Tag Gardening Activities for Folksonomy
Maintenance and Enrichment

Katrin Weller & Isabella Peters
Heinrich-Heine-University Düsseldorf
Institute for Language and Information
Dept. of Information Science

Graz, 04. September 2008
Motivation

How to combine the dynamics of freely chosen tags with the steadiness and complexity of controlled vocabularies?
Two Sides of the Same Coin

- Aim 1: Optimizing folksonomies for specific application scenarios.

- Aim 2: Maintaining and enriching knowledge organization systems (KOS, e.g. ontologies or thesauri) with folksonomy terms.

→ Tag Gardening!
The Two Sides of Tag Gardening

Complexity in structure

- Ontology
- Thesaurus
- Classification
- Nomenclature
- Folksonomy

Extent of captured knowledge domain

Broadening ontologies with folksonomy terms

Enriching folksonomies in structure
Some Problems of Unstructured Folksonomies

• Spelling variants, translations, abbreviations and synonyms have to be considered for query formulation and indexing.

• Tags serve a variety of functions, not just content description (e.g. “toread”, “@Henry”).

• Tags enable browsing by popularity – but not navigation by meanings and semantic interrelations.

• Even on a personal level unstructured tags may become unmanageable.
Source:
http://flickr.com/photos/jup3nep/2473905053
Tag cloud of the user „MarmaladeToday’s“ for 1.157 bookmarks. Source: del.icio.us (5.03.2008).
Some approaches: Adding structure to folksonomies
Some solutions: Adding structure to folksonomies
The Folksonomy Tag Garden
Tag Gardening

• Term coined by James Governor:
  „Like plants or animals, tags evolve in an emergent fashion, open to hybridisation. Stewardship can help grow and put roots down. Helping the darwinian process is tag gardening.“

• We define tag gardening as any activity to edit, reengineer, manipulate or organize tags – in order to make them more productive and effective.

→ Tag gardening is performed on top of existing folksonomies. Users may tag as usual, afterwards gardening activities may be performed for optimization and user support.
Levels of Tag Gardening

Document collection vs. single document level:
Simple form: editing the tags of one single document.
Complex task: handling all tags in a folksonomy system.

Personal vs. collaborative level:
Personomy level: single users edit the personal tags they use within a system.
Folksonomy level: enabling a user community to collaboratively maintain all tags in use.

Intra- and cross-platform level:
Usual case: consider only tags within one platform.
Broader view: for some cases the use of consistent tags across different platforms will be useful.
Tag Gardening Activities:
1. Weeding

- Removing “bad tags”: spelling variants (plural vs. singular, conflation of multi-word tags) and spam through “pesticides”.
- Aim: enhancing recall and a consistent indexing vocabulary.
- Achieved by
  - type-ahead functionality during indexing,
  - editing functionalities for tags after the application (remove, change, etc.),
  - Natural Language Processing of index tags and search tags,
  - indexing and retrieval tutorials or guidelines for users,
  - authorized users as gardeners

→ Simplest form of tag gardening
Tag Gardening Activities: 2. Seeding

- Extending the folksonomy with rarely used “seedlings” if high-frequency tags do not sufficiently discriminate resources.
- Aim: enhancing precision and expressiveness of the folksonomy.
- Achieved by
  - displaying an inverse tag cloud during indexing or particular “green house“ areas where the seedlings may develop and grow,
  - discrete tag suggestions during indexing.
Tag Green Houses

- Problem: high-frequent tags provide high recall – but low precision and low degrees of discrimination.

- Seedlings or „Baby tags“ as additional entry points to explore document collections.
- Promoting baby tags via alternative display options like
  - „New tags“ / trends
  - „infrequent tags“ / inverse tag cloud

Tag „Design“ on Del.icio.us.
Tag Gardening Activities: 3. Landscape Architecture

- Shaping the folksonomy into “flower beds”, distinguishing similar looking “plants”, identifying their “species”, branding each species with labels and giving additional information regarding their application area.
- Aim: enhancing precision and expressiveness of the folksonomy by adding semantics. For query expansion along semantic relations, for enhanced navigation, as basis for semantic-oriented display.
- Achieved by
  - conflations of multi-language tags,
  - summarization of synonyms,
  - distinction of homonyms,
  - establishment of semantic relations,
  - field-based tagging
Synonym interrelation and distinction of homonyms.
Hidden Semantic Relations in Folksonomies
Tag Gardening Activities: 4. Fertilizing

- Combination of folksonomies and KOS during indexing and retrieval.
- Aim: enhancing precision and recall and the expressiveness of the folksonomies by adding semantics, for query expansion during retrieval via semantic relations, for enhanced indexing functionalities, for enhanced navigation within the folksonomy, as basis for semantic-oriented displays.
- Achieved by
  - semantic-oriented tag suggestions during indexing and retrieval (→ tag suggestions not based on tag popularity to avoid “success breeds success-effect”)
  - establishment of semantic relations by mappings to KOS (after indexing)
Tag Gardening Activities: Fertilizing Type 1 and 2

- **Fertilizing Type 1**
  Fertilizing a folksonomy with semantic structures from other knowledge organization systems.

- **Fertilizing Type 2**
  Fertilizing a structured knowledge organization system (ontology, thesaurus) with user-generated terminologies.
Tag Gardening in Practice

Possibilities

- Editing and deleting functionalities for tags on personal and document level.
- Detecting and labeling semantic relations.
- Use of power tags as candidates on document collection level.
- Co-occurrence computations as starting points.
- Tools for personal tag gardening on persononomy level.
- Collaborative tag gardening in (small) communities.
- Power users as chief gardeners in communities.
Gardening support: Power tags and co-occurrences as candidates and starting points

Steps:

• Detecting power tags on the resource level
• Computing co-occurrences of power tags with the whole tag collection
• Result: distribution and power tags on co-occurrence level

→ Power relations as candidates for new folksonomy structuring approaches

Data retrieved 2008-05-15 from del.icio.us
Personal Tag Gardening

- Our approach: TagCare
  (www.tagcare.org, currently under development)

- „take care of your tags“
  TagCare imports personal tags from different tagging platforms and allows to manage them centrally for better personal overview.

Personal structured tag vocabulary for:
- Personal information management (PIM)
- Consistent tagging across platforms (functionalities under development).
- Search in folksonomy-based systems (e.g. use pre-cooked personal synonym lists).
Personal Tag Gardening

Use of TagCare – first version

- Adding, deleting and editing of tags
- Hierarchical structuring of tags
- Interrelating of synonyms
- Labeling of otherwise related tags
- Statistics on tag frequencies
Tag Gardening Community

Solution for small, closed communities and development of shared vocabularies:

Community tagging plus single information architects.

Information architect:
- builds a structured thesaurus and enhances it with tags.
- establishes structures between loose tags (emergent semantics).

Community:
- applies tags and performs minor editing activities.
Collaborative Tag Gardening

Challenges and open questions

- Who may perform editing actions? Who is an authorized user?
- Which tags are spam?
- How can users collaborate, which aspects are determined collectively?
- Can tagging guidelines be applied?
- …
Challenges and open questions

- The three elements of folksonomies (resources – people – tags) constitute a three-dimensional knowledge space: domain – community – expressiveness.

- Conflicts are caused, if all three elements should be considered to large extents simultaneously.
Conclusions & Outlook

• Folksonomies may be enriched with semantics to achieve a combination of vocabulary dynamics and structure.
• Aim: improved information retrieval on personomy and collection level.
• Manual and semi-automatic approaches may be combined.
• Tag gardening may first be applied to small communities or small application domains.
• Solutions for bigger communities are needed, “chief gardeners” as first approach.

• Future work: development of TagCare, analysis of types of semantic relations for folksonomy enrichment, tagging for support of ontology or thesaurus development.
Thank You

Best regards from Düsseldorf!

Contact: katrin.weller@uni-duesseldorf.de
isabella.peters@uni-duesseldorf.de
http://www.phil-fak.uni-duesseldorf.de/infowiss/