0.5031 | 0.2830 | 0.3063 | 0.3587 |
| T-Random | 0.4904 | 0.2690 | 0.2877 | 0.3495 |
| T-rank | 0.4875 | 0.2669 | 0.2870 | 0.3496 |
| T-Fresh(1,1,30) | **0.5051** | **0.3229** | **0.3347** | **0.3729** |
| **Freshness** | | | | |
| Method | P@10 | NDCG@3 | NDCG@5 | NDCG@10 |
| BM25 | 0.3138 | 0.2137 | 0.2379 | 0.2805 |
| PageRank | 0.3325 | 0.1946 | 0.2345 | 0.2838 |
| BuzzRank | 0.3327 | 0.2043 | 0.2234 | 0.2797 |
| TemporalRank | **0.3473** | 0.2312 | 0.2510 | 0.2992 |
| TimedPageRank | 0.3398 | **0.2443** | 0.2514 | 0.2972 |
| T-Random | 0.3316 | 0.2054 | 0.2403 | 0.2879 |
| T-rank | 0.3356 | 0.2269 | 0.2498 | 0.2950 |
| T-Fresh(1,1,30) | 0.3412 | 0.2411 | **0.2662** | **0.3076** |

# 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

| Good | Bad |
|---|---|
| Nice background informations<br><br>Good explanations of the ideas | Wrong formulas<br><br>Improvable detailed explanations |

Thank you for your attention!