Sharing and Integrating Ontologies

John F. Sowa

16 March 2010

Foundations for Interoperability

Interoperability requires precise documentation of the syntax, semantics, and pragmatics of all the interactions among interoperable systems.

Formal ontologies are metadata about the things, events, properties, people, and information involved in the design, implementation, and use of those systems.

Typical documentation for application programs:

> **This file contains the billing transaction codes (types of records) that are to be interfaced to General Ledger for the given month. For this process the following transaction codes are used: 32 - loss on unbilled, 72 - gain on uncollected, and 85 - loss on uncollected. Any of these records that are actually taxes are bypassed. Only client types 01 - Mar, 05 - Internal Non/Billable, 06 - Internal Billable, and 08 - BAS are selected. This is determined by a GETBDATA call to the client file. The unit that the gain or loss is assigned to is supplied at the time of its creation in EBT.**

Ontologies, Terminologies, and Lexical Resources

Ontology: The study of what exists and how it can be formally described and axiomatized.

Terminology: A classification of the technical terms used in any branch of science, engineering, business, and the arts.

Lexical resource: A dictionary, thesaurus, grammar, or other representation of the vocabulary, syntax, or semantics of some natural language.

These are three related, but different kinds of representations.

WordNet, for example, is a valuable lexical resource.

But aligning two ontologies to WordNet does nothing to assure that corresponding categories have equivalent definitions.

The World as Interface

From a book by Otto Rössler with the above title:

Exophysics: The many, often incompatible theories about invisible entities such as atoms, molecules, fields, etc.

Endophysics: The many, often incompatible views of the world by all the people and beasts who live in it.

Interface: The phenomena that people and other animals see, feel, interpret, and manipulate.

Consistency with the phenomena at the interface is the ultimate criterion of accuracy for both exophysics and endophysics.

Ambiguities in ordinary language result from reusing the same words in multiple versions of exophysics and endophysics.

Any formal theory can only express one version at a time.

Methods of Reasoning

Theorem proving is only one of many ways of using logic.

The most common way of using logic in computer systems is to evaluate the truth of a statement in terms of a model:

1. Database: The tables of a relational DB or the networks of an object-oriented DB are isomorphic to a Tarski-style model of the subject domain.

2. Query processing: An SQL query or a path-based query is evaluated to truth or falsity in terms of the given DB.

3. Constraint checking: For a DB update, each constraint must be evaluated to True before the update is permitted.

For these purposes, first-order logic is highly efficient.

In the worst case, FOL takes polynomial time. With indexing, evaluating a statement can often be done in logarithmic time.

The world economy depends on this method of using FOL.

Consistency Check

Using a theorem prover to check the consistency of an FOL theory can take exponential time or even be undecidable.

But any theory that has at least one model is guaranteed to be consistent.

Finding a model for an FOL theory is as difficult as theorem proving: it can be undecidable or take exponential time.

Fortunately, most ontology projects begin with at least one typical example, and the description of that example can serve as a model.

Simple and efficient consistency test:

1. **Represent the known example in a relational DB.**
2. **Translate each axiom of the theory to an SQL query.**
3. **Check whether all the queries are true of the given DB.**
4. **If so, the theory is both consistent and true of the example.**

Three Kinds Acyclic Graphs

An acyclic graph has no cycles.

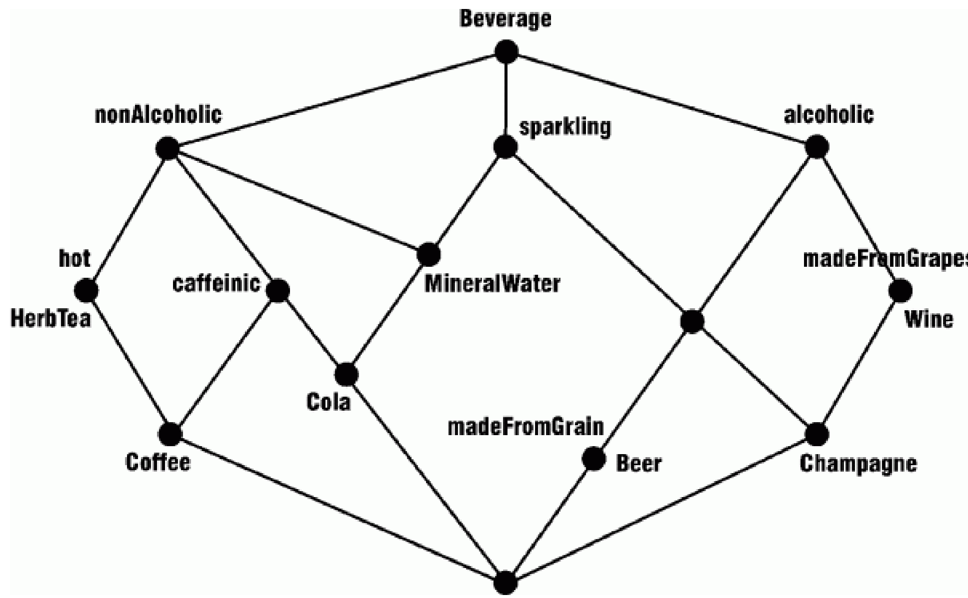
A tree is an acyclic graph in which each node has exactly one parent.

A lattice is an acyclic graph that looks like a tree from both ends.

Lattices are often used to show multiple inheritance.

The word *hierarchy* is sometimes used for trees and sometimes for any kind of acyclic graph.

Generating Lattices Automatically

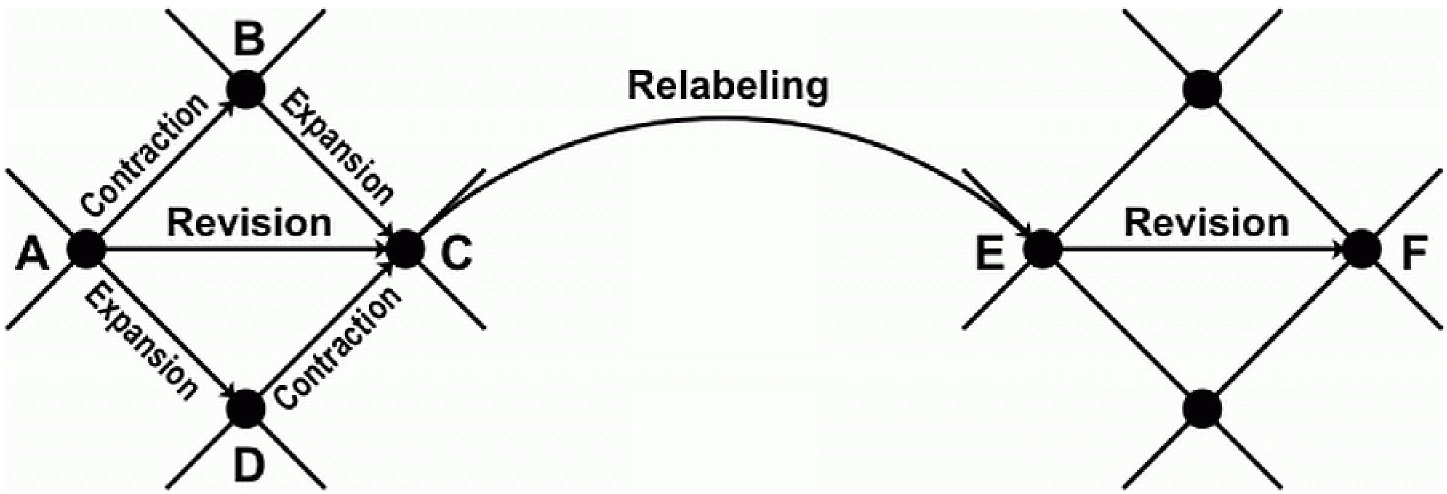


This lattice was generated from a table of concepts and attributes by the method of Formal Concept Analysis (FCA).

Automated methods guarantee that multiple inheritance is consistent.

Lattice of Theories

For any given logic, the set of all possible theories expressible in that logic forms a lattice.



The ordering is defined by specialization and generalization.

Adding axioms to a theory creates a more specialized theory.

Deleting axioms creates a more generalized theory.

Multiple Incompatible Theories

Different versions of exophysics may use incompatible ontologies.

Example of a three-dimensional ontology:

- A person is born at a time t1.
- A person dies at a time t2.
- For any time between t1 and t2, the person is alive at some location x and has attributes that may change from time to time.

Example of a four-dimensional ontology:

- A person corresponds to a 4D volume.
- The times t1 and t2 are the lower and upper bounds on the time coordinates of that volume.
- Some attributes of the person may be true or false at different points in that volume.

Reconciling Ontologies

All theories of exo- and endo- physics that have any claim to be true about the world must be consistent (at some level of approximation) with the observable phenomena.

Any data gathered from observation must be either consistent with or irrelevant to each theory that is claimed to be true about the world.

Different theories that are true of the world may also use terms and variables that are only indirectly related to the observable phenomena. Those terms might not be expressible in or consistent with the terms used in other theories.

But every conclusion derived by a true theory and expressed only in observable terms must be consistent with or irrelevant to all other true theories.

Hierarchy of Certified Theories

A finite subset of the infinite lattice of all possible theories.

Each theory in the hierarchy must be tested and certified to be consistent and used successfully in one or more applications.

Not all relationships are known or displayed in the hierarchy.

But any that are discovered can be recorded and saved.

Related Readings

A Guided Tour of Ontology,
http://www.jfsowa.com/ontology/guided.htm

Controlled Natural Languages for Semantic Systems,
http://www.jfsowa.com/talks/cnl4ss.pdf

Fads and Fallacies About Logic,
http://www.jfsowa.com/pubs/fflogic.pdf

Conceptual Graphs,
http://www.jfsowa.com/cg/cg_hbook.pdf

ISO/IEC standard 24707 for Common Logic,
http://standards.iso.org/ittf/PubliclyAvailableStandards/c039175_ISO_IEC_24707_2007(E).zip