Collection and integration of concept systems for semantic web computing

- Plan and a preliminary consideration -

(Position Paper)

Hideaki Takeda[1]

1. Introduction

The amount of information on the Internet has been increasing with the accelerating speed. The problem we are facing is how to ensure the quality of information within the enormous amount of information. The movement of Semantic Web is the straight answer to the problem, i.e., providing the knowledge level markup to documents which are understandable intelligently both by human and machines. Since markup tags are based on ontologies, shared understanding is ensured.

The mechanism is rational, but then the problem is shifted to "how to provide ontologies" or "how to construct semantic descriptions"?. Simply providing some "good" ontologies is not sufficient, because our semantics in the real world is dynamic in nature. Our semantics is spread from surely shared one to very domain dependent or private one. Meaning sometimes comes from the domain-dependent semantics and becomes public. Meaning in public is sometimes transformed into the domain-specific one. Meaning is thus dynamic, so we should support such dynamics of meaning to realize truly useful knowledge-level descriptions.

2. Collection and integration of concept systems

Towards support of the dynamics of semantics, we started a project called "Collection and integration of concept systems" (CICSS). The goal is to provide an infrastructure that can be used to generate new conceptualization for new requirements for conceptualization by mixing and combining existing concept systems. To realize the goal, the mission of the project is very simple, i.e., just collecting various concept systems in various fields or domains and relating them loosely.

One or more concept systems usually exist in each field / domain ether implicitly or explicitly. They are very good sources of knowledge for that domain. All concept systems show their view of categorization at least. Since the most primary nature of knowledge is categorization. Dealing with concept systems as categorization is expected to explicate the most basic level of knowledge.

We propose concept systems repository as the workspace for the above research. The basic functions of concept systems repository (CSR) are as follows (Fig. 1):

- Import concept systems
- Relate concept systems
- Retrieve concepts or concept systems

The strategies to create CSR are (1) treating concept systems as a whole not as individual concepts, and (2) Emphasizing variety of concept systems not hiding it.

The tasks to realize CSR are twofold. One is modeling concept systems to treat concept systems in a system, and the other is the algorithm to find relationship among them. We will consider the latter task more in the next section.

3. A preliminary consideration for algorithms to find relationship among concept systems

Finding relationship among different concept systems is a very difficult problem. We do not have the exact answer yet, but show some preliminary consideration and efforts for it.

We adopt the instance-based or extension-based approach for discovery of relationship among concept systems. Definition-based or intension-based approach is not applicable in principle. If definitions from two concept systems are comparative, it tells us that they are essentially the same conceptualization. If they are formed by the different conceptualizations, definitions are not comparative. Since the instance-based approach does not care definition, it is suitable for our purpose. Of course the instance-based has other problems. One is how to identify instances, i.e., even definitions of instances can be different. The other problem is how to approximate instances, i.e., since instances can be enormous in nature, we will need some methods to say "almost the same".

We have two trials based on this line.

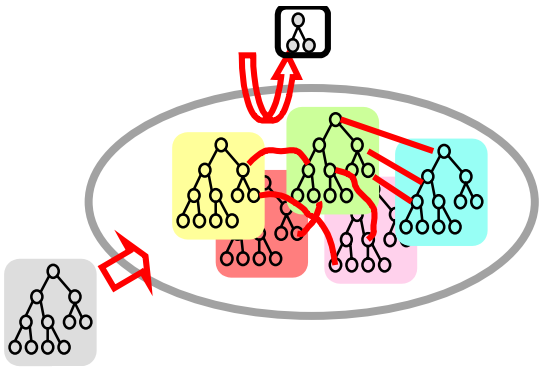

Fig 1: Concept Systems Repository

[1] National Institute of Informatics / The Graduate University of Advanced Studies, 2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 Japan. e-mail: takeda@nii.ac.jp

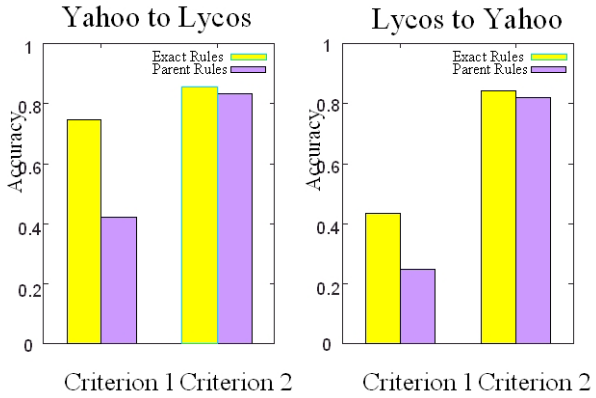

Fig 2: Results of mapping of categories between two directories (Literature case)

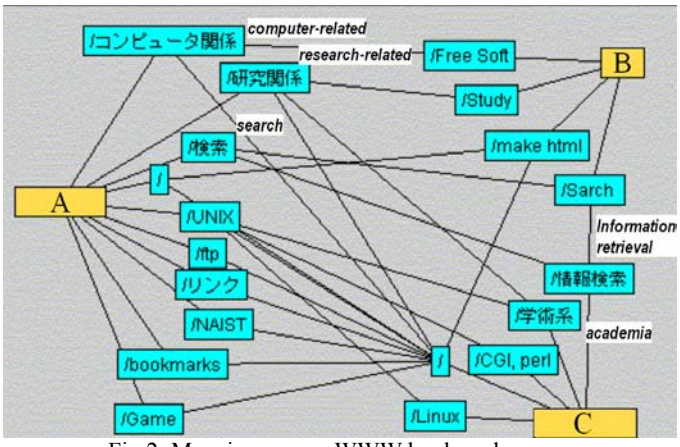

Fig 2: Mapping among WWW bookmarks

3.1 Alignment of Internet directories

One is a method to align different Internet directories like YAHOO!, Lycos, and so on. The Internet directories are not strictly defined but very elaborative and practical concept systems for the Internet. In this work, we proposed the method to generate mapping rules from categories in one Internet directory to the other by the statistical method[1][2][3]. The basic idea is to find similar categories by evaluating how much they share instances (URLs). An example of the results is shown in Figure 2 (see [2] for details). The overall result is very hopeful, around 80% of instances can be mapped correctly from one directory to the other.

3.2 Finding relationship among WWW bookmarks

The other work is to find human relationship through WWW bookmarks. WWW bookmarks are results of implicit or explicit efforts to represent personal views for the Internet [4][5]. Regarding WWW bookmarks as personal concept systems, we can find relationship among them with the similar method to the previous work. We call it *shared topics network* because concepts represent some topics which the user are interested in and they are shared among users by the relationship generated by this method. Figure 3 is an example of the generated shared topics network with three users. We can find some common relationship like (search, IR) and (academia, research-related), and community-dependent relationship like (Unix, academia).

As the evaluation as recommender systems is good enough. Topics found by the system were apparently more acceptable than pages themselves.

4. Concluding remarks

We described a project called collection and integration of concept systems (CICSs). The goal is to provide a repository of concept systems that can be used to yield new conceptualization. In some sense, it is similar to Cyc, but Unlike Cyc, we do not create but collect knowledge.

Our preliminary consideration leads to adopt the instance-based approach for discovery among concept systems, because concepts under different conceptualization cannot be compared directly.

References

[1] Ryutaro Ichise, Hideaki Takeda, and Shinichi Honiden. Rule induction for concept hierarchy alignment. In Proceedings of the IJCAI-01 Workshop on Ontology Learning (OL-2001), pages 26–29, 2001. (also available at CEUR Workshop Proceedings Vol-38 http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-38/).

[2] Ryutaro ICHISE, Hideaki TAKEDA, and Shinichi HONIDEN. An alignment algorithm between concept hierarchies. Technical Report NII-2001-001E, National Institute of Informatics, Tokyo, Japan, May 2001.

[3] Ryutaro Ichise, Hideaki Takeda, and Shinichi Honiden. Automated alignment of multiple internet directories. In *Poster Proceedings, The Tenth International World Wide Web Conference*, pages 194–195, 2001.

[4] Hideaki Takeda, Takeshi Matsuzuka, and Yuichiro Taniguchi. Discovery of shared topics networks among people --- a simple approach to find community knowledge from www bookmarks ---. In Proceedings of the Pacific Rim International Conference of Artificial Intelligence (PRICAI 00), Lecture Notes in Artificial Intelligence, No. 1886, pages 668–678, 2000.

Experimental Results for a method to discover of human relationship based on WWW bookmarks

[5] Masahiro HAMASAKI, Hideaki TAKEDA, Fifth International Conference on Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies (KES-2001), 2001.