Exploring Measures of Inter Tagger Consistency
Margaret E. I. Kipp <kipp@uwm.edu>
Information Organization Research Group (IOrg), School of Information Studies, UWMilwaukee

INTRODUCTION

Studies of inter indexer consistency have traditionally examined the consistency of indexing documents between 2 or 3 indexers. Leonard (1977) and Markey (1984) examined the results of multiple inter indexer consistency studies, examining not only the levels of consistency which varied widely but also the level of indexing exhaustivity (number of terms assigned), data collection method and vocabulary size. The majority of inter indexer consistency studies show high levels of inconsistency between indexers.

Social Tagging

Studies of social tagging show convergence of popular terms in frequency graphs, but an examination of user tag lists shows that divergence of opinion continues (Kipp 2009). Many measures of inter indexer consistency can be modified to examine tagging with an arbitrary number of indexers or taggers, allowing the calculation of inter tagger consistency.

METHODOLOGY

• Data was collected as part of a larger study examining patterns in convergence and divergence of tags on delicious.com (see Kipp 2009).
• This study used a number of inter indexer consistency measures to examine inter tagger consistency on delicious.com.
• All measures except the Pairwise Jaccard used a calculated centroid for comparison. Two centroids were used: one composed of all tags, the other of the top 25 tags.

INTER TAGGER CONSISTENCY MEASURES

A. Salton’s Cosine
  cosine similarity - weighted and unweighted terms and a centroid
B. Jaccard (aka Hooper and Rolling’s)
  ratio of the intersection of tag lists to the union with a centroid
C. Inter-indexer Consistency Density
  (Wolfram and Olson 2007)
  Euclidian distance measure of difference with a centroid
D. Pairwise Jaccard (Kipp 2009)
  compares tag lists to each other

Let I represent a tagger’s tag list as a vector, C a centroid, A and B two sets of tags, N the frequency of use of the tag and M the number of indexers.

Unweighted Cosine

Weighted Cosine (same formula)
  Two weight calculations were used:
  tf-idf weight: \( \frac{N}{\log M} \)
  tf-idf weight: \( \log \frac{N}{\log M} \)

Jaccard = \( \frac{A \cap B}{A \cup B} \)
  Exclusive Jaccard = \( \frac{|A \setminus B|}{|A \cup B|} \)

Inter Indexer Consistency Density (ICD)

\[ ICD = \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \left( 1 - \frac{1}{N} \right) \]

RESULTS

<table>
<thead>
<tr>
<th>43folders (highly tagged)</th>
<th>webmd</th>
<th>gtd_tiddly (lightly tagged)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cosine full</td>
<td>0.034</td>
<td>0.045</td>
</tr>
<tr>
<td>cosine partial</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td>cosine full tf-idf</td>
<td>0.012</td>
<td>0.016</td>
</tr>
<tr>
<td>cosine partial tf-idf</td>
<td>0.12</td>
<td>0.1</td>
</tr>
<tr>
<td>cosine full df</td>
<td>0.48</td>
<td>0.56</td>
</tr>
<tr>
<td>cosine partial df</td>
<td>0.48</td>
<td>0.56</td>
</tr>
<tr>
<td>jaccard full</td>
<td>0.0014</td>
<td>0.0024</td>
</tr>
<tr>
<td>jaccard exclusive full</td>
<td>0.0014</td>
<td>0.0024</td>
</tr>
<tr>
<td>jaccard partial</td>
<td>0.077</td>
<td>0.072</td>
</tr>
<tr>
<td>jaccard excl. partial</td>
<td>0.087</td>
<td>0.082</td>
</tr>
<tr>
<td>pairwise jaccard</td>
<td>0.079</td>
<td>0.11</td>
</tr>
<tr>
<td>pairwise excl. jaccard</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td>ICD</td>
<td>0.021</td>
<td>0.03</td>
</tr>
</tbody>
</table>

DISCUSSION AND CONCLUSIONS

Results of inter indexer consistency studies vary depending on the indexers and use of controlled vocabulary or natural language, showing ranges from 4% to 82% (Markey 1984; Leonard 1977).

Many inter tagger consistency values in this study fall into this range (4-82%) and are thus consistent with previous findings. Some of the consistency in tagging on delicious.com may be influenced by the incorporation of suggested terms into the interface, acting as a form of semi-controlled vocabulary.

REFERENCES