OntoWhat?

Kurt W. Conrad

President, The Sagebrush Group

conrad@SagebrushGroup.com

XML 2002

Baltimore

December 12

What’s an Ontology?

- That’s like asking what “is” is
- Depends on whom you ask
- And what their motivations are

Popular Definitions Focus on "Technologies"

- RDF
- Topic Maps
- Taxonomies
- Controlled vocabularies
- UML
- UBL
- DAML, DAML-S, DAML+OIL
- DTDs and Schemas
- UDDI
- WSDL
- OWL
- Ontolingua
- Ontology Markup Language
- Other lexicons and policy languages

All roads lead to KIF

- Stanford Knowledge Systems Laboratory
- Gruber definition
  - “An ontology is the specification of a conceptualization”
- Focused on AI and intelligent agents
  - De-emphasize traditional algorithmic software
  - Articulate patterns and rules
  - Let generalized inference engines decide how to act in new, “unanticipated” contexts

See Parallel Business Drivers

- Infoglut
- Complexity
- Uncertainty
- Maintainability
- Integration
- Web Services
- Semantic Web
- Multiple languages
  - Human languages
  - Computer languages
    - Database schemas
    - APIs
    - Function libraries
  - XML tag sets

The Bottom Line for Business

- Intelligent behavior
- Improved organizational performance
- Very strong organic and adaptive undercurrents

What's it Take to Specify a Conceptualization?

- Syntax
  - A symbol set
- Semantics
  - The definitions and meanings of the symbols
- Relationships
  - How the concepts relate to each other
- Rules / Axioms
  - Constrain use and/or model real world

“Standard” Ontological Models

- Represent meta-ontologies:
  - Conceptual structures that are designed to express and encode domain-specific ontologies
- Vary in their fundamental expressiveness
- Optimized around a set of human and automated behaviors
- Implicitly constrain expressiveness

Topic Maps

- Emphasize
  - Syntax (topics)
  - Relationships (occurrences and associations)
- De-emphasize rules and semantics

RDF

- Emphasize
  - Rules and relationships
  - Subject / predicate (object) model
- Little standardization of
  - Syntax (terms)
  - Underlying semantics (definitions)
- Tends to push the semantic problem off
  - Meaning can be found at the end of a URL

Taxonomies

- Emphasize syntax
  - Controlled vocabulary
- Meaning and relationships often implied through hierarchy
  - Routinely ambiguous

DTDs and Schemas

- Bound syntax
- Formalize a subset of the machine processing semantics
- Less emphasis on rules and relationships
  - Deal mostly with tree structures

The More “Serious” Encoding Standards

- Use logical constructs to specify
  - Semantics
  - Rules
  - Relationships
- Some argue that first order logic is absolutely required for adequate expressiveness

Working Definitions of Ontology

- Formalize semantics
- Formalize relationships
- Provide a model
  - At whatever level of detail or completeness

But this Whole Approach Breaks Down Rather Quickly

- Semantic specification trap
- Whole systems perspective
- Origins of conceptualization

Semantic Specification Trap

- Fundamental problem:
  - We express meaning with yet another set of symbols
  - These symbols have their own meanings
- These are some of the fundamental issues in language and linguistics
- We can never fully specify semantics
  - Humans are the final arbiters and repositories of meaning
  - Is meaning ultimately tied to desire and intent?

Whole Systems Perspective

- A formalized ontology is a Knowledge Artifact
- It is used by an agent to drive or otherwise influence behavior (actions and/or decisions)
- Computer systems aren't the only agents that interact with ontologies
  - Individual Agents
  - Collective Agents (organizational and social agents)
- Each agent type has different knowledge requirements and capabilities

Origins of Conceptualization

- Computer systems don’t conceptualize
- Individuals do
- We are the source for new concepts and new meanings
  - Create new terms
  - Assign new meanings to existing terms
- We generally make a mess of things

What Is “Is”?

- We perceive what is, automated systems don’t
- Human ontologies (world views) are incredibly rich in tacit and implicit knowledge
  - Ontologies exist in their richest form within our minds
  - Formalized ontologies represent a subset of concepts
  - Digitized ontologies represent an even smaller subset
- Computer will always be playing catch up
  - Algorithmic behavior perishes and loses relevance
  - Optimal (intelligent) behavior changes

What's an Ontology?

- In the real world, it's more than just "the specification of a conceptualization"
- It has to include the conceptualization, itself

Rant!

- Yet another example of the IT community co-opting an important term and damaging the underlying concept
- Purer definition: “The science of being”
- Understanding “what is” represents the increasingly critical boundary between the real world and our digital representations
- Being needs to be understood in its fullest and richest form, not just in terms of what can be expressed to a computer

Individual Agents

- Able to handle ambiguity
  - Find and create meaning
  - Derive meaning from context
- Exhibit bounded rationality
  - Operates on basis of perceived understanding, not absolute understanding
- Common sense seems to be key
  - Meaning appears to be somewhat independent of syntax/symbols

Individual Ontologies

- The most interesting aspects are those that most resist codification
  - The synthesis of prior knowledge into a value system
  - Ability to understand the bounds of a context, switch contexts, and apply concepts across contexts
  - The filters and triggers that we use to decide whether we understand something
  - The rules that impact knowledge selection and transformation behaviors
- These are the critical aspects of our world views that make us truly intelligent

Do We Agree On What “Is” Is?

- Composite agents comprise
  - Individual agents
  - Automated agents
  - Other composite agents
- Range from the more engineered to the more organic (organizational and social agents)
- My reality isn’t your reality. We will probably never fully agree
  - Inefficient to engineer for a single individual
  - Formalized ontologies tend to have greater value when reflect the consensus of a group

Semantic Sufficiency

- Alignment Theory provides an answer to the Semantic Specification Trap
- Won't get (and don't need) perfection
  - Alignment, agreement, understanding, meaning
- Only need semantic sufficiency
  - Varies among agent types
  - Varies across specific agents
    - Prior knowledge and training
    - Vocabularies and symbol sets
    - Behavioral intent

Collective Ontologies

- Require some level of negotiation
- Require sufficiency/overlap
- Agreement is key
- Agreement has bounds
  - Heavy reliance on context-specific implicit knowledge
- Sustained agreement requires re-interpretation

Example: Company A

- Uncontrolled product terminology
  - Hampered usability of web site
  - Precluded effective searching
- Identified
  - Preferred terms and alternatives for same concept
  - Normative definitions for key technology concepts
- Used a consensus model with small teams
- Published normative taxonomies for use across organization

Example: Company B

- Indexing biomedical research documents
- Dealing with four key ontologies
  - Software developers
  - Indexing trainers
  - Indexers
  - Customers
- None of these ontologies are fully understood or formalized
- Long-term alignment and maintenance issues

Company C

- Engineered a set of controlled vocabularies to support indexing and other automated processes
- Each concept is modeled only enough to allow individuals to differentiate and apply the right one

Formalization Options

- Range
  - Informal
  - Semi-formal
  - Formalized
- Different cost / benefit ratios
  - Formalizing using first order logic is non-trivial
  - Natural human language is the richest way to express semantics, but it still isn't complete
  - Optimal formalization varies (e.g., by level of automated behavior)

Ontological Engineering

- Two primary activities
  - Development of common definitions
  - Formalization through expression
- Two dominant strategies
  - Formalized methodology
    - Represents an effort to avoid negotiations
  - Negotiation-based approaches

Typical Methodology

- Determine scope
  - Concepts & range of behaviors
  - Determine level of formalization
- Determine decision making model
- Analysis and negotiation phase
  - Terms, definitions, rules, relationships
  - Capture supporting rationale
- Codification and delivery
  - Natural language definitions
  - Other representations

Semantic Extraction Tools

- The holy grail
  - Unstructured data in
  - Meaningful structures out
- I’ve never seen them do quite enough

Ontological Maintenance

- Bounded utilization scenarios
  - Can’t express what want to
- Unbounded utilization scenarios
  - Look for changes in usage patterns
  - Semantic extraction tools may have a role here
  - May be able to associate patterns with an engineered ontology to look for pattern shifts

The Dark Side of the Force

- Increasing interest in ability to predict behavior
  - Values assessment
- Interest in ability to control human behavior
  - Imperatives unreliable
  - Values normalization viewed as more reliable
    - Behavior-based safety
    - Innovation management
    - Knowledge sharing
- It's a short hop from alignment to engineering
  - Individual ontologies
  - Collective ontologies

In Conclusion…

What's an Ontology?

A knowledge artifact of specialized content which is used by an agent as one input to the higher-order processes that rationalize data and information in such a way as to drive supposedly intelligent behavior.

- The characteristics of the optimal ontology vary by agent type and purpose
- Automated agents require explicitly formalized ontologies
- But formalized ontologies represent only one of the many types that impact organizational performance

- Formalization is inherently problematic
  - Individuals drive conceptualization
  - The tacit and implicit dimensions resist complete formalization
  - Most semantic formalization represents the mapping of the properties of interest into an engineered symbol set that is machine processable
  - The smaller the scope of the conceptual domain, the more complete the formalization can be

- From a whole-systems perspective, intelligent system design involves
  - Recognizing the limits of ontological engineering
  - Formalizing and automating only those conceptualizations which are sufficiently understood and stable
  - Carefully distributing responsibility for ontological alignment across the various individual, automated, and collective agents that comprise the system

- To be successful, an ontological engineering initiative must deal with:
  - Conceptualization within Individual Agents
    - Learning and adapting
    - Use of language as a tool and enabler of conceptual change
  - Social factors which impact and drive conceptual alignment within a community of shared understanding
  - Representational issues
    - Some level of formalization is likely necessary to document the critical aspects of the shared ontology

- Even if we can't achieve perfection, ontological engineering can provide value
  - Insights into the conceptualization process
  - A more complete and comprehensive framework for formalizing
    - How to formalize, generally
    - How to improve the quality of formalization (or pick the right balance)
    - Where formalization issues point back to more fundamental conceptualization issues that need to be resolved

- Paradoxically, in most cases, quality of the conceptual model is less important than the expression itself
  - Will likely never be perfect
  - Can still function as an explicit alignment point
- Parallels some of the lessons of XML
  - Generic markup isn’t a complete specification
  - But often provides an adequate basis for alignment

A Blatant Marketing Pitch

- The Ontolog Forum is an open forum for discussing
  - Ontological Engineering
  - Business Ontologies
- Spun off and loosely associated with the OASIS UBL effort
- Archive can be found at ontolog.cim3.org/forums/ontolog
- To join, give me a business card or email one of the conveners
  - Kurt Conrad conrad@SagebrushGroup.com
  - Peter Yim yimpp@cim-oem.com
  - Leo Obrst lobrst@mitre.com