Evaluation of Reader Perception by Using Tags from Social Bookmarking Systems

Stefanie Haustein[1,2], Isabella Peters[1] and Jens Terliesner[1]

[1] *[isabella.peters | jens.terliesner]@uni-duesseldorf.de*
Heinrich-Heine-University Düsseldorf, Department of Information Science,
Universitätsstr. 1, 40225 Düsseldorf (Germany)

[2] *s.haustein@fz-juelich.de*
Forschungszentrum Jülich, Central Library, 52425 Jülich (Germany)

Introduction

The traditional way of evaluating scientific journals is citation analysis. Recent studies have emphasized the importance of including the readers' perspectives by analyzing download and click rates. Against the background of global download statistics still being inaccessible, Haustein et al. (2010) introduce social bookmarking data as an alternative source to measure journal perception. Social bookmarking services do not only allow users to store and share scientific literature on the web but also to index them with freely chosen tags. It is assumed that, if cumulated on journal level, these tags reflect a reader-specific view on journal content, which differs from traditional indexing methods. This contribution aims to follow up on this assumption and investigate the readers' tagging behavior in greater detail. In order to discover differences or similarities in contrast to common indexing methods, tags are compared to title and abstract terms, author keywords, indexer-generated Inspec subject headings and automatically generated KeyWords Plus^{TM} from Web of Science (WoS). Data is cleaned extensively and similarities are computed on the level of single documents to gain exact results.

Database

The data of this analysis is based on a previous study by Haustein et al. (2010) and Haustein & Siebenlist (to be published), which examines the application of social bookmarking data from CiteULike, Connotea and BibSonomy to journal evaluation. For this study the initial bookmarking data of 10,280 documents published in 45 physics journals is limited to a subset of 724 articles, for which all other necessary conventional indexing terms are available. Indexer-generated subject headings are taken from Inspec and automatically generated KeyWords Plus^{TM} as well as titles and abstracts downloaded from WoS. To discover differences and similarities of different indexing methods on document level, tags and terms are connected to their specific articles via DOI.

Methods

Preprocessing and Cleaning

Due to the uncontrolled nature of tags and the different spelling variants of terms in titles, abstracts and keywords, data cleaning and transformation has to be applied. In addition to common cleaning methods (e.g., Noll & Meinel, 2007) we take the preprocessing one step further and unify variants as far as automatically possible: British English suffixes are transformed into American English by applying a rule-based algorithm and all terms are stemmed using the Porter 2 algorithm to unify tags and indexing terms. For comparison of tags with titles and abstracts, tags were split at the separating character (i.e. hyphen) to allow for a matching of single-word terms of title and abstracts. When comparing tags to author keywords, subject headings and KeyWords Plus^{TM}, hyphens and underscores are deleted within tags and blanks within keywords in order to unify different spellings like *complex_network*, *complex-network*, *complexnetwork* and *complex network*. The combination of all processing methods reduces spelling variations for tags by 8.4% from 1,743 to 1,596 unique terms. Stemming and unification of BE and AE alone cause 6.1% improvement. The same cleaning methods are applied to the other terms. Especially abstract and title terms can be improved by these methods: unique term quantity is reduced by 30.5% and 19.8%, respectively.

Measuring Term Similarities

In contrast to previous studies, which compare tags to other indexing terms over whole datasets (e.g., Lin et al., 2006; Kipp, 2005), we follow the more exact approach to compare similarities and differences on the level of single documents. Thus, for each of the 724 documents the number of cleaned unique tags, author keywords, KeyWords Plus^{TM}, Inspec, title and abstract terms is determined. Three measurements are used to determine similarities between index terms for each document. The arithmetic means of the resulting 724 similarity values can be seen in table 1. First, the percentage of overlap is computed in contrast to the total number of unique tags per document on the one hand and to the number of the particular meta terms on the other, in order to detect the share of common tags from each of the perspectives. The overlap-tag ratio lists the percentage of overlapping tags in contrast to all unique tags assigned to the particular document and is defined as

$$overlap\ tag\ ratio = \frac{g}{a}$$

where *a* stands for the number of unique tags per document and *g* represents the overlap between tags and other indexing terms per document. Most tags are represented in the abstracts, which is to be expected, since the number of abstract terms is much greater than that of the other metadata. The overlap-analyzed term ratio calculates the same overlap from the other perspective.

$$overlap\ analyzed\ term\ ratio = \frac{g}{b}$$

where *b* stands for the number of unique terms per document and *g* represents the overlap between both sets per document. On average, 24.5% of title terms are used for tagging articles. Strikingly, only 3.4% of indexer terms are adopted. To combine both measurements, the similarity between the readers' point of view on the one hand and author, intermediary and automatic indexing perspective on the other hand is calculated by cosine.

$$cosine\ similarity = \frac{g}{\sqrt{a \cdot b}}$$

where *a* stands for the number of unique tags per document, *b* for the number of unique terms and *g* represents the overlap between tags and terms per document.

Table 1. Mean similarity measures comparing reader with author, intermediary, automatic indexing, title and abstract terms.

similarity of tags and:	mean overlap-tag ratio	mean overlap-analyzed term ratio	mean cosine similarity
author keywords	11.8%	10.4%	0.103
Inspec subject headings	13.3%	3.4%	0.062
KeyWords Plus^{TM}	2.9%	3.0%	0.026
title terms	36.5%	24.5%	0.279
abstract terms	50.3%	4.8%	0.143

Results and Conclusions

On average, there is hardly any overlap between reader and professional and automatic indexing methods. The mean cosine value is highest for title terms, abstracts and author keywords (table 1). The very low cosine values imply that social tagging represents a user-generated indexing method and provides a reader-specific perspective on article content, which differs extremely from conventional indexing methods. Lin et al. (2006) suspect that this is due to different goals of professional indexers, who want to index and cover all topics of a document using controlled vocabularies and users, who seem to seek out the subject they are interested in and add a tag rather than represent the document completely.

The results confirm our basic assumption that journal and article evaluation can profit from the application of user-generated tags for content analysis, as they add a third layer of perception besides the author and indexer perspectives. Due to the dynamic nature of social bookmarking and tagging, these descriptions evolve in realtime. As shown in figure 1, tag clouds can offer direct channels to the readers' opinions and depict trends in the language of a specific discipline.

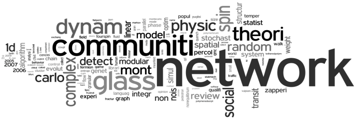

Figure 1. Tag cloud depicting the reader perspective on *Journal of Statistical Mechanics.*

Acknowledgments

Peters and Terliesner are supported by the German Research Fund DFG (STO 764/4-1).

References

Haustein, S., Golov, E., Luckanus, K., Reher, S., & Terliesner, J. (2010). Journal evaluation and science 2.0. Using social bookmarks to analyze reader perception. In *Book of Abstracts of the 11th International Conference on Science and Technology Indicators, Leiden, the Netherlands* (pp. 117-119).

Haustein, S. & Siebenlist, T. (to be published). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*.

Kipp, M. E. I. (2005). Complementary or discrete contexts in online indexing: A comparison of user, creator, and intermediary keywords. *Canadian Journal of Information and Library Science*, 29(4), 419-436.

Lin, X., Beaudoin, J., Bul, Y., & Desai, K. (2006). Exploring characteristics of social classification. In *Proceedings of the 17th Annual ASIS&T SIG/CR Classification Research Workshop, Austin, Texas, USA*.

Noll, M. G., & Meinel, C. (2007). Authors vs. readers. A comparative study of document metadata and content in the WWW. In *Proceedings of the 2007 ACM Symposium on Document Engineering, Winnipeg, Canada* (pp. 177–186).