User Goals and Motivation in Social Tagging Systems

Markus Strohmaier

Knowledge Management Institute
Graz University of Technology, Austria

e-mail: markus.strohmaier@tugraz.at
web: http://www.kmi.tugraz.at/staff/markus
Observations (in 2006/07)  
Knowledge Transfer and Social Media

- Knowledge transfer effectiveness is related to the **participants** of knowledge transfer, and their **goals**
- Analyzing the goals of knowledge transfer participants is **critical to KM** [9], but **large-scale access to data about user goals is difficult**
- **Social tagging systems increasingly used** for knowledge transfer

→ **“Agent-Oriented Engineering of Social Software”**

**TransAgere: A 2,5 year basic research project (2008-2010)**
- Focusing on knowledge transfer and the role of agents and goals in social software
- Currently 2 PhD students and 2 Master student
- Funded by the FWF Austrian Science Fund

Q: What are the **goals** and **motivations** of users on the web?
Collaborations

Graz University of Technology
- Mark Kröll
- Christian Körner
- Roman Kern
- Hans-Peter Grahsl

Klagenfurt University
- Mathias Lux

University of Southern California
- Andrew Gordon

University of Kassel
- Andreas Hotho

University of Toronto
- Eric Yu
17-Sep-82 14:59 Joseph Ginder at CMU-10A (*%)

I believe that the joke character should be % rather than *.

16-Sep-82 00:15 Jeff Shrager at CMU-10A

Just signifying that a message is a joke is certainly not sufficient.

One can develop a taxonomy of bboard message types along several different dimensions. Also, where a continuum is preferable to a taxonomy (such as where humor value is at issue) one can similarly use a scale to indicate where along that scale this message lies. Suppose that all dimensions are referred to by a ten point scale (we’ll use all integers here although one can certainly imagine reals in the case of fine grain continuous scales).

Some dimensions will be bitwise encoded as well.

Here is a sample of a coding scheme:

COMMUNITY: (this is a binary scale with a bit position for each department totalling about 32 bits)

TOPIC: (two digits 00-99)

(00) Political, (01) Scientific, (02) Computer, (03) Meta, etc

FLAME VALUE: (continuous 0.0-10.0)

HUMOR VALUE: (0.0-10.0)

BORDOM VALUE: (0.0-10.0)

INFORMATIONAL CONTENT: (-10.0 (for queries) to 10.0 (for their answers))

Note that some of these scales are purely according to the opinion of the author. Thus, we provide, also, a confidence scale: to go along with each continuous scale (to be enclosed in parens after the value).

(-) Proposed

(*/%&/#) Proposed

*: Proposal

Reply

Joke
Motivating Example: Intent vs. Content

**Intent**
*(What goals it aims at / helps to achieve)*

- find a physician
- organize a high-school reunion
- contact an old friend
- organize a marketing campaign
- find others who share the same family name
- find my way to an address
- ...

**Content**
*(What it is)*

With the advent of user-participation, a **unique opportunity** to make user **goals** amenable to scientific investigations has been created.

The web as a lab to study human intent.

---

Markus Strohmaier

2009
Traces of User Intent on the Web

[Battelle 2007]

A) Search Query Logs

B) Question Answering

C) Human Life Goals

D) Short Messages on the Web
An *Intentional Web*

Goal-Oriented Analysis and Engineering of Social-focused Web Applications

The process of web science

- How do intentional structures, relationships and communities form?
- What mechanisms aid reasoning about user goals?
- What are the goals of users on the web?

Web: Science → Engineering

- How can intentional structures be integrated into web applications?
- What forms of goal-oriented evaluation methods are adequate?

Emergent Goals → Intentional Structures

Issues → Idea → Design

Macro → Micro

Berners-Lee
Lovelace lecture 2007-03-13

Markus Strohmaier
2009
Agenda

Introduction

• **Social Tagging Systems**
  *A brief introduction to tagging research*

User Goals on the Web

• **Goal Acquisition via Tagging**
  *Making user goals more explicit during tagging*

• **Automatic Intent Annotation**
  *Automatically annotating textual resources with human intent*

• **Construction of Large Scale Goal Graphs**
  *Constructing goal graphs from search query logs*

User Motivation on the Web

• **Automatic Detection of Tagging Motivation**
  *Detecting user motivation in tagging systems*
Social Tagging Systems - Example

User

Resources

Tags

- Radish - Indoor Solar-powered Calendar Display
- MUJI USA ONLINE STORE
- agile approach | World Bank Open API 2.0 Launched
- Reverend
- Pogoplug — Access your external hard drive from anywhere on the Internet
- visualizing sorting algorithms
- Times

- hw
- tobuy
- data
data
tobuy
storage
- ir
- python
- tobuy
- storage
- graphics
- rss
- osx

Markus Strohmaier 2009
Folksonomy: Definition

[Fotho et al 2006]

A **folksonomy** is a tuple $F := (U, T, R, Y)$ where

- the three disjoint, finite sets $U$, $T$, $R$ correspond to
  - a set of persons or users $u \in U$
  - a set of tags $t \in T$ and
  - a set of resources or objects $r \in R$

- $Y \subseteq U \times T \times R$, called set of *tag assignments* \\

A **personomy** $P_u$ of user $u$ is the restriction of $F$ to $u$

Folksonomy Analysis

Three possible two-mode networks (or bipartite graphs):
- $M_{UT}$, $M_{TR}$, $M_{UR}$

Example: $M_{TR}^T$

<table>
<thead>
<tr>
<th></th>
<th>Tag 1</th>
<th>Tag 2</th>
<th>Tag 3</th>
<th>Tag 4</th>
<th>Tag 5</th>
<th>Tag 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>URL 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>URL 2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>URL 3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- Co-occurrence Analysis / Similarity Measures

How similar are tags w.r.t. co-occurrence with resources?

Raw Similarity: $M_T = M_{TR} * M_{TR}^T$

others: Dice, Cosine, Jaccard
Ontology Construction from Folksonomies

Related Concepts

Super/Subconcepts

<table>
<thead>
<tr>
<th>Broader</th>
<th>Narrower</th>
</tr>
</thead>
<tbody>
<tr>
<td>rss</td>
<td>atom</td>
</tr>
<tr>
<td>cmyk</td>
<td>rgb</td>
</tr>
<tr>
<td>cell</td>
<td>umts, wcdma, ev-do</td>
</tr>
<tr>
<td>phone</td>
<td>cell</td>
</tr>
<tr>
<td>ajax</td>
<td>json</td>
</tr>
<tr>
<td>xml</td>
<td>xslt</td>
</tr>
<tr>
<td>rdf</td>
<td>owl</td>
</tr>
<tr>
<td>flickr</td>
<td>gmail, picasa</td>
</tr>
<tr>
<td>ruby</td>
<td>rails</td>
</tr>
<tr>
<td>mac</td>
<td>iphoto</td>
</tr>
<tr>
<td>java</td>
<td>j2ee</td>
</tr>
<tr>
<td>google</td>
<td>gds</td>
</tr>
<tr>
<td>search</td>
<td>a9, engine</td>
</tr>
<tr>
<td>linux</td>
<td>ubuntu, gnome</td>
</tr>
</tbody>
</table>

Raw Similarity:

$$M_T = M^T_{TR} * M^{-T}_{TR}$$

How do user goals and motivations relate to folksonomies? [Mika 2005]
Variability in Folksonomies
[Strohmaier 2008]

Types of Users U
• at least four user roles including 1) resource author, 2) resource collector 3) indexer or tagger and 4) searcher [Voss 2007].

Types of Tags T
• For example, types of tags include: 1) Identifying what a resource is about 2) Identifying what it is 3) Identifying who owns it 4) Refining categories 5) Identifying qualities or characteristics 6) Self reference 7) Task organizing [Golder und Hubermann 2006]

Types of Objects or Resources R
• Different „Objects of sociality”: movies (youtube), URLs (delicious), photos (flickr), music (last.fm), etc.

Different Motivations: „Why do users tag?“
Examples: categorize resources, describe resources, etc

Different User Goals: „What do users want while tagging?“
Examples: buy a car, book a trip, prepare a presentation, etc.
TransAgere: An *Intentional* Perspective on Web

Lift „traces of intent“ on the web to the level of explicit knowledge

1. **Build goal extraction and acquisition algorithms** for the web
   To make implicit intentional knowledge more explicit

2. **Build large-scale goal graphs** that enable
   - Analysis of user goals on the web
   - Intent estimation
   - Reasoning about user goals

3. **Demonstrate how knowledge about user goals can enhance (social) web applications**
   - Intentional query expansion
   - Intent detection and tagging recommendation
   - Goal oriented clustering of users and resources
Agenda

Introduction
- Social Tagging Systems
  *A brief introduction to tagging research*

User Goals on the Web
- Goal Acquisition via Tagging
  *Making user goals more explicit during tagging*
- Automatic Intent Annotation
  *Automatically annotating textual resources with human intent*
- Construction of Large Scale Goal Graphs
  *Constructing goal graphs from search query logs*

User Motivation on the Web
- Automatic Detection of Tagging Motivation
  *Detecting user motivation in tagging systems*
Traces of User Goals in Social Tagging Systems

How can the degree to which users make their goals explicit be increased during tagging?

Observation: Tags exhibit a low degree of „intentional explicitness“

<table>
<thead>
<tr>
<th>Amazon</th>
<th>Youtube</th>
<th>Cnn</th>
<th>Wikipedia (en)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books to read (4)</td>
<td>To consume (2)</td>
<td>Up to date (1)</td>
<td>To watch (4)</td>
</tr>
<tr>
<td>Books to buy (2)</td>
<td>To watch (2)</td>
<td>On dock to read (1)</td>
<td>To read (2)</td>
</tr>
<tr>
<td>To read (2)</td>
<td>Films to see (1)</td>
<td>To test (1)</td>
<td>To print (2)</td>
</tr>
<tr>
<td>Want to read (2)</td>
<td>To see (1)</td>
<td>For me to read (2)</td>
<td></td>
</tr>
<tr>
<td>Technologies to evaluate (1)</td>
<td></td>
<td>Technologies to evaluate (1)</td>
<td></td>
</tr>
<tr>
<td>To process (1)</td>
<td></td>
<td>To be seen (1)</td>
<td></td>
</tr>
<tr>
<td>Want to see (1)</td>
<td></td>
<td>To George (1)</td>
<td></td>
</tr>
<tr>
<td>To write (1)</td>
<td></td>
<td>To visit (1)</td>
<td></td>
</tr>
</tbody>
</table>
Relating Content and Intent via Tagging

Proposal: Purpose tags

Definition: Purpose tags capture the goals of users while they tag resources e.g. capture that a resource \( r \) helps to achieve goal \( g \)

Intuition:
1) Goal-Resource relations are meaningful on multiple levels (requirements/documents, bugs/code, ..)
2) Folksonomy analysis can help to uncover latent intentional structures

Approach: Encourage purpose tags: When assigning such tags, users tag (a resource) with a specific purpose or goal in mind.

Example: When bookmarking the website www.facebook.com, a user might tag the resource with a purpose tag \( T_p \) “organize a high school reunion”.
A Goal Oriented Social Tagging Prototype

To Join the community: Please Register by using the sidebar

This community provides 147 URL's for 98 goals

[avoid the Euro 2008 in Graz] [find running sushi restaurants in graz] [second hand gewand kaufen] [get educated in graz] [learn about four elements festival] [learn about graz] [find english movies to watch in graz] [find a place located in Graz] [Watch a movie in Graz] [buy a car] [cycling in graz] [find public viewing places in Graz] [swimming in graz] [find events about Graz] [find a pharmacy in graz] [watch the Euro 2008 in Graz] [find asian food in graz] [get graz news] [wo und wie ein kind taufen lassen] [find events in Graz]

with students Andreas Haselsberger and Christoph Ruggenthaler
1. Can folksonomy analysis help uncover latent intentional structures?

**Results:**
- 2 weeks
- $|R| = 123$
- $|T| = 75$
- $|U| = 19$

$$M_T = M_{TR} * M_{TR}^T$$
Formal Concept Analysis: Inferring Hierarchical Relations from Folksonomies
[Wille 2005]

Sub/Superconcept Relation

• $A \subseteq G$, $B \subseteq M$

  $A' := \{m \in M \mid \forall g \in A: (g, m) \in I\}$

  $B' := \{g \in G \mid \forall m \in B: (g, m) \in I\}$

  Intent: all attributes shared by all objects of $A$
  Extent: all objects having all attributes of $B$

\[
\begin{align*}
\text{URL 1} & \quad \text{Tag 1} & \quad \text{Tag 2} & \quad \text{Tag 3} & \quad \text{Tag 4} & \quad \text{Tag 5} & \quad \text{Tag 6} \\
A_1 & \quad \times & \quad \times & \quad \times & \quad \times & \quad \times & \quad \times \\
A_2 & \quad \times & \quad \times & \quad \times & \quad \times & \quad \times & \quad \times \\
B_1 & \quad \text{A}_{1'} & \quad \text{A}_{1'} & \quad \text{A}_{1'} & \quad \text{A}_{1'} & \quad \text{A}_{1'} & \quad \text{A}_{1'} \\
B_2 & \quad \text{A}_{2'} & \quad \text{A}_{2'} & \quad \text{A}_{2'} & \quad \text{A}_{2'} & \quad \text{A}_{2'} & \quad \text{A}_{2'} \\
C_1 & \quad \text{A}_{1'} & \quad \text{A}_{1'} & \quad \text{A}_{1'} & \quad \text{A}_{1'} & \quad \text{A}_{1'} & \quad \text{A}_{1'} \\
C_2 & \quad \text{A}_{2'} & \quad \text{A}_{2'} & \quad \text{A}_{2'} & \quad \text{A}_{2'} & \quad \text{A}_{2'} & \quad \text{A}_{2'} \\
\end{align*}
\]

$(A_1, B_1) \leq (A_2, B_2) :\iff A_1 \subseteq A_2 (\iff B_1 \supseteq B_2)$.

The orange concept is a subconcept of the blue concept, if its extent is contained in the blue one. (equivalent to the blue intent is contained in the orange one)
Social Construction of Hierarchical Goal Graphs

1. Can folksonomy analysis help uncover latent intentional structures?

Based on Formal Concept Analysis [Wille 2005] visualized with ConExp

Partially Ordered Sets over a Bi-Partite Graph

Markus Strohmaier
2009
Agenda

Introduction
• Social Tagging Systems
  A brief introduction to tagging research

User Goals on the Web
• Goal Acquisition via Tagging
  Making user goals more explicit during tagging
• Automatic Intent Annotation
  Automatically annotating textual resources with human intent
• Construction of Large Scale Goal Graphs
  Constructing goal graphs from search query logs

User Motivation on the Web
• Automatic Detection of Tagging Motivation
  Detecting user motivation in tagging systems
Can we automatically infer categories of human intent from natural language text?

**Intent annotations** aim to describe which goals are referenced in given textual resources (such as “find a job”, “get a degree”).

**Problem**: Goals are rarely made explicit in natural language text

**Approach**: link text with a taxonomy of human goals via indicative actions

**Datasets**: transcripts of political speeches because

1) clear focus on discussing, conveying or achieving goals
2) are less affected by noise
3) contain a broad variety of intentions.

⇒ 44 speeches by B. Obama and J. Mccain from 2008
Profiles of People’s Interests

- Knowledge about a person’s interests can be used to create an informative profile.

- From knowing people’s goals and interests one can infer:
  - Their opinions
  - Their relationship with other people
  - Their attitude towards life

- Acquiring the data represents the easy part:
  - Weblogs
  - Transcripts of political speeches

- Creating an interest profile out of it, the more challenging part.
The idea is to:

1.) collect a list of representative actions that hint towards goal categories (→ Knowledge Base)
2.) based on the identification of actions, goal categories are assigned
In order to avoid wrinkles, you need to moisturize inside and out.

Yahoo! BOSS Phrase search results:

- "in order to avoid wrinkles" you need to moisturize inside and out.

Knowledge Base:

- Looking Young
  - You need to moisturize inside and out.
  - But the biggest reason women have such high risk of vitamin D deficit according to Holick, women are encouraged to avoid all sunlight and skin cancer.

Political Speeches:

- 21 Speeches by B. Obama
- 23 Speeches by J. McCain

Yahoo! BOSS Phrase search results:

- "in order to avoid wrinkles" you need to moisturize inside and out.

Helping Others

- Being intelligent
- Being respected
- Being responsible

- Easy Life
- Ethical Feels

- Pursuing ideals
- Seeking help

Indicative Actions

- Avoid wrinkles
- Age well
- Be vibrant with Energy
- Looking Vital

Knowledge Base Construction

- Taxonomy of Human Goals:
  - Abbreviation
  - Full label
    - Achieving salvation
    - Arts
    - Appetite
    - Attracting sexually
    - Avoiding failure
    - Avoiding guilt
    - Avoiding rejection
    - Essential for
    - Necessary for
    - ...
Figure 2: Intent profiles of Obama and McCain, generated from and averaged over 44 speeches (April and June 2008).
Automatic Intent Annotation: Results

Based on iTAG

Based on counting words and stop word lists

Intent Tag Clouds

Helping Others Being better than others Being Creative Being Honest Being intelligent Being Respected Being responsible Being socially attractive Belonging Career Charity Content with myself Easy Life Ethical Feeling Safe Good Parent Leader Money Peace of Mind Personal Growth Pursuing ideals Seeking Fairness Seeking Justice Self esteem Teaching

Aspirations Attracting sexually Avoiding stress Being better than others Being creative Being curious Being Free Being Likeable Bob Charity Content with myself Control over others Easy life Education Ethical Exercising Exploring Feeling Safe Helping Others Knowing many others Leader Money Pursuing Ideals Seeking justice Teaching

Traditional Tag Clouds

america american americans better care change country economy energy government great health iraq make need new oil people president security tax time way world years

america american americans bush care change children country don energy families health help jobs just know make mcain need new people president tax time washington work workers working world years
Intent Annotation: Evaluation

Evaluation Setup:
- two human subjects annotate the speeches
- union of annotations for Obama’s speeches produced by the two human subjects to mitigate data sparsity

Evaluation Results:
- Out of the top 25 intent categories produces by iTAG, the annotations produced by human annotators agreed with the automated iTAG approach in 10 cases (40%).
- Up to 40% recall (10 relevant annotations), the iTAG approach achieves a precision of 50% and above.
Agenda

Introduction
• Social Tagging Systems
  *A brief introduction to tagging research*

User Goals on the Web
• Goal Acquisition via Tagging
  *Making user goals more explicit during tagging*
• Automatic Intent Annotation
  *Automatically annotating textual resources with human intent*
• Construction of Large Scale Goal Graphs
  *Constructing goal graphs from search query logs*

User Motivation on the Web
• Automatic Detection of Tagging Motivation
  *Detecting user motivation in tagging systems*
Search Query Logs

Implicit tagging of resources via queries

Search Query Logs have a structure similar to folksonomies

- Users U
- Queries (or Tags) T
- Click-through (or Resources) R

Do search queries contain explicit statements of user goals (e.g. “buy a car”)?
Can folksonomy analysis be applied to develop goal graphs of queries containing goals?

We have developed an

- Automatic method to identify goals and construct goal graphs from search query logs
- Based on 2 datasets (Search Query Logs)
  - AOL (>20mio queries) and Microsoft Research (>15mio queries)

- Results:
  - Identification of 70,000 queries that contain explicit statements of goals (based on AOL dataset)
  - Precision of entries: ~70%
  - Folksonomy analysis to construct a large scale goal graph

- Details in [Strohmaier 2008, Strohmaier 2008a]
A „long tail“ of user goals

Population-specific goals in the „head“

Table 6: The 20 most frequent queries in the result set.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Query</th>
<th>#Users</th>
<th>Nr.</th>
<th>Query</th>
<th>#Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>add screen name</td>
<td>205</td>
<td>11</td>
<td>cancel sql service</td>
<td>54</td>
</tr>
<tr>
<td>2</td>
<td>create screen name</td>
<td>137</td>
<td>12</td>
<td>pump my myspace</td>
<td>53</td>
</tr>
<tr>
<td>3</td>
<td>rent to own</td>
<td>120</td>
<td>13</td>
<td>cancel sql account</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>listen to music</td>
<td>108</td>
<td>14</td>
<td>“deleted”</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>pump my space</td>
<td>102</td>
<td>15</td>
<td>“deleted”</td>
<td>48</td>
</tr>
<tr>
<td>6</td>
<td>pump my ride</td>
<td>97</td>
<td>16</td>
<td>how to lose weight</td>
<td>47</td>
</tr>
<tr>
<td>7</td>
<td>assist to sell</td>
<td>93</td>
<td>17</td>
<td>how to get pregnant</td>
<td>47</td>
</tr>
<tr>
<td>8</td>
<td>wedding cake hoppers</td>
<td>64</td>
<td>18</td>
<td>change my password</td>
<td>46</td>
</tr>
<tr>
<td>9</td>
<td>skating with celebs</td>
<td>58</td>
<td>19</td>
<td>discover credit card</td>
<td>46</td>
</tr>
<tr>
<td>10</td>
<td>lose weight fast</td>
<td>56</td>
<td>20</td>
<td>check my computer</td>
<td>43</td>
</tr>
</tbody>
</table>

Figure 3: Rank-frequency plot of queries containing unique user goals.
Search Query Log Analysis - Results

M. Strohmaier, P. Prettenhofer, M. Kroell, Acquiring Explicit User Goals from Search Query Logs
In Proceedings of the International Conference on Web Intelligence WI’08, Agent and Datamining Interaction Workshop, Sydney, Australia, 2008

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Verb: get</th>
<th>Verb: make</th>
<th>Verb: change</th>
<th>Verb: be</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>get pregnant*  (141)</td>
<td>make money*  (87)</td>
<td>change my password (100)</td>
<td>be anorexia* (28)</td>
</tr>
<tr>
<td>2</td>
<td>get rid of ants (28)</td>
<td>make your own website (43)</td>
<td>change my screen name (38)</td>
<td>be pregnant* (19)</td>
</tr>
<tr>
<td>3</td>
<td>get out of debt planner (19)</td>
<td>make money at home (41)</td>
<td>change screen name (32)</td>
<td>be bulimic (12)</td>
</tr>
<tr>
<td>4</td>
<td>get rich or die from (12)</td>
<td>make money fast (39)</td>
<td>change my aol password (28)</td>
<td>be rich* (11)</td>
</tr>
<tr>
<td>5</td>
<td>get rid of love handles (17)</td>
<td>make money online (34)</td>
<td>change password (24)</td>
<td>be emo (8)</td>
</tr>
<tr>
<td>6</td>
<td>get married (15)</td>
<td>make the band (30)</td>
<td>change my aol profile (21)</td>
<td>be funny* (8)</td>
</tr>
<tr>
<td>7</td>
<td>get rich* (15)</td>
<td>make money from home (25)</td>
<td>change your name (21)</td>
<td>be happy* (8)</td>
</tr>
<tr>
<td>8</td>
<td>get rich with trump (15)</td>
<td>make new screen name (24)</td>
<td>change* (20)</td>
<td>be sexy* (7)</td>
</tr>
<tr>
<td>9</td>
<td>get out of debt* (15)</td>
<td>make up (23)</td>
<td>change my email address (17)</td>
<td>be in love* (7)</td>
</tr>
<tr>
<td>10</td>
<td>get rid of moles (14)</td>
<td>make out (21)</td>
<td>change aol password (14)</td>
<td>be an actress (7)</td>
</tr>
</tbody>
</table>

Table 8: The 10 most frequent verb phrases containing the verbs “get”, “make”, “change”, and “be” are listed.

Goals marked with (*) are also included in ConceptNet Commonsense Knowledge Base v2.1
[H. Liu and P. Singh 2004]
Idea III

Approach: Treat the set of all queries \(\{q_{-n} \ldots q_i \ldots q_n\}\) (\(n=0\)) within the \(n^{th}\) environment of the explicit intentional query \(q_i\) as tags for \(q_i\).

With \(n=6\), this approach results in tagging “how to have good breast milk” with the following tags (excerpt):

- Breast milk
- [Yellow breast milk]
- [Breast feeding and going back to work]
- [Nestle formula]
- [Free nestle formula]
- [Good start]
- [What fenugreek]

http://www.verybestbaby.com
http://www.breastfeeding.com
Excerpt of Goal Graphs Constructed from Search Query Logs

Based on the AOL Search Query Log, ~ 20 mio queries [Pass 2006]
Goal Graph with ~ 70,000 goals, precision of entries: ~ 70%

Available via a RESTful Web-Service
SearchGoalNet – A RESTful API

joint work with C. Körner

Example: .../GoalNetService/goals/car?id=dummy
Replace dummy with your username

/goals?id=dummy returns all goals in json format (Value: "Goals", Key: Array of Strings representing the goals)
/tags?id=dummy returns all tags in json format (Value: "Tags", Key: Array of Strings representing the tags)

/goals/making quick cash?id=dummy returns all tags connected to the goal in json format (Key: tag, Value: Weight, Key: tag, Value: Weight, etc.)
/tags/car?id=dummy returns all goals connected to the tag in json format (Key: goal, Value: Weight, Key: goal, Value: Weight, etc.)

/rawRelatedGoals/making quick cash?id=dummy returns all goals which share one or more tags with the goal in json format (Key: goal, Value: array of key, value pairs - goal and number of shared tags)

/predict/car dealer?id=dummy&text=DICE&graph=DICE&alpha=0.5 returns all goals which are predicted by the goal prediction in json format (Key: query, Value: array of key, value pairs - goal and calculated weight)

/relatedGoals/making quick cash?id=dummy&text=DICE&graph=DICE&alpha=0.5 returns all goals which are predicted by the goal prediction in json format (Key: query, Value: array of key, value pairs - goal and calculated weight)

/URLofGoal/how to make money from stocks?id=dummy returns all URLs which are associated to a given goal in json format (Key: goal, Value: all urls in apostrophes, separated by commas)

Available Similarities:
- DICE
- COSINE
- OVERLAP
- ADAPT_OVERLAP
- JACCARD

The provided alpha range has to be between 0.0 and 1.0.
Prototype: Intentional Query Expansion

Greasemonkey: joint work with F. Wöriser

Evaluation Setup:
5 human annotators
35 short queries each (1-2 words)

Evaluation Task:
"Would goal G be a goal of a user who issues query Q?"

Answer 1: Yes, G is a plausible or potential goal for Q
Answer 2: No, G is not a relevant goal for Q

Evaluation Results:
Pairwise average interrater agreement: kappa = 0.67
Precision: 72% (based on a majority vote among 5 annotators)
Agenda

**Introduction**
- Social Tagging Systems
  *A brief introduction to tagging research*

**User Goals on the Web**
- Goal Acquisition via Tagging
  *Making user goals more explicit during tagging*
- Automatic Intent Annotation
  *Automatically annotating textual resources with human intent*
- Construction of Large Scale Goal Graphs
  *Constructing goal graphs from search query logs*

**User Motivation on the Web**
- Automatic Detection of Tagging Motivation
  *Detecting user motivation in tagging systems*
Current Tagging Research

Collective Evolution [Chi 2008]

Individual Evolution [Golder Huberman 2006]

Why do users tag differently?

Hypothesis:
Different Motivations for Tagging

Fig. 3. Two extreme users’ (#575 and #635) tag growth. As they add more bookmarks, the number of tags they use increases, but at very different rates.
Motivation for Tagging: Categorization vs. Description

• **Categorizing Resources:**
  – Motivation: Users want to construct and maintain a *navigational aid* to the resources (URLs, photos, etc) being tagged.
  – Tagging Approach: Tags capture aspects of resources that are important to the *mental model* of the user (e.g. ‘family photos’, ‘trip to Vienna’).
  – Implications: Because tags are very close to the mental models of users, they can act as suitable facilitators for *navigation* and *browsing*.

• **Describing Resources:**
  – Motivation: Users want to *accurately describe* the resources being tagged.
  – Tagging Approach: Tags *capture the content* of the resources (e.g. ‘car’, ‘sunset’)
  – Implications: Because tags focus on content, they can act as suitable facilitators for *retrieval* and *searching*.

In the „real world“, tagging is likely motivated by a combination of both, but we currently lack robust metrics and automatic methods to detect the degree to which users are motivated by one or the other without direct interaction.
Automatic Detection of Tagging Motivation based on Persononomies

Reminder: A *folksonomy* is a tuple $F := (U, T, R, Y)$

A *personomy* $P_u$ of user $u$ is the restriction of $F$ to $u$

<table>
<thead>
<tr>
<th></th>
<th>Tag 1</th>
<th>Tag 2</th>
<th>Tag 3</th>
<th>Tag 4</th>
<th>Tag 5</th>
<th>Tag 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>URL 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>URL 2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>URL 3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

$\rightarrow$ Automatically detecting tagging motivation based on $P_u$ only?
Preliminary Experimental Setup

Datasets (2009):
2000 bookmarks from
- several Bibsonomy persononomies (bookmarks, papers)
- 2 Flickr persononomies (tags, photo albums)
- ESP Game (pictures)
- 1 delicious personomy (bookmarks) each.

Preliminary Metrics:
- Tag Vocabulary size
- Tag Entropy
- Percentage of Tag Orphans
- … (e.g. Tag Overlap)
Preliminary Results: Vocabulary Size

- Growth of tag vocabulary $|T|$
Preliminary Results: Tag Entropy

Tag Entropy \( H(T) = -\sum_{i=1}^{n} p(t) \cdot \log_{2}p(t) \)
Preliminary Results: Orphan Tags

Percentage of Orphans $O(T) = |O|/|T|$
Implications

Tag Recommender Systems

- “Categorizers”, will more likely reject tags that are recommended from a larger user population
- They are primarily interested in constructing and maintaining their own individual tag vocabulary

Search in Tagging Systems

- Tags by categorizers are more subjective, whereas tags by describers are more objective.
- Tags produced by “Describers” can therefore be suspected to be more helpful for search because they focus on the content of resources

Knowledge Acquisition from Folksonomies

- Mixture of “Categorizers” and “Describers” might influence the set of possible folksonomies
Summary

Introduction
• Social Tagging Systems
  A brief introduction to tagging research

User Goals on the Web
• Goal Acquisition via Tagging
  Making user goals more explicit during tagging
• Automatic Intent Annotation
  Automatically annotating textual resources with human intent
• Construction of Large Scale Goal Graphs
  Constructing goal graphs from search query logs

User Motivation on the Web
• Automatic Detection of Tagging Motivation
  Detecting user motivation in tagging systems
Conclusions and Outlook: An *Intentional* Perspective on Web

Lift „traces of intent“ on the web to the level of explicit knowledge

- „Traces of intent“ *at large*, finally!
  - (Semi-)publicly accessible information about user intent

- Expressed in different *degrees of explicitness*
  - Traces of intent are informal, inaccurate, defeasible, implicit, …

- Goal extraction and acquisition methods are *immature*
  - Using natural language processing and network techniques

- Evaluation of large scale goal graphs is *difficult*
  - Scalability is an issue (e.g. goal graphs with >50,000 goals)

- **Reasoning** on the basis of statistically generated goal graphs
  - Only statistical assumptions about validity
Thank you!