Ontology for People
- Vagueness and multiplicity -

Hideaki Takeda[1,2]

[1] National Institute of Informatics
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
takeda@nii.ac.jp
http://www-kasm.nii.ac.jp/~takeda/
[2] Research into Artifacts, Center for Engineering (RACE), The University of Tokyo
Kashiwanoha 5-1-5 Kashiwa Chiba 277-8568, Japan

ABSTRACT. In this paper, I argued ambiguity of representing concepts. Ontologists in computer science, especially in Semantic Web dismissed the importance of variety of conceptualization. But richness of our conceptualization is crucial for representing ontology and can contribute various use of ontology in computer systems. In particular I picked up multiplicity of ontology in this paper. I show two types of treatment of multiplicity in ontology. One is representation of multiple and shared ontologies by combining fragments of ontologies. The other is representation of multiplicity of concept. Variety of concept representation by people are aggregated and represented as a *virtual concept*.

1 Introduction

Ontology in information systems becomes so popular mainly thanks to Semantic Web. It started from Tom Gruber's introduction of "ontology" in artificial intelligence [Gruber 1993]. He intended ontology as a basic layer of knowledge in the context of knowledge sharing, in particular knowledge sharing among computer agents. He defined ontology as "A specification of a representational vocabulary for a shared domain of discourse — definitions of classes, relations, functions, and other objects — is called an ontology." It is in spirit a return to traditional approach in artificial intelligence like [Genesereth *et al.* 1987]. We cannot represent grand truth in computer as the traditional AI approach presumed implicitly, but we can deal with truth-like representation on condition that people *agree*. Its simplicity and orthodoxy are embraced by AI researchers, and so are Web people later on. Now ontology languages RDFS and OWL become one of standard languages for Web.

Spreading out Semantic Web concept and its applications revealed that simplicity in Gruber's definition does not fit concept sharing in the real world, in particular in a large scale like Web. I believe that it comes from vagueness and multiplicity.

In this paper, I firstly discuss what concepts people want to express and share to each other by observing social software systems on Web. Then I show my approach how to deal with it with computer, in particular representation and interpretation of multiple ontologies.

2 Semantic Web and Social Software Systems on Web

Frankly speaking, it looks difficult that Semantic Web applications become popular. Aside from Semantic Web, social software like social tagging becomes so popular and recognized one of the important services on Web.
For example, people share their tags in social tagging systems. We can say that people share concepts through such systems. But sharing is very loose. What they can do when adding a new instance for a tag is checking the labels of the tag and understanding the existing instances for the tag at most. Even they skip to check existing tags but just add new tags and instances by themselves. So mixture of concepts in a single tag often happens like “apple” tag both for fruit and a computer company. Despite such flaws, social tagging systems are accepted as information sharing tools by people.
The other example of social software system is Wikipedia, volunteer-based encyclopedia. Everyone can create a new article and edit any articles. There are some interesting points from ontology viewpoint. Here we can assume that an article and a category represent an instance and a class respectively. Each article is identified only by its name. There is no need to provide definition of categories or classes in advance. Categories can be assigned to articles but it is not necessary and can be assigned more than one. Some groups of articles have common formats called “infobox” which represents something like attribute and its value. But it is also arbitrary.
People wish to share concepts but definition of concepts in Semantic Web way does not fit with what people want to express and share.
Our recognition of concept is very vague. We can indicate a concept but it does not imply that we can define or even represent it. There are many different points in comparison to concept in ontology in computer systems. I pick up two sub problems here.
One sub problem is *the problem of definition*. Except scientific or engineering concepts, concepts exist without or preceded their definition at most. In social tagging systems, definition is abandoned. Instead only extension is given, i.e., a set of instances (URLs). In Wikipedia, definition follows the existence of articles like adding categories and infobox.
The other problem is *multiplecity of concept*. Even if concepts are definable, it does not imply the definition is unique. Instead it is natural to have multiple definitions for a single concept, since concepts can be used in different contexts. In particular, we are assuming concepts for sharing among people, difference of context should be taken into accounts. In social tagging systems, people do not matter if there is disagreement or inconsistency of usage of individual tags. Rather they enjoy vague recognition of tags because it can aggregate more information. In Wikipedia, multiplicity of concept is acceptable but should be resolved in some way. The name multiplicity is solved by naming convention such as adding parenthesis expression. Multiplicity of class hierarchy is represented by multiple categories. Multiplicity of the content itself is represented as the structure of page after discussion by people.
In Semantic Web, the importance of these problems has been dismissed since it started from simplicity of ontology definition by Tom Gruber. Multiplicity of ontologies is now becoming a hot topic such as ontology alignment or ontology matching. But still ontol-

ogy is believed to be simple and unique ideally behind ontology alignment and ontology matching.
I believe that multiplicity of concept is necessary to represent richness of our conceptualization and can contribute various use of ontology in computer systems.
In the following sections, I introduce our attempts to deal with multiplicity. The first attempt is done before the rapid rise of Web [Takeda *et al.* 1995] and the second is our on-going research topic [Shakya *et al.* 2008].

3. Multiple Ontologies for Agents

Knowledge sharing with computer agents was a possible solution before the rise of Web. In this approach, heterogenous agents communicate to each other to solve problems. KIF [Genesereth *et al.* 1992] and KQML [Finin *et al.* 1995] are defined for inter-agent communication languages. Although KQML performative has an attribute to specify ontology, ontology in a multi-agent system is shared without inconsistency. But it is natural that heterogenous agents use heterogenous ontologies. Then the problem is how to realize communication between agents with different ontologies.
We tried to solve this problem within the logical treatment as an extension of Ontologinua [Gruber 1992]. We divide the logical theory which represents ontology into fragments of the logical theory (called *aspect*) and represent heterogenous ontologies with combination of fragments of the logical theory.

3.1 Multiple Aspects

It is often assumed that a single ontology is shared among agents. But it is not easy to build large ontologies in fact. One of the reasons is that we often confuse concepts from different conceptualization when we try to build large ontologies. For example, Figure 1 shows how concept "temple" is modeled differently. One may think temple as item in textbook of history, so "founded year", "sect", "historical events", and so on are used with "temple". One may think temple as place for religion, so "doctrine", "parishioners", "branch temples", "chief priest" are concepts related to "temple". Of course, some concepts are appeared in more than one aspect like "name" and "location".
Mixture of concepts from different conceptualization often confuses us. Some different concepts in different aspects may denote the same fact, for example "chief priest" in *aspect-for-temple-as-religious-place* and "representative" in *aspect-for-temple-as-public-organization* may be the same. On the other hand, concepts used in more than one aspect may denote different facts, for example "name" is used in common, but its meaning is either religious name or historical name or official name with according to aspect it is used. In such cases, the more concepts are collected, the less clear are meanings of concepts.

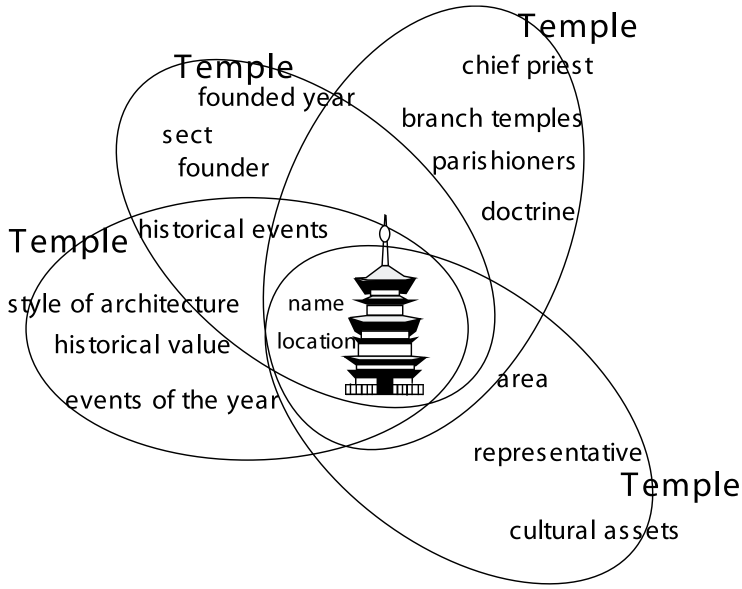

Fig. 1: What concept is good for "temple"?

Each concept is meaningful if and only if concept is used in proper way, that is, concept is used with concepts which come from the same conceptualization. We call this unit {\it aspect}. We can say that an aspect is a consistent view for conceptualization.
Then ontology can be composed of some aspects.
We use various aspects, for example in engineering we use aspects like dynamics aspects, kinematics aspects. To model the commonsense world of traveling, we may use aspects like traffic aspect or geography aspect.

3.2 Representation of Multiple Aspects

There are two issues on aspect. One issue is what should be in aspect. Aspect should have a vocabulary to describe phenomena in its domain. It should also have a theory which associates concepts in its vocabulary. And the theory should be consistent. In the other words, aspect is what we can conceptualize the world without inconsistency.
The other issue is how to compose aspects from other aspects. We provide two types of basic connections among aspects. One is *combination* aspect. This is just integration of aspects for different domains. For example, one of the ways to build *aspect-for-travel* is to combine *aspect-for-hotel* and *aspect-for-traffic*. In this case, concepts like "railway" which is in aspect-for-traffic do not exist in aspect-for-hotel, because domain of modeling is different from each other. In aspect-for-travel, concepts like "tour" are defined using concepts from both aspects.
The other is *category* aspect. This is collection of aspects which share domains but come from different conceptualization.

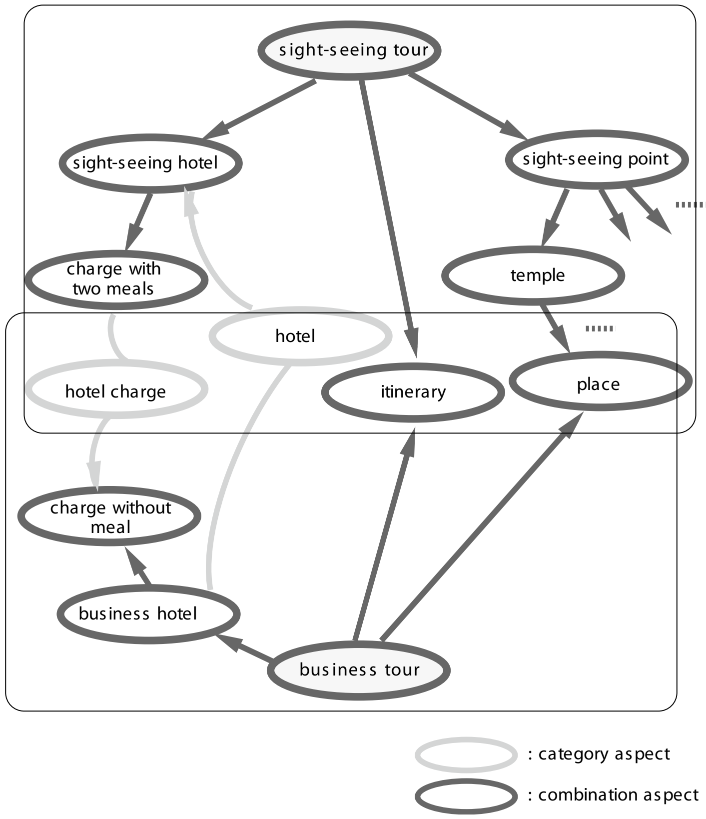

Fig 2: Multiple and Shared Ontologies

When a temple is modeled differently we have just shown, we can assume there is a *category-aspect-for-temple*. This aspect has some specific aspects for temple like *aspect-for-temple-as-history-textbook* and *aspect-for-temple-as-religious-place* as component. Since component aspects share domains, it is reasonable (but not mandatory) that there are relations among concepts in different component aspects. Such relations are contents of the category aspects.

Since combination and category aspects can use other combination or category aspects as component, we can construct large aspects from relatively small aspects. We call such relatively large aspects as ontologies.

Figure 2 is an example how different ontologies can be defined with sharing aspects. Two aspects can share aspects in their constitution, or be connected by category aspects. We call these two aspects are compatible. That is, they may share or transfer information to each other.

3.3 Logical Treatment of Multiple Aspects

We formalized aspect in a logical framework. The detail is shown in [Takeda *et al.* 1995]. The basic idea is use of modal logic. Combination aspect is represented as conjunction of logical theories of individual aspects. Category aspect is represented as possibility of conjunction of logical theories of individual aspects. In both cases, we can include some inter-aspect theory to describe relationship among individual aspects. Then we can prove some theorems to describe relationship among such composite aspects.

3.4 Computational Treatment of Multiple Aspects

We also provided a language of aspects called **ASPECTOL** (Aspect-based Ontology Description Language), which is an extension of Ontolingua-like ontology definition [Gruber 1992].
It extends Ontolingua to express the fragments of ontology as aspect. The other extension is to provide translation formula. A translation formula is defined between two aspects in a category aspect, and is defined as *define-translation* which describes logical relation between concepts in both aspects. It can represent a class of direct relations between formulae in two aspects which we have discussed in the previous section. A left hand side of an implication formula is a formula of the aspect of the first argument and a right hand side is a formula of the aspect of the second argument. The detail is also shown in [Takeda *et al.* 1995].

4. StYLiD: Semantic Web Approach for Multiple Ontologies

Many researchers in Semantic Web community are now aware that handling of multiple ontologies is crucial for Semantic Web. The problem is often referred as *ontology integration* or *ontology alignment* or *ontology matching*. There are many studies such as [Wache *et al.* 2001] [Noy 2004] [Euzenat *et al.* 2007] [Ehring 2006] and workshops [Shvaiko *et al.* 2006] [Shvaiko *et al.* 2007]. As we discussed in the previous section, different ontologies can be formed based on different conceptualization. As long as ontologies are based on a unique conceptualization, they can be either mapped, or aligned or even integrated in principle. But if not, such an operation is not possible *per se*. There need extra knowledge.
We are interested more in conceptualization to form ontology. So we are now developing a system called StYLiD ("Structure Your Linked Data") for sharing *a wide variety* of structured information, i.e., representation of concepts and their instances [Shakya *et al.* 2008].

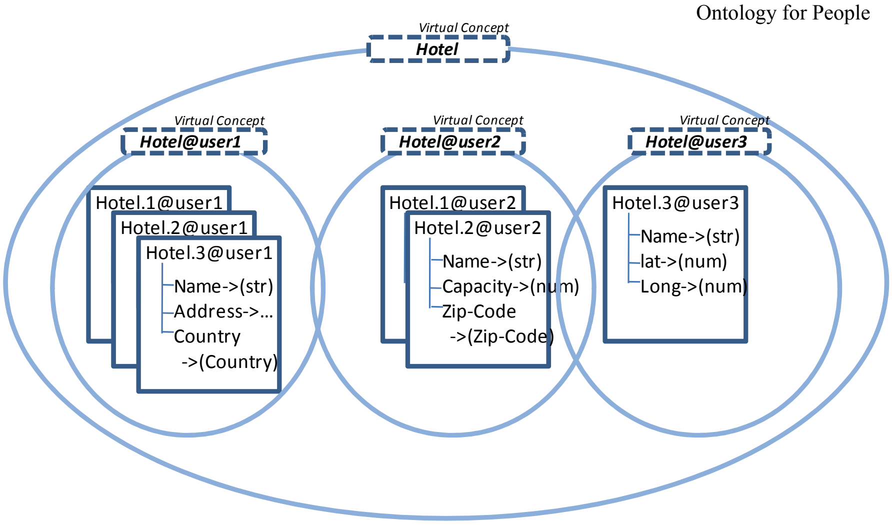

Fig3 : Variation of Concept and Virtual Concept

4.1 The Basic Idea

As we mentioned in Section 2, people are willing to share information via Web. Social tagging and Wikipedia are good examples to demonstrate their behavior. Some people just want to contribute other people or society but many people join to such activities because they are fun to themselves. Representing own knowledge and organizing collected information by own are beneficial for themselves. Thereafter they recognize the importance of information sharing. Information sharing from the beginning is difficult for people to join [Takeda *et al.* 2005].
StYLiD[4] is a service of sharing of structured information. It is similar to Freebase[5] but a variety of structure for each concept is allowed. Any user can organize information what they like. She can either use concept structure defined by other or modify it or even create a new structure. All the variations for a single concept is maintained and treated as a *virtual concept*. User can browse or search information either by virtual concept or by individual concept. User can use virtual concept to specify range of property of a concept. The system solves ambiguity when it is needed.

----
[4] The early prototype is accessible from http://dutar.ex.nii.ac.jp/stylid
[5] http://www.freebase.com/

4. 2 Consolidation of User Defined Concepts

Concepts having the same name defined by different users are grouped together by the system. This forms a single virtual concept which subsumes all the grouped concepts. This consolidated concept can be used to retrieve all instances though different users have different definitions for the concept.

All the concepts contributed by different users are visualized together as a concept cloud. When the user hovers over any concept, the attributes and description are shown so that it can be understood at a glance. Clicking on a concept shows all its instances. Better concept definitions will satisfy more users and will have more instances. Popularity of concepts is visually highlighted by increasing size. Stable definitions will gradually emerge from the vast cloud of concepts as more instances are contributed.

A consolidated concept can be expanded into a sub-cloud which shows all the versions defined by different users. Further, in the sub-cloud, multiple versions defined by the same user are sub-grouped together. The sizes of the different versions in a sub-cloud add up to form the size of the consolidated concept. Clicking on a consolidated concept shows instances of all versions of the concept. Similarly, we can see all instances of the multiple versions defined by a user.

5. Conclusion

In this paper, I argued ambiguity of representing concepts. Ontologists in computer science, especially in Semantic Web dismissed the importance. They simplified conceptualization too much according to AI tradition. But I believe that richness of our conceptualization is crucial for representing ontology and can contribute various use of ontology in computer systems.

References

[Calvanese *et al.* 2001] D. Calvanese, D. G. Giuseppe, and M. Lenzerini. Ontology of integration and integration of ontologies. In Proceedings of the 2001 Description Logic Workshop (DL 2001), (2001)

[Ehring 2006] M. Ehrig: Ontology Alignment: Bridging the Semantic Gap, Springer, (2006)

[Euzenat *et al.* 2007] J. Euzenat and P. Shvaiko: Ontology Matching, Springer, (2007)

[Finin *et al.* 1995] T.Finin, Y. Labrou, and J. Mayfield: KQML as an agent communication language, Proceedings of the 3rd International Conference on Information and Knowledge Management (CIKM'94), (1995)

[Gruber 1992] T. R. Gruber: Ontolinua: A Mechanism to Support Portable Ontologies, technical report, Knowledge Systems Laboratory, Stanford University, Stanford, United States (1992)

[Gruber 1993] Thomas R. Gruber. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199-220, 1993.

[Genesereth *et al.* 1987] Genesereth, M. R., & Nilsson, N. J. (1987). Logical Foundations of Artificial Intelligence. San Mateo, CA: Morgan Kaufmann Publishers.

[Genesereth *et al.* 1992] M. R. Genesereth and R. E. Fikes. Knowledge Interchange Format, Version 3.0 Reference Manual. Technical Report Logic-92-1, Stanford, CA, USA, (1992).

[Noy 2004] N.F. Noy, Semantic integration: a survey of ontology-based approaches. SIGMOD Rec. 33, 4 (Dec. 2004), 65-70, (2004)

[Shakya *et al.* 2008] A. Shakya, H. Takeda and V. Wuwongse: StYLiD: Structure Your own Linked Data, The Second International Conference on Weblogs and Social Media, Poster (submitted).

[Shvaiko *et al.* 2006] P. Shvaiko, J. Euzenat, N. Noy, H. Stuckenschmidt, R. Benjamins, M. Uschold ed., International Workshop on Ontology Matching (OM-2006), CEUR-WS Vol. 225. (2006)

[Shvaiko *et al.* 2007] P. Shvaiko, J. Euzenat, F. Giunchiglia, and B. He, ed., International Workshop on Ontology Matching (OM-2007), (2007)

[Takeda *et al.* 1995] H. Takeda, K. Iino and T. Nishida: Agent Organization and Communication with Multiple Ontologies, International Journal of Cooperative Information Systems, Vol. 4, No. 4, pp. 321–337 (1995).

[Takeda *et al.* 2005] H. Takeda and I. Ohmukai: Building semantic web applications as information/knowledge sharing systems, in End User Aspects of the Semantic Web, Colocated with ESWC 2005, Heraklion, Greece (2005)

[Wache *et al.* 2001] H. Wache, T. Vogele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S. Hubner, Ontology-based integration of information - a survey of existing approaches. In H. Stuckenschmidt, ed., IJCAI-01 Workshop: Ontologies and Information Sharing, 108—117 (2001)