Social Web Search

Filippo Menczer
Informatics and Computer Science

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

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

• Part 1: GiveALink.org

• Part 2: 6S
donate your bookmarks

share the missing links
donate your bookmarks

share the missing links

personalized recommendations
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
Markines et al, WWW 2006
Markines et al, WWW 2006
Markines et al, WWW 2006
Markines et al, WWW 2006
Markines et al, WWW 2006
I donated my bookmarks to science. Did you?
I donated my bookmarks to science. Did you?
I donated my bookmarks to science. Did you?
I donated my bookmarks to science. Did you?

news

research

fun

CNN.com

AAAI

σ_A + σ_B = σ

★☆☆☆☆

★☆☆☆☆

★★★★☆
\[ \sigma(x, y) = \frac{1}{N} \sum_{u=1}^{N} 2 \log \left( \frac{|F_u[a(x,y)]|}{|R_u|} \right) + \log \left| F_u(x) \right| + \log \left| F_u(y) \right| \]
I donated my bookmarks to science. Did you?
I donated my bookmarks to science. Did you?
I donated my bookmarks to science. Did you?
Fraction of URLs
Degree
$s_{\text{min}} = 0$
$s_{\text{min}} = 0.0025$
$s_{\text{min}} = 0.0055$
$\Pr(k) \sim \exp(-0.002 \cdot k)$
$\Pr(k) \sim k^2$
Fraction of URLs

Degree

$s_{min} = 0$
$s_{min} = 0.0025$
$s_{min} = 0.0055$

$Pr(k) \sim \exp(-0.002*k)$

$Pr(k) \sim k^{-2}$
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
User Study

Markines et al, AAAI 2006
Centrality & Prestige
\[ c(x) = \frac{1}{|U|} \sum_{y \in U} \left[ 1 + \min_{x \sim y} \sum_{(u,v) \in x \sim y} \left( \frac{1}{\sigma(u,v)} - 1 \right) \right]^{-1} \]

Centrality & Prestige

\[ p_{t+1}(x) = (1 - \alpha) + \alpha \cdot \sum_{y \in U} \frac{\sigma(x, y) \cdot p_t(y)}{\sum_{z \in U} \sigma(y, z)} \]
Ranking & Recommendation by Novelty

\[ \nu(x, y) = \left[ 1 + \min_{z \in U} \left( \frac{1}{\sigma(x, z)} + \frac{1}{\sigma(z, y)} - 2 \right) \right]^{-1} \]
\[ \eta(x, A) = \max_{y \in A} \left[ \sigma(x, y) \cdot \log \left( \frac{N}{N(y)} \right) \right] \]
REST (XML) Web service

Navigation
I donated my bookmarks to science. Did you?

**givealink.org**

Bookmark Donations for Science


similarity: 0.00177462

**REST (XML) Web service**

**Navigation**
3 Ways to Share

1. Donate Bookmarks
   - **Donate as registered user** to get personalized recommendations and search results based on your bookmarks
   - **Donate anonymously** to help us without giving us your email address; we will not be able to personalize your searches, nor give you personalized recommendations

2. Import Bookmarks from del.icio.us
   - username: fil@indiana.edu
   - password: ********

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**

Shared data
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/
6S: Collaborative Peer Search

- WWW
- bookmarks
- Crawler
- Index
- Peer
- local storage

A

query

B

query

C

query

query

query

query

query

WWW2004
WWW2005
WTAS2005
P2PIR2006
6S: Collaborative Peer Search

Data mining & referral opportunities
6S: Collaborative Peer Search

Data mining & referral opportunities

Emerging communities

WWW2004
WWW2005
WTAS2005
P2PIR2006

Crawler

Index

local storage

bookmarks

query

hit

query

query

query

query

hit

hit
Algorithm 3: Reinforcement
**Query Routing**

Let's consider a query for a library book named "Lama". The query routing process involves calculating the relevance of the query to a peer's profile, and then forwarding the query to the most relevant peers.

The formula for calculating the relevance score, \( \sigma(p, \text{Query}) \), is given by:

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

where
- \( \alpha \) is a weight parameter,
- \( w_{p,i}^f \) is the weight of term \( i \) in the focused profile of peer \( p \),
- \( w_{p,i}^e \) is the weight of term \( i \) in the expanded profile of peer \( p \).

### Focused Profile

<table>
<thead>
<tr>
<th>Peer</th>
<th>Term</th>
<th>( w_{p,i}^f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>lama</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Expanded Profile

<table>
<thead>
<tr>
<th>Peer</th>
<th>Term</th>
<th>( w_{p,i}^e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>library</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>book</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The N top ranked among known peers are selected as neighbors and sent the query. Forwarding the query to these peers.
Simulating 500 Users

ODP (dmoz.org)

Simulating 500 Users

ODP (dmoz.org)

Simulating 500 Users

ODP (dmoz.org)

P@10
Small-world
Semantic Similarity
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
Questions?

http://homer.informatics.indiana.edu/~nan/6S/

6S-Six Degree of Separation Search System

Project Abstract

6Search (6S), a collaborative peer network application, is aimed to address the scalability limitations of centralized search engines. Each peer within the 6S network crawls the Web in a focused way, guided by its user's information context. This way better (distributed) coverage can be achieved. Each peer also acts as a search "servent" by submitting and responding to queries to/from its neighbors. This search process has no centralized bottleneck. Peers depend on a local adaptive routing algorithm to dynamically change the topology of the peer network and search for the best neighbors to answer their queries.

The key idea of the proposed peer search network is that the flooding problem can be alleviated by intelligent collaboration between the peers. This should lead to an emergent clustered topology in which neighbor communities tend to form according to clusters of peers with shared interests and domains. In fact we predict that the ideal topology for such a network would be a "small world".

Acknowledgments:

We are grateful to the Nutch Organization for its open source search engine code. This work was supported in part by NSF Career Grant IIS-0348940 to FM.

Members:

- Menczer, Filip
- Akvipt, Rui
- Maguitman, Ana Gabriela
- Shin, Yeongsang
- Wu, Le-Shin
THANK YOU

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