

Title:

Annotation sémantique de dossiers patients électroniques

Information :

Encadrants : Sandra Bringay (LIRMM, UPV) – sandra.bringay@lirmm.fr
Clement Jonquet (LIRMM, UM) – jonquet@lirmm.fr
Andon Tchechmedjiev (LIRMM) - andon.tchechmedjiev@lirmm.fr

Profil : Computer science or informatics master students

Contexte : [Project SIFR](#) (Semantic Indexing of French Biomedical Data Resources)
[Project PratikPharma](#) (Practice-based evidences for actioning Knowledge in Pharmacogenomics)

Ou : Laboratoire d'Informatique, Robotique, & Microelectronique de Montpellier ([LIRMM](#)).
Une mission à Stanford (Center for Biomecial Inforamtics Research) est à prévoir.

Quand : Février 2017- Juillet 2017

Mots-clés :

Annotation sémantique, ontologies biomédicales, traitement automatique du langage naturel, fouille de texte, web sémantique, dossiers patients électroniques.

Technologies :

BioPortal (SIFR & NCBO), NLP tools (Tree tager, MGREP, UNITEX), Semantic Web technologies (RDF, OWL, SKOS, SPARQL), Web application technologies (Java, JEE, Ruby, Ruby on Rails, Web services API).

Résumé :

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. En collaboration avec l'université de Stanford, nous développons au sein du projet SIFR un workflow d'annotation sémantique de texte avec des ontologies : le SIFR Annotator (<http://bioportal.lirmm.fr/annotator>). Au sein du projet PratikPharma (en collaboration avec le LORIA et l'HEGP), nous allons utiliser cet outil pour annoter des dossiers patients électroniques (en français) afin d'en extraire des informations pharmacogénomiques (relations ternaires variant génétique–médicament–effet indésirable) et les représenter avec des ontologies du domaine.

Le travail du stage consiste dans un premier temps à faire un état de l'art sur l'extraction de connaissances à partir de dossier patients et en particulier faire le point sur ce qui est fait dans le contexte francophone en utilisant des ontologies. Nous nous intéresserons de près aux méthodes du traitement de la langue utilisables pour le français sur ce genre de données. Dans un second temps, le stage consistera à mettre en œuvre certains des propositions dégagées de cet état de l'art au sein du SIFR Annotator et si l'accès aux données le permet, expérimenter et évaluer avec les données du projet PratikPharma. Nous pourrions par exemple nous pencher sur l'outil Unitex (<http://www-igm.univ-mlv.fr/~unitex>) comme alternative au composant utilisé actuellement dans le workflow.

Références:

Clement Jonquet, Amina Annane, Khedidja Bouarech, Vincent Emonet & Soumia Melzi. SIFR BioPortal : **Un portail ouvert et générique d'ontologies et de terminologies biomédicales françaises au service de l'annotation sémantique**, In *16th Journées Francophones d'Informatique Médicale, JFIM'16*. Genève, Suisse, July 2016. pp. 16. [\[PDF\]](#)

Clement Jonquet, Nigam H. Shah & Mark A. Musen. **Un service Web pour l'annotation sémantique de données biomédicales avec des ontologies**, In *13èmes Journées Francophones d'Informatique Médicale, JFIM'09*. Nice, France, April 2009. Informatique et Santé, Vol. 17 [\[PDF\]](#)

Nigam H. Shah, Nipun Bhatia, Clement Jonquet, Daniel L. Rubin, Annie P. Chiang & Mark A. Musen. **Comparison of concept recognizers for building the Open Biomedical Annotator**, *BMC Bioinformatics*. September 2009. Vol. 10 (9:S14), BMC. [\[PDF\]](#)

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 & Mark A. Musen. **BioPortal: ontologies and integrated data resources at the click of a mouse**, *Nucleic Acids Research*. May 2009. Vol. 37 ((web server)), pp. 170-173. [\[PDF\]](#)

Adrien Coulet, Nigam H. Shah, Yael Garten, Mark A. Musen, and Russ B. Altman. Using text to build semantic networks for pharmacogenomics. *Journal of Biomedical Informatics*, 43(6):1009–19, 2010.

Gunnar Declerck, Cedric Bousquet, Iulian Alecu, Agnes Lillo-Le Louet, and Marie-Christine Jaulent.

Knowledge-based approaches in pharmacovigilance: Lessons and prospects for personalized medicine. In *Handbook of Personalized Medicine - Advances in Nanotechnology, Drug Delivery and Therapy*, pages 727–78. Pan Stanford Publishing, 2014.

Y Garten, A Coulet, and RB Altman. Recent progress in automatically extracting information from the pharmacogenomic literature. *Pharmacogenomics*, 11(10):1467–89, 2010.

Peter B. Jensen, Lars J. Jensen, and Soren Brunak. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet.*, 13(6):395–405, 2012.

Hong-Guang Xie and Felix W Frueh. Pharmacogenomics steps toward personalized medicine. *Personalized Medicine*, 2(4):325–37, 2005.

Candidature:

Contactez les encadrants.