Thanks to:
Kristin Olofsson
Roy Gundersen
Isabel Filella Ramírez
Lluis Anglada
Michel Wesseling
Susanna Peruginelli
Irene Veerman
Jeroen Hoppenbrouwers
Peter Noerr (virtual)

Workshop Social Tagging & Subject Search Engines
at ELAG 2007, Barcelona

Did you share your tag today?

Tagging is…
• Labelling content with open-ended strings
• Uncontrolled vocabulary
• No typing (author, title, …), just values
• Unorganised tags (no taxonomy, relations, …)
• Uncontrolled user activity (no pressure, few guidelines)
• Uncontrolled user group (volunteers)
• Mostly directly by the content end users or author(s)
• Without explicit system/community-wide goal
… but therefore very popular!

Tagging is mostly popular with newer, web-based systems:
• No legacy software/data/habits to cope with
• Millions of users, 1% active is already a significant number
• By definition interactive, so input is familiar
• Near-zero content support staff (just technicians and lawyers)

The challenge
To (possibly) incorporate tagging in the existing library systems, on top of existing collections.

Why do people tag?
Authors want to be found and read/viewed.
Readers want to tag material that they want to find back in the future.
Where authors and readers meet, we have a community of interest. These communities of interest drive tagging.
Tags need to be shared to be useful.

Be very careful to add “useful features” to the basic tag system: any addition has the potential to break the weak chain which makes people tag at all.

Focus on creating cross walks between initially separate tag systems may be the best first start.

There might be value in a specific proper noun tag set. Proper nouns are quite sharable between domains.

Tag sets may be visualised as tag clouds.
Possible natural work division:

- **Authors** tag when they *create* a new content object.
- **Readers** tag when they *appreciate* a found object.
- **Communities** tag when they *see* a useful structure. (KM)
- **Professionals** tag when on behalf of communities. (IM)

‘tags on tags’ or relationships between tags may contribute the most to structuring the area. No technology yet.

---

Questions:

- Simple (no-brainer, no-worker, no-problemo, el-cheapo, …)!?
- No pre-typing (no closed vocabulary, not even types)
- No difference between concepts and relationships?
- Implicit relationship by the fact ‘this tag is not related to any concepts, but only to tags’??
- How to visualise relationships, or present the tagging interface for structural tags???

Would knowledge structures be feasible at all?

---

Subject Specific Search Engine

Create quality by controlling the content
Involve information professionals (IM)
Involve researchers and practitioners

http://www.focus.eu

6,300 objects / 20 institutes / 20 individuals

---

Tag Pyramid: From Folksonomy to Taxonomy

Philosophical summit
classifications
thesaurus

Pragmatical application
social bookmarking

Organisational division
WorldCat / Google scholar

Community structure
Dictionaries / wikipedia

---

Make tag sets available to external sites etc. (such as Google which uses these to boost your page rank).

Immediate, direct, relevant feedback when typing tags is very good to keep consistency and selecting relevant tags. Importing tag sets from elsewhere might be nice to bootstrap systems. But the imported sets must be on-topic. Experiment!

Browser plugins to put the tags locally, and upload them to a generic central place where others can download them. Replace search engine/portal by a distributed system in the browsers, with very limited infrastructure?
Use typing for searching. Use the tag cloud for exploring. Exploring is interesting for people new to the domain -- not people new to the system.

Operations on the tag set, such as morphological unification, are rejected when people have typed their own tags. Only do this for community-level tags. Look at ownership.

Tag ownership: meta-tag. Defines who is allowed to change the tag (spelling, etc.).

Take existing resources to automatically build structures.

Multilinguality: suggest words in other languages when relevant (for example, with few hits).

Structure in general may be of use when searchers get stuck, to suggest alternatives. People hate ‘zero hits’.

Tags can also be of use to describe media types, etc.

Portals for beginners probably never want to fire queries; they want to suggest actual content. Actual projects do ‘simulated’ searches and offer students a progressively larger collection, ending with the real live collection.

Thank you