FolkTrails: Interpreting Navigation Behavior in a Social Tagging System

Thomas Niebler¹, Martin Becker¹, Daniel Zoller¹, Stephan Doerfel¹, Andreas Hotho²

¹ Data Mining and Information Retrieval Group, University of Würzburg, Germany
² KDE Group, University of Kassel, Germany, L3S Research Center, Hannover, Germany

Motivation
The Navigation Behaviour of users on Web 2.0 systems is largely unexplored. This is however important to effectively analyze and improve website structures for better content provision.

In this work, we focus on the analysis of navigation in Social Tagging Systems, specifically in BibSonomy. Social Tagging Systems have been in the focus of research for many years now and interesting results about the generated content entities have been presented, exhibiting the great semantic exploitability of such systems. Still, how users navigate such systems is largely unknown.

We provide and analyze several hypotheses about the incentives of user navigation in the Social Tagging System BibSonomy, thus shedding light on how users consume the provided content.

Hypotheses about Navigation

Uniform Hypothesis (uniform)
The Uniform Hypothesis serves as the baseline. It assumes that users navigate randomly.

User Consistent Hypothesis (user)
Many people make use of BibSonomy as a place to save their resources. Thus, they navigate mainly on their own pages.

Page Consistent Hypothesis (page)
In previous studies it was found that users often stay on the same pages, e.g. due to pagination effects.

Category Consistent Hypothesis (cat)
A relaxation of the page consistent hypothesis, we assume that users tend to stay on pages with the same category, e.g. only on user pages.

Folksonomy Consistent Hypothesis (folk)
The Folksonomy Hypothesis is based on the Folksonomy structure; we expect users to only follow links provided by the Folksonomy.

Semantic Navigation Hypothesis (tfidf)
Because tagging data contains a lot of semantic information, we expect users to prefer pages, which are semantically stronger related.

Datasets
We analyze the large set of content and log data from the Social Tagging System BibSonomy, spanning from 2006 to 2012.

User and Content Dataset
We use the Folksonomy data from nonspammers with their respective resources and tags. 17,932 users were explicitly classified as nonspammers. They created 466,777 bookmark posts and 2,410,844 publication posts using 65,228 disambiguation, which have been visited at least twice.

Request Log Dataset
The BibSonomy log files include all HTTP requests to the system. After filtering, the remaining dataset contains 183,419 distinct visited content pages, i.e. pages which show at least one resource. We recorded 327,060 transitions between these pages. 123,452 transitions were self-transitions (i.e., transitions from a page to itself) and 281,308 transitions, where the logged in user owns both the source and the target page.

Request Log Subsets
We additionally investigated navigation outside, i.e., where users navigate outside their own pages. The subset of those requests counts 42,183 requests, and effects of usage continuity. The subset of requests by short-term users, i.e., who used the system less than half a year, is comprised of 48,221 requests.

Results
We analyzed navigation on the overall request log dataset and on two subsets: Outside navigation and short-term usage.

Overall Request Log Dataset
- All of the basic hypotheses explain the observed transitions better than the baseline, which indicates at least some structural properties explaining the observed transitions.
- The good performance of the semantic hypotheses indicates that semantic similarity of pages is a strong factor for navigation in BibSonomy. However, users tend to mostly navigate on their own pages.
- Overall, the combination of the user consistent and the semantic hypothesis performs best, indicating that navigation on BibSonomy can mainly be explained by semantic navigation within the resources of a specific user.

Request Subset: Usage Continuity
Since we expect users to adapt to systems they are using, we investigate if their navigation behaviour changes over time.

- The semantic hypothesis performs significantly better than the folksonomy hypothesis. We may thus observe a learning process: Short-term users are not as adapted to the folksonomy structure as long-term users and thus rely on their semantic intuition.
- The page consistent and folksonomy consistent & user consistent hypotheses explain navigation equally well. This might be due to increased use of pagination as well as a lack of personal resources.

Conclusion
- In this work, we studied a large dataset of webserver logs from the social tagging system BibSonomy in order to analyze and understand components of user navigation in social tagging systems.
- Our results show that there is a strong semantic component inherent in user navigation. On subsets of the set of request logs, we call Tetra show that users navigating outside their own resources largely follow the folksonomy structure.
- Overall, we were able to gain new insights into the underlying processes of navigation in tagging systems, which can be extended and leveraged in the future, for example, by considering new hypotheses, improving navigation experience or extracting the latent semantic information.