SnoopyTagging: Recommending Contextualized Tags to Increase the Quality and Quantity of Meta-Information

Wolfgang Gassler, Eva Zangerle, Martin Bürgler, Günther Specht
Databases and Information Systems, Institute of Computer Science
University of Innsbruck, Austria
{firstname.lastname}@uibk.ac.at

ABSTRACT

Current mass-collaboration platforms use tags to annotate and categorize resources enabling effective search capabilities. However, as tags are freely chosen keywords, the resulting tag vocabulary is very heterogeneous. Another shortcoming of simple tags is that they do not allow for a specification of context to create meaningful metadata. In this paper we present the SnoopyTagging approach which supports the user in the process of creating contextualized tags while at the same time decreasing the heterogeneity of the tag vocabulary by facilitating intelligent self-learning recommendation algorithms.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: On-line Information Services; H.2.8 [**Database Management**]: Database Applications—*Data Mining*

General Terms

Algorithms, Performance, Human Factors, Experimentation

Keywords

Tagging, Recommender Systems, Heterogeneity, Structure

1. INTRODUCTION

Within the last decade, the web transformed to a participatory, collaborative medium. One of the most popular paradigms on the web is the *tagging* of online resources. It enables users to annotate online resources such as images or bookmarks with keywords aiming at creating a categorization to simplify the search and retrieval of resources. Especially in large and flexible collaborative systems, the definition of a fixed set of categories in advance is not feasible. Hence, tagging is very import to bring some kind of order to the existing chaos. One major feature of tags is the simple usage as tags may be chosen freely without any restrictions. Thus, tags can be used to describe different aspects of online resources by the users. Consider the example of an image tagged with "Robert Capa". By just considering the tag, it is not clear whether Robert Capa is pictured on the photo, whether Robert Capa is the photographer or the picture shows Robert Capa's house.

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

In this paper we introduce SnoopyTagging, an approach aiming at solving this problem by adding contextual meta-information to tags while at the same time increasing the homogeneity of the resulting tag-vocabulary by facilitating a recommender system. As a showcase example, we implemented our approach on top of Flickr.

2. SNOOPYTAGGING

The SnoopyTagging Concept is based on Structured Tags in combination with a self-learning Recommendation Engine. The paradigm of Structured or Contextualized Tags, which was first introduced as "The Poor Man's RDF" [1], represents the basis of the Snoopy approach. Structured Tags consist of two parts which are divided by a delimiter sign (a colon in our case): `context:tag`. The first part of a Structured Tag defines the context of the actual tag, whereas the actual tag is specified after the delimiter sign. As most of today's tagging platforms allow colons in tags, the backward compatibility of SnoopyTagging is guaranteed. Structured Tags enable the user to provide tags with context, e.g. the tag `photographer:Robert Capa` expresses that Robert Capa is the photographer of a certain photo.
The disadvantage of freely chosen tags is the increasing latent heterogeneity of the resulting folksonomies. The reason for this behaviour was already described by Furnas *et al.* [1] who showed in the 80's that the chance of two humans choosing the same term for the same object is only about 20%. This fact is crucial in online mass-collaboration tagging systems, as there are thousands of different users from different social levels and backgrounds who add tags in different domains and settings. Consider e.g. the tags `takenBy:Robert Capa` and `photographer:Robert Capa` which are differentiated during the search process although they are semantically equivalent. We tackle this problem by introducing a self-learning recommendation system which aims at providing suitable tags and contexts to homogenize the used vocabulary and avoid the use of synonyms.

3. RECOMMENDATIONS

In order to create and maintain a homogeneous set of both contexts and tags, we make use of our developed Snoopy approach [2] and implemented *SnoopyTagging*, a first prototype based on the Flickr platform. The Snoopy approach basically aims at dealing with the problem of heterogeneous

[1] `http://weblog.scifihifi.com/2005/08/05/meta-tags-the-poor-mans-rdf/` by Buzz Andersen 2005

vocabularies in semistructured knowledge bases by providing the users with suitable recommendations. SnoopyTagging extends the concept to Structured Tags and recommends both the context and the actual tag. As these recommendations are computed based on all tags which have previously been entered by users, recommendations are strongly tied to the community and its vocabulary. We distinguish between two types of recommendations:

(i)Additional Structured Tags are recommended based on co-occurrence analysis [2] of the already entered tags and leads to the suggestion of further applicable contexts. If e.g. the tag `photographer:Robert Capa` was already specified, the contexts `camera:?` and `location:?` may be recommended to the user during the tag insertion process as these tags co-occur on many already tagged photos. Hence, this type of recommendations aims at encouraging the user to (a) use Structured Tags and (b) enter more (meta-) information.

(ii) The extended auto-completion feature recommends context and tags during the typing of the user. This intelligent feature recommends the re-usage of contexts and tags. Thus it avoids the insertion of additional synonyms [2] and leads to a more homogeneous vocabulary. E.g. if `loc` has been entered, the key `location` and according values, e.g. `Vienna` are recommended.

The SnoopyTagging approach incorporates the user co-occurrence space to personalize both types of recommendations. Once a user entered a context a, the system computes those tags co-occurring with the context a within both the set of global tags (of all users) and user-specific tags (tags used by the current user). As only the top-5 recommendations are shown in the interface, a ranking function is applied. The ranking is based on the co-occurrence rate of the already entered tags on the respective resource and the remaining tags within the data sets (ranking of the global set: $rank_{global}$, ranking of the user-specific set: $rank_{user}$). For the final set of recommendations which is based on both previously computed sets (user and global co-occurrence), we propose to use a hybrid ranking algorithm $rank = \gamma * rank_{global} + (1 - \gamma) * rank_{user}$ with the factor γ which defines the weight of the final rank between the user-specific and global set of recommendations. In our experiments a value of $\gamma = 0.1$ turned out to be beneficial.

4. PRELIMINARY EVALUATION

Based on the implemented prototype for SnoopyTagging[2] we conducted user tests. In total, 20 voluntary test users (originating from various backgrounds and having diverging levels of computer skills) took part in this evaluation. The users were asked to use the SnoopyTagging prototype to upload and subsequently tag arbitrary photos.

Overall, 98% of all tags the users facilitated were structured tags, even though the insertion of simple tags was still possible. This number is a strong indicator for the acceptance of the concept of contexts. 350 recommendations (233 contexts, 117 tags) were accepted for the creation of 310 Structured Tags. Such a high percentage of Structured Tags and accepted recommendations can be attributed to the user-friendly SnoopyTagging system which allows users to easily create Structured Tags, and the acceptance of the underlying recommender system.

During the user experiments, all recommendations, user inputs and user recommendation acceptances were logged. Based on this usage data, we evaluated the performance of the hybrid ranking function. For this purpose, we reconstructed the test run and varied the γ-values responsible for the weighting of the ranking. Our evaluations showed that the recall value reached about 79% when displaying five recommendations to the user (recall@5) and can be obtained by setting $\gamma = 0.1$. This strong emphasis on personalized recommendations can be explained by the limited amount of users and the resulting size of the folksonomy. The global set of recommendations will increase with the amount of users as there are more appropriate tags to recommend. However, there is no Web 2.0 platform which advertises Structured Tags and therefore it is not possible to observe this behaviour in a large community in a real environment.

As for the homogeneity within the tagging vocabulary resulting from the experiments, a total of 170 distinct Structured Tags were entered, where only 37 different contexts were used. This fact is remarkable considering the circumstance that users were allowed to enter photos about an arbitrary topic and shows the frequent re-usage of contexts by recommendations.

There are many other approaches dealing with the recommendation of appropriate simple tags for online resources, like e.g. for Flickr images [4] or social bookmarks [3]. However, currently there are no other approaches facilitating recommender systems in order to encourage users to use contextualized tags for online resources and decrease the heterogeneity of Structured Tag vocabularies.

5. CONCLUSION

In this paper we presented SnoopyTagging, a prototype for tagging online resources (Flickr images). SnoopyTagging facilitates Contextualized Tags aiming at adding context to simple tags and relies on a recommender system to homogenize the tag vocabulary. The conducted evaluation showed the acceptance of the concept of contexts by the users. Our experiments also showed the frequent re-usage of tags and the creation of a homogeneous tagging vocabulary due to the strong acceptance of the underlying recommendation engine.

6. ACKNOWLEDGMENTS

This research was partially funded by the University of Innsbruck (Nachwuchsförderung 2011).

7. REFERENCES

[1] G. Furnas, T. Landauer, L. Gomez, and S. Dumais. The vocabulary problem in human-system communication. *Comm. of the ACM*, 30(11):971, 1987.

[2] W. Gassler, E. Zangerle, and G. Specht. The snoopy concept: Fighting heterogeneity in semistructured and collaborative information systems by using recommendations. In *Intl. Conf. on Collaboration Technologies and Systems*, Philadelphia, PE, 2011.

[3] M. Lipczak. Tag recommendation for folksonomies oriented towards individual users. *ECML PKDD Discovery Challenge*, pages 84–95, 2008.

[4] B. Sigurbjörnsson and R. Van Zwol. Flickr tag recommendation based on collective knowledge. In *Proc. of the 17th WWW Conference*, pages 327–336. ACM, 2008.

[2](available at `http://dbis-snoopy.uibk.ac.at`)