

Upstream knowledge design configurations to construct user-relevant semantic help tools

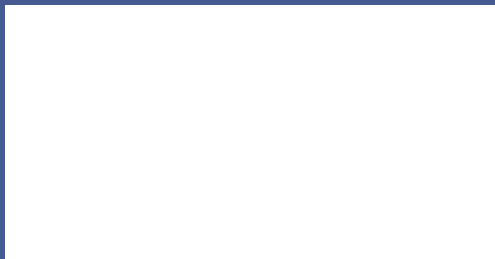
Workshop

**Semantic Help: Status,
Research Objectives, First
Implementation Steps**

6th International Conference
on Electronic Government

DEXA 2007, Regensburg,
Germany, 3-7 September 2007

Dr. Olivier Glassey
Observatoire Science
Politique et Société
University of Lausanne



Dr. Pierre Rossel
Ecole polytechnique
fédérale of Lausanne
(EPFL)



Overview

- Navigation problems have increased as the number of web sites, pages and their more or less consistent cross-referencing links rapidly generated a complex situation for the user of e-administration services
- Ontology-based research triggers pathways to cope with in this increasingly complex world when it comes to both machine to machine and human to machine communication
- After years of such developments, new questions now arise, in particular linked to the very notion of semantic performance in the context of configuring more sensitive forms of helping interactions with the users
- In our contribution, we would like to **explore how user-driven classifications of online contents could contribute to design technologically, logically and socially more robust help tools.** Our focus is clearly set beyond the too simplistic and fruitless opposition between folksonomy and ontology
- In the last part of the presentation we will examine potential research areas which could foster the cross-fertilisation of both domains

Strating point: user complexity

The building of a truly comprehensive semantic description of the e-Gov domain is in itself a complex task but when one considers the idea of building help tools. This is only half of the problem (or of the solution) when it comes to addressing the design of an effective help system

The other half of the challenge is defined by how to deal with users of such tools, considering that they will have at disposal:

- **Heterogeneous forms of knowledge**
- **Heterogeneous skills**
- **Heterogeneous needs and expectations**
- **Heterogeneous representation of the world(s)**
- **Heterogeneous contexts of use**

Future users of such tools do not (and will not) possess a singular, common and stabilized set of ontologies by which they search information, formulate their queries and process the information they receive back

In order to build a robust help system we have to address at the same time both the complexity of machine to machine communication and human to machine interaction.

A possible solution: Folksonomy or the crowdsourcing of classification

Folksonomy:

- belongs to the family of what could be broadly described as “stakeholder-distributed classifications”
- combines ancestral spontaneous classification practices, well described in many ethnographical works (Berlin 2002), with information technologies supporting collective referencing
- replaces the top-down classification method proposed by taxonomy by a bottom-up approach in which classification categories are built freely **by and for the users in actual practice**
- emerges directly from users contributions who label subjectively their actions and queries with tags over the contents that they produce (narrow folksonomy) or access to (broad folksonomy).

Tags as an entry point

- “Tags are a little bit like keywords, but they're chosen by you, and they do not form a hierarchy. You can assign as many tags to a bookmark as you like and rename or delete the tags later. So, tagging can be a lot easier and more flexible than fitting your information into preconceived categories or folders” (del.icio.us)
- To establish these tags, users can tap into **a natural thesaurus: their own imagination and experience**
- As such, aggregation of tags could be considered as hints of people interests (ex: popularity translated in fonts hierarchy tags cloud)
- The analysis of the content of a set of tags and the topology produced by users driven multiple tags combination could provide a way to better understand and build solution upon user's mental model at play in information processing

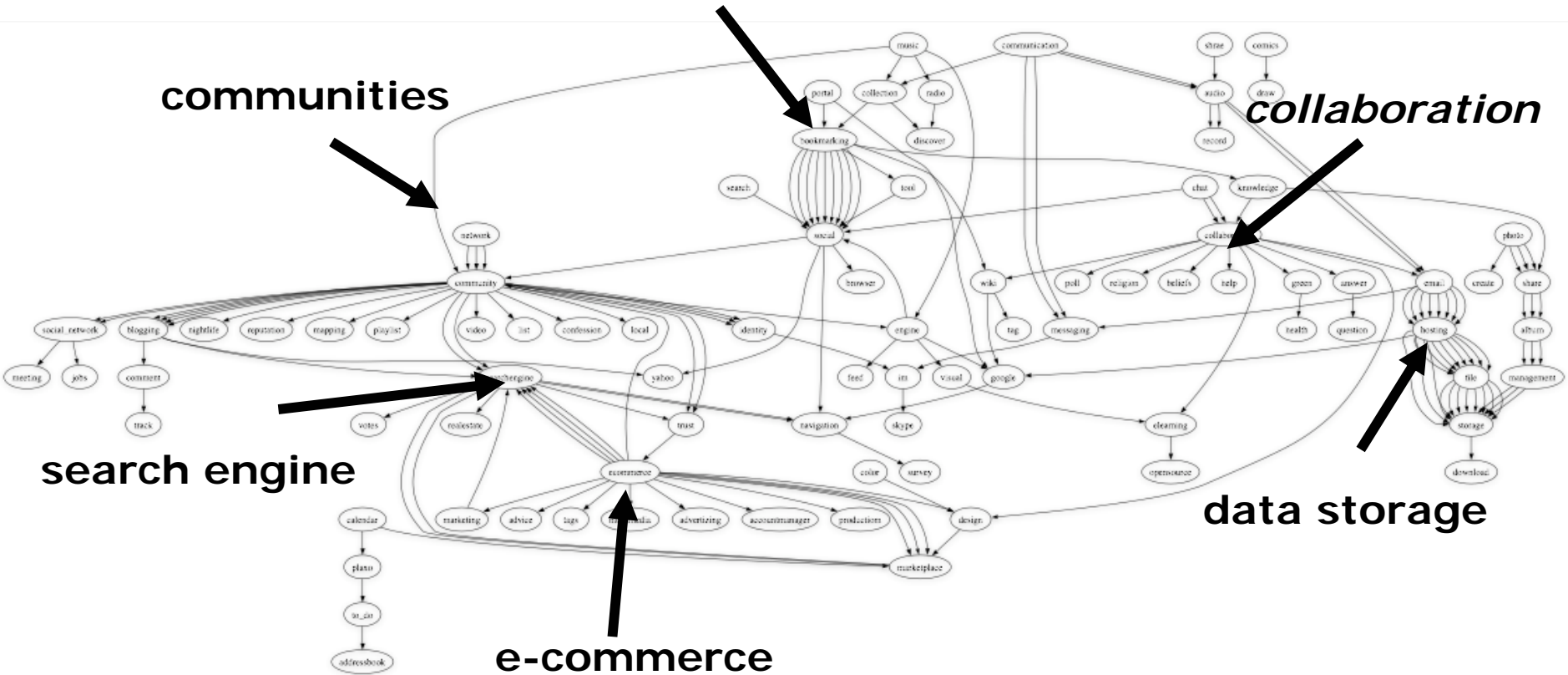
ajax apple architecture art article audio blog blogging blogs book books
business collaboration comics community computer cool css culture database design
development diy download downloads education email english environment fashion fic
finance firefox flash flickr food free freeware fun funny game games google
graphics green gtd hardware health history home howto humor icons illustration
imported inspiration internet japan java javascript jobs kids language learning library
lifehacks linux list mac maps marketing math media mobile money movies music
news online opensource osx outsourcing photo photography photos photoshop
php plugin podcast politics portfolio productivity programming psychology python
rails recipe recipes reference religion research resources rss ruby rubyonrails science
search security seo sga shopping slash social software tech technology
tips tool tools toread travel tutorial tutorials tv twitter ubuntu usability
video videos web web2.0 webdesign webdev wedding wiki windows
wordpress work writing youtube

Tags cloud of the most
popular tags on del.icio.us

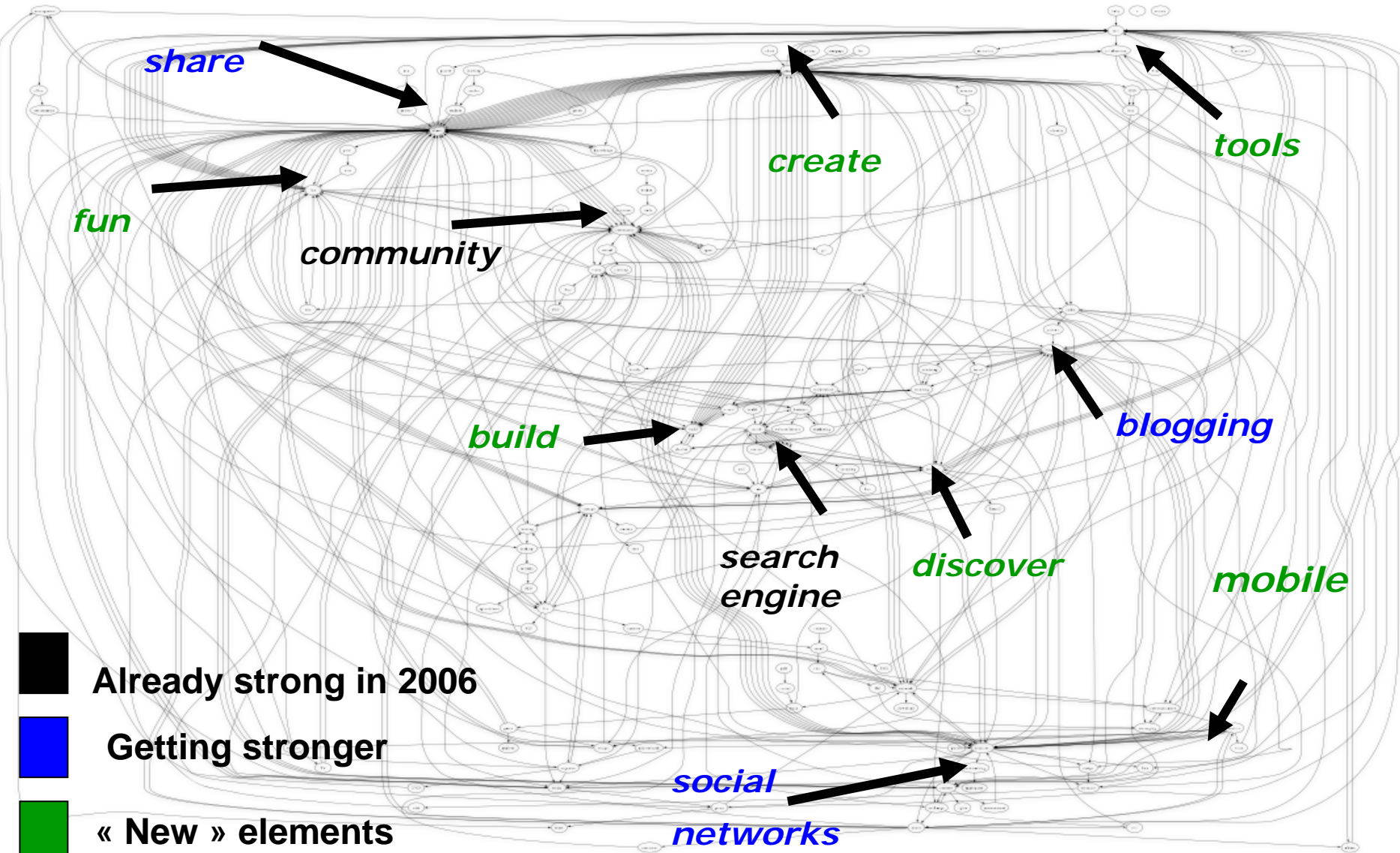
Ex/1: The folksonomic map of items used to describe

125 start-ups in 2005-2006

social bookmarking



Ex/2: The folksonomic map of items used to describe 125 Web 2.0 start-ups in early 2007



- Already strong in 2006
- Getting stronger
- « New » elements

Folksonomy: shortcomings

- **Lack of precision**
 - No control over accuracy of queries
- **Lack of hierarchy**
 - Horizontal set of tags (Tag soup)
 - No simple tools to reduce complexity
- **Lack of coherence**
 - Overlapping tags
 - Many ambiguity
 - No semantic control over tag production
- **Lack of consistency**
 - No semantic control over tag production
 - Heavily depend on the community of contributors
- **Open to manipulation**
 - Group of users with specific interest could game the process

Folksonomies could be perceived just as special, highly simplistic cases of ontologies with extremely basic semantics and almost no constraints

BUT....

Folksonomy: potential

- **Familiar:**
- close to users experience
- taps in existing cognitive processes without adding much in cognitive costs,
- lower barrier of access
- **Facilitate:**
- low cost solution and distribution
- **Exploratory:**
- fosters positive serendipity
- provide alternative paths and options to explore contents and seek information
- follows desire lines,
- **Flexible:**
- supports multiple views, (and even multiple language)
- allows for hybrid classifications
- **Emergent:**
- adapts to fast evolving contexts and contents
- Allows for detecting and seamlessly incorporating new trends !!!!!!! (imp.)

Taxonomy and Folksonomy an opposition?

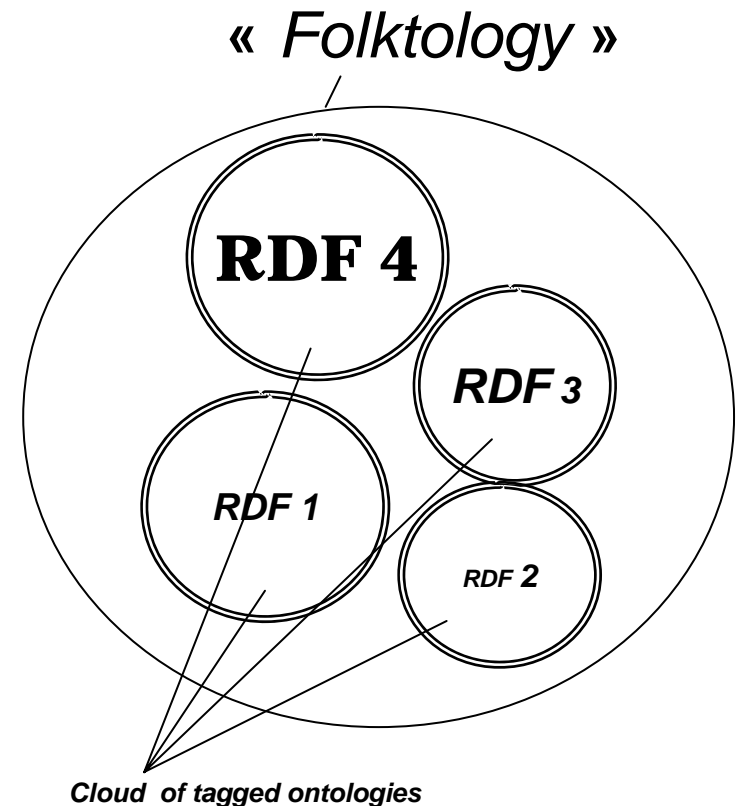
- while many IT-commentators present a binary distinction between the top-down classification and bottom-up initiatives, real cases offer a somewhat less clear dichotomy
- taxonomies are rarely pure in the sense that even the more comprehensive classification systems are subject to local interpretation and are not immune to vernacular understanding
- folksonomies are not free from taxonomy and control. The infrastructures of the tagging systems are often controlled at technological and even at semantic levels by small groups of technicians/supervisors who act as moderators
- many of the most popular webs services use an *ad hoc* combination of both classifications. Their success is even often built on the articulation of both classification (Google, Amzon, etc...)

Mashing-up ontology and folksonomy?

- The benefit from tagging resides in the adaptative nature of the resulting taxonomies
- The benefit of ontologies resides in the rich and close to unambiguous semantics they define
- Tagging systems are useful when all is needed is the ability to link items to topics;
- Ontology are useful when what is needed is to rigorously define or understand what is meant or not mean, through particular classes, fields and relationships, something that is essential for good machine data processing
- Folksonomy are better for sharing of information between people
- « Classical » Semantic web is better for communication between machines
 - Semantic Help: could/should benefit from interfacing both perspectives

Possible strategies: (1) Folksonomy of Ontology

- Aim: Benefits from ontology that make the data semantic explicit but also allows for the emergence and co-existence of a diversity and sometimes conflicting set of ontologies to mimic real-life blurred semantic situations
- "Folktology users could instantly propose or modify ontological classes and properties in the same manner they do with tags in tagging systems. The most popular ontological construct (the most - instantiated classes or slots on classes, for example would rise to the top and self amplify while less initiated ones would fall to the bottom »
(*"Folktology"*, Spivacks, 2005)
- Pandora box issues



Possible strategies: (2) Tagging in RDF

“**Ontologies** should not be understood as a narrow form of specification (as taxonomic classification) or as a very specific methodology for agreeing on a conceptualization (centrally controlled categorization) but **as a mechanism for making semantic agreement**” (Gruber, 2005)

“Ontologies could also be perceived as “**conceptual specification** that enable multiple , independently developed database of categorized artifacts to interoperate and for agent to reason about the difference among the vocabulary used in each of those independent databases” (Gruber, 2005)

“TagOntology” project is about identifying and formalizing a the activity of tagging. It’s aim is not to build a common folksonomy but to allow tagging to be part of RDF:

Tagging (object1, tag1, tagger1, + or -, source1)

- Open possibilities for application such as collaboratively filtered searches (*find thing matching “Q” that my tagging buddies think matches “Q”*)
- Semi-structured query (*all “Q” tagged with “useful”*)
- Issue: tagging dynamics are difficult to formalize

Other possibilities to explore

- **Tapping from other core vocabularies:**

A potential good candidate as starting point for such exploration: the SIOC project (Semantically Interlinked Online Communities) which provides methods for interconnecting discussion methods such as blogs, forums and mailing lists to each other.

The SIOC ontology, an open-standard machine readable format for expressing the information contained both explicitly and implicitly in internet discussion methods.

- **Working at the query level:**

How to interface folksonomic triggered queries with ontology driven answers. This could lead to explore how to add a folksonomic level in query language for RDF (e.g. SPARQL).

Such a narrow approach would have the advantage of supporting the exploration of the possibility to build links between ontology and folksonomy without actually trying to redesign one or the other, which is still the main drawbacks of the other existing propositions

Next steps

- **Assess existing solutions** on ontology-folksonomy interfacing (with also an eye on folksonomy-to-folksonomy and ontology-to-ontology integrating tools).
- **Explore with actual cases** how folksonomy could benefit to the semantic level of help tools in specific situations:
 - **Close to universal administrative interactions vs. Specific situations** (i.e. problems and queries that convey a quite substantial tacit knowledge that cannot be turned into fully generic forms)
 - **Limited sporadic queries vs. Regular, repeated, queries** (for which there is a possibility/need of a learning process by letting the user building their own pathways)
- **Exemplify** how folksonomic approaches could also contribute to the rethinking of other components of already existing facilitating architectures (interface, queries and result display, life event organization of content etc.)