Markus Heckner, Michael Heilemann, Christian Wolff
Personal Information Management vs. Resource Sharing: Towards a Model of Information Behavior in Social Tagging Systems
Overview

- Introduction: tagging and indexing
- Previous studies
- A novel model of tags (How do tags differ for different media?)
- Tags usage and media (How do tags differ for different media?)
- An empirical study on tagging motivations: Why do people tag?
- Towards a model of tagging as information interaction behavior

Tag cloud generated with wordle, www.wordle.com
Heckner et al., Personal Information Management vs. Resource Sharing, 2
Social Tagging

- indexing by professionals is costly and cannot be done for all digital information questions
- non-textual media are hard to index automatically (images, video, music)
- "The Potential Ease of Tagging" vs. the "Analysis-Paralysis" found in Categorization (Sinha 2005)
- Def. Tagging: "Annotate a particular resource, such as a web page, a blog post, an image, a physical location, or just about any imaginable object with a freely chosen set of keywords (= "tags") (Marlow et al., 2006, p. 31)
Tagging: A fourth layer of indexing?

1. Author keywords
2. Intellectual indexing by information professionals
3. Automatic indexing
4. Tagging – “fourth layer of indexing”?

- Tagging as a possible solution to the vocabulary problem?
- Can tags “identify qualities or characteristics” of resources (Kipp 2007)?
- Do tags carry aspects beyond topicality (e.g. time, task, emotion)?

“Cognitively and functionally different representations of information objects may be used in information retrieval to enhance quality of results” (polyrepresentation continuum, Larsen et al. 2006)

Heckner et al., Personal Information Management vs. Resource Sharing, 4
Empirical Research on Tagging

- Statistical analysis on tag cloud development (Golder and Hubermann 2006, Hammond 2005)
- Tag data mining (Begelman 2006, Halpin 2007)
- Some case studies on tagging, focusing on a small amount of examples Bar-Ilan 2006
- Work on using tagging in enterprises (Maßun 2009, Farrell & Lau 2007)
A Model for Tags

- Empirical study on tagging behavior in Connotea
  1. Explorative creation of a category model
  2. Explanatory case study: Applying and verifying the category model

- Emerging tag models for
  - Language (Linguistic Tag Category Model, LTCM)
  - Function (Functional Tag Category Model, FTCM)
  - Relationship Tag – Text (Tag-to-text Category Model, T2TCM)

HECKNER, MÜHLBACHER & WOLFF 2008

Heckner et al., Personal Information Management vs. Resource Sharing, 6
Linguistic and functional categorisation

- **numbers**: 72%
- **acronyms**: 12%
- **adjectives**: 15%
- **not content-related**: 6%
- **content-related**: 94%

Very little time or task-related tags

“content related”: 96%
“general content description”, not: method, codes, categories

HECKNER, MÜHLBACHER & WOLFF 2008
Heckner et al., Personal Information Management vs. Resource Sharing, 7
Comparison of Tagging Systems

- **Connotea**: Scientific articles
- **Flickr**: photos
- **YouTube**: Videos
- **Delicious**: Bookmarks

- **media type influences tagging**
  - number and language of tags
  - different function of tags
- **data acquisition via system APIs or RSS feeds**
- **TACS** – web-based tool for tag analysis

Heckner et al., Personal Information Management vs. Resource Sharing, 8
Motivation – Why do people tag?

- Tagging studies without direct user contact / interviews so far
- study using the *Mechanical Turk* service ("artificial artificial intelligence") as a research instrument for contacting “real taggers” (http://www.mturk.com)

HECKNER, HEILEMANN & WOLFF 2009

Heckner et al., Personal Information Management vs. Resource Sharing, 9
Scientific crowdsourcing using MTurk

- Why use MTurk?
  1. Mechanical Turk surveys are comparatively cheap, medium scale surveys incorporating more than one hundred test subjects are possible.
  2. Vital importance of reaching real users of social tagging platforms

- Problems
  - Verification of users’ authenticity
  - Biased population in MTurk?
  - Does not work for all platforms

Heckner et al., Personal Information Management vs. Resource Sharing, 10
Assumption: Different motivations for tagging

- Organization of one’s own digital content, i.e. personal information management
  - Delicious
  - Connotea
- Media and information sharing:
  - Flick
  - YouTube

Heckner et al., Personal Information Management vs. Resource Sharing, 11
Questionnaire Design: Question Types

- Online questionnaire posted as a „human intelligence task“ (HIT) on MTurk
- Rating on a 7-point Likert scale + open questions
- Describing most recent system experience (critical incidence technique, Flanagan 1954)
- Measuring personal information management with two items, resource sharing with three items

Heckner et al., Personal Information Management vs. Resource Sharing, 12
Questionnaire Design: Structure

- General information (demographics, 7 questions)
- General motivation (e.g., “I use XXX to share bookmarks with my friends or family”, 10 questions)
- Tagging Motivation and Understanding (e.g., “I tag, because it helps me finding things using my own words”, 14 questions)
- Social Bookmarking and Search (e.g., “Tags of other users help me to discover their bookmarks”, 5 questions)
- Recent usage (open questions, 4 questions)
- Open questions (e.g., “I like XXX, because …”, 4 questions)

Heckner et al., Personal Information Management vs. Resource Sharing, 13
Demographics

- 142 participants (68 female, 71 male and 3 unspecified)
- 48 Flickr, 47 Youtube, 32 Delicious, 15 Connotea
- Age distribution:
  - 18-25 (66)
  - 26-35 (46)
  - 36-45 (20)
  - 46-55 (9)
  - >=56 (1)
Results

- motivation analysis
- perception of tagging
- Social Tagging and perception of IR

Heckner et al., Personal Information Management vs. Resource Sharing, 15
Results for Motivation

- quantitative analysis
  - systems differ significantly (one-way-Anova) for both dimensions
  - YouTube is significantly weaker for PIM

- qualitative analysis
  - shows much stronger differences (see below)
  - no additional / other type of motivation detected
Users’ explicit rating of sharing (green) vs. PIM (orange)
User motivations emerging from qualitative judgements (sharing (green) vs. PIM (orange))

Heckner et al., Personal Information Management vs. Resource Sharing, 18
Examples for users’ PIM motivations

- “To keep an archive of useful info for work and hobbies.” (Delicious)
- “I add bookmarks to my delicious in order to keep the all organized. And since they are online, I can keep track of cool websites from anywhere.” (Delicious)
- “To be able to find the item again, not search through 3 pages of Google results hoping to come on the exact link again.” (Connotea)
- “Finding it again, grouping it together with similar papers, organizing/clustering publication space.” (Connotea)
Examples for users’ sharing motivations

- “I uploaded images of my progress on a long distance driving trip, to keep my friends and family updated on that progress.” (Flickr)
- “I just wanted to add pictures which can prove useful for others as wallpapers.” (Flickr)
- “I uploaded a movie about my nephew to share with my other relatives.” (YouTube)
- “To share a video of my daughter’s piano recital with family who were not able to be there.” (YouTube)

Heckner et al., Personal Information Management vs. Resource Sharing, 20
Perception of Tagging

- **ease of tagging**: Connotea users perceive tagging as most easy, followed by YouTube, Delicious, and Flickr (not significant)

- **tagging as a useful feature**: Connotea users agree very strongly (M = 6.8 !), YouTube, Flickr and Delicious users to a lesser degree (significant: A Tukey post-hoc test shows that Connotea significantly differs from Flickr and YouTube)

- **tagging as a means of classification**: Fair agreement for Connotea, to a lesser degree for Flickr, YouTube and Delicious (differences not significant)

Heckner et al., Personal Information Management vs. Resource Sharing, 21
Two dimensions of social interaction (Shneiderman, 2002)

Activity dimension

- Collect
- Relate
- Create
- Donate

Social Spheres

- Self
- Family & Friends
- Colleagues & Neighbours
- Citizens & Markets
Social Spheres of Media Sharing

Friends and Family

Citizens and Markets

Colleagues and Neighbors

Heckner et al., Personal Information Management vs. Resource Sharing, 23
Social Tagging and perception of IR

Search other collections

1. Flickr
2. Youtube
3. Delicious
4. Connotea

Tags perceived as helpful for IR

Search own collection

 Heckner et al., Personal Information Management vs. Resource Sharing, 24
Towards a Model of Tagging Behavior

- Ben Shneiderman’s approach towards social software
  - „Leonardo’s Laptop“
  - relationship spheres
  - classification of activities
  - Activities and relationship table (ART)

- Cool & Belkin’s classification of information interaction

Heckner et al., Personal Information Management vs. Resource Sharing, 25
Activity types

Information Behaviors Facet  
(Cool & Belkin 2002)
- Create
- Disseminate
- Organize
- Preserve
- Access (Method, Mode)
- Evaluate
- Comprehend
- Modify
- Use

Activity Dimensions  
(Shneiderman 2002)
- Collect
- Relate
- Create
- Donate

Social Spheres:
- Self
- Family and Friends
- Colleagues
- Citizens and Markets

Heckner et al., Personal Information Management vs. Resource Sharing, 26
<table>
<thead>
<tr>
<th>Information Behaviour in Social Tagging Systems</th>
<th>Information Behaviour Classification (Cool and Belkin 2002)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“select item”</td>
<td>access</td>
</tr>
<tr>
<td>“tag (specific sharing)’”</td>
<td>create with intention to disseminate to one or more selected users</td>
</tr>
<tr>
<td>“tag (unspecific sharing)’”</td>
<td>create with intention to disseminate to the whole community</td>
</tr>
<tr>
<td>“tag (PIM)”</td>
<td>create with intention to organize or preserve</td>
</tr>
<tr>
<td>“add to system”</td>
<td>disseminate</td>
</tr>
<tr>
<td>“point to questions from other users”</td>
<td>disseminate Communication Behaviour: medium (text, video, image), mode (mediated), mapping (one-to-one or one-to-many)</td>
</tr>
<tr>
<td>“point to own questions”</td>
<td>disseminate Communication Behaviour: medium (text, video, image), mode (mediated), mapping (one-to-one or one-to-many)</td>
</tr>
<tr>
<td>“later re-retrieval”</td>
<td>access (method: searching, mode: specification)</td>
</tr>
<tr>
<td>“retrieve”</td>
<td>access (method: searching, mode: (internal / user’s) specification)</td>
</tr>
<tr>
<td>“worth viewing?”</td>
<td>evaluate takes place before accessing</td>
</tr>
<tr>
<td>“follow link from email”</td>
<td>access (method: searching, mode: (external) specification)</td>
</tr>
<tr>
<td>“browse”</td>
<td>access (method: scanning, mode: recognition)</td>
</tr>
<tr>
<td>“read”</td>
<td>comprehend</td>
</tr>
<tr>
<td>“view”</td>
<td>comprehend</td>
</tr>
</tbody>
</table>
Open Questions / Future research

■ how to **measure retrieval quality** for tagging systems – comparisons with other approaches?
■ what are **system and UI-related influence factors** for tagging strategies (e.g. tag avoidance or phrasal tags)?
■ in what situations do users want to tag actively? (in what won’t they?)
■ what is the **role of context and system type**?
■ study **tagging for other media types** (e.g. music)

Heckner et al., Personal Information Management vs. Resource Sharing, 29
Some credits

- The work presented here has been done together with Dr. Markus Heckner (now with Accenture)
- Additional collaborators have been
  - Michael Heilemann
  - Tanja Neubauer
  - Dr. Susanne Mühlbacher


Heckner et al., Personal Information Management vs. Resource Sharing, 31
Bibliography II


Heckner et al., Personal Information Management vs. Resource Sharing, 32


