The Semantic Web: Vision or blindness?

Diana Santos
SINTEF ICT, SET

Main arguments

- NL as an ages-old knowledge representation language
- Current difficulties with NER and the Web
- The problem of practical feasibility

- Examples
- The ontological status of the SW

The hype and its translation

- *The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and to work in cooperation*
  (Tim Berners-Lee, James Hendler, OraLassila, The Semantic Web, *Scientific American*, May 2001, W3C Semantic Web home: http://www.w3.org/2001/sw/)
  i.e.
- Will computer scientists devise a language better than natural language in a decade?

A more limited version

- SW just for people, organizations and locations
- yellow pages, company registry, maps
- for every person ever existing (mentioned on the Web), a different identifier
- for every "location", a decision: *I love Norway*, *Norges mest kjent pianist*
- and the temporal dimension?

Knowledge is dynamic

- knowledge is not static
- and has a infinite (?) structuring capacity
- different roles
- different views
- different opinions
- different classification hypotheses

- what about if the webpages are rotten? Not maintained properly

Essential properties of natural language

- vague
  - no need to further specification
- flexible
  - allows creative use
- context dependent
  - meaning is always a blending of situation and system
- no obligatory formal distinction between classes and instances
  - Tweety is a bird; a sparrow is a bird

Practically feasible?

- The problem of consensus
- Given a straightforward categorization task, two persons almost always disagree
- how to encode a shared conceptualization if there is none?

(Subtle?) distinction

- Wishful thinking (wouldn't it be wonderful?)
- Attainable goal (generally with an estimate, a budget)

- An obvious parallel in machine translation
  - IF it were possible to create an interlingua THEN it would make sense to compare advantages and benefits of that approach compared to the transfer approach
  - Assuming something impossible did not take us anywhere

Case studies

- Tsujii: biomedical ontologies
- Tourism and trivial disagreement
- NER: its challenges and problems
- Malucelli & Oliveira: closing the circle
- Ontology merging: isn't this enough?
- Publications: what about SW for them?
- What do practitioners do, 7 years after?

Real applications which use ontologies...

- are not so happy about the ontology world
- *an organised collection of terms enriched with relations is more useful for text mining than formal ontologies*
- *complementary approach:* text-centred
- Tsujii, Jun-ichi & Sophia Ananiadou. "Thesaurus or logical ontology, which one do we need for text Mining?", *Language Resources and Evaluation* **39**,1, Feb 2005, pp.77-90.

People do not agree in simple tasks

- 2 subjects were asked to annotate 30 texts with destination descriptions, using a manually pruned ontology with 682 concepts
  - A: 436   B: 392 categorizations
  - 277 proper nouns classified by both, with 59 concepts
  - Kappa: 63.48% → Upper bound F-measure: 62.09
- Cimiano, Philipp & Steffen Staab. "Learning by googling". *SIGKDD Explor. Newsl.* **6**, 2 (Dec. 2004), pp. 24-33.

NER and QA: a different perspective

- *Where is the Taj Mahal?*
- 20% of the occurrences of country names represent concepts, not a geographical location nor an organization (official or people)
- Attempts to go beyond structured information (for example Wikipedia)

Closing the circle: back to NLP

- instead of full ontology mapping,
- *if all the previous steps are not enough [...] it looks in texts for new meanings and characteristics...*
- Malucelli & Oliveira. "Ontology Services to Facilitate Agents' Interoperability". In: *6th Pacific Rim International Workshop On Multi-Agents, Prima 2003,* Springer-Verlag, 2003, pp.170-181.

Ontology merging

- survey of 35 works
- *we hope that this article contributes to a better understanding of the emerging field of ontology mapping*
- Kalfoglou, Y. & M. Schorlemmer. "Ontology mapping: the state of the art", *The Knowledge Engineering Review* **18**(1), 2003, pp.1-31.

A possible easy world?

- semistructured
- ages with many publication standards
- ages with specialized tools and environments
- references on the Web (made up):
  - 70% of are ill-formed
  - 20% are wrong
- solutions like Citeseer provide alternative textual fragments for you to choose

Seven years later
(3 Keynote speakers at the Semantic Web conference 2005)

- Mismatch between computational biology and the SW
  Goble, C. & C.Wroe. "The Montagues and the Capulets", C*omparative and Functional Genomics* 5 (8) 2004, pp. 618-622
- Demonstration: a meeting organizer
  Ferrucci, D. & A. Lally. "Building an example application with the Unstructured Information Management Architecture", *IBM Systems Journal* **43** (3), 2004, pp. 455-475.
- Talks about design goals for the SW
  Wietzner, D. "Semantic Web Public Policy Challenges: Privacy, Provenance, Property and Personhood"

3

An example UIM application

- UIM: unstructured information management
- input: emails; output: a "Meeting finder" with an XML interface
- find all UIMA-related meetings in building H1 during which refreshments would be served
- simple, purely pedagogical...

Not science; ill-motivated technological research

- the SW is not a scientific theory, since it does not allow any facts to falsify it (Popper)
- the SW may be framed as technological research: create systems that bring it about
  - but: it identified the wrong problem - we need to make information LESS precise, not more, if we want to cope with its huge amount
- what is the ontological status of the SW?
  - dream, goal, project, community, misunderstanding, term,?

My view

- AI has long ago turned into IA
- artificial intelligence -> intelligent aid
- New disciplines have the bad property of trying to reinvent the wheel and not look at the body of knowledge already amassed (in the present case in philosophy and linguistics and even AI), even if they borrow freely their terms from them
- (semantics, ontologies, etc.)

Happy ending?

Misguided as the name *Semantic Web* may be,

- it may help people understand that they need to look in the semantics and philosophy disciplines as well as NLP
- it may help give us a lot of funding, because it is a buzzword.