Postdoc position available in Montpellier, France
(Multilingual semantic annotations of Electronic Health Records)
Fall 2016

Title:

Multilingual semantic annotations of Electronic Health Records and Pharmacogenomics data with ontologies.

Information:

Type:	Postdoc position
Employer:	University of Montpellier
Context:	PratikPharma project (**Pr**actice-based evidences for **acti**oning **K**nowledge in **Pharma**cogenomics) – ANR project
When:	September 2016 – for 24 months (12 months renewable)
Where:	LIRMM, Montpellier (requires collaboration with Stanford University, USA.))

Keywords & background:

semantic annotation, biomedical ontologies & terminologies, multilingual context (French & English), text mining, semantic web, natural language processing, medical informatics, knowledge extraction.

Abstract:

The goal of the PractiKPharma project (http://practikpharma.loria.fr) is to validate or moderate Pharmacogenomics state-of-the-art knowledge on the basis of practice-based evidences, i.e., knowledge extracted from Electronic Health Records. During this project, we will extract state-of-the-art knowledge from (English) structured and unstructured descriptions in reference databases (e.g., PharmGKB) and literature (e.g., PubMed) as well as extract observational knowledge from (French) EHRs. Part of this multilingual knowledge extraction process will be based on semantic annotation (using relevant biomedical ontologies) of plain-text data. We plan to reuse and enhance tools developed in the context of the NCBO (www.bioontology.org) and SIFR projects (www.lirmm.fr/sifr).

We are seeking a motivated, curious and interested postdoc candidate to design and develop the semantic annotation workflows and help the project to annotate HER and Pharmacogenomics data. The postdoc develop new methodologies to capture the context of French clinical narrative, such as negations, specific sections, modality and word sense disambiguation. To support this French-English context we will also investigate the generation and use of multilingual ontology mappings (mostly reusing LIRMM's YAM++ approach).

Context:

Pharmacogenomics (PGx) studies how individual gene variations cause variability in drug responses and constitutes a basis for implementing personalized medicine i.e., a medicine tailored to each patient by considering her/his genomic context [8]. PGx data is often not yet validated because most of it results from studies that do not fulfill statistics validation standards and are difficult to reproduce because of the rarity of gene variations studied. The goal of the PractiKPharma project is to validate PGx state-of-the-art knowledge publicly available (in English) on the basis of practice-based evidences, i.e., knowledge extracted from EHRs (in French). The project is mainly funded by the French ANR and leaded by A. Coulet (INRIA Nancy) in collaboration with HEGP (Georges Pompidou European Hospital, Paris), CHU Saint-Etienne and the LIRMM.

Units of knowledge in PGx typically have the form of ternary relationships gene variant–drug–adverse event, and can be formalized to different extents using biomedical ontologies. To achieve our goal, we will (1) extract state-of-the-art knowledge from PGx databases, (such as PharmGKB [3]) and literature (such as PubMed); (2) extract observational knowledge (i.e., knowledge extracted from observational data) from French EHRs, (3) to compare knowledge units extracted from these two origins, to confirm or moderate state-of-the-art knowledge, with the goal of enabling personalized medicine. (4) Finally, we will emphasize newly confirmed knowledge by investigating omics databases for molecular mechanisms that underlie and explain drug adverse events use biomedical Linked Open Data [1].

PractiKPharma
Practice-based evidences for actioning Knowledge in Pharmacogenomics

The clarification of the validity of PGx data will have an important impact on clinical care. This would permit the definition of guidelines that enable clinicians to implement personalized medicine by choosing and dosing drugs more precisely. Concretely, this will reduce the toxicity and increasing the efficacy of prescribed drugs, consequently reducing cost and improving quality of care.

Postdoc description:

Within this project, we (task leaded by LIRMM) will design and implement a semantic annotation workflow that will use the relevant reference ontologies to facilitate the knowledge extraction process. We plan to reuse and improve the outcomes of the National Center for Biomedical Ontology (www.bioontology.org - NCBO) and the Semantic Indexing of French Biomedical Data Resources (www.lirmm.fr/sifr - SIFR) projects that have developed tools for semantic annotations respectively for English and French data [5] [4].

The biggest improvements will come from introducing more natural language processing (NLP) mechanisms in the annotation process. We will specifically work on better capturing the context of French clinical or patient narrative. The new workflow improvements will include:

- to identify polysemic terms during the annotation process and choose the proper concept (e.g., cell, the biological component or telephone).

- to detect negation (e.g., the patient does not have the symptom) for instance using state-of-the-art results in the domain such as NegEx [2] already tested by NCBO team, and inventing others.

- to detect, to some extent, elements of context, by detecting modulators words (hypothetically, strongly) to associate annotations with different levels of importance and detect time information.

- the publication and link of the annotations produced by our workflow in the Web of data using standard semantic Web technologies & the new W3C Web Annotation Model standard.

- to improve the annotation scoring, which allows to rank annotations by importance, related to the context in which the annotation has been done and to the frequency of the matches [6].

- to make the workflow handles multilingual data and offers annotations with multilingual ontologies by leveraging multilingual mappings. For this purpose, multilingual mappings will have to be generated between English ontologies and their French counterparts.

The annotation of EHRs data will have to be run in-house at the HEGP in collaboration with the team there. The work done on text mining and semantic annotation in French, will be generalized to English and improvements of the SIFR annotator will be incorporated into the NCBO annotator when possible in collaboration with Stanford University. Multiple visits are planned. The work on ontology alignment will capitalize on the work of the OpenData team at LIRMM (e.g., on YAM++ [7]) and involve Zohra Bellahsene & Konstantin Todorov.

The ontologies used will include: SNOMED-CT, NCIt, HDO, HPO, MESH, LOINC, RXNORM, WHO-ATC, MEDRA

For reference, SIFR Annotator code and API are available here:

https://github.com/sifrproject & http://data.bioportal.lirmm.fr/documentation

Expected profile:

We are seeking a motivated, curious and interested postdoc candidate to design and develop the semantic annotation workflow and bring interesting new ideas to the project team. A computer science or bioinformatics PhD degree is required. Besides an important motivation for the research questions, we are also looking for someone with some good technical skills and motivation for concrete outcomes. The best candidate will demonstrate good programming skills in addition of a good track record of publications. The supervision will be done mostly remotely as Clement Jonquet is currently visiting scholar at Stanford University. The candidate will demonstrate aptitudes or matches with some of the following aspects:

- Research experience that match with the proposed subject.

- Experience with semantic some of the technologies involved: Java/JEE, Ruby/Rails, RESTful web services, XML/JSON.

- Experience with the semantic web vision (& technologies OWL, RDF, SPARQL)

Postdoc position available in Montpellier, France
(Multilingual semantic annotations of Electronic Health Records)
Fall 2016

PractiKPharma
Practice-based evidences for actioning Knowledge in Pharmacogenomics

- Experience in biomedical informatics (knowledge extraction, use of ontologies, BioPortal)
- Good track records in terms of publications and communication of his/her work.
- Excellent remote working capabilities (emails, trackers, collaborative tools, etc.)
- Perfect English oral and writing skills.
- Few knowledge in French language with objective to learn the language during the contract.
- Multiple project meetings are planned in France (Paris or Nancy),
- International trips accepted (collaboration with Stanford) and the being eligible for a visa for the USA.
- Excellent writing skills as reports, publications, and technical notes will always be necessary.
- Autonomy and initiative, take on technical decisions within the project and justify choices.
- Friendly person to join a small research team in Montpellier.

Application:

For more information about this position, please contact Clement Jonquet (jonquet@lirmm.fr). To apply, please send an email including links to (NO ATTACHED DOCUMENTS) the following:

- a motivation letter describing an explanation of YOUR interest for the position;
- a curriculum vitae describing your experience and the matches with the expected profile;
- copies of diplomas (PhD) and other relevant certificates (MSc grades, PhD jury evaluation);
- names and contact details of referees.

Contract:

- The postdoc will be hired by the University of Montpellier (social security, etc. included).
- Salary will be around 2000€ net per month depending on experience.
- Contract should start September 1st 2016 or October 1st 2016.
- The contract will be for 1 year and renewable another year.

References:

[1] C. Bizer, T. Heath, and T. Berners-Lee. Linked Data - The Story So Far. *Semantic Web and Information Systems*, 5(3):1–22, 2009.

[2] W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Biomedical Informatics*, 34(5):301–310, October 2001.

[3] M. Hewett, D. E. Oliver, D. L. Rubin, K. L. Easton, J. M. Stuart, R. B. Altman, and T. E. Klein. PharmGKB: the pharmacogenetics knowledge base. *Nucleic acids research*, 30(1):163–165, 2002.

[4] C. Jonquet, A. Annane, K. Bouarech, V. Emonet, and S. 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, July 2016.

[5] C. Jonquet, N. H. Shah, and M. A. Musen. The Open Biomedical Annotator. In *American Medical Informatics Association Symposium on Translational BioInformatics, AMIA-TBI'09*, pages 56–60, San Francisco, CA, USA, March 2009.

[6] S. Melzi and C. Jonquet. Scoring semantic annotations returned by the NCBO Annotator. In A. Paschke, A. Burger, P. Romano, M. Marshall, and A. Splendiani, editors, *7th International Semantic Web Applications and Tools for Life Sciences, SWAT4LS'14*, Berlin, Germany, December 2014.

[7] D. Ngo and Z. Bellahsene. YAM++ : A Multi-strategy Based Approach for Ontology Matching Task. In A. ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Acquin, A. Nikolov, N. Aussenac-Gilles, and N. Hernandez, editors, *18th International Conference on Knowledge Engineering and Knowledge Management,EKAW'12*, p. 421–425, Galway, Irland, 2012. Springer.

[8] H.-G. Xie and F. W. Frueh. Pharmacogenomics steps toward personalized medicine. *Personalized Medicine*, 2(4):325–337, 2005.