Academic Knowledge Ontologies and a Systems Solution

Akifumi Tokosumi, Naoko Matsumoto, Miho Tomioka, and Klaus Voss
({akt, matsun, miho0903, voss}@valdes.titech.ac.jp)
Department of Value and Decision Science, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro, Tokyo 152-8552, Japan

Abstract

Two academic areas (a) cognitive science—an established but actively changing field—and (b) large-scale knowledge resources—an emerging discipline—are chosen as targets for academic domain ontology design research. Among the findings obtained from designing domain dictionaries readable for both humans and machines, this paper discusses (a) the interfacing of local knowledge to global knowledge, (b) the creative nature of academic concepts, and (c) an agent-based approach to support the creation of academic concepts.

Domains of the Ontologies and Research Goals

In this paper, we describe findings and insights concerning the nature of knowledge within academic domains that we have gained through our work on constructing two academic ontologies:
(a) cognitive science—an established but actively changing field, and (b) large-scale knowledge resources—an emerging discipline.

The goals of our ontology construction work, conceived of as a human and machine readable collection of lexical items and concepts, are:
i) to support communication between researchers in related fields,
ii) to support new students, and
iii) to support translation.

We summarize our findings into the characteristics of academic knowledge.

Locality and Globality of Academic Knowledge

Theory dependency is an especially prominent feature of concepts within the two areas under consideration. While it is almost impossible to provide a single universal definition for any concept, we found that it is possible to find working definition for practical purposes, once a theoretical perspective is identified. Because, in general, scientific papers include a large number of implicit assumptions that depend on a particular theoretical viewpoint and social background which are rarely ever stated explicitly, it requires domain expertise to find a suitable definition for the field (Tokosumi, et al., 2006).

This is especially problematic in two typical cases: (i) communication between researchers in neighboring fields where shared knowledge may not be sufficient, and (ii) in new students. Analyzing both (i') the interdisciplinary gap, and (ii') the expertise gap, we have found that both can be reduced in terms of the locality of academic concepts.

The Private and Public Nature of Ontologies

It is essential for an academic ontology to appropriately factor in the bidirectional nature of the concept revision process. Scientists are, by definition, innovators of concepts. The process of proposing new or updated concepts varies at several levels, such as the individual researcher, the laboratory, scientific communities, and scientific journals. In this process, concept definitions need to be circulated simultaneously within subsets of the scientific community and within groups of individual researchers.

In this paper we address this process by proposing a framework to support the creation of both private user- and common group-dictionaries or -documents with the purpose of developing a public concept, while simultaneously maintaining a private (individual) concept definition.

Both cognitive science and large-scale knowledge resources are both fundamentally interdisciplinary research fields. By its very nature, interdisciplinary research must handle varying interpretations from different communities. This means that interdisciplinary research provides an ideal context for examining the issues of domain specific language usage in different communities, because a specific meaning cannot be defined unless the points of reference are known, which will vary between communities.)

Ontology for Comprehension and Creation

We propose two definitions, to differentiate between ontologies for creation and ontologies for understanding:
i) the understanding/interpretation problem: when a concept is already established in the literature,
ii) the creation problem: when a new research concept is proposed.

Ontology Construction Environment

The development of an ontology is an ongoing process without precisely defined termination criteria (Noy & Musen, 2000). Accordingly, flexibility is a crucial issue when an ontology is being developed for a heterogeneous or simply large community. This is particularly true in the case of academic research, where the specific domains are constantly evolving due to scientific advancements. The attempt to fix a self-contained depiction of a given field of research is an oxymoron, because no ultimate authority exists capable of declaring the current state of the art. Thus the guiding principle in developing an ontology for a scientific domain should be not to impose artificial constraints on the scientific community.

From Knowledge Base to Ontology

The actual building of exhaustive ontologies leads to well known technical problems, related to:
- terminology, when an established nomenclature does not exist in the field, or has to be mapped between related fields (Nature Editorial 1997, Nature Opinion 1999);
- the scope of the scientific domain in question; and
- the encoding of formal representations in the ontology.

These issues can only be resolved through a process of intense communication between the potential users of the ontology.

To support this process, we propose an agent-based system, which simply seeks to synchronize the development of user-defined ontologies at various levels, such as individual researcher, group (by interest and/or affiliation), scientific community and which depicts the as-is-state of a discussion at any point in time.

Existing systems for ontology development impose artificial constraints to the extent that they require the user to learn to build and edit a specific formal representation, which is static in nature. This poses a problem particularly for students new to the field, and the process is also time consuming and does not scale well for large communities.

Representation of Concepts by Agents

The proposed system organizes the building of an ontology in the form of a network of concepts and categories. Each concept is represented by software agents, and a concept-agent is defined by:
- a label (concept name or category);
- keywords (belonging to the category);
- a set of documents (papers, book-chapters etc.) belonging to the category and describing the keywords;
- an automatic generated classifier.

Although labels, keywords and documents are in natural language, the underlying agent technology constructs a machine-readable classifier (e.g. by naïve Bayes algorithms). Agents of different users can autonomously scan and compare their concepts and on recognition of a similar or related agent propose to the user to modify their concepts (modify keywords, set of documents), to merge agent with other agent, or to reject the other agent.

This procedure allows for the circulation of concepts almost in real-time. Each user can trace back on whether an agent-/concept has been derived from existing concepts and whether it is an refinement or a new, alternative definition. In this way, new users can start with a small set of broad definitions, and proceed to find related sets of agents trained by expert users.

Agents can be completely anonymous in the sense that the process of ontology-building and comparison of concept-design resembles the Delphi-process (Malsch, Luehrs, & Voss, 2001). Due to the underlying agent technology, the process scales to large numbers of users and documents. Likewise, the system greatly enhances the awareness of users for evolving concepts in their field, since the agents can scan other agents and notify the user whenever they detect related agents. Users, groups and communities can easily maintain different versions of concept-agents for each category or domain.

Finally, agents can be assigned to represent a class or subclass respectively, in order to establish a network of categories and concepts with a shallow hierarchy as an implementation of a dynamic ontology.

Given an initial knowledge base in form of a set of documents, and an initial, small set of broad, undisputed master- categories, the community can proceed to copy, edit and create new subcategories, assign concepts (agents) to these categories, and subsequently exchange and refine their own personal concept definitions. New documents that are not initially part of the knowledge base can be added as personal concept-agents and enter the process of exchange with other personal agents. The exchange and reuse of user agents facilitates the sharing of knowledge through user refinement of agents.

As a result, each user maintains a refined version of the initial knowledge base, which is organized in the form of an ontology reflecting the user's perspectives and containing modifications based on interaction with other user with trusted expertise. Such user ontologies can be consolidated into a single ontology by selecting agents from among all users that represent similar concepts.

Acknowledgements

This research is part of the "Framework for Systematization and Application of Large-scale Knowledge Resources" (LKR) 21st Century COE Project at the Tokyo Institute of Technology. The authors would like to thank Professor Furui and the COE21-LKR for their continued support.

References

Malsch, Th., Luehrs, R., & Voss, K. (2001). Internet, Discourses and Democracy. Terano, T. et al. (eds): *New Frontiers in Artificial Intelligence. Joint JSAI 2001 Workshop Post-Proceedings.* Heidelberg: Springer-Verlag. 67-74.

Tokosumi, A., Joyce, T., Matsumoto, N., Noda, K., & Murai, H. (2006). Cognitive Principles for a Sensibility-Aesthetics Ontology. *Proc. for the Second International Symposium for Large-Scale Knowledge Resource.* 21-22.

Noy, N. F. & Musen, M. A. (2000). PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment. *AAAI/IAAI 2000,* 450-455.

Nature Editorial. (1997). Obstacles of Nomenclature. *Nature,* 389:1-1.

Nature Opinion. (1999). Wanted: A New Order in Protein Nomenclature. *Nature,* 401: 411.