



Panel Discussion P3:

Will the Semantic Web scale?

**New York Sheraton
May 19th, 2004**

Proposers:

- Raphael Volz
- Carole Goble
- Rudi Studer

Panelists:

- Dr. Cathy Marshall
- Prof. Dr. Alon Y. Halevy
- Prof. Dr. Jürgen Angele
- Prof. Dr. Ian Horrocks

Organizers:

- Raphael Volz
- Daniel Oberle

Panelist 1



Dr. Cathy Marshall
Microsoft Corporation
Texas A+M University

**Why the Semantic Web
won't scale**

the scaled semantic web seen as mass-market product



“the Flowbee uses the suction power of your household vacuum to draw the hair up to the desired length, and then gives it a perfect cut.....every time.”

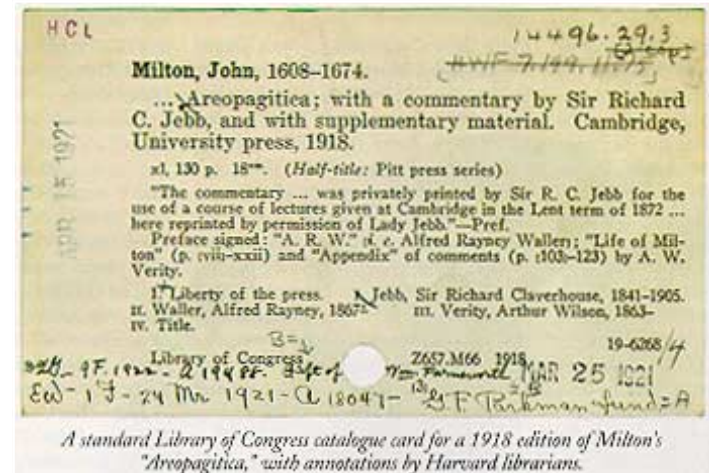
Three important questions:

- Will it really work?
- Who needs it?
- Is it safe?

will it work?

evaluating the semantic web as metadata

- compare the semantic web to a widely adopted metadata scheme like the MARC record used for library cataloging
 - MARC practitioners are members of a community and are trained to create metadata
 - MARC reduces interpretive load by careful choice of attributes, authority lists, & cataloging rules (AACR, e.g.) to constrain values
 - MARC records are controlled for interoperability and consistency in various ways (e.g. by clearinghouses like OCLC)
 - so... on-line catalog (OPAC) users know what to expect



will it work?

evaluating the semantic web as metadata

- by contrast, the semantic web is subject to the following pitfalls as it scales:
 - social structures for creating universal semantic web metadata are missing (local culture/practices/needs prevail)
 - semantic web metadata requires substantial interpretation of domain knowledge; underlying assumptions about use are highly situated
 - no way of ensuring interoperability, consistency, accuracy
 - e.g. EVLIS PRESLEY memorabilia on eBay
 - e.g. HTML visual mark-up
 - so... semantic web users are guaranteed to be surprised



**a beehive is a hairstyle.
Or is it?**

who needs it?

the semantic web is expensive

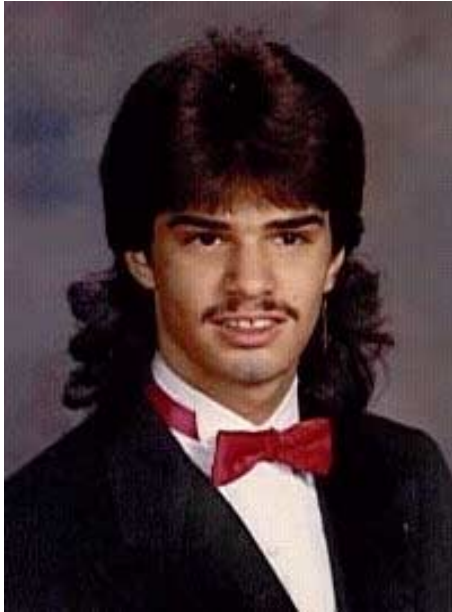
- metadata is expensive
 - often professional metadata creators have to choose among standards
 - e.g. OAI v. Semantic Web
 - cost may not be borne by the parties who benefit from the semantic web
 - e.g., retailers with on-line catalogs
- a Google-like approach works well enough much of the time
 - social evaluation through links
 - the human reformulates and supplies the missing bits (see Marcia Bates' "berry-picking" interpretation of IR)
 - highly robust
 - demonstrated scalability



canonical mohawk from google image search; better than telling my intelligent agent "find me pictures for my talk"

finally: is it safe?

the semantic web raises trust issues



unsafe Flowbee use:
the mullet

- how will porn sites and creative spammers use the semantic web?
 - e.g. "Re: The information you requested"
 - e.g. "V.i.a.ggg.r.a"
 - e.g. clever phishing techniques
 - e.g. phony metadata
- how can mildly deceptive semantic web schemes get the best of people in a commercial situation?
 - e.g. shipping and handling costs

Panelist 2



Prof. Dr. Alon Y. Halevy
University of Washington
Nimble Technologies (ex)
Transformic, Inc.

Will the Semantic Web Scale?



Need Two Definitions

- Scale
- Semantic Web



Two Comparison Points

- How pervasive is database technology?
 - Not as much as you'd expect. Most people are intimidated. They go for spreadsheets and structured files.
- Enterprise Information Integration:
 - A very recent industry sector. And it has been a very rough ride / hard sell.

Why? The Structure Chasm

Authoring

Writing text

Creating a schema

Querying

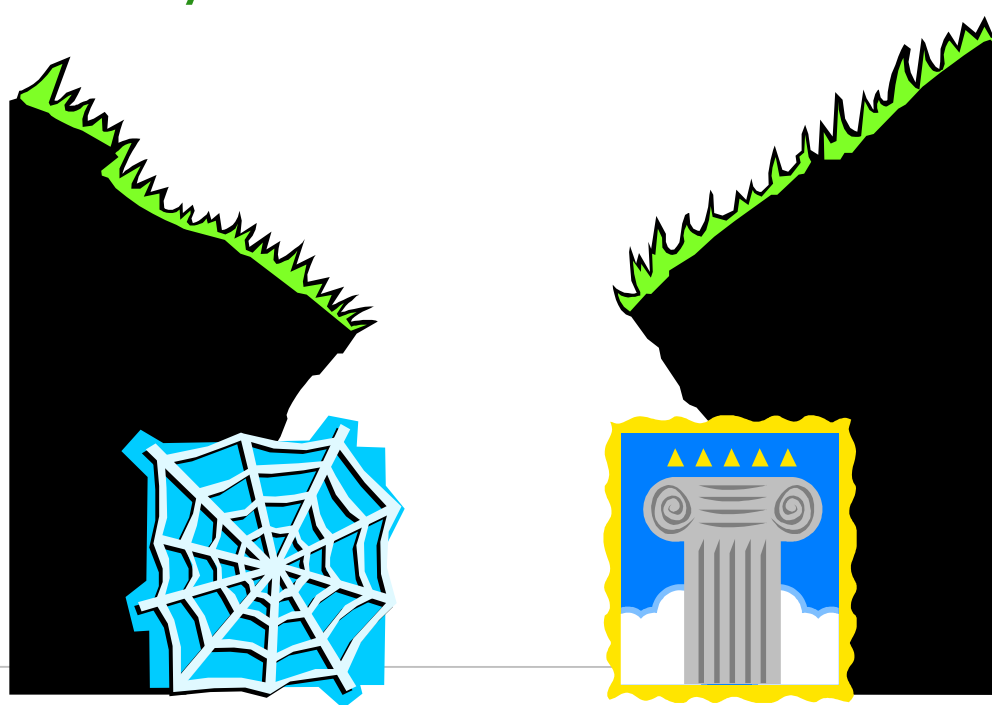
keywords

Using someone else's
schema

Data sharing

Easy

Committees, standards



Why? The Structure Chasm

Authoring

Writing text

Creating a schema

Querying

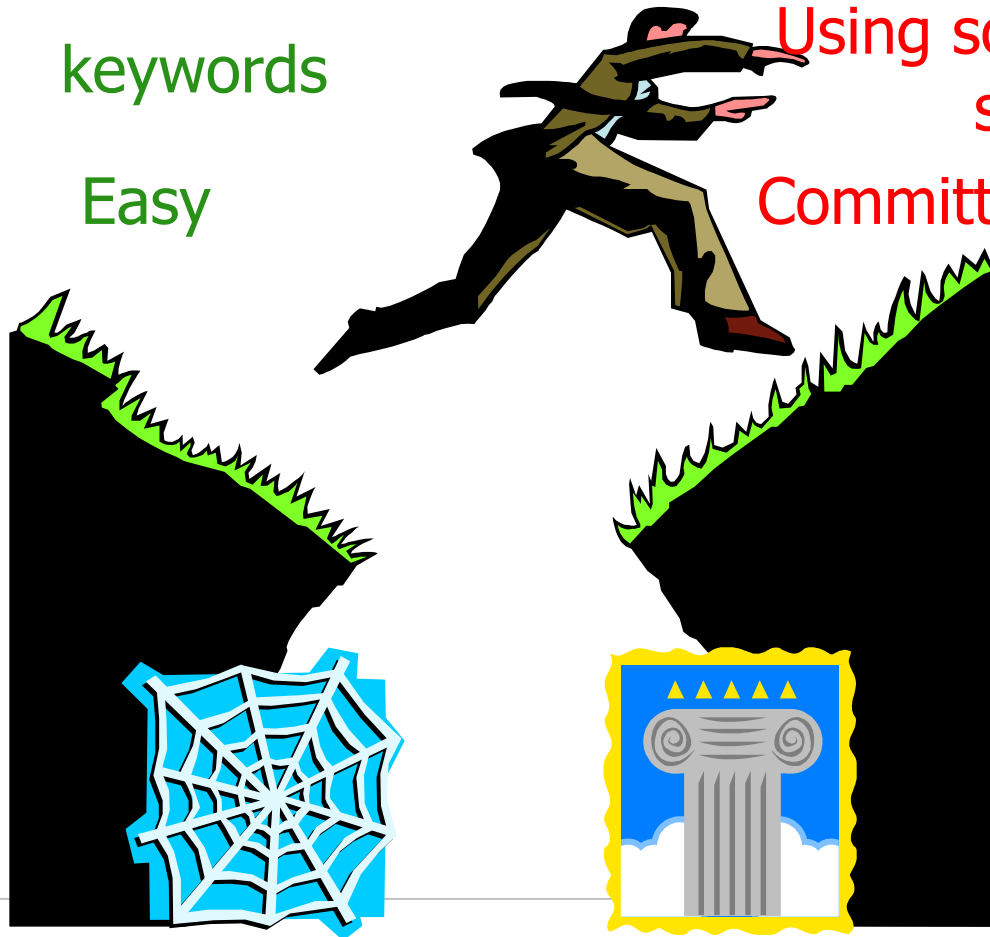
keywords

Using someone else's
schema

Data sharing

Easy

Committees, standards





(My) Conclusions

- It's a people issue:
 - People need clear return on their investments.
 - It has to be dead easy:
 - **K**ep **I**t **S**imple, **S**tupid
- When it's time to scale computationally, we'll figure it out
 - And hopefully, there will be some database people in the room.

Panelist 3



Jürgen Angele



Does the Semantic Web scale?

Ontologies are a success story !

- Large ontologies in the Web
 - Mesh (Pharmacy)
 - Gene (Biology)
 - Wordnet (Linguistics)
 - Ontologies in inhouse applications
 - Deutsche Telekom (ontology based search)
 - Audi (test car configuration)
 - Vulcan (chemistry expert system)
- clear benefit for application in the next generation web

BUT

- OWL does NOT scale conceptually !
 - people do NOT understand DL
 - no tools (editors) to hide DL appropriately
 - OWL misses appropriate expressiveness
- Instead
 - people are used to think frame-based
 - and rule oriented
 - and constraints oriented

AND

- OWL does NOT scale technically !
 - current inference engines too slow
 - no instance reasoning with appropriate performance
- Instead
 - technologies with appropriate performance:
(deductive) databases



SO

Semantic Web is a real great opportunity

BUT

OWL is a step in the wrong direction

Panelist 4

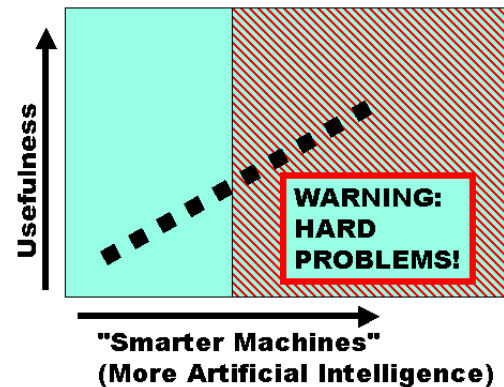
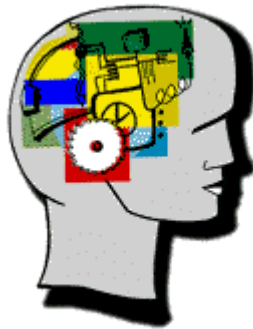


Prof. Dr. Ian Horrocks
University of Manchester
Network Inference

Will the Semantic Web scale?

Will the Semantic Web Scale?

- Not clear what “The” Semantic Web is/will be
 - If it means “semantics + web = AI”, then answer is a definite **NO**



- If it means “semantics + web + AI = more useful web”, then answer is a definite **MAYBE**

Semantic Web Vision

- Current vision includes (at least):
 - Adding semantic annotations to web resources
 - Using ontologies to provide vocabulary for annotations
 - Exploiting semantics to **improve** (machine) “understanding” of web content
- What does it mean to “understand” web content?
 - Ability to derive additional (implicit) meaning (i.e., *reasoning*)
 - Treating (annotated) web as huge KB and reasoning over it clearly *wont* scale (and issues of trust, consistency, etc.)
 - But identifying (small) relevant/interesting subsets and reasoning over them *might* scale

Cost-Benefit Analysis

- **Costs**
 - Development of ontologies
 - Time consuming and costly for useful (high quality) ontologies
 - Adding annotations to resources
 - Perhaps the most serious potential bottleneck
 - But many/most annotations will be automatically generated
 - Exploiting (reasoning with) annotations in applications
 - Developing software to reason with annotations is non-trivial
- **Benefits**
 - Improved accessibility, visibility & utility of resources for/to automated processes
 - Not clear if all providers of web content will want this!
 - Improved sharing and interoperability of resources

Can Data be Managed Efficiently?

- Problems are inherently intractable *in the worst case*
 - But may be manageable in typical cases
- Ontologies: some evidence for scalability
 - Not clear if large or small ontologies will predominate
 - High (but manageable) development cost for large ontologies
 - Existing ontologies with 10s/100s of thousands of classes
 - High integration cost for small ontologies
 - Active research area; still an open problem
- Instance data: jury still out, but promising (early) results
 - Using database and LP technology
 - Hybrid database/reasoning techniques