

| | |
|--------------|---|
| Title | オントロジーを用いた生物医学文献からの知識抽出手法に関する研究 |
| Author(s) | 亀谷, 聡 |
| Citation | |
| Issue Date | 2003-03 |
| Type | Thesis or Dissertation |
| Text version | author |
| URL | http://hdl.handle.net/10119/442 |
| Rights | |
| Description | Supervisor:佐藤 賢二, 知識科学研究科, 修士 |

Ontology-based Knowledge Extraction from Biomedical Literature

Satoshi Kamegai

School of Knowledge Science,
Japan Advanced Institute of Science and Technology
March 2003

Keywords: ontology, natural language processing, information extraction, text mining

Genome projects are now determining whole DNA sequences and genes of various species. The next target is the analysis of the function of each gene. To understand the role of each gene, we must clarify when it is expressed and how its product interacts with other materials.

In the biomedical literature, information about functions and interactions of genes and proteins are described, and their number is very large. To extract these information from the biomedical literature, an intelligent information extraction system is needed. In this context, bio-ontology that arranged the technical terms in biology and genomics is being built. In addition, research is also being conducted which aims at extracting relevant information, such as a protein-protein interaction, from biomedical literature. However, since the former approach is built with the help of domain experts, it only covers tens of thousands of terms at most. The latter one extracts only a narrow variety of the useful information because it uses template matching on limited verbs showing interaction information.

Therefore, first we tried to extend Gene Ontology based on the extensional ontology built in previous work. However, we found that it is difficult to link Gene Ontology with the extensional ontology. We try to automatically extract verbs and templates showing interaction information using extensional ontology only. In comparison with related works, the experimental results showed that our method could extract verbs and templates which were humanly prepared in related works. Furthermore, it was shown that it can extract wide variety

of such verbs and templates without utilizing any background knowledge. As next steps, we will need to evaluate verbs and templates discovered by our method. Since various information is being stored electronically in recent years, we think that our research can be applied to not only life science but also various field.

Publication

Satoshi Kamegai, Kenji Satou, Akihiko Konagaya. “*Toward Ontology-based Knowledge Extraction from Biomedical Literature*”, Genome Informatics 2003, UNIVERSAL ACADEMY PRESS,INC. TOKYO,JAPAN.