ONTOLOGY : THE WAY OUT OF WORLD WIDE "WEB"

Aparajita Suman

Keywords : Ontology, Information Retrieval

1. Introduction

The problem of today's world in not the dearth of information rather it is the absence of reliable tools for retrieval of relevant information. The reasons are the unique features of Internet domain such as its size, distributed nature and rapid evolution. Most of the search engines can index and search the documents, however, the different available techniques (exact word search, truncation search, and Boolean search, proximity search, and other techniques like fuzzy, soundex, ranking of query results, case sensitive/ insensitive etc.) are nowhere near the way information gets organized and interpreted by human mind. The ideal approach will be to appreciate and understand the way information gets logically organized in human mind.

Of late, the concept of ontology has been borrowed from philosophy and is being suggested as a mechanism for denoting the existence of a document on the World Wide Web. But, ontology *per se* doesn't suffice to express the problems associated with the retrieval of information from web. The technology needs to be redefined for the purpose of knowledge organization on WWW and digital libraries.

2. Ontology

The concept has been derived from philosophy; ontology is the study of what exists and what we must assume to exist in order to achieve a cogent description of reality. But the same definition is now being applied to define the existence of information resources on WWW.

a. Concept and theory

The reference back to Newell **[1]** is encountered frequently in discussions of ontology in knowledge engineering. In a presidential address to the American Association for Artificial Intelligence, he considered computer system levels, e.g. currents and voltages at the circuit level, changing to bits at the logic level, to data structures (variables, arrays, etc) at the symbol level. He regarded the representations in knowledge bases (production rules, frames, semantic nets, etc) as being higher structures at the symbol level: above them was a knowledge level, comprising the conceptual knowledge that was embodied in the representations. The same is evident in Ranganathan's distinction between the 'idea' level (corresponding to Newell's knowledge level), the 'verbal' level (the verbal expression of a concept) and the 'notational' level (the symbolic representation of the idea or concept in classificatory language).

Finally the concept of Ontology emerged from different disciplines of study and it was concluded that 'the study of ontology, as a branch of philosophy dealing with the nature of reality, can be of benefit to the knowledge construction process in yielding high-value knowledge bases' **[2].**

Definition : Uschold **[3]** writes that an ontology: . . . is often conceived as a set of concepts (e.g. entities, attributes, processes), their definitions and their inter-relationships . . . An ontology may take a variety of forms, but will necessarily include a vocabulary of terms and some specification of their meaning (i.e. definitions). It may be:

- highly informal, expressed in natural language,
- semi-formal, expressed in a restricted and structured form of natural language,
- semi-informal, expressed in an artificial formally defined language,
- rigorously formal, with meticulously defined terms with formal semantics and theorems.

b. Tools and techniques

Although sufficient number of ontology tools and techniques are available but they are not very popular for the lack of defined standards. Defining the criteria for the design and evaluation of ontologies, the methods and tools for their development, is the only way out of the present web of information where the hapless user gets caught.

3. Building ontologies

Various researchers have given different methods of building ontologies. The method proposed by Uschold and Gruninger **[4]** is being discussed here:

- The domain to be covered by the ontology must first be decided, and a specification document may be drawn up. Concept terms are collected by scanning the literature of the domain, and by consulting domain experts.
- The collected terms may then be grouped, perhaps according to work areas in the domain, 'such that terms are more related to other terms within a group than they are to terms in other groups'.
- Next step is of producing definitions. Here, the idea of a 'meta-ontology' is useful. Meta-ontology is a statement of the categories of term required in the ontology. It is recommended that one should not decide on a particular meta-ontology in the beginning rather it should be finalized only after careful consideration of the concepts and their interrelations.

Following are the basic principles for 'ontological engineering' **[5]**:

1. Not unique: There is no unique ontology of the world or even of a narrow domain. Ontologies are not natural entities to be discovered, they are artifacts to be constructed and their structure depends on their intended use.
2. Task specific: An ontology is always built for a specific task meaning thereby an ontology built for natural language processing may not be suitable for some other task such as planning, design, or other reasoning. The type of knowledge that ontology contains differs from task to task.
3. Usability: It must be easy to browse the ontology, to find the right concept to map a word and to see the definition/description of that concept in its entirety. This should help in visualization of the organization (taxonomic or otherwise) of that concept relative to other concepts in the ontology.
4. Modularity: It must be possible to add new concepts and conceptual relations incrementally.
5. Grain size: Fine-grained decomposition of concepts is the basic element for building a precise ontology where each concept is clearly distinguishable from others. The choice of granularity must be determined on the basis of utility to the task involved.

6. Redundancy: The classification of concepts in an ontology is necessarily redundant. Multiple dimensions of classifying a set of concepts are often overlapping. It is futile to try to eliminate such ontological redundancy entirely from ontology of non-trivial size.

4. Application areas

An ontology may be regarded as a database with information about what categories and/or concepts exist in the world/domain, what properties they have, and how they relate to one another. The principal reasons for using an ontology could be to provide a grounding for representing **[6]**:

- Text meaning in an interlingual setup;
- To enable lexicons for different languages to share knowledge;
- To enable source language analyzers and target language generators to share knowledge; and to resolve semantic ambiguities and interpret non-literal language by making
- Inferences using the topology of the ontology to measure the semantic affinity between meanings.

An ontology can also be of great value in a variety of other tasks such as database merging or integration of software or business enterprise models. Essentially, an ontology is invaluable wherever a 'semantic wall' is to be scaled, i.e. a situation where there are two or more systems that overlap conceptually, but have different knowledge representations. Thus, it may be used to translate between a pair of natural languages or a pair of database schemas, or to integrate different models of the same domain or similar phenomena in the world.

Ontologies are believed to provide a sufficient common structure for application in following primary areas:

- Natural language understanding and generation;
- Semantic database integration, consistency-checking, and data mining;
- Semantic information retrieval;
- Ontology-constrained simulation;
- Building and utilizing user models;
- Knowledge sharing by groups working independently.

4.1 Adding Semantics to Web using Ontology

The theory of ontology can help in studying the reusable knowledge components and semantically marking up web pages using terms from an explicit ontology. This will not only improve retrieval but will help in integration of the data from many pages at the same time. Further, one needs to study the way information organization is being practiced in the libraries world over also, the same concept can be applied in the case of digital libraries/ depositories/ portals also.

This workbench should not be created from scratch, but instead by integrating the technology components that are currently available.

5. Conclusion

To summarize, new development of tools at the 'knowledge level' shows the growing understanding of the importance of semantic analysis in information processing. This could lead to solution to the problems with which information scientists have for so long been struggling. It is to be hoped that all involved will

continue to learn from each other's experience. The cooperative venture could lead to 'an open, multifunctional, multilingual system for integrated access to knowledge about concepts and terminology'. This will provide a helpful integrating background to work at the knowledge level.

6. Reference

1. A. Newell, The knowledge level, Artificial Intelligence 18 (1982) 87–127.
2. N. Guarino, Formal ontology, conceptual analysis and knowledge representation, International Journal of Human-Computer Studies 43 (1995) 625–640.
3. M. Uschold, Building Ontologies: Towards a Unified Methodology. Presented at Expert Systems '96 Conference and available from Internet site: ftp://ftp.aiai.ed.ac.uk/pub/documents/1996/
4. M. Uschold and M. Gruninger, Ontologies: principles, methods and applications, Knowledge Engineering Review 11(2) (1996) 93–136.
5. K. Mahesh, Ontologies for Natural Language Processing. Available from Internet site: http://cr1/nmsu.edu/users/mahesh/
6. Vickery, B.C. Ontologies. University College London, UK. Journal of Information Science, 23 (4) 1997, pp. 277–286 Available from Internet site www.let.uu.nl/~Paola.Monachesi/ personal/papers/vick97-onto.pdf
7. Opdahl, A.L. and G. Sindre, A Taxonomy for Real-World Modeling Concepts. Information Systems, 1994. 19(3): p. 229-241
8. Oscar, Corcho et al. Methodologies, tools and languages for building ontologies. Where is their meeting point? Data & Knowledge Engineering 46 (2003) 41–64.
9. Wand, Y. and R. Weber. An ontological analysis of systems analysis and design methods. in Proceedings of the IFIP WG 8.1 Working Conference on Information Systems Concepts: An In-Depth Analysis. 1989:Namur/Belgium, Amsterdam/North-Holland
10. Wand, Y. and R. Weber, An Ontological Model of an Information System. IEEE Transactions on Software Engineering, 1990. 16(11): p. 1282-1292

About Author

Ms. Aparajita Suman is presently working as Information Officer, Virtual Information Center, ICICI Knowledge Park, Hyderabad. She has served DRDL, Hyderabad and Rajiv Gandhi University of Health Sciences, Bangalore also in the capacity of JRF and Project Assistant respectively. Ms. Aparajita is M.Sc. in Zoology and ADIS (Associateship in Documentation and Information Science) from DRTC, Indian Statistical Institute, Bangalore.

E-Mail : aparajita@icicikp.com