Tagging and the social web

Paola Monachesi
Overview

• Tom Gruber on: Where the Social Web Meets the Semantic Web. ISWC 2006
  – [http://videolectures.net/iswc06_gruber_ws wms/](http://videolectures.net/iswc06_gruber_ws wms/)

• Main points of the discussion
  – Overview of the papers
  – Overview of the group discussion

• Practicum: tagging news
Shirky

- Ontologies: what kinds of things can we say exist in that domain, and how can we say those things relate to each other?
- Categorization: hard to agree, change
- Yahoo vs. Google
- Ontological classification works well when there are constraints on the domain and on the people
- His conclusion: ontologies don’t work well on the web
Shirky

• Tagging allows for classification without a top-down scheme

• Advantages:
  – No professionals but many people
  – Distinctive points of view remain
  – Tagging gets better as more people do it
  – Users and time are added dimensions
  – Filtering comes after tagging
  – Merges are probabilistic
Shirky

- Semantics is in the users, not in the system.
- System doesn’t know the meaning of the tag, he only knows they are related because many people say so...
- Users decide
Discussion Q1

• Shirky says: in the digital world, we don't need the shelf. Why does he say it? Do you agree?
  – Resources cannot be confined in a single place, limitation of categorization
  – Books vs. Links
  – No categories, no structure, useless collection of data
  – Categories help disambiguate
Discussion Q1

- Classification can be useful if you don’t know what you are looking for
- The shelf is a personal thing
- Machine readability vs. Tagging
- CD example - categorization make us loose information
- Search through categorization: you are sure you have retrieved all there is available, this is not the case with tagging (Paola)
Discussion Q2

- Shirky provides some criteria to assess whether ontological categorization might work. Do you agree with his criteria and do you agree with his conclusion that ontologies won't work on the web? Explain.
  - Shirky is against top down categorizations but ontologies and SW techniques go beyond that and they can be integrated with tagging
  - Experts to create the initial structure, taggers can contribute (but what about Wikipedia but cf. Wouter’s comment that there could be an expert view on tagging?) Check: http://www.ontoprise.de/en/home/products/ontostudio/
Discussion Q2

• Ontologies could also be dynamic and evolve.
• Ontology of everything is impossible. Check:
  – http://www.cyc.com/
  – http://commons.media.mit.edu/en/
  – http://www.loa-cnr.it/DOLCE.html
  – http://www.ontologyportal.org/
• Application can play a role/= domain
• Tagging requires low level of training - collective intelligence vs. Ontologies
Discussion Q2

• When is a domain stable? (cf. Ferdy hard to say...) Ontologies can work for specific domains and we can decompose the web. Problem: connection of the ontologies

• Conclusions: criteria make sense to almost everybody, but the conclusion that ontologies for the web don’t make sense is not shared.
Discussion Q3

- Can we substitute the structure we have through ontological classification with the number of people that categorize a resource in a certain way (i.e. user participation)? What are the advantages and the disadvantages to move from binary categorization to a probabilistic world, as Shirky puts it?
  - + we include users view, scalable, flexible, allow uncertainty, cheap, average’s view but also minority view, consensus in approaching the truth, less static, less biased,
Discussion Q3

- no experts (but what about wikipedia?), reliability linked to number of people tagging, expert view might be the minority, hard to reason with, cold start, spammers, diversity of the group will play a role (bias if the group is uniform), domain can play a role (medicine vs. books), opinions will be included (is this really negative?), no room for learning
Discussion Q3

- Right=accepted by high number of people (not always true)
- Use the tags to create the ontology (Ruud- Ferdy) ==> not trivial (relations are different and not expressed) ==> Combine by creating a probabilistic ontology (Wouter-Paola)
Discussion Q4

- Do we need to design a specification of tag concepts that might enable services for analyzing and reasoning over tag data across applications, as Gruber suggests? Do we need to understand and formalize the action of tagging? What do you think about his approach?
  - + would allow reasoning across applications, useful for search engines, structure could be unseen but drive the tagging process, translation process
  - - unclear if needed, feasibility, same problems as with the ontology, against the spirit of tagging
Discussion Q4

– Tag semantics unclear (Fredrik)
– OpenID (http://openid.net/what/) (Ferdy)
– Ambiguity (Wouter)
– Communication patterns better than formalization of tagging (Bodo)