

Titre :

Comparison and convergence of biomedical ontology repositories platforms: BioPortal vs. CISMEF

Information :

Encadrant : Clement Jonquet (LIRMM, UM2) – jonquet@lirmm.fr

Spécialités : IFPRU (DIWEB, I2A, GL)

Nombre d'étudiants : 2-3

Mots clés :

Ontologies, Web application, Knowledge representation, Service-oriented architecture, Web services, biomedical data.

Résumé (français) :

Les terminologies et ontologies biomédicales jouent un rôle clé dans l'interopérabilité sémantique des données des sciences du vivant en servant de dénominateur commun. Pour construire des applications cliniques, médicales ou industrielles, il est crucial que les chercheurs convergent vers un ensemble de méthodes et de formats interopérables pour le traitement des données. L'Université de Stanford et le CHU de Rouen ont développé des portails pour les ontologies/terminologies biomédicales (e.g., édition, navigation, visualisation, annotation de données, indexation, etc.) qui assistent les professionnels de santé et les chercheurs en médecine dans la construction de système à base de connaissances. L'objectif sur le long terme est la convergence des recherches menées à Stanford avec les initiatives similaires en France dans le domaine des terminologies/ontologies médicales. Le travail consiste à faire une comparaison exhaustive des fonctionnalités des plateformes (accompagné de test d'évaluation) et s'intéresserez à l'interopérabilité des services fournis par les 2 plateformes.

Présentation du contexte :

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. It is important that different service provider work to converge toward unified set of tools and formats to process biomedical data in the design of clinical, research or industrial applications related to biomedicine. The *Stanford Center for Biomedical Informatics Research* (BMIR) group at Stanford University (<http://bmir.stanford.edu>) and the *Catalogue et Index des Sites Médicaux de langue Française* (CISMeF) group at Rouen University Hospital (<http://www.cismef.org>) have both invested lot of efforts in developing terminology/ontology-based tools and services to assist health professionals and users in their search for electronic information available on the Web and in the use of ontologies.

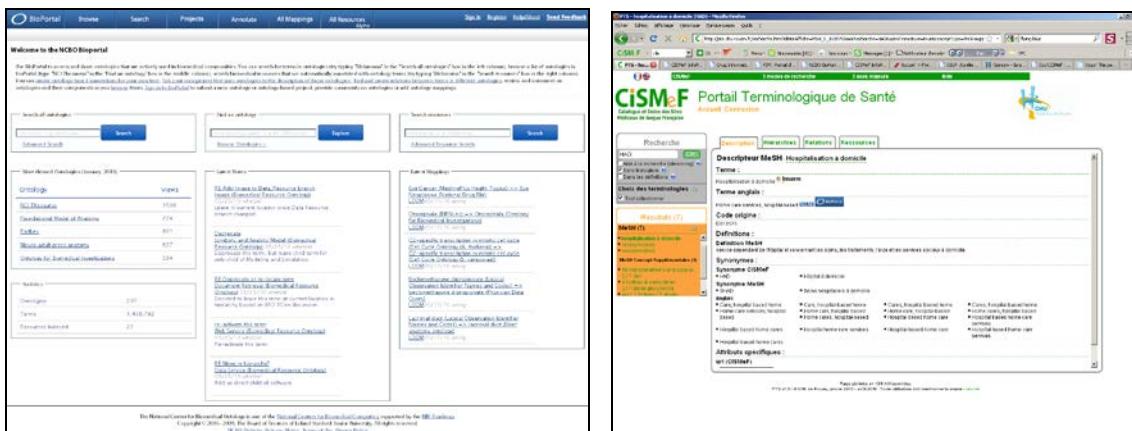
The two groups have developed Web-based portals, respectively the *NCBO Bioportal* [5, 7] and the *CISMEF Health Multi-Terminology Portal* [2], that offer a variety of services to search or index biomedical data as well as searching, exploring, annotating and visualizing the available standards ontologies. Despite each portal specific aspect (e.g., use of French language) they share a common objective while not being interoperable. As of today, a person who likes to use both portals will face double amount of work. Furthermore, there are no means allowing an easy automatic composition of the services provided by each portal.

The project aims to precisely compare the features provided by the 2 portals and draw the lines for their convergence and interoperability.

Présentation du sujet :

Public biomedical data is already enormous and is expanding very fast. Biomedical data integration and interoperability is necessary to enable cross data search and query as well new scientific discoveries [4]. These data are often unstructured and are available in different formats (database, documents, etc.) preventing discoveries that could be made by merging them. To address this problem, the biomedical community has turned to ontologies and terminologies to describe their data and turn them into structured and formalized knowledge [1, 6]. Ontologies formalize the knowledge of a domain by means of concepts, relations and rules that apply to that domain [3].

Both BMIR and CISMEF have both invested lot of efforts in developing terminology/ontology-based tools and services to assist health professionals and users in their search for electronic information available on the Web and in the use of ontologies. Despite their specificities, the two portals share an overall common objective which is to provide to the biomedical community a common interface to explore health information leveraging the knowledge represented in ontologies and terminologies. However, the two portals are not interoperable and are sometime repetitive. As of today, a person who likes to use English or French bio-ontologies will face a “cornelian choice” to pick up a portal for his tasks. Furthermore, sometime he may have to use both portals to obtain satisfying results (e.g., using French terms to search English data and vice-versa). This situation results in a loss of time and no capitalization (code reuse, service architecture, unique identifiers) on the service providers side as well as a loss of time and a much complex environment to get familiar with (several formats, APIs, user interfaces, tools) on the service users side.



Screenshots of BioPortal (<http://bioportal.bioontology.org>) and CISMeF Health Multi-Terminology Portal (<http://pts.chu-rouen.fr>).

The project aims to review the portals functionalities (mainly at the API level) and elicitate the possible convergence between the two portals that will facilitate the use of each portal as well as the composition of services provided by each of them. Such convergence could span from using a common (SKOS) model for the platforms' data to reuse specific features such as visualization, or mappings from one platform to the other. Suggestions in terms of which features of one portal could be replaced and/or completed by the equivalent, but better, one in the other portal, will be expected.

If the project is successful and time allows, you will review the scientific literature in the domain of multilingual knowledge representation in order to propose and set up a state-of-the-art model respecting the best practices in the Semantic Web community to address multilingual issues in BioPortal and CISMeF.

Références :

- [1] Olivier Bodenreider and Robert Stevens. Bio-ontologies: Current Trends and Future Directions. *Briefings in Bioinformatics*, 7(3):256–274, August 2006.
- [2] Stéfan J. Darmoni, Suzanne Pereira, Saoussen Sakji, Tayeb Merabti, Élise Prieur, Michel Joubert, and Benoit Thirion. Multiple Terminologies in a Health Portal: Automatic Indexing and Information Retrieval. In C. Combi, Y. Shahar, and A. Abu-Hanna, editors, *12th Conference on Artificial Intelligence in Medicine, AIME'09*, number 5651 in Lecture Notes in Computer Science, pages 255–259, Verona, Italy, June 2009. Springer.

- [3] Tom R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, June 1993.
- [4] Maurizio Lenzerini. Data Integration: A Theoretical Perspective. In L. Popa, editor, *21st ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, PODS'02*, pages 233–246, Madison, WI, USA, June 2002.
- [5] Natalya F. Noy, Nigam H. Shah, Patricia L. Whetzel, Benjamin Dai, Michael Dorf, Nicholas B. Griffith, Clement Jonquet, Daniel L. Rubin, Margaret-Anne Storey, Christopher G. Chute, and Mark A. Musen. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 37((web server)):170–173, May 2009.
- [6] Daniel L. Rubin, Nigam H. Shah, and Natalya F. Noy. Biomedical ontologies: a functional perspective. *Briefings in Bioinformatics*, 9(1):75–90, 2008.
- [7] Patricia L. Whetzel, Natalya F. Noy, Nigam H. Shah, Paul R. Alexander, Csongor Nyulas, Tania Tudorache, and Mark A. Musen. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research*, 39((web server)):541–545, June 2011.