On Understanding the Sharing of Conceptualizations

Steffen Staab, Klaas Dellschaft
What is an ontology?

Gruber 93 (slightly adapted by Borst):

An Ontology is a

formal specification ⇒ Executable, Discussable
of a shared ⇒ Group of persons
conceptualization ⇒ About concepts
of a domain of interest ⇒ Between application and „unique truth“

Where do we get the ontologies from?
Collaboration and Knowledge

Degree of Collaboration
(“shared“)

none to low

high

none to low

Formalize

Share

Folksonomies, Text memes

Ontologies

Keywords, Text

Knowledge Base, Logical Schemata, Most “Ontologies“

Degree of Formalization

ISWeb - Information Systems & Semantic Web
Steffen Staab
staab@uni-koblenz.de
Understanding Conceptualization
3 of 38
Language Games

Features
- For playing
- Round shape
- Certain sizes
- Not for eating
- Not liquid

Ball

Duck
Narrow Folksonomy

One resource tagged by at most one user
Broad Folksonomy

Search results for laksa

Related tags: recipes, food, cooking, asian, laksa, recipe, singapore, vegetarian, curry, noodles

Chubby Hubby » Laksa fisherman’s pie, a post inspired by Adventures of an Italian Food Lover

BBC - Food - Recipes - Pumpkin and tofu laksa

Recipe LAKSA

Seafood and Coconut Laksa from Delia Online

Laksa - Wikipedia, the free encyclopedia

yum..yum..that’s laksa???

One pot laksa noodles in sauce recipe - vegetarian - other vegetarian | cuisine.com.au

Sydney’s best laksa - Good Living - Entertainment - smh.com.au

KUIDAORE: Laksa Lemak

Laksa Me melbourne, Australia eating

« previous | next »
Broad Folksonomy

Add your own tag-annotated bookmark
In delicious

Search Results

Search results for "laksa"

Crab cakes with green chilli mayo
Filed Under: Cool foods and Recipes; Simple and All posts

Last weekend, S and I hosted several friends for dinner. Our menu consisted of a few old favourites and a few new ones. We started our dinner with Teage Ezard's crab cakes with green chilli mayonnaise. This was followed by sakura sbi pasta, and then a pan-fried snapper fillet served with laksa sauce, shelled edamame […]

(to view associated entry please click here)

Laksa fisherman’s pie, a post inspired by Adventures of an Italian Food Lover
Filed Under: Cool foods and Cookbooks and Recipes: Time-consuming and All posts

Food author Faith Heller Willinger has had the great fortune of calling Florence home for the last thirty years. (If only we were all so fortunate!) In her latest release, Adventures of an Italian Food Lover, she has created a gorgeous hybrid food lover's tome which is part cookbook, part travel guide and part old-school […]

(to view associated entry please click here)
Collaboration and Knowledge

Degree of Collaboration ("shared")

<table>
<thead>
<tr>
<th>Degree of Collaboration</th>
<th>None to Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>None to Low</td>
<td>Keywords, Text</td>
<td>Knowledge Base, Logical Schemata, Most &quot;Ontologies&quot;</td>
</tr>
<tr>
<td>High</td>
<td>Folksonomies, Text memes</td>
<td>Ontologies</td>
</tr>
</tbody>
</table>

Conceptualization: An epiphenomenon of sharing and explication/formalization!

Formalize

Share
Folksonomies serve as our fruit flies! Not, the real thing, but part of the real world for which we have data!
Dynamics of Knowledge Sharing in Folksonomies
Better understanding of the tagging process

- Cooperative classification of resources
- Which factors influence the tagging process?
  - Background knowledge of the user?
  - Tag assignments of other users?

Hypothesis: Tagging involves imitation of other users AND selection of tags from background knowledge of users.
Overall Scheme

User interface

Something else?

Conceptualization

Own Knowledge

Shared terminology

Model of User Interface Influence

Joint Stochastic Model

Model of Own Knowledge

Model of Sharing

Tagging Behavior

Comparison of Statistics

Simulated Tagging Behavior

Comparison of Statistics

Model of Own Knowledge

Model of Sharing

Tagging Behavior

Joint Stochastic Model

Model of User Interface Influence

Something else?

User interface
Components of Analysis

- Properties of Tag Streams
  - Stream view of Folksonomies
  - Co-occurrence streams
  - Resource streams

- Dynamic model for Tagging Systems
  - Simulating background knowledge
  - Simulating tag imitation

- Simulation Results
  - Co-occurrence streams
  - Resource streams

Observations in the real world

Stochastic models of influence

Which models best fit the reality?
Properties of Tag Streams
Stream Views of a Folksonomy

- Folksonomies:
  - Vertices: Users, tags, resources
  - Edges: Tag assignments
- Postings:
  - Tag assignments of a user to a single resource
  - Can be ordered according to their time-stamp
Co-occurrence Streams

- Co-occurrence Streams:
  - All tags co-occurring with a given tag in a posting
  - Ordered by posting time
- Co-occurrence stream for 'apple':
  - \{mackz, r1, \{apple, tree\}, 13:25\}
  - \{klaasd, r2, \{apple, mac, ibook\}, 13:26\}
  - \{mackz, r2, \{apple, macintosh, stevejobs\}, 13:27\}
  - tree, mac, ibook, macintosh, stevejobs

| Tag | |Y| |U| |T| |R| |
|-----|---|---|---|---|---|
| ajax| 2.949.614 | 88.526 | 41.898 | 71.525 |
| blog| 6.098.471 | 158.578 | 186.043 | 557.017 |
| xml | 974.866 | 44.326 | 31.998 | 61.843 |
Co-occurrence Streams – Tag Growth

linear growth

# of tag assignments

# of distinct tags

Ajax
Blog
XML
Co-occurrence Streams – Tag Frequencies

power law
Resource Streams

- Resource Streams:
  - All tags assigned to a resource
  - Ordered by posting time
- Resource stream for 'r2':
  - \{mackz, r1, \{apple, tree\}, 13:25\}
    \{klaasd, r2, \{apple, mac, ibook\}, 13:26\}
    \{mackz, r2, \{apple, macintosh, stevejobs\}, 13:27\}
  - apple, mac, ibook, apple, macintosh, stevejobs
Resource Streams – Tag Frequencies

![Graph showing tag frequency distribution for different resource streams. The graph plots tag rank against relative tag frequency. The data points are marked with lines for three different resource streams: http://www.googleguide.com/, http://www.pandora.com/, and http://www.netvibes.com/.

Key:
- Solid line: http://www.googleguide.com/
- Dotted line: http://www.pandora.com/
- Dash-dotted line: http://www.netvibes.com/]

Understanding Conceptualization
Simulating the Evolution of Tag Streams
Which of my concepts represent this web page? How do I tag this web page?

Inspiration for conceptualization from:

1. Most popular tags
2. Most recently used tags
3. Tags used for this resource
4. Tags co-occurring with similar text documents
5. Creating completely new tags
6. …

Which combination of inspirations develop the same statistics as the one observed for delicious?
The Delicious User Interface

- Imitating previous tag assignments:
  - Recommended tags: Intersection of tags of a user and tags already assigned to the resource.
  - Your tags: Tags of the user.
  - Popular tags: 7 most popular tags assigned to the resource.
Simulating a Tag Stream

- Start with empty tag stream
- Each simulation step appends a new tag assignment
- Simulation of a single tag assignment:

\[ p(w|t) \]: Probability of selecting word \( w \) for topic \( t \). Modeled by word distributions in a topic centered text corpus.

\[ n \]: Number of visible previous tags.

\[ h \]: Maximal number of previous tag assignments used for determining ranking of the \( n \) distinct tags.
Modeling Background Knowledge

$P_{BK}$: Probability of selecting from background knowledge

- $p(w|t)$: Probability of selecting word $w$ for topic $t$. Modeled by word distributions in a topic centered text corpus.

- $p(w|r)$: Probability of selecting word $w$ for resource $r$. 
Modeling Tag Imitation

- $P_I = 1 - P_{BK}$: Probability of imitating a previous tag assignment
- $n$: Number of visible top-ranked tags
- $h$: Maximal number of previous tag assignments used for determining ranking of the $n$ distinct tags
Simulation Results
Tag growth:
- Influenced by $P_{BK}$ and $p(w|t)$

Tag Frequencies:
- Influenced by $P_{BK}$, $p(w|t)$, $n$, $h$
- $n$: Semantic breadth of a topic (blog: 100 tags, ajax: 50 tags, xml: 50 tags; Cattuto et al. 2007)
- $h$: No hint for realistic values. Good guesses may be 500 and 1000.
Co-occ. Streams – Simulated Tag Growth

![Simulated Tag Growth Graph]

- 0.6 < 1 < 0.9, 0.4 > BK > 0.1
- Ajax
- Blog
- XML

# of distinct tags

# of tag assignments

1e+06
1e+05
1e+04
1e+03
1e+02
1e+01
10
1

Simulating Resource Streams

- $P_l$ and $P_{BK}$: Values comparable to co-occurrence streams
- $p(w|r)$: Approximated by $p(w|t)$
- $n$: 7 tags are visible (cf. Delicious user interface)
- $h$: Smaller value than for co-occurrence streams
Res. Streams – Simulated Tag Frequencies

- http://www.netvibes.com/
- $I=0.6$, $BK=0.4$, $n=7$, $h=300$
- $I=0.9$, $BK=0.1$, $n=7$, $h=300$
Res. Streams – Simulated Tag Frequencies
Conclusions

- **Imitation AND background knowledge** are needed for explaining properties of tag streams
- Probability of imitating previous tag assignments: ~70-90%

<table>
<thead>
<tr>
<th>Model</th>
<th>Frequency Rank</th>
<th>Co-occurrence Streams</th>
<th>Resource Streams</th>
<th>Tag Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polya Urn Model</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>fixed size</td>
</tr>
<tr>
<td>Simon Model</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>linear</td>
</tr>
<tr>
<td>YS Model w/ Memory</td>
<td>+</td>
<td>o</td>
<td>o</td>
<td>linear</td>
</tr>
<tr>
<td>Halpin et al. Model</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>linear</td>
</tr>
<tr>
<td><strong>Our Model</strong></td>
<td><strong>+</strong></td>
<td><strong>+</strong></td>
<td></td>
<td><strong>power-law</strong></td>
</tr>
</tbody>
</table>
Overall Scheme

User interface

Conceptualization

Something else?

Own Knowledge

Shared terminology

Model of User Interface Influence

Joint Stochastic Model

Model of Own Knowledge

Model of Sharing

Tagging Behavior

Comparison of Statistics

Simulated Tagging Behavior

Model of Sharing

Steffen Staab
staab@uni-koblenz.de

Understanding Conceptualizations

38 of 38
Thank You!