

# Semantic Technology

Chris Welty  
IBM Research

# What are semantic technologies

- Dates back to the 60s, 70s, 80s, 90s
    - STRIPS, SNePS, CG, KL-ONE, NIKL, CLASSIC, LOOM, RACER, etc...
  - Today we have standards
    - Common Logic, IKL
    - RDF, SKOS, OWL, RIF
    - ODM, PRR
- } Normative XML syntaxes

# What can you do with Semantic Technology?

- Build information systems
- Thesauri, terminologies
- Learning, testing, training systems
- Event processing, back-office systems
- Software design automation, architecture
- Web services, Planning/scheduling
- Intelligence analysis
- ...what do you need?

# Can't Databases Do that???

- No
- Yes
- Well.....

# Advantages of Semantic Technology

- Consider Software Architecture
  - More declarative, open
  - Better abstraction
  - Cheaper maintenance
  - Better integration
  - ...by making the *semantics* explicit
    - At least a little...

# Common Logic

- Standard (ISO/IEC 24707:2007) syntax and semantics for First Order Logic (FOL)
- XML and “KIF style” syntaxes
- “Web savvy” (can use URIs)
- A few implementations, still in nurturing stage
- A context-logic extension proposed (IKL)

# Semantic Web

## ■ RDF

- A language for semantic graphs
- The nodes are anywhere in the web
- The arcs are labeled

## ■ OWL

- A language for giving more semantics to RDF graphs
- *classes* of nodes
- Constraints, equality, negation

## ■ RIF

- Rules for extending graphs automatically

# Can't UML do that???

- There is overlap between UML and OWL
  - Classes, relations, constraints
- But there are significant differences
  - OWL has a full model-theoretic semantics
  - OWL is designed for specifying *information* systems
  - OWL limited to consistent, sound, and computable reasoning

# ODM

- Interoperability standard between several semantic technologies
- Facilitates MOF-enabled tools to work with RDF, OWL, CL, ...
  - Editors, translators
  - Visualization
  - Gives semantic technologies a UML “flavor”

# The Semantic Web Vision

- ~80% of web pages are generated from *back end databases*
- Publish the semantics (schema?) as well as the data
- URIs provide a web-based form of identity
  - It's the semantic WEB, not the SEMANTIC web
- **NOT:** humans will markup their web pages with semantics
- **NOT:** NLP will populate the SW from web pages

# Errors by analogy

- The web: just hypertext
- The web: bad UI design
- The semantic web: just semantic technology
- The semweb: bad KR design

# History of Hypertext

- 1945: Vannevar Bush's *Memex*
  - *Associative Indexing* and links
- 1965: Ted Nelson coins hypertext
  - “Nonsequential writing”
- 1967: Andries van Dam's Hypertext Editing System (sponsored by IBM).
- 1985: Janet Walker's Symbolics Document Examiner
- 1987: Bill Atkinson's Hypercard on the Mac
- 1991: Tim Berners-Lee proposes HTTP, HTML, & URL
  - Genesis c. 1989
- 1993: Mark Andreessen releases Mosaic for Mac, Unix, Windows...

# Hypertext Research

- Dating back at least to the late 60s
- Many foci
  - Technology (mouse, software, protocols)
  - User interaction
  - Aesthetic
  - Post-modern
  - *Engineering*
- Largely ignored by web developers
  - Especially in the early days of the web (93-96)

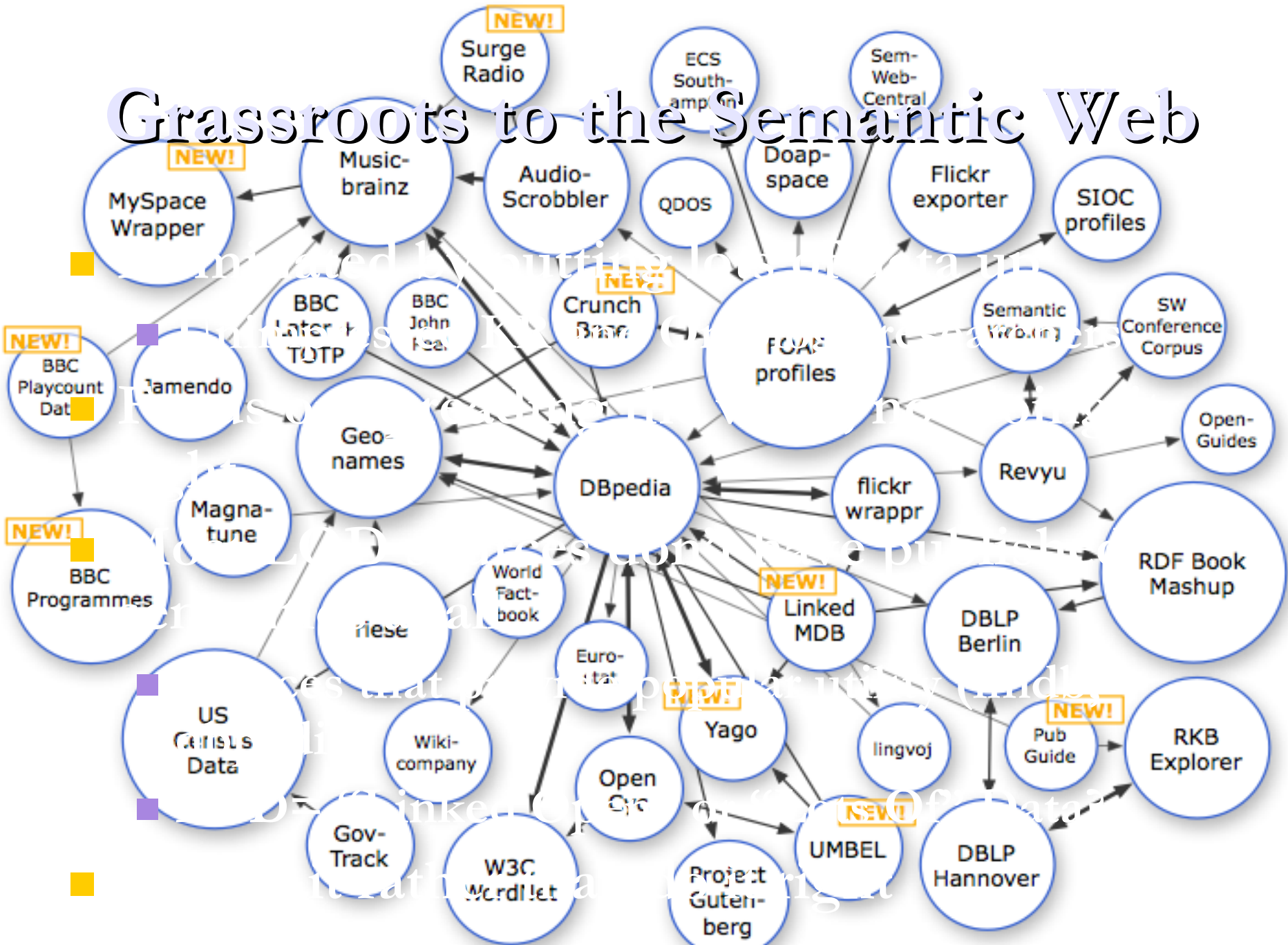
# Grassroots to the Web

- Early web dominated by “what it looks like” in Mosaic
  - Unimpressed UI and Hypertext researchers
- Focus on spreading the word, not doing it right
- Many early web pages didn’t have links in text at all
  - “Catalog” pages with lists of links
  - “Text” pages with few or no links
  - *Embedded images more interesting than links*
- Just do it rather than do it right
- But...
  - The web became *serious*
  - Then research started to matter
  - **Tooling for web/UI design became important**

# Ontology Research

- Dating back...
- Multiple foci
  - Technology (logics, reasoners...)
  - Meta-physics (what there is)
  - Knowledge Acquisition
  - NLP
  - Engineering
- Largely ignored by SW developers
  - Web 2.0, groundswell
  - Specifically criticized by some SW pundits

# Grassroots to the Semantic Web



# A little semantics...

- The SW catchphrase
  - “A little semantics goes a long way”
- Sometimes strengthened
  - A lot of semantics is too much
  - 80/20 rule
- Double-edged sword
  - FOAF doesn't look like even 1%
  - The simplicity of FOAF hides any serious value proposition for SW
  - SW not for people, for data
  - Reasoning? Quality?

# Wherefore Reasoning?

- Very hard to “sell” OWL reasoning
- Many users want very simple reasoning
  - Simple subclass
  - Simple range/domain *constraints*
  - Simple **rules** (q.v. RIF)
- Some users want more than OWL
  - But just to express their semantics, *not in run-time system*
- Reasoning supports quality
- Improving precision?
- Improving recall?

Must be measured.

# Getting it right

- Does quality matter?
- Good quality ontologies cost more
  - **Required** for some applications
- Improvements in quality can improve performance [Welty, et al, 2004]
  - 18% *f*-improvement in search
  - Cleanup cost ~1mw/3000 classes
  - BUT ... low quality ontology still improved base

# Where's it going?

- Real business uses will require reasoning & quality
- Reasoning & Quality drive need for tooling
- W3C not about APIs

# Tooling

- Protégé (Stanford & Manchester)
- ODM metamodel & profiles
- Eclipse-based
  - ATL
  - IBM STK (Semantic Technology Toolkit)
    - alphaworks
- VOM
  - UML-profile plugin (e.g. Rose, MagicDraw, RSA...)
  - Integration with reasoning services
- ...others (growing)