How can social tagging benefit information access?

Toine Bogers Royal School of Library & Information Science Copenhagen, Denmark

> India-Norway WWCT workshop October 2, 2011

Outline

- Introduction
- Social tagging for
 - Search
 - Browsing
 - Recommendation

Social tagging

- Social tagging is collectively describing (tagging) items/resources by assigning keywords (tags)
 - Collaborative version of controlled vocabularies
 - The resulting item taxonomy is called a folksonomy ('folk' + 'taxonomy')
 - Emergent network of users, items, and tags

Domains



Publications about social tagging



"social tagging" OR "collaborative tagging" OR "social bookmarking"

Research directions

- Two main directions
 - Why and how do people tag?
 - What can we use the tags for?

Why do people tag?

| | | FUNCTION | |
|----------|------------------|---|--|
| | | Organization | Communication |
| AUDIENCE | Self | Retrieval & sharing | Context for self, memory aid |
| | Family & friends | Contribution, attention, ad hoc photo pooling | Content description, social signaling |
| | Public | | |

How do people tag?

- Web pages (e.g., Delicious)
 - Topic, usage context, type
- Images (e.g., Flickr)
 - Topic, location, opinion/quality, usage context, time
- Music (e.g., Last.FM)
 - Type, opinion/quality, author/owner

Bischoff et al. (2008)



Research directions

- What potential do tags have for improving search?
 - Based on an analysis of social tagging systems and tagging behavior
- How should we integrate tags into search algorithms?

Potential of tags

- Heymann & Garcia-Molina (2008)
 - Analyzed a large crawl of Delicious
 - Question: can social tagging improve search?
 - Around 12.5% of Web pages in Delicious are not found in search engines
 - Pages in Delicious are newer on average than those indexed by search engines

Potential of tags

- Tags occur in the text of the bookmarked page
 50% of the time
- Tags occur in 16% of the titles
- Tags and query terms show significant overlap
- Tags describing Web pages are overwhelmingly objective (90% vs. 10% subjective tags)
- Problem: remains untested!

Integrating tags in search

- What can we use tags for?
 - Mostly work on improving search on social bookmarking websites
 - Documents
 - Clustering ambiguous search results
 - Queries
 - Disambiguating troublesome queries
 - Personalized query expansion using tags

Future work

- What is missing?
 - Direct comparison of different approaches
 - On same data, with same queries, etc.
 - Can tags contribute to actual Web search?
 - Evaluation with real users on real websites
 - Are the gains good enough for everyday use?



Research directions

- How do people navigate social tagging websites?
 - Browsing vs. search
- How do we add structure to the sea of tags?
 - Identifying synonymous or related tags
 - Generating tagging hierarchies

- Garama & De Man (2008)
 - Influence of social tagging on image search
 - Controlled user-centered evaluation
 - Broad vs. narrow folksonomy (Delicious vs. Flickr)
 - Crawled 165,000 different images with tags and surrounding text
 - Single unified interface for both systems with 54 participants

- Browsing vs. searching a folksonomy
 - Contextual information search
 - Tag search
 - Tag browsing using dynamic tag clouds
 - ★ Regenerate similar to faceted browsing

contro panel

Step 1: Your task is to rediscover the image you saw on the previous page by clicking on the keywords below. Try to find the target image in your search results by clicking on the most relevant keywords. When the first keyword has been chosen, results will show up on the right hand side of the screen. You may select relevant images from this results list. On the left hand side is a list of keywords to choose from if you want to refine your search. To delete one of the words in your search query, you can click on the particular word in the keywords bar in the control panel. Selections you have made in your results will be kept in memory during the task and will all show up in the next step, except when you deselect the image(s). When you believe the target image is in your selection, press 'Done'.

Reset | linux = ubuntu

baghira beryl beta breezy breezybadger compiz computer computers dapper dapperdrake debian desktop desktopeffects distorted drake drop ecoprint fawn feisty firefox flash flickr fonts free freesoftware geek geeky gentoo gimp gnome gnu **GNULINUX** hkopp hosting ie imac install installer installfest kde kopp kubuntu laptop lived mac macbook macintosh macos macosx macro medellin microsoft mozilla mysgl nvidia nx opengl openoffice **OPENSOURCE** os parallels

penny portatil powerbook proprietary gemu screencapture

screenshot

screenshots server setup shadow shipit snapshot softwarelibre sticker suse synaptic sysadmin taks transparency tux ubuntulinux A substance Discussion or set



Done

- Findings
 - Searching faster than browsing using tag clouds
 - Exploratory tasks
 - Search faster, but browsing more successful & satisfactory
 - Known-image tasks
 - Search faster, more successful and more satisfactory than browsing using tag clouds

Tag hierarchies

- Heymann & Garcia-Molina (2006)
 - Simple yet robust method for generating tag hierarchies
 - Generate tag similarity graph
 - Convert similarity graph into hierarchy
 - ★ Most central tags at the top of the hierarchy



Tag hierarchies



Heymann & Garcia-Molina (2006)

What have we learned?

- Navigation
 - Tags good for exploratory tasks
 - Search better for locating specific information
- Structure
 - Simple, effective algorithms for generating tag hierarchies

Future work

- What is missing?
 - Realistic studies of user navigation behavior in different social tagging domains
 - Web pages, images, music
 - In controlled and in real-world settings
 - Do tag hierarchies and disambiguation improve the browsing experience of real-world users?
 - Does tagged browsing promote serendipity?

Recommendation

Recommendation

- What is recommendation?
 - Identifying sets of items that are likely to be of interest to the user
 - No explicit information need
 - "People who bought this, also bought..."
 - Two types of algorithms
 - Memory-based
 - Model-based

Research directions



User-based CF

- User-based collaborative filtering (CF)
 - Determine the k most similar users based on overlap in items added/used/bought
 - Look for new items to recommend among them



User-based CF

- How can we incorporate tags?
 - Calculate user similarity based on tag vocabulary overlap between users
 - Does not work as well as usage data...



Item-based CF

- Item-based collaborative filtering (CF)
 - Determine the k most similar items for the items added by the active user
 - Item similarity based on overlap in users
 - Recommend the new items most similar to the user's items

items



users

Item-based CF

- How can we incorporate tags?
 - Calculate item similarity based on tag vocabulary overlap between items
 - Works better than item-based CF with usage data!
 - Works better than user-based CF with either!

Fusion

- What works even better?
 - Fusing different data sources





Fusion

- What works even better?
 - Fusing different data sources
 - Fusing different algorithms
 - The more different the individual algorithms and data sources, the better!
- Also seems to hold for tag recommendation!

Future work

- What is missing?
 - Online, user-centered evaluation with real users
 - Which recommendations do the users accept and why?
 - Can we use tags to better explain why recommendations were made?
 - How do tag suggestions affect the folksonomy on the social tagging website?



References

- Ames & Naaman (2007). Why We Tag: Motivations for Annotation in Mobile and Online Media. In: Proceedings of CHI 2007, pp. 971-980, ACM Press
- Au Yeung et al. (2008). Web Search Disambiguation by Collaborative Tagging. In: Proceedings of ESAIR '08, pp. 48-61
- Bao et al. (2007). Optimizing Web Search using Social Annotations In: Proceedings of WWW 2007, pp. 501-510, ACM Press
- Bischoff et al. (2008). Can All Tags be Used for Search? In: Proceedings of CIKM 2008, pp. 203-212, ACM Press

References

- Bogers & Van den Bosch (2009). Collaborative and Contentbased Filtering for Item Recommendation on Social Bookmarking Websites. In: Proceedings of the ACM RecSys '09 workshop on Recommender Systems and the Social Web, pp. 9-16
- Bogers (2009). Recommender Systems for Social Bookmarking, Ph.D. thesis, Tilburg University
- Carman et al. (2008). Tag Data and Personalized Information Retrieval. In: Proceedings of SSM '08, pp. 27-34, ACM Press
- Clements et al. (2008). Detecting Synonyms in Social Tagging Systems to Improve Content Retrieval In: Proceedings of SIGIR '08, pp. 739-740, ACM Press

References

- Heymann & Garcia-Molina (2006). Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. Technical Report 2006-10, Infolab, Stanford
- Heymann et al. (2008). Can Social Bookmarking Improve Web Search? In: Proceedings of WSDM '08, pp. 195-206, ACM Press