Applying Semantic Web Technologies to Bioinformatics

BioSem

Introduction

In 1989 Sir. Tim Berners-Lee, while working at CERN, proposed a method for managing documents using several technologies: HTTP, URI and HTML. The first web server was created in 1991 and the web as we know it today was created. The first Web also known as Web1 is based on organising documents and linking them together. While this was a big step in knowledge management it did and does not manage the actual knowledge itself, but rather the documents. In 1991 The Semantic Web was introduced by Sir. Tim Berners-Lee, James Hendler and Ora Lassila in an article published in The Scientific America magazine. The difference between this vision and the latter is the way the knowledge is viewed. The Semantic Web brings structure to the actual knowledge hidden within the Web pages, thus shifting the emphasis from documents to the actual content of web pages. In the semantic web, agents can take advantage of the logic connecting the contents of web pages and automatically carry out sophisticated tasks in behalf of the users. The semantic web is based on current web technologies and some new ones. Figure one shows the different layer of the semantic web.

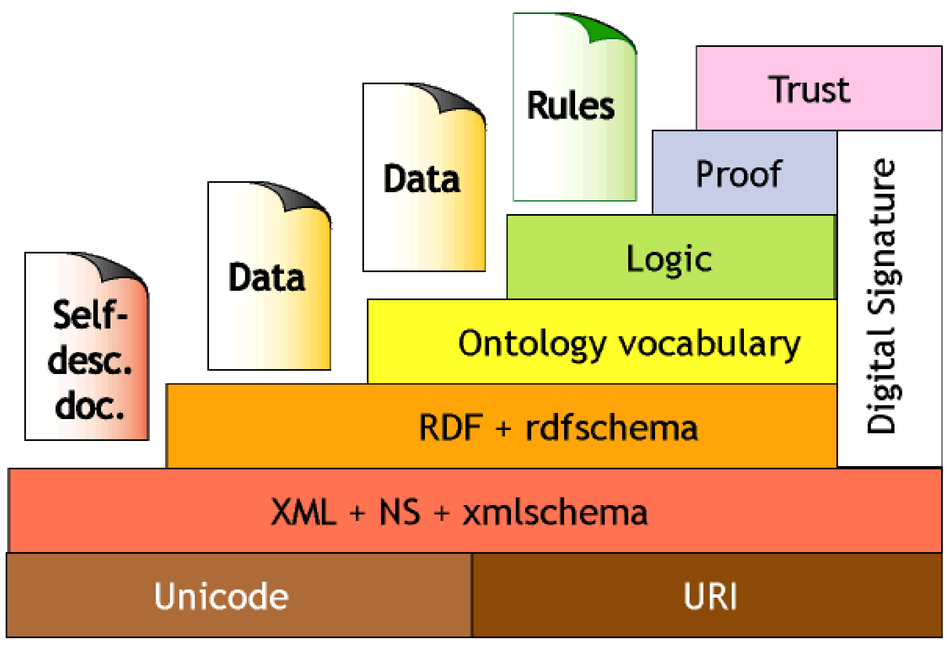


Figure 1: Semantic web layers

The Ying-Yang Framework: An agent Society

Agents are autonomous programs that can act on behave of other programs or persons to perform a particular task. Agents can either act alone or work together in a multi agent society. This social gathering of agents can be on a single machine or on the entire globe, creating a matrix of agents, which its sole responsibility is to perform one task.
The Ying-Yang framework developed by Dr. Ky Van Ha, is a java package for developing multi agent Societies. The framework can support autonomous, none-autonomous, non-mobile or mobile agents. A society is based on one or more domains that are managed by the manager, each domain is controlled by the domain manager. Communication in the society is done by using the KQML, Knowledge Query Machine Language.

RDF: Resource Description Language

RDF is a graph based decription language that can logically describe,logically connect terms to one another and finally encapsulate the domain knowledge. Such knowledge is called an ontology of the domain.

BioSem

BioSem is an attempt to create an information portal where researchers and students in bioinformatics can search for different terms used in their field of study and get relevant information about the term. The information will include description, relations, synonyms, web page and scientific papers.

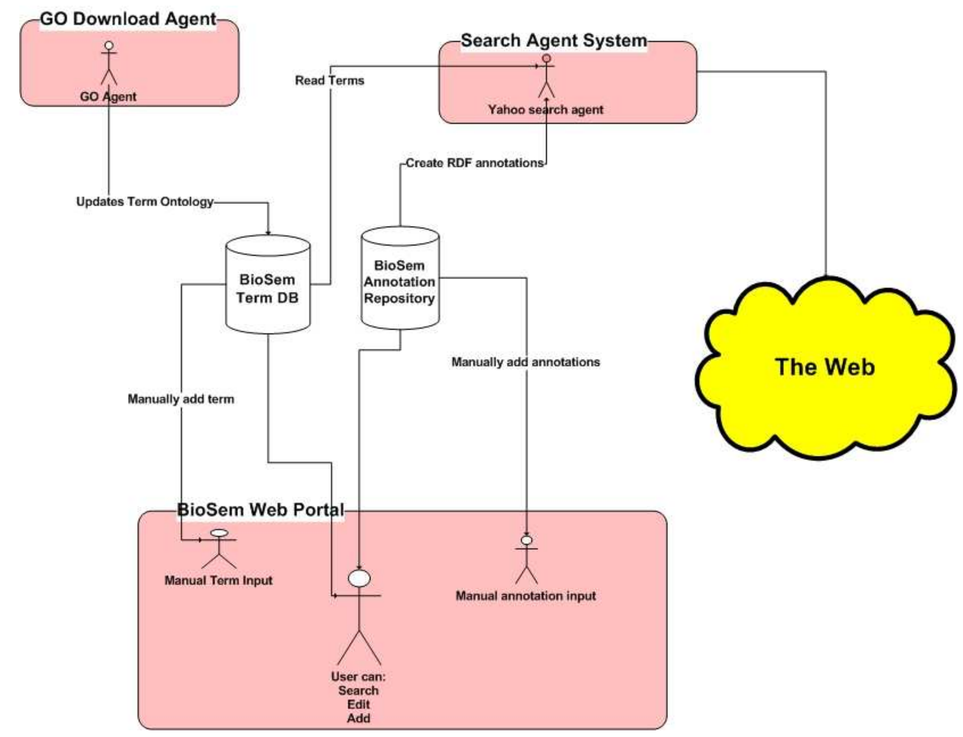


Figure 2: BioSem Architecture

The application is composed of three different parts:

(I). The Go agent is responsible for downloading the Gene Ontology and updating the BioSem database with the new terms, the whole process is done automaticlly once a week.

(II). The yahoo search agent system uses the terms stored in the database to search the internet for relevant webpages, images and scientific papers; the results from each search is annotated using RDF and stored in an repository.

(III). Through the web portal different actions can take place: searching, introducing new terms, creating users and annotation. Searching is simply entering a term and viewing information about the term, e.g. term's name, definition, relationships and synonyms. By choosing a term the system will perform a semantic search in its repository and display the different links, web pages, papers and images. The term semantic search here applies to searching for the term itself, searching for terms that are related to the term and terms that are synonym to the term. The user can search for a term by using a wild card, viz. the user can search for terms that begin, end or having some characters.

BioSem has two methods for annotating:

(I) The search agent automaticlly creates annotations and stores them in an RDF repository.

(II) An expert user can create annotations manually through the bioSem portal. New terms can be added to the database and links can be given to the program to be annotated semi-automaticlly.

Project details

The biosem application has been developed by Ehsan Sadeghi as part of a master degree in informatics at Østfold university college. The master thesis has been supervised by Dr. Ky Van Ha.