IPRO 327 Project Plan
Fall 2008

Semantic News Aggregation Sofware

Advisor :Wai Gen Yee

1.0 Objectives

- Design and build software to provide Web searchable news data and related semantic information from current events. This information will be extracted from online news sources, processed by our software to add semantic information, and then organized and presented in a logical manner through a Web interface.

2.0 Background

A. The Semantic Web uses metadata to add meaning to information. The idea is to switch from the paradigm of information that is understandable by humans only, to information that is understandable by humans and computers. This way, computers will be able to connect relevant information together though well defined logical relations.

B. Semantic Web technologies and standards have been created by various organizations or groups in preparation for the next Web paradigm. There are still various issues to be resolved before Semantic Web technology will be ready to put into wide scale use. Current and future trends in Semantic Web technologies and standards must be taken into account and incorporated into our software. For this reason it is important for us to keep up with emerging trends in the development of these technologies to ensure that our software will be compatible with other semantic software in the future.

C. One important technology we will use is RDF, or resource description framework. RDF represents information as a triple, which contains essentially a predicate, subject, and object. This is a basic assertion model that allows us to define facts about data and relationships between things. RDF is a structured as a very generalized model so that in the future it will be the common framework with which all application will be able to understand data from other applications.

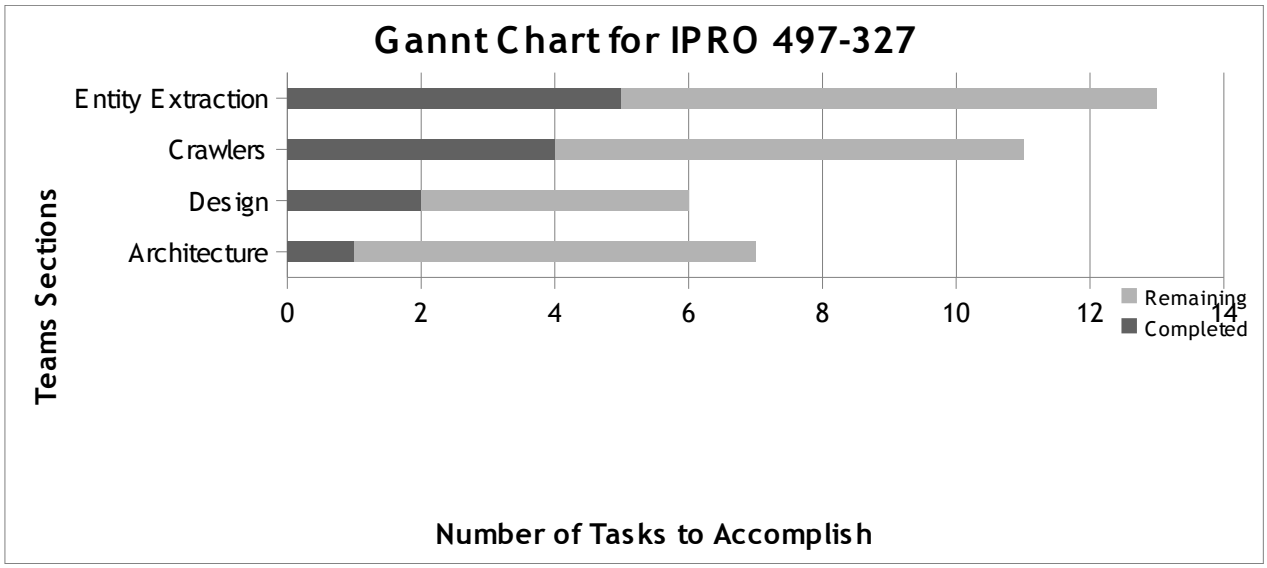
3.0 Methodology

A. Components of our news data aggregation software can be described as four distinct components: web crawling, entity extraction, software design, and software architecture. Initially each member of our group is assigned to a specific team and is responsible for completing the objectives of their respective team. Once each team has working implementations of their goals, we will collaborate to integrate each component into one fluid piece of software.

B. In order to accomplish the tasks laid out in Part A, each team will gather information on the Semantic Web and related technologies. Each team will research intensively whatever aspect they are assigned to, and will present information on their research and results to the rest of the team.

C. In addition to team presentations we have created a Wiki page for our project where we can gather all information we uncover as we go along. This Wiki page will help us track our progress, goals, research papers, Semantic Web applications, tutorials, and various other information. Each team member updates their progress every week with what they have accomplished during the week, what their current goals are, and any other pertinent information regarding the status of their work.

D. We have adopted a general model for solving problems that each team will implement to complete their portion of the project. This plan involves defining the problem they are trying to solve, researching possible solutions, creating metrics for evaluating different solutions, testing solutions, and finally implementing a solution.



4.0 Project Budget

A. The expected budget for this IPRO is $0. The main aspects of this project are research, design, and implementation, which will not cost any money. There may be unforeseen expenses later in the project such as printing brochures or paying for dedicated web hosting.

5.0 Team Structure and Assignments

A. List of team members:

Name	Major	Skills, Strengths, Experience
Nick Bathum	Computer Science (4th year)	Coding, Previous successful IPRO
Evan Estola		
Jaeyeon Kim		
Jay Mundrawala		
Chris Osswald		
Pete Schmitz		
Cameron Zangenehzadeh		

B. Team structure:

Name	Team
Nick Bathum	Crawling
Evan Estola	Architecture
Jaeyeon Kim	Entity Extraction
Jay Mundrawala	Entity Extraction
Chris Osswald	Design
Pete Schmitz	Crawling
Cameron Zangenehzadeh	Design

* Denotes team leader