Use Case 1 in Health – Business Cases

Integrated access to Biological Data

KW Partner: UPM

1. Overview

Challenge
To provide an unified point of access to different biological data repositories accessible through the Internet, corporate databases, results of experiments, health cards, medical literature sites and so on.
Solution
Application of semantic technologies to solve the inherent features of the biology field: huge quantity of dispersed, distributed and autonomous data with great difficulties to be integrated due to differences in terminology, syntax and semantics.
Why a Semantic solution
Ontologies describe the vocabulary of the data stored at each repository. Annotations describe the data and link it with a corresponding ontology. Ontology merging and mapping techniques allow integration of repositories in a consistent and unified way.
Key Business Benefits
Aid to the researchers in the biological field, providing a unique point of access to biological data. For example, when a researcher wants to compare the results of an experiment with the genome annotation database.
Business Partners
Life science companies.

Keys components

Existing Software
No existing software.

Research and development
Study and selection of ontologies already existing in the biological field.
Merging and mapping of ontologies
Annotation using selected ontologies

Technology locks
Knowledge extraction
Ontology based reasoning

Currently, a great diversity of biological data exists in repositories: databases accessible through Internet, corporate databases and experiment results among others. Equally there exists a great diversity of ontologies for modelling this data. Therefore the situation that the researchers has to face with is a lot of dispersed data and different disconnected and non-user-friendly tools to access such data, therefore the researches have to confront great difficulties to aggregate all the data to carry out the research tasks in an integrated way (Figure 1).
Up to now ontologies in biology were considered as mere guides for data structure, with their only purpose being to access the more adequate documents and articles to the researchers' interests. This new vision will allow for combining and associating existing ontologies in the biological field and an integrated modelling of the biological data sources (genomics, proteomics, metabolomics and systems biology). Once modelled, the annotations, intelligent agents, semantic web agents and the semantic grid will offer a centralised access point to extract and generate knowledge from the biological data repositories.

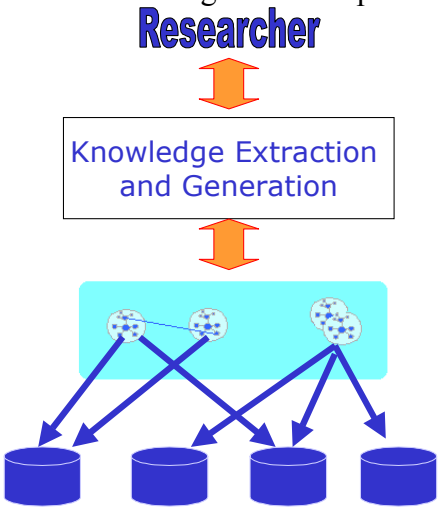

Figure 1 – Integration of Biological Data Repositories

All rights reserved © Knowledge Web

2. Current Practices and Technologies

2.1 Current business practices

Current practices by biologists seeking data are guided by the available tools and their capabilities, as given in section 2.3.

2.2 System requirements Analysis

Some requirements identified from the use case:

- Generation and extraction of knowledge from biological data by means of ontologies, ontologies existing in the knowledge domain, combining them (ontology merging) and/or associating them (ontology mapping) is to be exploited by means of annotations, intelligent agents, semantic web services and/or semantic grid.
- Using standards for the semantic web providing a unified entry point to different biological data repositories in the most automated way possible.

Ontology Mapping allows the determination of which concepts in some ontology A are the same as in another ontology B. Detecting common concepts is allowing the "jump" between ontologies. The ontology merging could be used by a company that wants to use de facto standard ontologies associating them to specific company ontologies. In this way, proprietary repositories can be "linked" with public ones.

Ontology Merging allows to sum one ontology C with another ontology D to obtain a more complete ontology. Due to the fact that a protein could be implied in cell signaling as in a biological process, summing up two ontologies, one describing cell signaling and other describing biological processes can give us a general overview of a protein function.

Both mapping and merging are shown in Figure 2 below.

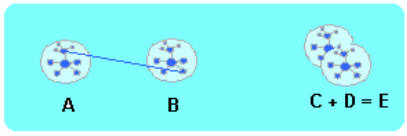

Figure 2 – Ontology mapping and merging

Once we have merged or mapped the ontologies in the above task we have to be capable to link these resulting ontologies with public or proprietary data repositories through semantic annotation. Deep annotation is a framework taken into account at this stage (Figure 3).

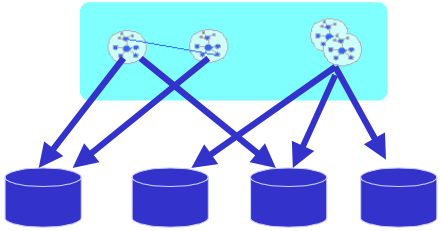

Figure 3 – Deep annotation to link data repositories to ontologies

The owners of the databases have the knowledge, so they are the ones that should annotate the databases with the appropriate ontology.

All rights reserved © Knowledge Web

If the biological data repositories are linked to the ontologies by annotation, now we have to offer to the researchers appropriate tools to extract knowledge from these sources. In this task existing semantic and artificial intelligence technologies applicability and requirements detection will be carried out.

2.3 Review of the current systems

Currently there are several Data Mining tools focused to solve the immediate and concrete problems that the researchers in the biological field face day by day.

Entrez is a database search engine that provides a search that combines documents containing nucleotide or protein sequences, 3D structures and their respective references in MedLine; its power resides in the numerous cross references that it offers between the different databases, along with a computerised system for similarities between documents, that allows to provide the documents set most similar to the required one.(http://www.ncbi.nlm.nih.gov/Entrez/).

At the SRS tool there is no limit to the number of databases and applications that can be accessed. It also allows creation and saving of the users' own intuitive views for displaying data which can be made up of as many different databases and fields as wished. The most frequent database queries can be set up to be available every time the user logs into the system. To make results more meaningful, filtering of application results is possible using predefined queries. Work can be published to SRS for all appropriate colleagues to access in a read-only mode.

TAMBIS aims to aid researchers in biological science by providing a single access point for biological information sources round the world. The access point will be a single interface (via the World Wide Web) which acts as a single information source. It will find appropriate sources of information for user queries and phrase the user questions for each source, returning the results in a consistent manner which will include details of the information source.

With the current tools it is still needed to understand the structure of the different databases, as with SRS, so the inherent problems of terminology, syntax and semantics are still present. It is difficult to know if a table in one database called Organisms is the same table called Species in another database. With TAMBIS it is necessary to define a wrapper service for each database to translate the queries so no automation is provided. To design the wrapper it is necessary to study the database before including it into the system.

All rights reserved © Knowledge Web