KBB: A Knowledge-Bundle Builder for Bio-Research

Bio-researchers struggle to manage vast amounts of data coming from thousands of sources, from highly technical publications in online repositories. Researchers need to search, retrieve, filter, extract, integrate, organize, and share bio-information in an efficient, user-friendly, timely, and high-precision manner. This project will deliver tools designed specifically to assist bio-researchers in this daunting task.

This proposal addresses this problem by creating "knowledge bundles" (KBs), which are data representations that allow bio-researchers to perform their information extraction work. A tool called a "knowledge bundle builder" (KBB) will help researchers develop KBs in a synergistic and incremental manner and reuse them in multiple tasks. These KBs can support both individual and team work, and even larger-scale bio-research repositories to be shared widely among the research community.

For example, researchers exploring the relationship between polymorphism and lung cancer must navigate a dizzying array of information sources containing data about SNPs, alleles, and human subjects. These sources include database-generated semi-structured sources and unstructured scientific literature. The KBB will enable researchers to navigate, distill, collect, analyze, and share relevant information to answer research questions.

This approach centers on an ontology-based representation for the conceptual structure of information. The proposers' previous research has demonstrated that ontologies can be highly effective tools for extracting and organizing domain knowledge. They have also implemented a number of tools to support ontology building, information extraction, information integration, and queries over extracted information. However, these tools are research prototypes that need additional work and significant enhancement in order to serve the needs of scientific research practitioners in general and bio-researchers in particular.

During the course of this project, the research team will develop a bio-research support system and evaluate it using (1) standard measures such as precision and recall and (2) performance and effectiveness assessments in the field by expert users. Field-testing will be continuous, occurring throughout the project timeline, not just after a lengthy development phase. The research team is interdisciplinary in nature, involving experts in database systems, conceptual models, information systems, computational linguistics, software development, and bio-research. Team members are affiliated with Brigham Young University and the Mayo Clinic.