Introduction (10N9)

The "ontology" umbrella covers different types of digital artifacts created and used in various communities to represent data entities and their relationships for purposes of annotating datasets, supporting natural language understanding, integrating information sources, semantic interoperability and serving as background knowledge in various applications. (10NA)

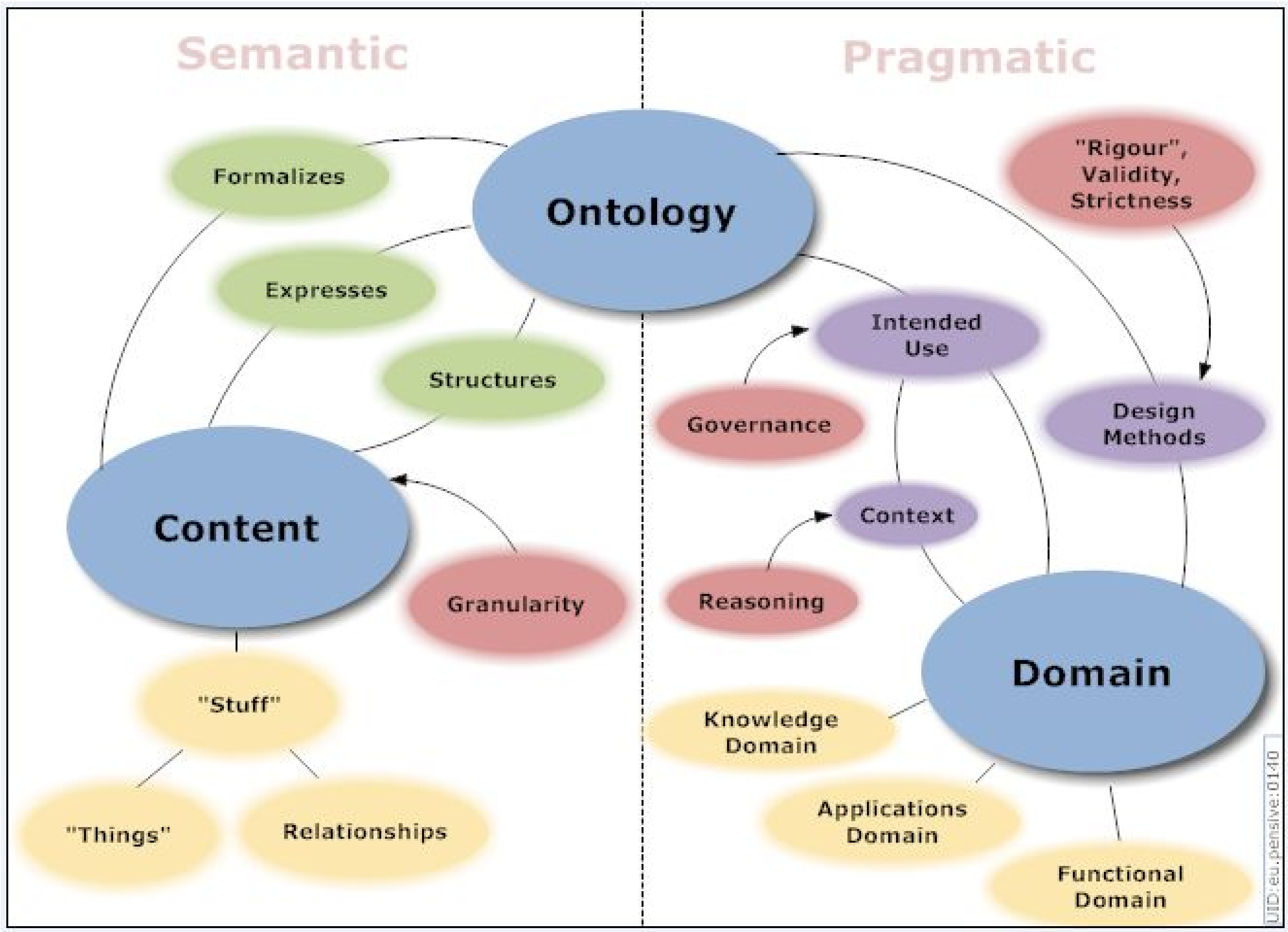
The Ontology Summit 2007 "Ontology, Taxonomy, Folksonomy: Understanding the Distinctions," co-organized by NIST and Ontolog Forum and co-sponsored by 50 institutions, brought together various communities (computer scientists, information scientists, philosophers, domain experts) with different understandings of what an ontology is to foster dialog and cooperation among diverse communities involved with ontologies. (10NB)

In practice, the term ontology spans a spectrum from formal upper-level ontologies expressed in first order logic (e.g., Basic Formal Ontology (BFO) and DOLCE) to simple lists of user-defined keywords used to, for example, annotate resources on the Web. The latter are called "folksonomies" playing an increasingly important role in Web 2.0. Taxonomies and controlled vocabularies (e.g., MeSH) in the middle of the ontology spectrum are typically organized hierarchically and most often used for information indexing and retrieval. Somewhere in between are ontologies representing other kinds of relationships among entities (e.g., functional, physical), often based on formalisms such as frames or description logics. Examples of such ontologies in the biomedical domain include the Foundational Model of Anatomy, SNOMED CT and the NCI Thesaurus. (10NC)

The goal of the Ontology Summit was not to establish a definitive definition of the word "ontology", which has proved extremely challenging. Analogously, the goal is not to organize ontologies along any particular single dimension either. Rather, we propose to identify a limited number of key dimensions and configurations along which ontologies can be characterized while simultaneously providing operational definitions for identified dimensions. The relative position of ontologies in the space defined by these dimensions, the "Framework", illstrates the similarities and differences between ontologies. The Framework has been applied to the characterization of a dozen ontologies, whose descriptions were collected through a survey. (10ND)

The Framework Dimensions (10NG)

A major goal of Ontology Summit 2007 was bringing together communities working on ontology-like activities to encourage cooperation. With this in mind, the summit has attempted to characterize what an ontology is to construct a typology. The framework of dimensions is comprised of two groups of dimensions: semantic and pragmatic. Semantic dimensions include expressiveness, structure, and representational granularity. Pragmatic dimensions include intended use, automated reasoning, and prescriptive vs descriptive specifications. See diagram. <PLEASE PLEASE INSERT THE DIAGRAM AS LARGE AS POSSIBLE DIRECTLY INTO THE TEXT SO PEOPLE WITH LOW VISION OR SMALL SCREENS CAN SEE IT AMONGST THE TEXT> (10NH)



Expressiveness is a property of Knowledge Representation Languages (KRL) which describes the extent and ease a KRL can describe increasingly complex semantics, for example propositional logic, description logic(s), first order logic, sorted logics, modal logics, and others. (10NI)

Structure is a property of an ontology that records how elaborate or well organized semantics are encoded. The expressiveness of a KRL encoding an ontology may be more or less expressive than the knowledge representation language itself. Thus, a simple taxonomy, for example a tree structure, may be encoded in RDF/S, a description logic language such as OWL-DL, or first order logic, such as Common Logic. Viewed from a graph theoretic perspective, level of structure might be a simple set of terms (glossary), tree structures (taxonomy), directed acyclic graphs, a partial orders (faceted classification schemes), arbitrary directed graphs (e.g., RDF) and other structures in the future. (10NJ)

Granularity concerns the level of detail which an ontology is specified and the ways data may be processed and displayed. A crude measure of granularity is the number of concepts (nodes or circles) and the number relation instances (connecting lines known as links or edges in graph representations). However, this simple explanation does not account for the fact some ontologies have larger scopes (domains or areas of expertise and recorded knowledge) than others. A coarse grained ontology might be suitable for use as an upper ontology or broad subject index, while a fine-grained ontology (such as SNOMED CT with 300K concepts) may be better suited for encoding medical diagnoses. (10NK)

Intended use is the dimension that records the original purpose(s) of an ontology. Intended use includes semantically informed search, specification for databases and data entry, integration across multiple sources, agent communication languages, controlled vocabularies for recording medical diagnoses, and many others. (10NL)

Automated reasoning is a dimension which records the extent it is anticipated an ontology will be used by automated reasoning software to answer questions and similar tasks. If so, one would expect the ontology is likely be encoded as using some form of logic. (10NM)

Prescriptive vs. Descriptive characterizes whether the intent of the ontology developer is simply to describe contemporary semantic usage without regard as to the scientific correctness of the encoded knowledge; e.g., a whale might be described as a large fish in common parlance. Examples of such descriptive ontologies include folksonomies and most linguistic ontologies. Alternatively, ontologies intended as normative prescriptive documents, correctness is of considerable concern; e.g., a whale is a mammal not a fish. Other prescriptive ontologies include medical diagnostic terminologies, legal or regulatory ontologies, accounting ontologies, mathematical or engineering ontologies. (10NN)

Additional findings (10QE)

Governance (10QF)

A new dimension, **Governance,** addresses how decisions concerning the structure and content of ontologies are made. There was wide spread agreement at the summit that the practice of ontology with legal or regulatory implications must defer to existing legal, regulatory, and professional organizations concerning natural language definitions of domain specific entities and semantic relationships. Ontology development should be viewed as an effort to organize and formalize concept definitions and relationships defined by working institutions, not an attempt to replace existing definitions with de novo definitions generated by computer scientists working in isolation. It was also observed it is necessary to record the provenance of every definition incorporated into an ontology over time, e.g., controlling legislation, changing regulations, updating standards, etc. where definitions are taken. (10PB)

Folksonomies and Formal Ontologies (10NO)

The relationship between social tagging and folksonomies versus more traditional structured / formal ontologies such as taxonomies and axiomatized ontologies was also discussed. Until recently, these efforts have been viewed as competitive approaches. The consensus of the Ontology Summit was that social tagging efforts presents an opportunity to serve as large scale corpora for inferring and validating formal ontologies, akin to the use of large text corpora in computational linguistics studies. Correspondingly, formal ontologies can be used to inform social tagging by providing improved tag sets, and multi-faceted tagging. (10NP)

Ontologies as designed artifacts (10NQ)

Some members of the community argued the position that ontologies should be considered a type of designed artifact, and that ontological engineering should become a discipline complementary to software engineering and virtually any discipline dealing with data and information exchange. A state objective of this communique is that this course of instruction and learning will increasingly become a required educational component of relevant curricula. (10NR)

Design methodologies (10QG)

After looking at a large number of intended uses for ontologies, another discovery was the idea that a basic spectrum of design methodologies may be able to be applied. The method vary from strong software engineering design lifecycle issues with formalized requirements, evaluation, and verification ; all the way to "no-design" methodologies where folksonomies emerge from the local behavior of thousands of individual users. Design methodologies are strongly related to intended use. For instance, methodologies for managing controlled vocabularies and taxonomies are social in nature intended to capture generalities about the general meanings of words in a culture or domain. Whereas, the "verification" of ontologies is related to the role of reasoning or types of computational services to be enabled by particular ontologies. For example, if an ontology is used for data integration, the verification of the consistency and completeness of data metamodels are important, as opposed to a domain where the meanings of terms has legal consequences and verification is directly tied to capturing the provenance of design choices, rooted in authority or appropriate processes. (10QH)

Survey (10NS)

To elicit distinctions between various kinds of ontologies, an interactive survey was designed and posted on the Web. Respondents were invited to identify the community they represent and to describe the value of ontologies, as well as issues with ontologies in their community. The last section of the survey invites the respondents to describe and characterize ontologies or related artifacts in use in their community. (10NT)

Over fifty respondents from 42 communities submitted entries to the survey. The best represented communities were Formal ontology, Applications development, Standards development, Web 2.0 and Biomedicine. 41 terms were identified as closely related to ontology, including formal ontology, upper ontology, concept system and controlled vocabulary. 70 ontologies from a variety of domains were characterized including formal ontologies (BFO, DOLCE, SUMO), biomedical ontologies (Gene Ontology, SNOMED CT, UMLS, ICD ....can link to these definitions also?.......), thesauri (MeSH, National Agricultural Library Thesaurus), folksonomies (Social bookmarking tags like example X, Y, Z), general ontologies (WordNet, OpenCyc) and specific ontologies (Process Specification Language). The list also includes markup languages (NeuroML), representation formalisms (Entity-Relation model, OWL, WSDL-S) and appropriate ISO standards (ISO 11179). This sample clearly illustrates the diversity of artifacts collected under "ontology". (10NU)

Conclusions (10QI)

The Ontology Summit 2007 "Ontology, Taxonomy, Folksonomy: Understanding the Distinctions" collaboratively identified previously undefined, important dimensions whereby ontologies could be characterized and measured. It resulted in a Framework including six major dimensions, as illustrated in this diagram.

An interactive survey was realized. In the face-to-face meeting, the Framework was put to a stress test when participants used it to position a dozen ontologies along its dimensions. (10QK)

<...can you also show Denise Bedford's drawings as a sample ?>

We recognize the current Framework is preliminary. Specifically, work needs to be pursued to refine the dimensions to establish operational definitions and performance requirements for populating the Framework. We encourage professionals from various disciplines to contribute to this work by joining the Ontolog Forum Community of Practice. (10QJ)

Endorsed

Deborah MacPherson

Projects Director, Accuracy&Aesthetics

Specifier, WDG Architecture