Introduction   (10N9)

Under the appellation of "ontology" are found many different types of artifacts created and used in different communities to represent entities and their relationships for purposes including annotating datasets, supporting natural language understanding, integrating information sources, semantic interoperability and to serve as a 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 some 50 institutions, is an attempt to bring together various communities (computer scientists, information scientists, philosophers, domain experts) having a different understanding of what is an ontology, and to foster dialog and cooperation among these communities.   (10NB)

In practice, the name ontology covers a spectrum of useful artifacts, from formal upper-level ontologies expressed in first order logic (e.g., Basic Formal Ontology (BFO) and DOLCE) to the simple lists of user-defined keywords used, for example, to annotate resources on the Web. The latter are called "folksonomies" and play an important role in the Web 2.0. In between the two extremities of the ontology spectrum are taxonomies and controlled vocabularies (e.g., MeSH), often used for information indexing and retrieval, and whose organization is mostly hierarchical. Finally, there are ontologies which represent not only subsumption , but also 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 is not to establish a definitive definition of the word "ontology", which has proved extremely challenging due to the diversity of artifacts it can refer to. 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 along which ontologies can be characterized and to provide operational definitions for these dimensions. The relative position of ontologies in the space defined by these dimensions, the "Framework", is indicative of the similarities and differences between these 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)

One major goal of the Ontology Summit 2007 was to bring together the various diverse communities working on ontology-like activities so as encourage cooperative efforts. Toward this end the summit has attempted to characterize what is an ontology, e.g., to construct a typology of ontologies. The framework of dimensions is comprised of two groups: semantic dimensions and and pragmatic dimensions. Semantic dimensions include expressiveness, structure, and representational granularity. Pragmatic dimensions include intended use, use of automated reasoning, and prescriptive vs descriptive. See diagram.   (10NH)

Expressiveness is a property of the knowledge representation language which describes the extent and ease with which the KRL can describe increasingly complex semantics, cf. propositional logic, description logic(s), first order logic, sorted logics, modal logics, ...   (10NI)

Structure is a property of the ontology, which records how elaborate (or well organized) are the semantics encoded by the ontology. It may be the same as the expressiveness of the KRL in which the

ontology is encoded, or it may be less the expressiveness of the knowledge representation language. Thus a simple taxonomy, e.g., a tree, may be encoded in RDF/S, a description logic language such as OWL-DL, or first order logic, e.g., Common Logic. Viewed from a graph theoretic perspective level of structure might be either a simple set of terms (glossary), a tree structures (taxonomy), a directed acyclic graph, e.g., a partial order (faceted classification schemes), or an arbitrary directed graph (e.g., RDF).   (10NJ)

The granularity dimension concerns the level of detail at which the ontology is specified. A crude measure of granularity measure would be the number of concepts (nodes) and the number relation instances (links or edges in graph representations). However, this fails to recognize that some ontologies may have larger scopes (domains) than others. A coarse grained ontology might be suitable for use as an upper ontology, or a 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 which records the orginal purpose(s) of the ontology. These may include semantically informed search, data semantics specification for databases or data entry, data integration across multiple data sources, agent communication languages, controlled vocabularies for recording medical diagnoses, etc.   (10NL)

Automated reasoning is a dimension which records the extent to which it is anticipated that an ontology will be used by automated reasoning software, e.g., for question answering, etc. If so, then one would expect that the ontology would likely be encoded as using some form of logic, e.g., first order logic. (10NM)

Prescriptive vs. Descriptive is a dimension which characterizes whether the intent of the ontology developer is simply to describe contemporary semantic usage without much regard as to the scientific correctness of the encoded knowledge (e.g., a whale might (in common parlance) be described as a large fish. Examples of such descriptive ontologies include folksonomies and most linguistic ontologies. Alternatively, an ontology may be intended as a normative prescriptive document whose correctness is 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, etc.   (10NN)

Additional findings   (10QE)

Governance   (10QF)

The governance dimension addresses how decisions concerning the structure and (esp.) content of an ontology are made. There was agreement at the summit that ontology with legal or regulatory implications will need to defer to existing legal, regulatory, and professional organizations concerning the natural language definitions of entities and semantic relationships. Ontology development should be viewed as an effort to organize and formalize concept definitions and relationships which are conventionally defined by existing institutions, not as an attempt to replace existing definitions with de novo definitions generated by autonomous computer scientists. As a corollary, it was observed that it is necessary to record the provenance of every definition, etc. incorporated into an ontology, e.g., the controlling legislation, regulation, standard, etc. from which a definition is taken.   (10PB)

Folksonomies and Formal Ontologies   (10NO)

One of the issues discussed was the relationship between social tagging and folksonomies and more

traditional structured / formal ontologies such as taxonomies and axiomatized ontologies. Until recently these efforts have been viewed as competitive approaches. The consensus of the Ontology Summit was that social tagging efforts should be viewed as large scale corpora to be used for inferring and validating more formal ontologies, akin to the use of large text corpora in computational linguistics studies. In addition, more formal ontologies can be used to inform social tagging by providing improved tag sets, and faceted tagging.   (10NP)

Ontologies as designed artifacts   (10NQ)

Some members of the community argued that ontologies could be considered a type of designed artifact, and that ontological engineering should be thought of as a discipline complementary to software engineering and to virtually any discipline dealing with data and information exchange. We expect that this discipline will increasingly become a standard part of the relevant curricula.   (10NR)

Design methodologies   (10QG)

Another finding of the summit, after looking at a large number of intended uses for ontologies, is that there is also a spectrum of design methodologies. They vary from a strong software engineering design lifecycle with requirements, evaluation, and verification, all the way to a "no-design" methodology in which folksonomies emerge from the local behavior of thousands of individual users. Design methodologies are strongly related to the intended use. For instance, methodologies for managing controlled vocabularies and taxonomies are very social in nature and are intended to capture generalities about the meanings of words in a culture or domain. Also, the "verification" of ontologies is related to the role of reasoning or types of computational services to be enabled by the ontologies. For example, if an ontology is used for data integration, the verification of the consistency and completeness of data metamodels are important, whereas in a domain where the meanings of terms has legal consequences, verification is more about capturing the provenance of the design choices, rooted in authority or appropriate process.   (10QH)

Survey   (10NS)

In order to elicit the distinctions between various kinds of ontologies, an interactive survey was designed and posted on the Web in order to engage various communities. The respondents were invited to identify the community of which they are a representative and to describe the value of ontologies, as well as issues with ontologies in this community. The last section of the survey invites the respondents to describe and characterize the ontologies or related artifacts in use in this 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. Some 70 ontologies from a variety of domains were characterized in the survey, including formal ontologies (e.g., BFO, DOLCE, SUMO), biomedical ontologies (e.g., Gene Ontology, SNOMED CT, UMLS, ICD), thesauri (e.g., MeSH, National Agricultural Library Thesaurus), folksonomies (e.g., Social bookmarking tags), general ontologies (WordNet, OpenCyc) and specific ontologies (e.g., Process Specification Language). The list also includes markup languages (e.g., NeuroML), representation formalisms (e.g., Entity-Relation model, OWL, WSDL-S) and various ISO standards (e.g., 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" was an attempt to collaboratively identify important dimensions whereby ontologies could be characterized. It resulted in a Framework including six major dimensions, as illustrated in this diagram. An interactive survey also was realized to contribute to the description of existing ontologies. In the face-to-face meeting, the Framework was put to a stress test: the participants used it to position a dozen ontologies along its dimensions.   (10QK)

We recognize that the current Framework is still preliminary. In particular, work needs to be pursued on refining the dimensions and establishing operational definitions 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 by: (10OX)

- 33 individuals at the symposium endorsed the communique as finalized above. They did so as individual (and not as representatives of the organizations they are affiliated with.) The list of names will be posted shortly.   (10OY)
- We are inviting everyone on the [ontology-summit] list and those who had contributed to the summit proceedings (virtual discourse and/or at the face-to-face meetings) to confirm their endorsement of this Communique. Please review the above communique and confirm your endorsement by e-mail to <peter.yim@cim3.com> on/before Apr. 30, 2007 (so that we can add your name to the list too.)   (10QN)

This Communiqué (v 1.0.0 / 2007.04.24) was review, collaboratively edited, finalized and adopted by individuals present at the OntologySummit2007_Communique_Session.   (10QT)

Resources: (10PH)

- [OntologySummit2007_Communique/Draft First draft] proposed by OntologySummit2007_Communique the co-chairs (OlivierBodenreider & FrankOlken) - /Draft (10PI)