Semantic Technologies

Karin Breitman
Laboratório de Engenharia de Software - LES
PUC-Rio
Organization of the talk

- Colateral Effects
- Who gets the benefits?
- What the Semantic Web isn't?
- Syntactic Web x Semantic Web
- How will it work?
- Scenarios Technology
- Ontology
- Metadata
- Ontology
- Software Agents
- Inference Mechanisms
- Model Integration
- Trust

Semantic Web Technologies
The Web today

- The syntactic web
  - Developed as an information exchange medium for PEOPLE
  - Computers display information whereas interpretation is left for humans.

- 8 billion of pages
  - Web search engines
  - No “result interpretation” engines
Information for Human Consumption

Karin Koogan Breitman

Dr. Karin Breitman received her DSc. from the Departamento de Informática da Pontifícia Universidade Católica do Rio de Janeiro, where she is currently teaching and continues to work in her research. Her interests are software requirements engineering, scenario based software process and software evolution. She was part of the Program Committee of the last two editions of the International Conference on Requirements Engineering (ICRE) and the workshop de Engenharia de Requisitos (MER) since 1998. Dr. Breitman is currently serving as the South American publicity chair for the RE'03. She belongs to ACM, IEEE and the Brazilian Computing Society (SBC), where she is currently serving in the board of directors.

Courses

- Engenharia de Requisitos (INF 1620)
- FMs Graduação: Curso de Engenharia de Requisitos (INF 2044 & INF 1809)
- Engenharia de Requisitos (INF 1620)

Algumas Publicações
How a machine sees this page
Semantic Web

• “The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

Tim Berners-Lee, Hendler e Lassila
Scenario I - Search for “organ” (musical instrument)

- The Official US Government Web site for Organ and Tissue Donation...
  Official Organ Donation and Transplantation Web site of the US Department of Health and Human Services. Each day, about 70 people receive an organ transplant.

- United Network for Organ Sharing: Organ Donation and...
  UNOS oversees the national database of clinical transplant information and operates the computerized organ sharing system, matching donated organs to patients.

- Allen Organ Company - The Largest Builder of Church Organs in the...

- Organ Historical Society
  www.organsociety.org/ -

- Coalition on Donation -- Donate Life -- Homepage
  Send an e-postcard. Get the Facts Check out organ donation facts and transplantation statistics; review stories of hope and visit related links.

- TransWeb: All About Transplantation and Donation
  Questions and answers, myths, and other information about organ transplants and donation.

- Organ History
  The Pipe Organ, a description in two parts: "The organ and how it works" and "The History of the Organ".

- Organ System Pathology
  Organ System Pathology Images. Return to the WebPath main menu. Sections of the WebPath images are available for viewing below by organ system.

- Organ Pipe Cactus National Monument (National Park Service)
  ... park conditions. more -> Organ Pipe Cactus National Monument celebrates the life and landscape of the Sonoran Desert. Here, in this...
Scenario II - Scheduling medical appointment

- Lucy’s mom needs physical therapy sessions:
  - Lucy is driving her, needs to be near her place of work
  - Should be a good professional
  - Should belong to her medical plan
  - Should have openings compatible with Lucy’s agenda
Semantic Web Architecture (TBL)
Semantic Web

- **Universality**
  - hypertext link “anything can link to anything”
  - Web technology must not discriminate:
    - between the scribbled draft and the polished, performance,
    - between commercial and academic information,
    - among cultures.

- **Decentralized**
  - compromise: throw away the ideal of total consistency
  - allow exponential growth - unverified
How to get there?

- Metadata
- Ontology
- Agents
- Inference Mechanisms
- Model Integration
- Trust
Metadata

- Metadata - “data about data” - International Federation of Library Associations
- WWW Conference - agreement on semantics for internet resources
- Metadata Workshop - Dublin Core (subject, title, author, publisher, other agent, date, type, form, identifier....)
- Warwick Framework
- RDF
  - Extends the Warwick Framework
  - Defines an unified data model and syntax
RDF Model

Dublin Core:

- Subject
- Title
- Creator
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Relation
- Source
- Language
- Coverage
- Rights

<table>
<thead>
<tr>
<th>Subject (resource)</th>
<th>Predicate (property)</th>
<th>Object (value)</th>
</tr>
</thead>
</table>
Metadata

- Metadata alone is not enough!
- Communication
  - Polisemy, ambiguity and synonyms
  - Tacit Knowledge
Polisemy - organ

- The noun “organ" has 5 senses in WordNet.
  - 1. organ -- (a fully differentiated structural and functional unit in an animal that is specialized for some particular function)
  - 2. organ -- (a government agency or instrument devoted to the performance of some specific function; "The Census Bureau is an organ of the Commerce Department")
  - 3. electric organ, electronic organ, Hammond organ, organ -- ((music) an electronic simulation of a pipe organ)
  - 4. organ -- (a periodical that is published by a special interest group; "the organ of the communist party")
  - 5. organ, pipe organ -- (wind instrument whose sound is produced by means of pipes arranged in sets supplied with air from a bellows and controlled from a large complex musical keyboard)
  - 6. harmonium, organ, reed organ -- (a free-reed instrument in which air is forced through the reeds by bellows)
Ambiguity

- Increases the range of possible interpretations
- Different cultures may use the same expression with different meanings
- Appears as if there is understanding
  - Draw the curtains
  - Put the lights out
  - Dress the chicken
Synonyms

- **Common in Natural Language**
- **Have great impact on word-based searches:**
  - If we try harmonium instead of organ:
  
  - [Chandra and David's Indian Harmonium Page - Hand Pumped Indian](http://example.com) ... HARMONIUM. by David Courtney, Ph.D. ... Parts of Harmonium. There are a number of parts of the harmonium, here are some of the main ones: Portable harmonium. Body. ...
  
  - [Harmonium sur le web - [Translate this page]](http://example.com) Quoi de neuf? 13 septembre 2004 Serge Fiori autorise un hommage à Harmonium. Le groupe Premier Ciel est fier d'annoncer avoir obtenu ...
  
  - [The harmonium - Indian classical instrument](http://example.com) The harmonium is a Western instrument adapted for Indian music. It is mainly used in the accompaniement of Khyal, Thumri and Qawwali. ...
  
  - [Amazon.com: Music: Harmonium [ENHANCED]](http://example.com) ... Yes, Carlton is maturing and Harmonium brings her closer to where she's going, but she's still got a ways to go. ... Harmonium is beautiful. ...
Tacit Knowledge

- Too trivial to mention
- Embodied knowledge
- Always present in human conversation
- Must be made explicit to allow machine interaction

→ more expressive representations
What’s needed

- More expressive representations that:
  - Explicit domain concepts, axioms, properties and relationships
  - Make domain assumptions explicit
  - Share common understanding of the structure of information among people and machines (software agents)
  - Separate domain knowledge from operational knowledge
Ontology

- **Subject:** study of the categories of things that exist or may exist in some domain”
- **Product:** “catalog of the types of things that are assumed to exist in a domain of interest $D$

*ontos* (being) + *logos* (word)

- Plato - metaphysics
- Aristotle - 10 categories κατηγορία
- Philosophy - study of being $X$ study of various kinds of beings
Porphyry’s Tree

Supreme genus:

Differentiae:
- material
- immaterial

Subordinate genera:

Differentiae:
- animate
- inanimate

Subordinate genera:

Differentiae:
- sensitive
- insensitive

Proximate genera:

Differentiae:
- rational
- irrational

Species:

Individuals:
- Socrates
- Plato
- Aristotle
- etc.
Ontology x ontology [Guarino98]

- Philosophy
  - Discipline → Ontology

- Computer Science
  - Artifact → ontology
Approaches

- **AI - Knowledge Engineering**
  - Domain mapping, creation of large knowledge bases, mapping human knowledge
  - Upper Ontologies
  - Built by experts

- **Semantic Web**
  - Specific applications
  - Built by non experts
Upper ontologies

• Cyc upper ontology
  - Knowledge base with around 3000 terms
    (“capturing the most general concepts of human consensus reality”)

• SUMO

• John Sowa’s

• WordNet
  - On line lexical reference with over 42,000 links to words
    (nouns, verb, adjectives and adverbs)
    (“whose design is inspired by current psycholinguistic theories of human lexical memory”)
Criticism

- There is no consensus on a über upper ontology
  - Open questions
  - 3D Versus 4D
  - As software engineers:
    - Little chance of making a real contribution
    - Tackle practical problems

- Ontologies are not substitutes for human categorization
  - Wittgenstein
  - Rosch
Semantic Web

- Application rather than upper ontologies:
  - More restricted
    - Contextualized information
    - Relevant to the application

- Narrower Goals

- Smaller
Definition [Tim Berners-Lee]

“an ontology is a document or file that formally defines the relationship among terms”

- Maedche’s definition:

\[ O : = \{ C, \mathcal{R}, \mathcal{H}^C, \text{rel}, \mathcal{A}^O \} \text{ consisting of:} \]

\[ \mathcal{H}^C : \text{A concept hierarchy, } \mathcal{H}^C : \mathcal{H}^C \text{ is a directed relation } \mathcal{H}^C \subseteq C \times C \]
\[ \text{which is called concept hierarchy or taxonomy. } \mathcal{H}^C (C_1, C_2) \text{ means } C_1 \text{ is a subconcept of } C_2 \]

\[ \text{A function } \text{rel} : \mathcal{R} \rightarrow C \times C \text{ that relates the concepts non taxonomically} \]

\[ \text{A set of ontology axioms } \mathcal{A}^O, \text{ expressed in appropriate logical language.} \]
## Types of ontology

<table>
<thead>
<tr>
<th>Catalog/ID</th>
<th>Thesauri “narrower term” relation</th>
<th>Formal is-a</th>
<th>Frames (properties)</th>
<th>General Logical constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terms/glossary</td>
<td>Informal is-a</td>
<td>Formal instance</td>
<td>Value Restrs.</td>
<td>Disjointness, Inverse, part-of...</td>
</tr>
</tbody>
</table>
Ontology versus Metadata

- Combine resources with metadata,
- Use ontology(ies) to explicit concepts and relationships.
Ontology Construction

- Uschold
- Methontology
- Tove
- McGuiness & Noy (Ontology 101)
- Lexicon Based Construction Process:
  - Non experts
  - Mature process
  - Validated in real life projects
  - Lightweight
  - Focus in the “language of the problem”
    - Makes explicit the separation between domain specific terms and those from the minimal vocabulary
Ontology Construction

ontological construction process

Lexicon

Ontology
Language Extended Lexicon (LEL)

- A different type of lexicon.
- Oriented towards the language of the problem.
- Using both denotation and connotation.

Foundations:
- The Theory of Semiotics (Eco): the idea of symbols and their contextualization.
- Carnap’s Observational Language (Vienna School): the idea that terms in a language are anchored on observational phenomena.
- Sociology: the idea that language is a reflection of culture.
- Domain Languages (Neighbors): high level reuse is achieved by specific languages, similar to the concept of little languages (Bentley).
- Social-Aspects of Computing (Kling): the interplay between social actors and computers systems.
- Hyperties (Shneiderman): one of the first implementations of the hypertext.
Example - Lexicon Entry

Requester

Notion:
• person who invites attendees to a meeting.
• may be a participant.

Behavioral Response:
• defines the objective of the meeting, the subjects to be discussed, the attendee list and materials.
• records the objective and the attendee list in the agenda.
• organizes the meeting.
• decides changes in the meeting requirements.
## Types of entry

<table>
<thead>
<tr>
<th>Notion</th>
<th>BehavioralResponse</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject</strong></td>
<td>Who is the subject?</td>
</tr>
<tr>
<td><strong>Verb</strong></td>
<td>Who performs, when it happens e what procedures are involved.</td>
</tr>
<tr>
<td><strong>Object</strong></td>
<td>Define the object and identify other objects with which it relates to.</td>
</tr>
<tr>
<td><strong>State</strong></td>
<td>What it means and which actions gave rise to this state.</td>
</tr>
</tbody>
</table>

- **Subject**: Who is the subject? Which actions are performed?
- **Verb**: Who performs, when it happens and what procedures are involved. What are the impacts of the action in the environment (other actions that also occur) and what are the resulting states.
- **Object**: Define the object and identify other objects with which it relates to. Actions that can be applied to the object.
- **State**: What it means and which actions gave rise to this state. Identify other states and actions that may happen departing from the state object of the description.
Informações sobre o léxico

Nome: fast-food restaurant

Nação: It is a restaurant where customers place an order, pay and receive the order standing in front of a cashier.

Impacto: customer buys food at fast-food restaurant.

Sinônimo: fast-food

Cenários e termos do léxico que referenciam este termo

Cenários | Léxicos
----------|----------
menu option
employee
customer
attendant

Veja o código fonte.
Lexicon to ontology

![Diagram of lexicon to ontology conversion process]

 ontology construction process

Lexicon → Ontology
Process

Check list relation $R$

list

term

new concept $C$

verb

verify behavioral responses

new relation

situation

analyze behavioral responses

identify importance

identify disjoint relationship

object subject

identify generalization

ontology

Prediction

- (The Semantic Web will be composed of a) "great number of small ontological components consisting largely of pointers to each other" [Hendler01].

- As opposed to general, upper ontologies such as CYC and Wordnet
The problem is not ontology engineering

- A Model by any other name....
  - Model construction - ER, OO, Kaos, i*, ...
  - Difficulty is in the “good model”
    - Corresponds to reality
    - Good decomposition
    - Adequate terminology
    - Explicit important concepts
    - Validated with users …
  - Focus on generalization (subsumption)

- Implementation is straightforward
  \[ \text{DAML+OIL - OWL} \]
Semantic interoperability

- Bottleneck: Different ontologies must speak
  - 8 Billion pages - cannot expect everyone to commit to the same ontology
  - Give opportunities for “strange agents” to communicate:
    - Different implementations
    - Open Environment
    - Web Services paradigm
Existing approaches

- Automatic, semi-automatic or, even, manual.

- Ontology design community- syntactic matches, use of dictionaries, transformation:
  - Ontomorph
  - Chimaera
  - Protégé - Prompt
  - CATO
  - Bailin & Truszkowski
Ontology Interoperability

• **Merge:** results in a unique ontology that contains all the terms from merged original ontologies, without indication of their former origin.
  - Often, the ontologies cover similar or overlapping domains.

• **Mapping:** results in a formal structure containing expressions that link concepts from one conceptual model to another.
Ontology Interoperability

- **Alignment**: results in separate ontologies with links between them. The links allow ontologies to share terms.
  - Often, the ontologies cover complementary domains.

- **Integration**: results in a unique ontology created by assemblage, extension, specialization or adaptation of ontologies from different subject areas.
  -> It is possible to identify provenance.
CATO: an implementation
Result

- Part of the Aligned Ontology:

```xml
- <owl:Class rdf:about="file:firstOnto.owl#Book">
  - <owl:disjointWith>
    - <owl:Class rdf:about="file:firstOnto.owl#Manual"/>
  </owl:disjointWith>
- <owl:disjointWith>
  - <owl:Class rdf:about="file:firstOnto.owl#Proceedings"/>
  </owl:disjointWith>
- <owl:disjointWith>
  - <owl:Class rdf:about="file:firstOnto.owl#InConference"/>
  </owl:disjointWith>
- <owl:disjointWith>
  - <owl:Class rdf:about="file:firstOnto.owl#Conference"/>
  </owl:disjointWith>
- <owl:disjointWith>
  - <owl:Class rdf:about="file:firstOnto.owl#MastersThesis"/>
  </owl:disjointWith>
- <owl:disjointWith>
  - <owl:Class rdf:about="file:firstOnto.owl#InBook"/>
  </owl:disjointWith>
- <owl:disjointWith>
  - <owl:Class rdf:about="file:firstOnto.owl#PhdThesis"/>
  </owl:disjointWith>
- <owl:disjointWith>
  - <owl:Class rdf:about="file:firstOnto.owl#Misc"/>
  </owl:disjointWith>
- <owl:equivalentClass>
  - <owl:Class rdf:about="file:secondOnto.owl#Book"/>
  </owl:equivalentClass>
</owl:Class>
```

**Added Information**
Agents and the Semantic Web

- Rely on Semantic Web technologies to better function in Open Environments
- Autonomous behavior
- Interaction is only possible if software agents can communicate
  - Metadata
  - Ontology
One Scenario

- Scientist has a personal agent
  - Responsible for paper submission
  - Based on an ontology it decides which conference to submit to
  - Sends paper to chair agent
  - If received before deadline, chair agent notifies that paper was submitted.
  - Chair agent distributes the paper to reviewer agents, taking into consideration paper topics and author institution....
Inference Mechanisms

- Semantic Web Languages can be translated to a formal representation:
  - Deduce new logical sentences from existing ones → inference
  - First Order Logic:
    - Very expressive → widely used in KR
    - Reasoning in FOL → undecideble an intractable
  - Description Logics
    - Reduced form of FOL (to frame description)
    - Computability by limiting the expressiveness of FOL
    - Focuses on describing things and by determing the subsumption relationship
Example
Example

[Diagram showing an entity-relationship diagram with classes and relationships between entities such as Living, Animal, Mammal, Plant, Herbivore, and their associated properties and relationships like eats.]
**Example**

FaCT Reasoner

[Diagram of a conceptual graph with classes and relationships such as Animal, Plant, Living, etc., connected by arrows showing relationships like is-a and has-a.]
Inference Mechanisms

• Consistency checks
• Verification tools
• Viewpoints
Trust

• Are there “greener” ontologies?

• How can we trust semantic content?
  - Naive assumptions
    • Honesty,
    • Users are able to describe their domains adequately,
    • No Bias.
  - Provenance
  - Traceability
Web Services

- Authentication, Authorization and Integrity:
  - XML Signatures
  - SAML (Security Assertion Markup Language)
  - Single Sign On
  - WS- Security

Semantic Web Services
“Bringing the Web to its full potential…”
- **semantic web enabled web services**
  - swws.semanticweb.org

- **Ontology Adoption**
  - Guarantees same interpretation by service users and providers

- **Ontology mediation services**

- **DAML-S service ontology** - provides primitives for describing service concepts and relationships.
Scenario I Revisited

Music Instrument Classification - 1914
Erich Moritz von Hornbostel & Curt Sachs
Scenario I - Revisited

GOOGLE For: organ instrument wind

- **BUBL LINK: Wind instruments** ... Subjects: keyboard instruments, organ music, wind instruments DeweyClass: 786.5 Resource type: documents; Trevor Wye Corner: Articles on various aspects ... bubl.ac.uk/link/w/windinstruments.htm - 17k - Cached - Similar pages
- **Greek and Roman Pipe Organs** ... amongst wind instruments by Bellermain's "anonymus," (Anonymi de Musica, ... the idea of his organ from the Syrinx or Pandean pipes, a musical instrument ... Wind instruments, Woodwind, Organs Woodwind, wind instruments, organs, the flute: a web directory.
- **The Classical Free-Reed, Inc. Taxonomy of Musical Instruments** ... by the wind, can be convincingly classified as a free-reed instrument, ... the organ belongs to both the edge instruments and reed pipe instruments. ...
- **American Music - Software - Software Synthesis/Sampling - B4 Organ** ... B4 Organ VST Keyboard Native Instruments (Mac/Wind Item # TW-ninb42 -ZB Click for Larger Picture. This product is eligible for FREE SHIPPING! ...
- **St. John's Presbyterian Church - The Organ** ... The organist who plays an instrument with flexible wind must refine his or her playing technique in response to factors that simply do not arise with a ...
- **Rieger - Kloss [ The History ]** ... The actual first ancestor of the organ is the Chinese instrument "sheng". ... in organ playing was the introduction of spring and slider wind-chests. ...
- **organ: Information From Answers.com** ... Organs were the first keyboard instruments, even though technically they ... Other wind instruments that have no reservoir of gas but use a separate ...
Scenario II - Revisited

- Lucy’s mom needs physical therapy sessions:
  - **Software Agent working on Lucy’s behalf**
  - Lucy is driving her, needs to be near her place of work
    - metadata
  - Should be a good professional
    - Trust layer - not quite there yet.
  - Should belong to her medical plan
    - metadata
  - Should have openings compatible with Lucy’s agenda
    - software agent compares Lucy’s and Doctor’s agendas using concepts such as time, session, working hours.... negotiated between their ontologies - we assume both ontologies are consistent (inference mechanisms).
What the semantic web is NOT

- AI
  - Machine processable data and documents DO not imply that computers now magically gain understanding
    - Many techniques come from AI
    - Partial solutions are acceptable in the SW
    - “if AI’s goal is to build agents that possess intelligent equal (or superior) to that of human beings, the Semantic Web wants to help humans accomplish daily chores on the web”. Antoniou & Harmelen

- The semantic web is not a separate web.
  - Layered architecture
  - Refactoring

- Not every application in the semantic Web will need to explore its full potential
  - Solution complexity depends on the problem
  - W3C provided for three versions of OWL - lite, DL & Full
What to expect?

• Immediate Result:
  - reduce load on end users

• Collateral Effect
  - Esperanto versus Tower of Babel
  - perhaps, we will create knowledge representation models that can be globally shared...