



BIOINFORMATICS SEMINAR

**CBCB
SEMINAR**

10/08/2018

3:30 PM

**DELAWARE
BIOTECHNOLOGY
INSTITUTE**

bioinformatics.udel.edu

Xiang Li

M.S. Bioinformatics Student

UNIVERSITY OF DELAWARE

Integration of Heterogeneous Data for Protein Ontology Database Using Semantic Web Technology

As the volume and diversity of data and the desire to share those data increase, we inevitably encounter the problem of combining heterogeneous data generated from many different but related sources and providing the users with a unified view of this combined data set. This problem emerges in the life sciences research community, where research data from different bioinformatics data repositories and laboratories need to be combined and analyzed. The benefit of developing a data integration system is that it can facilitate information access and reuse by providing a common access point. It also provides users with more complete view of the available information. The Semantic Web provides technologies to make such integration possible:

- An abstract model for the relational graphs: RDF
- A query language adapted for the relational graphs: SPARQL
- Various technologies to characterize the relationships, categorize resources: RDFS (RDF Schemas), OWL (Web Ontology Language)

What's more, after integration of heterogeneous data, the application of database is still a problem. Therefore, in this project, the API for PRO database is built so that users can directly use the querying result in their projects and no need to rebuild database in local. In the recent years, with the development of API, make mutual reference between applications easier. There are many mature frames, like Django, Diesel and Flask, can be selected. In this project, we applied the Django framework to build the API for PRO. It can work now but we are still testing it.



**College of
Engineering**

CENTER FOR BIOINFORMATICS &
COMPUTATIONAL BIOLOGY