GROUP REPORT ON SOCIAL SEMANTIC WEB

By Evan Patton and Jie Bao
Tetherless World Constellation, RPI

Advanced Semantic Web (Spring 2009)
2009-03-17
OUTLINE

- Mining Semantics from Social Web Data
  - Partially based on
    - Xian Wu, Lei Zhang, Yong Yu: *Exploring social annotations for the semantic web*. WWW 2006: 417-426 and
  - Presenter: Evan Patton

- Semantic MediaWiki: Revisited (or Adding Semantics into a Social Web App)
  - Presenter: Jie Bao
MINING SEMANTICS FROM SOCIAL WEB DATA
FOLKSONOMIES

- Tags for user-generated content
  - YouTube, Flickr, etc.
- Represent a concept or property
- Variable with little expressed hierarchy (cats, dogs, animals)
Statistical measures can be used to determine relatedness of words

- LSA, PMI, NSS, etc.
- WordNet - matches synonymous words
- Context is lost due to word-by-word processing
SEMANTICS

- Can logically describe concepts and relationships between concepts
- Concepts are independent of the symbols used to describe them (truck, lorry; map to the same schema)
- Could be used to generate Web 2.0-style tag clouds
TOOLS

- RDF, RDFS, OWL -- large overhead, learning curves for site administrators (high barrier to entry)
- Maintenance of dynamic ontologies over time
- Assist user in describing new information and locating answers to structured queries
SOCIAL ANNOTATIONS

- Map users, objects, tags into conceptual space
- Use parametric models and MLE to find a fit to the n-dimensional conceptual space (EM)
  - Separable Mixture Model
- Mathematical model for determining groups of related tags
Figure 3: Conditional Distribution of Tag ’todo’ on dimensions of conceptual space
Figure 5: Conditional Distribution of Tag ‘xp’ on dimensions of conceptual space
NEXT STEPS

- Web 2.0 Tags ==> Concepts (Check)
- Concepts ==> Web 3.0 CPIs
CONCLUSIONS

- Most focus on Social Web and Semantic Web is focused on data mining
- Concept extraction never bothers to match semantic web concepts (future research?)
- Too much overhead
SEMANTIC MEDIAWIKI: REVISITED
• The 2007 JWS paper summarizes basic features of SMW
  – Annotation, Browsing, Query

• We will check SMW assumptions against our own practices.
• **Assumptions:**
  – User will use a special markup to add annotations to the wiki text
  – Primary goal is to enable text-based editing, but strongly structured content is allowed.
  – Page-centric knowledge organization fits wiki better (than viewing triple as primary units)
  – Formal semantics via a mapping to OWL DL

• **Approach:** introducing typed link.
  “London is the capital city of [capital of:UK]”
• Assumptions:
  – Users need browsing without disturbing normal reading

• Approach:
  – Factbox
  – Semantic Browse
• Assumptions:
  – Queried lists are more accurate, easier to create and easier to maintain than manually edited listings.
  – Tractable query language (P-Time)
  – Semantics in OWL DL
• Approach:
  – The ask query
    {{#ask: [[Category:City]][[capital of::UK]]}}
MORE HOPE

• Where SMW can help [1]:
  – Manually generated lists
  – Searching information
  – Inflationary use of categories
  – Inter-language consistency
  – External reuse.

• Wikis are now an IT code word for “zero-training” [2]

    SemanticWiki-Future--MarkGreaves_20090305.pdf
BEYOND 2007

• Many extensions
  – Semantic Template, Semantic Forms, Halo Extension…

• Many real-world applications
  – Chickipedia, Discoogle, Recipes Wiki, Creative Commons Wiki,…
TW WIKIS EXPERIENCE

Example Apps
- TW Wiki
- RPI Map
- IC Checking

Questions
- Do SMW assumptions hole?
- Does SMW suit building apps (e.g., blog), not only CMS?
- What’s missing?
TW WIKI
TW.RPI.EDU

+ • It is successful as a wiki
  • Lots of hands-on experience from it
    - Blog, mail archive, issue tracker, projects, publications, tasks, …
  • Semi-successful as a semantic wiki
  • Ease my own life enormously: universal access to my professional work

− • Few uses “semantics”
  • Profs get confused on adding papers, etc.
  • The majority are not lured by its addon services
  • Constant ontology war
  • OWL semantics is rather limiting
  • Privacy wins, openness out.
TW WIKI (2)

TW.RPI.EDU

+ -

- Many templates + many queries => slow
- Weak connectivity to other semantic apps.
- Turns out to be a project that can eat huge amount of time
  - All SE troubles, plus KE troubles!
• Linking live external data (majority of our efforts)
  – Bus, people, class, events, or satellites
• Map on wiki
  – Tetherless map extension
• Serves e-Science projects

+  

–

• UI (skin) design is time-consuming
• No clear benefits from “semantics” (but certainly from tagging), at least for now.
+ We can represent/edit OWL
+ We can represent rule
+ We can represent IC
+ We do add a CNL interface to ontologies on SMW

- But all are quite limited
- Page-centric organization is restrictive
- UI Nightmare: Semantic Form limited
- Due to considerable learning cost, developers may still prefer Java than MW+SMW+SF+String Function+ etc
LET'S GO BACK AND CHECK THE SMW ASSUMPTIONS
CHALLENGES

To build better connected data platform
import/export external data
balance latent and formal semantics
be part of “Linked Data” cloud

To facilitate the network effect of collaboration
replicate Wikipedia effect for knowledge base
support personalization, privacy protection
need best practices and guidelines

Interoperability

Collaboration

Usability

Methodology

To make it easy to be used by end users
how much training, documentation?
graphic UI for e.g. editing, presenting results
effective data access interface, e.g. search

To make it easy to develop and deploy
it is software + knowledge engineering
we need best practices and design patterns
identify the capabilities and limitations

Also see: Semantic Wiki mini-series Session-6 http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2009_03_05
CONCLUSIONS AND VISION
CONCLUSIONS

- Mining semantics from social web data
  - Most focus on Social Web and Semantic Web is focused on data mining
  - Concept extraction never bothers to match semantic web concepts (future research?)
  - Too much overhead
- Adding semantics to social apps (case study SMW)
  - SMW adds a small thing to MW: typed link
  - This turns out to be surprisingly powerful
  - Yet it also brings even more problems: interoperability, collaboration, usability, methodology, etc. Some defeats SMW’s original assumptions.
OUR VISION

- Start with email
- Tag important items in an email similar to inserting HTML links, formatting (client handles underlying RDF)
- Exposing information to trusted applications, contacts, could provide better interoperability experience
OUR VISION

- Keep Hope, not Hype
  - Social Semantic Web = Social Web + a little Semantic Web

- Make it simple
  - Remember: users are not domain experts, nor knowledge engineers.
  - Simple semantics (tagging + tree + typed link)
  - Easy UI (as easy as blog)

- Make it scalable
  - Only constant-time algorithms survive on web
OUR VISION

- Make it web-ish
  - Boost core web apps: blog, email, calendar, news, homepage, mail list, social network, music, video, shopping ...

- Make it social
  - e.g. tagging propagation in emails, linking SMW with twitter/facebook/linkedin

- Semantics is tool, not goal
  - RDF is a plus, not a must
Suppose you were in 1800 AD and were asked
Backup
Bike

Pros
• It can move

Cons
• Against “physics”
• Risky
• Slow (cf. horse)
• Uncomfortable (cf. coach)
• Hard to learn riding
Will it work?

• Web: Yes
  – Because: simple, useful, … (others)

• Social Web: Yes
  – Because: fun, network effect, … (others)

• Semantic Web: ?

• Social Semantic Web: ?
Evolution Matters

1.0
1818
Draisine
Karl von Drais
Germany

2.0
1830
two-wheel velocipede
Thomas McColl
Scotland

3.0
1860
pedal-bicycle
Pierre Michaux
France

... 
1870
high-wheel bicycle
James Starley
France

1885
safety bicycle
John Kemp Starley
England

1960s
racing bike
---
USA

Mid 1970s
mountain bike
---
USA