Social Web Search

Filippo Menczer
Informatics and Computer Science

Leveraging online social behavior for collaborative Web search
Nan: Networks & agents Network
Outline

- Part 1: GiveALink.org
- Part 2: 6S
social search

donate your bookmarks

share the missing links

personalized recommendations

social search
Today's Personalized Recommendations

kuro5hin.org || technology and culture, from the trenches
The Register: Sci/Tech News for the World
Auditor main - Hackingdefined
de.lirio.us: Store, share and tag your favourite links
Gizmodo, The Gadget Guide
anandgraves.com
ExtremeTech - Deep technology for enthusiasts and professionals
Questionable Content: New comics every Monday through Friday
Neil Gaiman - Home
eWEEK.com - Enterprise Technology News and Reviews

Click Here for New Recommendations
I donated my bookmarks to science. Did you?
\[ \sigma(x, y) = \frac{1}{N} \sum_{u=1}^{N} 2 \log \left( \frac{|F_u[a(x,y)]|}{|R_u|} \right) \]

\[ + \log |F_u(x)| \frac{|R_u|}{|F_u(y)|} + \log |F_u(y)| \frac{|R_u|}{|F_u(x)|} \]
I donated my bookmarks to science. Did you?
Fraction of URLs: De
gree
$s_{min} = 0$
$s_{min} = 0.0025$
$s_{min} = 0.0055$
$Pr(k) \sim \exp(-0.002*k)$
$Pr(k) \sim k^{-2}$
Related URLs sorted by Related URL List

1-10 of 1000 results for http://www.cnn.com

**BBC NEWS | News Front Page**
Find Similar Results

Los Angeles, California, national and world news, jobs, real estate, cars - Los Angeles Times
Find Similar Results

Find Similar Results

CNET News.com -- Technology news and business reports
Find Similar Results
User Study

Markines et al, AAAI 2006
Centrality & Prestige
Ranking & Recommendation by Novelty

$$\nu(x, y) = \left[ 1 + \min_{z \in U} \left( \frac{1}{\sigma(x, z)} + \frac{1}{\sigma(z, y)} - 2 \right) \right]^{-1}$$
\[ \eta(x, A) = \max_{y \in A} \left[ \sigma(x, y) \cdot \log \left( \frac{N}{N(y)} \right) \right] \]
# 3 Ways to Share

<table>
<thead>
<tr>
<th>Way</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Donate Bookmarks</td>
<td><strong>Donate as registered user</strong> to get personalized recommendations and search results based on your bookmarks. <strong>Donate anonymously</strong> to help us without giving us your email address; we will not be able to personalize your searches, nor give you personalized recommendations.</td>
</tr>
</tbody>
</table>
| 2. Import Bookmarks from del.icio.us | username: fil@indiana.edu  
password: ******  
[Import] |
| 3. Bookmarklet | Install the bookmarklet by dragging the link to the toolbar of your browser; then click when visiting a page you would like to donate: **Give This Link**  
Install the pop-up bookmarklet by dragging the link to the toolbar of your browser; then click when visiting a page you would like to donate: **Give This Link Pop Up** |
collaborative filtering, social semantic similarity, unlinked pages, multimedia content, trust

scalability, density

spam

Questions?
Outline

- Part 1: GiveALink.org
- Part 2: 6S
http://homer.informatics.indiana.edu/6S/
http://homer.informatics.indiana.edu/6S/
KDnuggets: Data Mining, Web Mining, and Knowledge Discovery Guide
Newsletter on the data mining and knowledge industries, offering information on data mining, knowledge discovery, text mining, and web mining software, courses, jobs, publications, and meetings.
http://www.kdnuggets.com/
Contributors: Main6S#0,wls_angel#3,wls_iceman#1
Similar pages (Power by GiveALink)

Data Mining - Home Page (Misc)
... in an unrestricted hands-on web. Software Analytical Processing (OLAP), Data Mining ...
http://www.the-data-mine.com/
Contributors: Unknown#7,wls_angel#3,wls_iceman#1
Similar pages (Power by GiveALink)

Data Mining Software in the Yahoo! Directory
... for IT professionals focusing on data mining, data analysis, and reports ... to Business > Computers > Software > Databases > Data ...
Contributors: Unknown#7,wls_iceman#1
Similar pages (Power by GiveALink)

UMBC Agent Web -- news and information on software agent technology
... UMBC Agent Web -- news and information ...
http://agents.umbc.edu/Topics/Related_Topics/Information_retrieval_and_knowledge_management/
Contributors: Unknown#7,wls_iceman#1
Similar pages (Power by GiveALink)

BUBL LINK: Information retrieval
... and Forward Knowledge Approach Data Mine: Data Mining and Knowledge Discovery ... the Gaps KD Mine: Data ...
http://bUBL.ac.uk/link/i/informationretrieval.htm
Contributors: Unknown#7,wls_iceman#1
Similar pages (Power by GiveALink)

Library and Information Science > Information Retrieval in the Yahoo! Directory
Done
6S: Collaborative Peer Search

Data mining & referral opportunities

Emerging communities

query

hit

query

hit

WWW2004
WWW2005
WTAS2005
P2PIR2006
Algorithm 3: Reinforcement
Query Routing

\[ \sigma(p, \text{Query}) = \sum_{i \in \text{Query}} [\alpha \cdot w_{p,i}^f + (1 - \alpha) \cdot w_{p,i}^e] \]

N top ranked among known peers are selected as neighbors and sent Query.

Forwarding the query

<table>
<thead>
<tr>
<th>Peer</th>
<th>Term</th>
<th>(w_{p,i}^f)</th>
<th></th>
<th>Peer</th>
<th>Term</th>
<th>(w_{p,i}^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>lama</td>
<td>.98</td>
<td></td>
<td>10</td>
<td>library</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>book</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simulating 500 Users ODP (dmoz.org)

Distributed vs Centralized
Small-world
Ongoing Work

- Improve coverage/diversity in query routing algorithm
- Spam protection: trust/reputation subsystem
- User study with 6S application
User study
Query network
Result network
THANK YOU

http://homer.informatics.indiana.edu/~nan/