Hybrid Approach for Corpus Based Bio-ontology Development

Rivindu Perera ✎
Informatics Institute of Technology
Email: rivindu.perera@hotmail.com

Udayangi Perera
Informatics Institute of Technology
Email: udayangi@iit.ac.lk

Bio-ontology development can be considered as a next generation requirement of providing access of health informatics for public. But still attempts and approaches taken to develop such upper merged ontology are restrained by the diverse set of features associated in biological corpora. Therefore, this paper proposes new approach to develop bio-ontology without any supervision from the controlling units. Basically this approach is empowered by two processes to extract the hidden biological knowledge to be exposed with a generic schema, relation extraction process and user-query based extraction process. Relation extraction process will analyze bio-logical corpus and will then extract semantic relations which then formalized to conceptual graphs. User-query based extraction process is used to speed up and direct the relation extraction process to knowledge which has a demand from user community. Therefore, user query based extraction process is designed to extract important terms from user queries which may be formed by users to get information about diseases, drugs or other medical information. These extracted terms are then passed to the relation extraction process to search corpus looking for an interesting text snippet to be extracted. During this search Bag-Of-Word model can also be incorporated to maximize the retrieval quality. Nevertheless, extracted relations are then transformed to a conceptual graph which will represent the agent-patient relationship with a connector element.