

Fundamental Research challenges generated by the Semantic Web

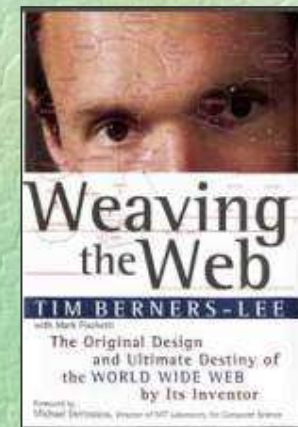
Frank van Harmelen
Vrije Universiteit Amsterdam

**or:
two dozen Ph.D. topics
in a single talk
(at least:-)**

This is **NOT**
a Semantic Web
evangelization talk



(I assume
you are already
converted)



Outline

- What's up in the Semantic Web
- Fundamental Research challenges generated by the Semantic Web ("two dozen Ph.D. topics")
- Three example topics in some more detail



What's up in the Semantic Web? ①

Science:

- ✓ steady progress on many fronts
- ✓ increasing contributions from DB, NL, ML, IR, KE
- ✓ emphasis on light-weight semantics:
 - ✓ RDF(S), OWL Lite/DLP
- ✗ Too much "old wine in new bottles"
 - ✗ not enough "SemWeb causes radical rethink"

What's up in the Semantic Web? ②

The 4 hard questions:

Q1: "where does the meta-data come from?"

- ✓ NL technology is delivering on concept-extraction

Q2: "where do the ontologies come from?"

- ✓ many handcrafted ontologies
- ✗ ontology learning remains hard
- ✗ relation extraction remains hard

Q3: "what to do with many ontologies?"

- ✗ ontology mapping/aligning remains VERY hard

Q4: "where's the 'Web' in the Semantic Web?"

- ✓ more attention to social aspects (P2P, FOAF)
- ✗ non-textual media remains hard

What's up in the Semantic Web? ③

Applications:

- ✓ healthy uptake in some areas:
 - ✓ knowledge management / intranets
 - ✓ data-integration (Boeing)
 - ✓ life-sciences (e-Science)
 - ✓ convergence with Semantic Grid
 - ✓ cultural heritage
- ✓ emerging applications in search & browse
 - ✓ Elsevier, Ilse, MagPie, KIM
- ✗ very few applications in
 - ✗ personalisation
 - ✗ mobility/context awareness
- ✗ Most applications for companies, few applications for the public

The Semantic Web is an integrator of many different CS fields

- Databases
- Natural Language Processing
- Knowledge Representation
- Machine Learning
- Information Retrieval
- Agents
- HCI
-

Methodological challenges

- Need >1 branch of the ACM Topic Hierarchy
- More “empirical” Computer Science
 - evaluation
 - besides theory and engineering
- Traditional metrics need reconsideration
 - e.g. precision and recall
 - “big-O” complexity mostly irrelevant
 - metrics from different fields:
 - databases
 - IR
 - HCI

Opportunities of the Semantic Web

- Content rich environment
- Standardisation, syntactic convergence
- Many hands/free labour
“millions of knowledge engineers”
- tested availability, an “on-line lab”
- partial solutions work

Outline

- What's up in the Semantic Web
- Fundamental Research challenges generated by the Semantic Web ("two dozen Ph.D. topics")
- Three example topics in some more detail



Take-home message

- The Semantic Web is mostly phrased as a technology challenge:

- This talk:

- The Semantic Web

radically breaks some of the assumptions

underlying many current CS results,

- and thus generates many

new fundamental research questions

“Intelligent” things we can’t do today

- **Search engines**

- concepts, not keywords
- semantic narrowing/widening of queries

- **Shopbots**

- semantic interchange, not screenscraping

- **E-commerce**

- Negotiation, catalogue mapping, personalisation

- **Web Services**

- Need semantic characterisations to find them

- **Navigation**

- by semantic proximity, not hardwired links

-

(slide often used by me)

Semantic Web Challenges for KR

■ it's large

- It's even larger
- Current RDF/OWL: 100k pages
large usage, but $\approx 0.01\%$ of the Web
- Anytime/good-enough inference

■ no referential integrity

- Semantic equivalent of "404 page not found"

■ many authors

- distributed authority
- trust

■ high variety in quality of knowledge

- Yahoo OpenCyc

■ Unpredictable use of knowledge

Semantic Web Challenges for KR

■ diverse vocabularies

- Ontology mapping
- Hardest and most urgent problem

■ Decentralised

- Inference may be cheaper than lookup!
- Distributed querying, peer-to-peer

■ high change rate

- time-dependent content

■ local containment of inconsistencies

- Modularisation
- Unbreakable inferencing

■ justifications as first order citizens

- Important for trust

Note: not just for KR

- Databases
- Natural Language Processing
- Machine Learning
- Information Retrieval
- Agents
- HCI
-

Challenges for DB research

- metadata management
 - storage, viewing, querying, updating, cleaning, warehousing
- heterogeneous schema integration
- partial schema's
- on-the-fly source mediation (P2P)
- semantic similarity for retrieval
- data integration of surface & deep web



(by **Dimitris Plexousakis**, ICS/FORTH)

Challenges for ML research

- curse of the representation language
- semi-automatic ontology construction
- scalability
- learning from multi-modal resources
- probabilistic knowledge on the Semantic Web
- non-stationary data ("streams")
- learning/predicting ontological drift

(by **Marco Grobelnik**, Ljubljana)

Outline

- What's up in the Semantic Web
- Fundamental Research challenges generated by the Semantic Web ("two dozen Ph.D. topics")
- Three example topics in some more detail

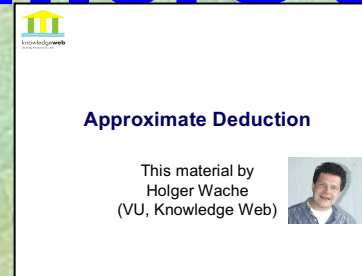


Three of these in more detail

- It's **large**: approximation
- It's **inconsistent**: containment
- It's **heterogeneous**: mapping


- Note:

- these are all **foundational questions**, important without the Semantic Web
- this could have been done for any item on the previous lists



Approximate Deduction

This material by
Holger Wache
(VU, Knowledge Web)



Reasoning with
Inconsistent Knowledge

This material with
Zhisheng Huang &
Annette ten Teije



This Material
Zharko Aleksovski & Michel Klein



PHILIPS



Remember the Take-home message

- The Semantic Web is mostly phrased as a technology challenge:

"Intelligent" things we can't do today

- **Search engines**
 - concepts, not keywords
 - semantic narrowing/widening of queries
- **Shopbots**
 - semantic interchange, not screenscraping
- **E-commerce**
 - Negotiation, catalogue mapping, personalisation
- **Web Services**
 - Need semantic characterisations to find them
- **Navigation**
 - by semantic proximity, not hardwired links
-

- This talk:

- The Semantic Web **radically breaks some of the assumptions** underlying many current CS results,
- and thus generates many **new fundamental research questions**

**Go forth &
investigate!**

