Investigating Knowledge Discovery with Ontologies in Scientific Domains

David Ramamonjisoa
Faculty of Software and Information Science, Iwate Prefectural University
e-mail: david@soft.iwate-pu.ac.jp
Address: 152-52 Sugo Takizawa Iwate, Japan 020-0193

This research has a purpose to discover new useful links between keywords or terms within the different sources of ontology database on the web. Keywords are described as methods or applications according to our criteria then classified and structured within the domain they belong. Implementation of this system is studied in Protégé tools and OWL. KAROKA Model is elaborated in [1]. The figure below (Figure 1) shows the KAROKA model and the collected keywords that are relevant to the discovery.

Figure1 : Karoka Processing Model

The figure Figure1 makes it easy to understand that method keywords are models based on theory rather than empiric. Those methods are fundamentals in the top and some new theories at the bottom. There are some newly discovered theories among them. For the applications keywords, the correlation or association of keywords give the most interesting insights or breakthrough. The diseasome [3] example is similar to our representation of the disease and other domains such as virology, microbiology and genetics. However, in KAROKA model, there is no explicit link between the keywords as genes, proteins and disorders as in diseasome except the cyan line which indicates genes, mutation and diseases. Keywords are grouped to its domain and arranged according to the hierarchy (from general to a specific one) from the most common (left) to the newest (right). Important application keywords are colored in red or blue. Those keywords are still in research fields in the laboratories. Keywords with black color are widely accepted and well established research topic keywords. Keywords (Gene names) in pink are the most cited in the medical publications. Similar study can be found in the work on mapping knowledge domains as in [4], [5].

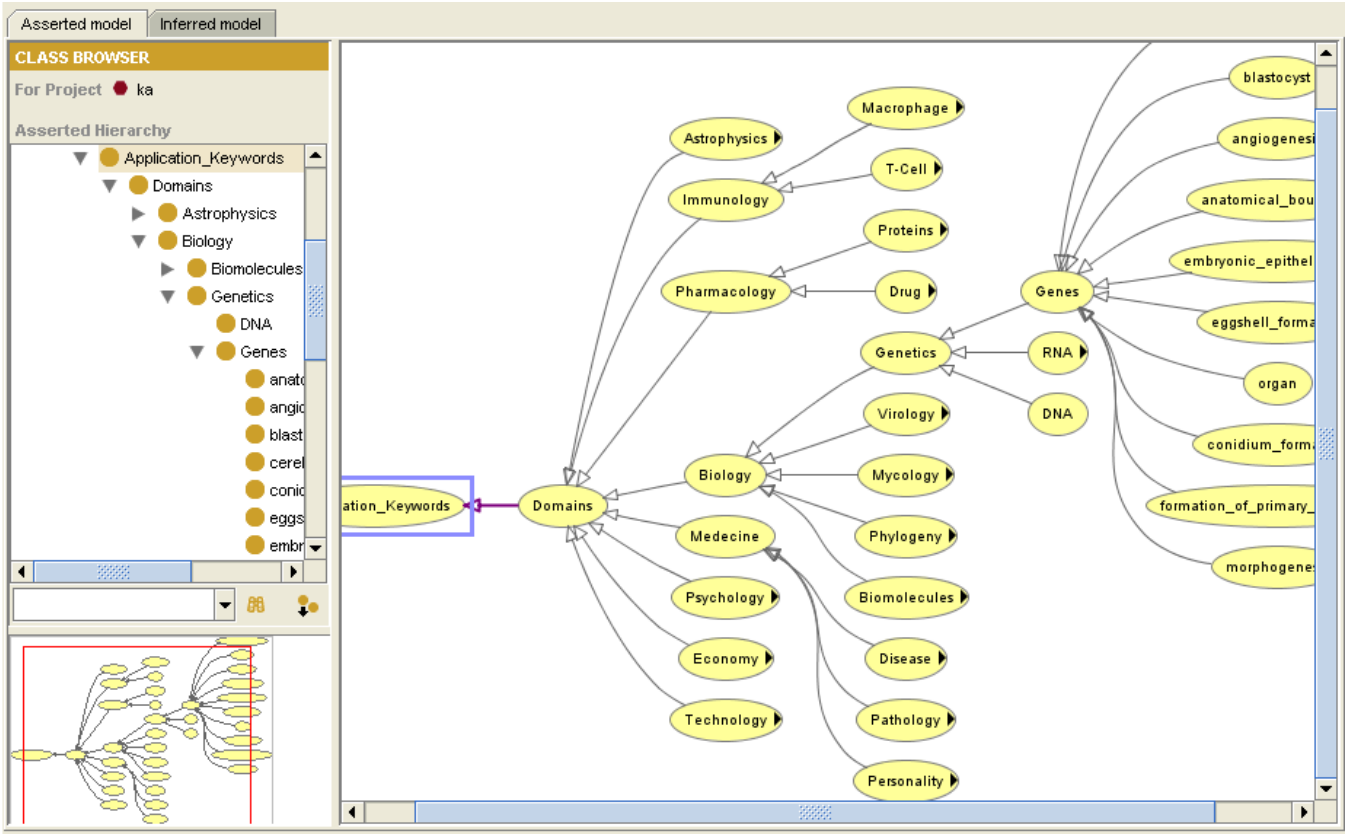

Figure2 : Prototype in Protégé Tools

The early prototype snapshot in Protégé tools is depicted in Figure 2. All application keywords are defined as classes.

1. REFERENCES

[1] Ramamonjisoa D. et al., *Research Topics Discovery from WWW by Keywords Association Rules*, In Proceedings of RSCTC 2000, pp.412-419.

[2] Ramamonjisoa D., *Designing and Implementing Knowledge Bases for Narrative Animations System*, In Proceedings of SIGGRAPH, August 2007.

[3] Kwang-Il Goh et al, *The Human Disease Network*, In Proceedings of National Academy of Science (PNAS) 2007, vol. 104, no. 21, pp. 8685—8690.

[4] Richard M. Shiffrin and Katy Borner, *Mapping Knolwedge Domains*, In Proceedings of National Academy of Science (PNAS) 2004, vol. 101, no. suppl. 1, pp. 5183—5185.

[5] Katy Borner, et al., *Visualizing Knowledge Domains*, in Annual Review of Information Science and Technology 2005, vol. 37, issue 1, pp. 179—255.