When Experts Agree: Using Non-Affiliated Experts to Rank Popular Topics

Krishna Bharat
George A. Mihaila

presentation by Na Dai
Motivation

• Queries in popular topics
  – Too many related web pages

• Tradition
  – Content analysis
  – Authority? Spam?

• Use link structure
  – Basis of web pages’ authorities
Hilltop Algorithm Overview

- Select experts
- Identify relevant links
- Rank target web pages

- Expert lookup
  - Special inverted indexing
  - Compute the best experts and matching info

- Target Ranking
  - Select subset of hyperlinks
  - Targets are at least pointed to by two non-affiliated experts

nad207@lehigh.edu
Expert Documents

• What makes a page an expert?
  – Unbiased
  – Point to numerous non-affiliated pages

• Detecting Host Affiliation
  – Constraints
    • Sharing the same first 3 octets of the IP address
      – Host 1’s IP: 192.128.122.3 vs. Host 2’s IP: 192.128.122.4
    • The rightmost non-generic token in the hostname is the same
      – www.ibm.com vs. ibm.co.mx
  – Affiliation relation is transitive
  – Host-affiliation lookup
Expert Documents

• Selecting the experts
  – Outdegree is greater than a threshold k
    • Non-affiliated
    • Belongs to one broad topic

• Indexing the experts
  – Key phrases:
    • Scope
    • Importance
Query Processing (Expert Score)

• Requirement: at least one URL whose qualified key phrases cover all the query terms

• Expert Score Computation (S0, S1, S2):
  – FullnessFactor(p, q): measure of the number of terms in p covered by terms in q

\[
S_i = \sum_{\text{key phrases } p \text{ with } k - i \text{ query terms}} \text{LevelScore}(p) \times \text{FullnessFactor}(p, q)
\]

\[
\text{Expert}_\text{Score} = 2^{32} \times S_0 + 2^{16} \times S_1 + S_2.
\]
Query Processing (Target Score)

- Requirement: targets should be pointed to by at least 2 experts from non-affiliated host

- **Target score computation:**
  - \( \text{Edge} \_\text{score}(E,T) = \text{Expert} \_\text{score}(E) \times \text{Sum}\{\text{query keywords } w\} \times \text{occ}(w,T) \)

  - Discard one edge when two affiliated experts point to the same target

  - Sum all the Edge_scores
### Evaluation (Recall)

<table>
<thead>
<tr>
<th>Organization</th>
<th>Query Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha Phi Omega</td>
<td>Best Buy</td>
</tr>
<tr>
<td>Dollar Bank</td>
<td>GroupLens</td>
</tr>
<tr>
<td>Mountain View Public Library</td>
<td>Macy's</td>
</tr>
<tr>
<td>MENSA</td>
<td>OCDE</td>
</tr>
<tr>
<td>Pizza Hut</td>
<td>Rice University</td>
</tr>
<tr>
<td>Stanford Shopping Center</td>
<td>Trek Bicycle</td>
</tr>
<tr>
<td></td>
<td>Digital</td>
</tr>
<tr>
<td></td>
<td>INRIA</td>
</tr>
<tr>
<td></td>
<td>Minneapolis City Pages</td>
</tr>
<tr>
<td></td>
<td>Moscow Aviation Institute</td>
</tr>
<tr>
<td></td>
<td>ONU</td>
</tr>
<tr>
<td></td>
<td>Pittsburg Steelers</td>
</tr>
<tr>
<td></td>
<td>SONY</td>
</tr>
<tr>
<td></td>
<td>Safeway</td>
</tr>
<tr>
<td></td>
<td>USTA</td>
</tr>
<tr>
<td></td>
<td>Vanguard Investments</td>
</tr>
</tbody>
</table>

- organization → query terms

![Graph showing recall values for different ranks with legend: HT, E1, E2, E3]
Evaluation (precision)

- Broad topics/popular topics → query term
- Relevance judgment (binary)
Conclusion

• Broad query → authoritative pages

• Experts are created for directing users to resources

• Unbiased recommendations

• Content match between queries and experts ensures hyperlinks being considered are on the query topic