Social Web Search

Filippo Menczer
Informatics and Computer Science

Leveraging online social behavior for collaborative Web search
Outline

• Part 1: GiveALink.org

• Part 2: 6S
\[ \sigma(x, y) = \frac{1}{N} \sum_{u=1}^{N} \left( \frac{2 \log \left( \frac{|F_u[a(x,y)]|}{|R_u|} \right)}{\log |F_u(x)| + \log |F_u(y)|} \right) \]
Fraction of URLs

Degree

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

$\Pr(k) \sim \exp(-0.002*k)$  
$\Pr(k) \sim k^{-2}$
Link Recommender

Enter URL or keyword to get related bookmarks. (More Options)

Search for: webir.org

Sort by Bookmark: • Similarity • Novelty • Prestige

Similar URLs sorted by Similarity

1-10 of 360 recommendations for http://www.webir.org/

http://www.parc.xerox.com/istl/groups/iea/dynamics.shtml
http://www.parc.xerox.com/istl/groups/iea/dynamics.shtml

http://www.springer.de/cgi-bin/search_book.pl?isbn=3-540-65112-8
http://www.springer.de/cgi-bin/search_book.pl?isbn=3-540-65112-8

Web Research Collections - Web Track
http://es.csiro.au/TRECWeb/

DIMACS Workshop on Internet and WWW Measurement, Mapping and Modeling
http://dimacs.rutgers.edu/Workshops/Internet/

http://www.ibm.com/java/fetuccino/
http://www.ibm.com/java/fetuccino/

Web Term Document Frequency Form
http://ellb.cs.berkeley.edu/docfreq/index.html

Finding Out About
http://www.cs.ucsd.edu/~rik/foa/

Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace
http://sherlock.berkeley.edu/asis96/asis96.html

Terabyte TREC Homepage
http://www-nlpir.nist.gov/projects/terabyte/

http://www.neci.nj.nec.com/homepages/lawrence/websize.html
http://www.neci.nj.nec.com/homepages/lawrence/websize.html
User Study

Precision vs. Recall plot for Google related and GiveALink results. The GaL and Google sets are shown with overlapping regions, indicating the relevance of the results.

Google related
GiveALink results

Recall

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

0.5

0.4

0.3

0.2

0.1

0

User Study
\[ c(x) = \frac{1}{|U|} \sum_{y \in U} \left[ 1 + \min_{x \sim y} \sum_{(u,v) \in x \sim y} \left( \frac{1}{\sigma(u,v)} - 1 \right) \right]^{-1} \]

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

\[
\nu(x, y) = \left[ 1 + \min_{z \in U} \left( \frac{1}{\sigma(x, z)} + \frac{1}{\sigma(z, y)} - 2 \right) \right]^{-1}
\]
\[ \eta(x, A) = \max_{y \in A} \left[ \sigma(x, y) \cdot \log \left( \frac{N}{N(y)} \right) \right] \]
collaborative filtering, social semantic similarity, unlinked pages, multimedia content, trust, scalability, density, spam
Questions?
Outline

- Part 1: GiveALink.org
- Part 2: 6S (next semester)