Tags for search

Drawbacks

Kaj Torrkulla, Samuli Ylenius, Bujia Guo
Recap

Problems

- Finding the right tag
- Ranking tags
- Categorizations tags

Solutions

- Self tagging
- Automatic tagging
Drawbacks

Self tagging

- Adding wrong tags
- Usually there aren't any restrictions when adding tags
- May leads to creating false connections between items

Automatic tagging

- Might not be the most relevant tag
Example Cohen’s kappa algorithm

In the article they used algorithm to evaluate inter-rater agreement.

They selected 75 tags per system and let students select the category for the tags after that they used the algorithm to computed the k-value

Results:

Flickr: 0.8 and 0.69; Del.icio.us: 0.67 and 0.46; AT: 0.83 and 0.7
Bischoff et al. (2008) Drawbacks

Manual labour

Usefulness

Accuracy
Heymann et al. (2008) Drawbacks

Accuracy compared to anchor tags (i.e. link text)

“A substantial proportion of tags are obvious in context, and many tagged pages would be discovered by a search engine.”

“It may be more efficient to train librarians to label domains than to ask users to tag pages.”

\[
\text{classify}(t_i, d_j) = \begin{cases} 
  t & : \frac{\text{tagged}(t_i, d_j)}{\text{total}(d_j)} > \tau \\
  \neg t & : \frac{\text{tagged}(t_i, d_j)}{\text{total}(d_j)} \leq \tau 
\end{cases}
\]
Heymann et al. (2008) Drawbacks

(a) Accuracy on Positive Examples

(b) Accuracy on Negative Examples
Heymann et al. (2008) Drawbacks

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Avg Accuracy (+)</th>
<th>Avg Accuracy (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.33</td>
<td>19.647</td>
<td>99.670</td>
</tr>
<tr>
<td>0.5</td>
<td>7.372</td>
<td>99.943</td>
</tr>
<tr>
<td>0.66</td>
<td>4.704</td>
<td>99.984</td>
</tr>
</tbody>
</table>

About 20% of the tags which were sampled were deemed by our users to “apply to the whole domain.”
Next week: Related works