

Titre :

Design and realization of a biomedical-ontology-based semantic distances Web service

Information :

Encadrant : Clement Jonquet (LIRMM, UM2) – jonquet@lirmm.fr
Spécialités : DECOL, AIGLE
Nombre d'étudiants : 3-4
Contexte: Projet SIFR (Semantic Indexing of French Biomedical Data Resources)
Ou: LIRMM, SMILE & TEXTE research team
Quand: 2nd semestre 2012-2013

Mots clés :

Web application, ontologies, knowledge representation, semantic Web, semantic distances, graph-based measures, database

Technologies :

Java, JEE, MySQL, Tomcat, ResTful web services (RestLet), XML, RDF, BioPortal Web services API + web client technologies (HTML5, JavaScript, CSS) if the number of students is enough

Résumé :

Conception et implémentation d'un service Web de distance sémantiques basées sur les ontologies biomédicales

Ce TER consiste à implémenter une application Web qui traite des ontologies biomédicales pour réaliser l'exécution de plusieurs distances sémantiques. Il s'agit à la fois de comprendre et implémenter les algorithmes de graphes sous-jacents aux distances et de réaliser le service web qui retourne les résultats de l'exécution de ces distances.

Contexte :

Ontologies allow representing data with clear semantics that can be leveraged by computing algorithms to search, query or reason on the data [3]. An ontology can be viewed as a **graph of concepts**.

The volume of data in biomedicine is constantly increasing. A key aspect in addressing semantic interoperability for life sciences is the use of terminologies and ontologies as a common denominator to structure biomedical data and make them interoperable. Especially, the community has turned toward **ontologies to design semantic indexes of** that leverage the medical knowledge for better information mining and retrieval [7].

In the context of the ANR SIFR project, LIRMM is **building ontology-based services to leverage biomedical ontologies and terminologies in indexing, mining and retrieval of French biomedical data**. We use the National Center for Biomedical Ontology (NCBO) *BioPortal* [1] which is a web repository of biomedical ontologies (<http://biportal.bioontology.org>), mostly in English and the **Health Multi-Terminology Portal (HMTP)** - <http://pts.chu-rouen.fr/> which provides an access to health terminologies and ontologies in French. Users can browse, search, and comment ontologies both online and via a web services application programming interface.

Among services available in SIFR's portfolio, we like to implement a **semantic distance web service**.

Présentation du sujet :

An ontology-based semantic similarity measure (or semantic distance) [4, 5] is a function that, given two ontology concepts, returns a **numerical value that reflects how similar these two concepts are**. Similarity measures are often based on the structure of the ontology and they have often been used in biomedical applications for semantic search, annotations or alignment [1, 6, 2]. Semantic similarities can be used to improve the concept search mechanism by reformulating user inputs, so similar concepts and synonymous can be incorporated to query results. Another potential application is query disambiguation, where similarities support search engines in deciding which responses should be considered, when concepts with several meaning are used in user queries.

Many measures have been proposed in the literature. We have previously selected a few distances to be used in order to provide new ontology-based services to the biomedical community. Therefore, this project aims to: (1) read the articles presenting the selected measures and understand the algorithms described; (2) implement these measures and run them over a set of ontologies that we have already prepared and harmonized; (3) to provide a web service API for users to re-use the similarity in their application. The web service API will at least provide the following functionalities:

- for 2 given concepts returns the distance between them.
- for 1 given concept and a maximum distance value, returns all the concepts distant from the given concept by a distance under the given value.

Two main programs will have to be written (in Java):

- a workflow that implements and runs the distance algorithms over the harmonized ontologies, available in a MySQL DB, to produce distance tables in the same DB.
- a web application that implements a RestFul Web service realizing the functionalities described hereafter. Technologies to use will certainly be: JEE (business logic), RestLet (rest API), XML/JSON (data). If the number of student allows, a Web user interface will be designed in addition.

Références :

- [1] Hisham Al-Mubaid and Hoa A. Nguyen. A Cluster-Based Approach for Semantic Similarity in the Biomedical Domain. In *28th IEEE Engineering in Medicine and Biology Society Annual International Conference*, pages 2713–2717, New York, NY, USA,, September 2006.
- [2] Jorge E. Caviedesa and James J. Cimino. Towards the development of a conceptual distance metric for the UMLS. *Biomedical Informatics*, 37(2):77–85, April 2004.
- [3] Tom R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, June 1993.
- [4] Wei-Nchih Lee, Nigam H. Shah andKaranjot Sundlass, and Mark A. Musen. Comparison of Ontology-based Semantic-Similarity Measures. In *American Medical Informatics Association Annual Symposium, AMIA'08*, pages 384–388, Washington DC, USA, November 2008.
- [5] Ted Pedersen, Serguei V. S. Pakhomov, Siddharth Patwardhan, and Christopher G. Chute. Measures of semantic similarity and relatedness in the biomedical domain. *Biomedical Informatics*, 40(3):288–299, June 2007.
- [6] Sylvie Ranwez, Vincent Ranwez, Jean Villerd, and Michel Crampes. Ontological Distance Measures for Information Visualisation on Conceptual Maps. In R. Meersman, Z. Tari, and P. Herrero, editors, *Workshop on Ontology Content and Evaluation in Enterprise, OntoContent'06*, volume 4278 of *Lecture Notes in Computer Science*, pages 1050–1061, Montpellier, France, November 2006. Springer-Verlag.
- [7] Daniel L. Rubin, Nigam H. Shah, and Natalya F. Noy. Biomedical ontologies: a functional perspective. *Briefings in Bioinformatics*, 9(1):75–90, 2008.