PROS: a Personalized Ranking Platform for Web Search

P.-A. Chirita, D. Olmedilla, W. Nejdl
L3S Research Center

3rd Int. Conf. on Adaptive Hypermedia
Outline

- Motivation
- Background
- PROS
- Experiments
- Conclusions & Further Work
Motivation

- Current web search engines return the same web pages to any user

- Existing ranking algorithms do not take into account user knowledge, preferences or interests (global ranking)
  - Only basic personalization techniques like presenting a user interface in Spanish if access if from a Spanish IP

- One ranking algorithm for the whole set of people using a web search engine does not seem to be enough
PageRank

- Page score based on the link structure of the web

- Used to provide relevant answers to user queries (it measures page popularity)

- It has a personalization vector

\[ PR(p) = (1-c) \sum_{q \in O_p} \frac{PR(q)}{O_q} + cE(p) \]

- Computationally expensive: not possible to make the whole computation for each user
Topic-sensitive PageRank

- Computes 16 PageRank vectors oriented on 16 different topics taken from the Open Directory Project (ODP)
- A similarity score with these topics is computed for each query
- The 16 vectors are then biased into one using the previous scores
- The resulting vector is used to rank the query answers
Scaling Personalized Web Search (I)

- Hubs: pages pointing to many important pages

- Compute one Personalized PageRank Vector for each user (PPV) →
  - Challenges:
    - Reduce storage required
    - Reduce time for computation

- Each PPV corresponding to a Preference Set P can be expressed as a linear combination of Basis Hub Vector

- Decomposes each Basis Hub Vector in two parts:
  - Hub skeleton vector (common interrelationships and precomputed)
  - Partial vector (unique values and computed at construction-time)
Scaling Personalized Web Search (& II)

- Hub pages
- Ordinary pages
- Hub Set (H)
- Preference Set (P)
Inferring a Preference Set

■ Personalization relies on user’s ability to choose a good Preference Set
  ▪ High quality hubs which match his preferences

■ This process can be automated:
  ▪ Information collected from the user can be used to derive his Preference Set

■ User does not even need to know what is a hub
PROS

A Personalized Ranking Platform

- Exploit user’s interests to automatically construct his Preference Set (profile)
- Use this Preference Set to compute the ranks
  - Different ranks to different users
- Sort results according to personalized ranks
Exploiting user information

- User’s interests are determined by
  - Most surfed pages
  - User’s bookmarks

- A proxy tracks user behaviour
  - Pages visited
  - Time user was looking at each page
  - Other information can be further develop
  - Current assumption
    - One user for IP address
HubFinder

- We get a set of pages from the user but
  - They are not highly ranked hubs

- HubFinder is an algorithm to find related web pages

- It allows pluggable filtering mechanisms

- We use HubRank to find highly rated hubs related to a given initial set of pages

- User web pages → set of related highly rated hubs
Summary of the process (I)

1. Crawl the Web
2. Calculate Ranks
3. Get bookmarks from user
4. Get user’s surfing information (Proxy)
5. Find related important hubs (HubFinder)
6. Construction of preference set & the hub set
7. Apply personalized ranks
Summary of the process (II)
Experiments (I)

- Crawl with 3,000,000 web pages
- 30 bookmarks
  - 15 on architecture
  - 7 on traveling
  - 6 on software
  - 2 on sports
- 78 selected surfed pages
- Computed 1300 pages as hub set
## Experiments (II)

<table>
<thead>
<tr>
<th>Query Keywords</th>
<th>PageRank</th>
<th>PPR</th>
<th>PROS</th>
</tr>
</thead>
<tbody>
<tr>
<td>architecture</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>building</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Paris</td>
<td>6</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>park</td>
<td>6</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>surf</td>
<td>3</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>23</strong></td>
<td><strong>5</strong></td>
<td><strong>22</strong></td>
</tr>
</tbody>
</table>
Experiments (& III)
Conclusions & further work

- Conclusions
  - Personalized Web Search Platform
  - Personalized ranking of Web Pages
  - Profile based on user preferences
    - Bookmarks
    - Most surfed pages

- Further work
  - Increase the size of the crawl
  - Increase user information:
    - E.g. user feedback
  - More experiments and evaluation
References

- C.J. Keith van Rijsbergen. 
  *Information Retrieval*

- S. Brin and L. Page
  *The Pagerank Citation Ranking: Bringing Order to the Web*

- Glen Jeh, Jennifer Widom
  *Scaling Personalized Web Search*

- Paul-Alexandru Chirita, Daniel Olmedilla, Wolfgang Nejdl
  *Finding Related Pages Using the Link Structure of the WWW*
  IEEE/WIC/ACM Int. Conference on Web Intelligence, Sep. 2004, Beijing, China
Thank You!

Questions •