

# Exploring Measures of Inter Tagger Consistency

Margaret E. I. Kipp <kipp@uwm.edu>

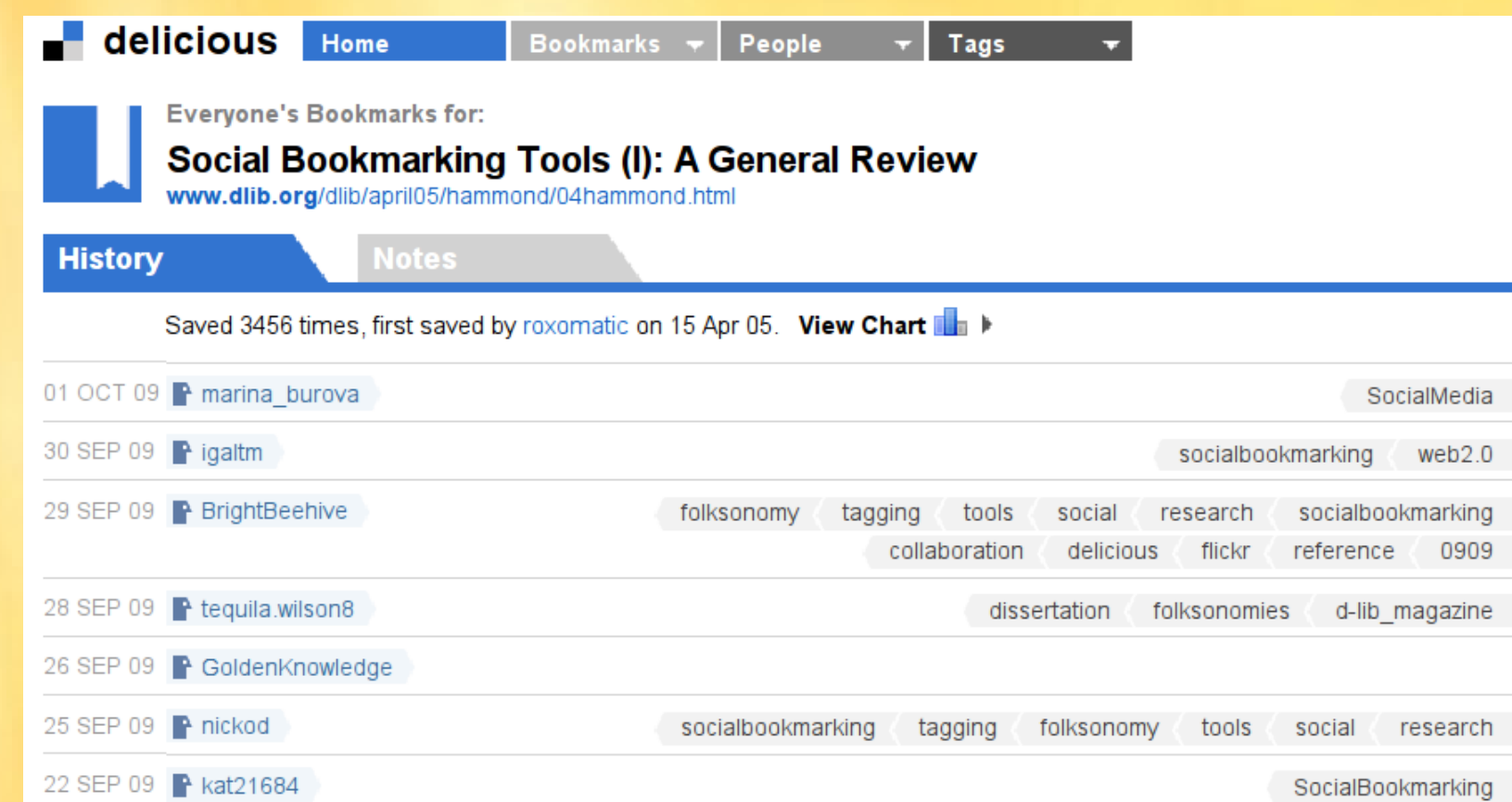
Information Organization Research Group (IOrg), School of Information Studies, UWMilwaukee

## 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 indexing 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

- Salton's Cosine cosine similarity - weighted and unweighted terms and a centroid
- Jaccard (aka Hooper and Rolling's) ratio of the intersection of tag lists to the union with a centroid
- Inter-indexer Consistency Density (Wolfram and Olson 2007) Euclidian distance measure of difference with a centroid
- 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.

$$\text{Unweighted Cosine} = \frac{I \cdot C}{\text{mag } I * \text{mag } C}$$

Weighted Cosine (same formula)

Two weight calculations were used:

$$\text{tf-df weight: } \frac{N}{M} \quad \text{tf-idf weight: } \log\left(\frac{M}{N}\right)$$

$$\text{Jaccard} = \frac{A \cap B}{A \cup B}$$

$$\text{Exclusive Jaccard} = \frac{A \cap B}{A \cup B - A \cap B}$$

Inter Indexer Consistency Density (ICD)

$$ICD = \frac{\sum_{i=1}^m \text{Sim}(I_i, C)}{m} \text{ where } \text{Sim}(I_i, C) = \begin{cases} \frac{1}{\text{Dist}(I_i, C)} & \text{for } \text{Dist}(I_i, C) \neq 0 \\ 1 & \text{for } \text{Dist}(I_i, C) = 0 \end{cases}$$

## REFERENCES

Kipp, Margaret E.I. 2009. Information Organisation Practices on the Web: Tagging and the Social Organisation of Information. PhD Thesis, University of Western Ontario, London, Ontario.

Leonard, L. 1977. Inter-indexer consistency studies, 1954-1975: a review of the literature and summary of study results. Occasional papers (University of Illinois at Urbana-Champaign. Graduate School of Library Science) 131.

Markey, Karen. 1984. Interindexer Consistency Tests: a Literature Review and Report of a Test of Consistency in Indexing Visual Materials. Library and Information Science Research 6(2): 155-177.

Wolfram, Deitmar and Olson, Hope. 2007. A method for comparing large scale inter-indexer consistency using IR modeling. Proceedings of the 35th conference of the Canadian Association for Information Science, Montreal, QC, May 10-12, 2007.

## RESULTS

	43folders (highly tagged)	webmd	gtd_tiddly wiki (lightly tagged)
cosine full	0.034	0.045	0.068
cosine partial	0.25	0.24	0.26
cosine full tf-idf	0.012	0.016	0.027
cosine partial tf-idf	0.12	0.1	0.11
cosine full df	0.48	0.56	0.63
cosine partial df	0.48	0.56	0.63
jaccard full	0.0014	0.0024	0.0053
jaccard exclusive full	0.0014	0.0024	0.0054
jaccard partial	0.077	0.072	0.083
jaccard excl. partial	0.087	0.082	0.095
pairwise jaccard	0.079	0.11	0.12
pairwise excl. jaccard	0.11	0.16	0.17
ICD	0.021	0.03	0.04

## 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.