Abstract

“Ontology is Overrated: Categories, Links and Tags” by Clay Shirky basically discuss about categorization issues. It thinks a lot of what people think about categorization is wrong, especially in the electronic world, such as Internet. It declares that ontology is overrated and the world is moving from binary categorization and into the probabilistic world.

Definition of Ontology

In the philosophical sense, ontology is the study of entities and their relations. Then the knowledge management and AI communities have a related definition, which is something like “an explicit specification of a conceptualization.” Ontological classification or categorization’s strategies are to cover possible cases in advance based on what people are primarily concerned with or the thing is widely used. For instance, library is best-known categorization system. If people own more than a few hundred books, they have to organize them somehow. But however they do it, the physical fact of books make some sort of organizational scheme a requirement, and hierarchy is a good way to manage physical objects. For example, a book can be about several things at once, but it has to be declared to be about some main thing.

Ontology is overrated

However, ontological classification or categorization is badly overrated in terms of its value in the digital world. In the real world things do not have clear edges. For instance, book can be entertainment or humanities. Yahoo has used a hierarchy to categorize the things. Yahoo decided to put book under humanities category, because they wanted to make assertions about what is “real”. Therefore, people are told which should be part of which. If people think book is entertainment, they may not find book there. In order to solve this problem,
Yahoo add extra link of book which should be under humanities. Yahoo insists on its view to the real world. In its hierarchy structure the elements within one category may link to others, but the effect was to override the users’ sense of what things ought to be. If people want something that has not been in the way Yahoo think about it, they are out of luck.

As mentioned before, library has physical constraints to classify and organize the books. However, on Internet there is no shelf. If there are enough links to all others in the hierarchy, the hierarchy is not needed anymore. Compared with Yahoo, Google decides what goes with what after hearing from the user, rather that trying to predict in advance what is people need to know. That means nobody gets to tell users in advance. At the moment that users are looking for something, Google will do its best to service it based on link structure.

**Ontological classification can be a good strategy**

However, ontological classification still can be a good strategy if the system has certain characteristics, such as small corpus, simple ad derivable categories, stable entities, and clear edges. The increasing complexity of problems addressed by massively distributed information systems has led to the development of a new generation of worldwide information systems. An ontology is a partial conceptualization of a given knowledge domain, and it has been defined in a formal, machine-process able language for the explicit purpose of sharing semantic information across automated systems, which can be shared by a community of users. It can enable systems and agents to understand the content of a web source and to integrate that understanding with the content of other resources. Therefore, actually ontology is a unique representational system that integrates within a single structure the characteristic of more traditional approaches, such as nested hierarchies, faceted thesauri and controlled vocabularies. Also ontology offers the semantic basis for metadata schemes and
facilitates communication among systems and agents by enforcing a standardized conceptual model for a community of users (Elin K 2003, p.19).

Contrast, the web have largest corpus, most naïve users and no global authority, so usually ontological classification does not work well there. Some of biggest problems are that categorists guess what their users are thinking and make predictions about the future. In order to avoid these problems, the systems can use tagging, which can enable a huge amount of user-produced organizational value at vanishingly small cost. Instead of paying a professional tag data once and once again, individuals are given the access to tag their stuff, which generates a lot more data. The merges from URLs are probabilistic, but not binary.

As the developments with ontology, it will be used primarily on the domains that contain the most valuable knowledge, such as the medical and business sectors. ONTOWEB has implemented an ontology-based web retrieval system for the web resources of international organizations such as the World Bank and the Organization for Economic Co-operation and Development (OECD). The study shows that ontology can be used not only to improve precision, but also to reduce the search time (Hyun H 2005, p.1167). Ontology provides us the best knowledge of the real world, but it can be very expensive. We need expert catalogers, authoritative source of judgment, coordinated users and expert users. The bigger the user base is, the more work that user education is.

**Conclusion**

In conclusion, catalogers need to think about what the real world is and what the strategies of the organizations are. It is not a simple question. In order classifying the things by the best strategy, the characteristics of the system need
to be understood. Also catalogers need to be aware that they may underestimate the difficulty of understanding what users are thinking, and they may overestimate the amount of which users will agree, either with one another or with the catalogers about the best way to categorize.
Bibliography

Elin K Jacob Apr/May 2003 *Ontologies and Semantic Web* Bulletin of the American Society for Information Science and Technology Vol.29, Iss. 4; Pg. 19

Hyun Hee Kim Sep 2005 *ONTOWEB: Implementing an ontology-based Web retrieval system* Journal of the American Society for Information Science and Technology Vol.56, Iss. 11, pg. 1167