Social tagging is a widespread activity for indexing user-generated content on Web services. This paper summarizes research on folksonomies and their retrieval effectiveness. A TREC-like retrieval test was conducted with tags and resources from the social bookmarking system delicious, which resulted in recall and precision values for tag-only searches. Moreover, several experimental tag-based databases (e.g., Power tags, Luhn tags) have been tested regarding their retrieval effectiveness. Test results show that folksonomies work best with short queries although recall values are high and precision values are low. Here, a search function “Power tags only” greatly enhances precision values.

### Abstract

Social tagging is a widespread activity for indexing user-generated content on Web services. This paper summarizes research on folksonomies and their retrieval effectiveness. A TREC-like retrieval test was conducted with tags and resources from the social bookmarking system delicious, which resulted in recall and precision values for tag-only searches. Moreover, several experimental tag-based databases (e.g., Power tags, Luhn tags) have been tested regarding their retrieval effectiveness. Test results show that folksonomies work best with short queries although recall values are high and precision values are low. Here, a search function “Power tags only” greatly enhances precision values.

1. **Extract documents and tags**
   - 1,989 resources consisting of the dcollections of the delicious-bookmarks
   - collected from delicious.com in October 2010
   - tags form the indexed databases for the retrieval test runs
   - each resource is tagged from at least 30 users
   - each resource was tagged with “folksonomy”, “seo” or “folksonomie”

2. **Adjust tags**
   - original tags (Information Retrieval)
   - tags unified, i.e. without special characters (informationretrieval)
   - tags unified and stemmed (informationretrieval)

3. **Extract Power tags and Luhn tags**
   - The algorithms developed for Power Tag and Luhn tag extraction work differently and depend on the tag distribution of the dcollection.
   - Baseline = F - measure of delicious - tags for expert queries

4. **Create information needs**
   - 55 information needs and search tasks have been created. They vary in their complexity as one can see in the following examples: simple lookup, complex lookup, exploratory search task.
   - Find a thesaurus
   - Find articles which present social bookmarking tools
   - Find articles which advise the combination of folksonomies and controlled vocabularies for indexing

5. **Judge relevance of documents**
   - 24 students and 9 people from staff of the department act as assessors
   - assessments are binary and are conducted manually
   - assessors had access to resources (i.e., websites). Tags were hidden for relevance assessments
   - if two assessors agreed in their decision the relevance judgment was saved immediately
   - this procedure results in 109,395 relevance judgments

6. **Create queries**
   - 25 students and 2 people from staff created queries for each information need
   - 730 queries
   - Boolean operators and brackets are allowed (they can be used in delicious)
   - minimum number of query terms: 1, maximum number of query terms: 13
   - average query length: 3 terms
   - experts from staff built queries for each search task with a maximum number of 5 query terms (simulation of real-world-users: they use 2 terms or less)

7. **Retrieval and evaluation**
   - The figures A to F visualize the results of the retrieval test and display average recall values, average precision values and values for F-measure for conducted search runs.
   - Figures A to C visualize the results of the expert searches, figures D to F show the results of simple lookup-search tasks requesting one query term.
   - Figures G and H visualize the relative benefit or loss while using Power tags and Luhn tags with expert queries (G) and one word queries (H).

Test results show that retrieval in folksonomies works best with very short queries. Here, a search function “Power tags only” greatly enhances retrieval effectiveness.