

REPORT

CSE 736 - DATABASE SEMINAR

SURAJ THYAGARAJAN PARAMASIVAM

Data Integration and Genomic Medicine

**Bio2RDF: Towards a mashup to build
bioinformatics knowledge systems**

Overview

The papers dealt with the ways and means of usage of data integration in genomic medicine. The authors dealt with the basic methods of data integration in the first paper and discussed its various usages in the field of genomic medicine. The authors had reviewed all the latest data integration systems and their pros and cons in relation to the field of genomic medicine. The authors had dealt with two axis of integration in the first paper. The strength of the paper lies in the fact that the authors were effectively able to demonstrate the various strength and weaknesses of various data integration techniques in the field of genomic medicine. The authors concluded that the peer data managements systems were the most preferred kind of data integration suited to the field of genomic medicine after careful analysis of the kind of data dealt with in the field of genomic medicine.

The second paper dealt with a specific implementation called the Bio2RDF which was based on peer data integration framework. The paper dealt with the construction of this implementation which included the framework under which this was built. The bio2rdf system is based on the semantic web approach , which makes it one of the most advanced kinds of data integration initiatives in the field. The Bio2RDF system is based on a definite ontology which was created specifically for the project with standard owl methodologies. The project also contained the RDFizing part which consisted of a series of steps including URI normalization which forms one of the most important parts of the projects in terms of its goals towards integrating biological data.

The second part of the paper deals with a use case testing for the Bio2RDF system. The authors had chosen the parkinson's disease for testing the system. The authors had come up with various different scenarios in which the system performed effectively and the authors had documented the results

The system is still a work in progress and the authors are trying to come up with better implementations of the system including many other sources of data and a larger amount of data. The project has an open source and can be found at bio2rdf.sourceforge.net. The authors provide open access and clear instructions to add new database sources to the system.

Detailed Overview

The two papers dealt during this seminar dealt with the relevance of data integration in the fields of genomic medicine. The first paper which deals with the various methods of data integration that can be used in the fields of genomic medicine. This is the strength and weakness of the paper. It is the strength of the paper since it has come up with all the possible data integration systems possible for the field of genomic medicine. The paper has completely dealt with all the possible methods of data integration with clear mention of the pros and cons of each technique. However, this paper has not dealt with the details of any specific implementation details of any specific system, which used these integration techniques. The second paper overcomes this weakness by dealing with a specific implementation.

Both the papers discussed have a good technical depth, with the second paper taking up a recent technology that has not developed to its fullest potential. The paper is a work towards a semantic mash up for the biological data. The paper has dealt with the state of the art technologies in producing semantic web.

The papers are technically sound and they are more oriented towards databases and computer sciences than to biology in general. The papers deal more with the computer science part of bioinformatics when compared to the other papers which have more medical and biology components to them.

The comments and questions raised during the presentation were about the usages of the various integration systems. Discussions were made on why a certain technology could or could not be used for data integration in genomic medicine. The discussions mainly concentrated on evaluating each data integration technique in depth to analyze why they could or could not be used in genomic data integration. The comments were mainly towards the use of data warehousing techniques for data integration in genomic medicine.