Semantic Indexing of French Biomedical Data Resources - mobility

HORIZON 2020

Semantic Indexing of French Biomedical Data Resources - mobility

Results in Brief

An open web tool for annotation of biomedical text in the French language

Considering the constant increase in the volume of biomedical data, there is an imminent need for using standard terminologies and ontologies (semantic indexing) to guide future data-driven scientific discoveries. An open web tool developed by the SIFRm European project provides annotation of biomedical text in French texts.



© kentoh, Shutterstock

Scientists employ ontologies and terminologies to index, mine and retrieve biomedical information. The use of ontologies and terminologies serves as a common denominator to structure biomedical data. However, most tools are in the English language and despite the large amount of clinical data produced in French, there is little readily available technology.

A biomedical annotation tool in the French language

Undertaken with the support of the Marie Skłodowska-Curie (MSCA) programme, the aim of the www.lirmm.fr/sifr (SIFRm) project was to build an ontology-based indexing workflow specialised for other EU languages, starting with French. "Our main goal was to make annotation of biomedical text data available at the click of a mouse to free researchers from the burden of dealing with terminologies and ontologies or natural language processing," explains the MSCA fellow Clement Jonquet.

SIFRm was a collaboration between Professor Cerri's team at the Laboratory of Informatics, Robotics and Microelectronics of Montpellier (LIRMM) in France and Professor Musen's team at the Stanford Center for Biomedical Informatics Research (BMIR) in the United States, renowned for the development of ontology-based services.

Researchers built the <u>SIFR Annotator</u>, a publicly accessible web service that enables the processing of biomedical text data in French. The annotator essentially tags raw text with relevant biomedical ontology concepts and semantically expands the annotations using the knowledge embedded in the ontologies. For instance, if a clinical note contains the sentence 'no sign of melanoma', semantic annotation will help to classify the patient as not relevant for cancer studies.

To support the service, the project has developed the <u>SIFR BioPortal</u> ontology repository. Similar to the <u>NCBO BioPortal</u> technology developed at Stanford University, SIFR BioPortal hosts different terminologies and ontologies in French, offering multiple ontology-related services to the community.

Annotating clinical data and agronomical entities

In collaboration with the <u>PractiKPharma</u> project, the SIFR Annotator has been enriched to process clinical data and contextualise medical conditions in clinical notes. Scientists developed specific features for the annotation of clinical text, addressing the need of the European Hospital Georges Pompidou and the Nancy University Hospital.

Furthermore, SIFRm generalised the scientific methods to build an open repository for agronomical ontologies called <u>AgroPortal</u>, a community effort initiated by the Montpellier scientific community and finalised through the mobility of the researcher to Stanford.

Based on the scientific outcomes and experience of the biomedical domain, scientists developed AgroPortal for agronomy and related domains such as food, plant sciences and biodiversity. "AgroPortal addresses the need for a common platform to host, serve and align semantic resources available in this domain, allowing their exploitation in agro-informatics applications," reports Jonquet.

The AgroPortal repository currently hosts over 110 vocabularies or ontologies and will be further enriched in the near future. The platform already has more than 190 registered users with frequent visits every month.

Overall, the SIFRm project provided the first openly accessible web tool to recognise entities and annotate and contextualise French biomedical text. The web service performs comparably well to other annotation platforms and is expected to improve the work of a wide range of scientists, including clinicians, health professionals and researchers.

Plans for future partnerships with hospitals and research centres in France will expand the use of the SIFR Annotator in biomedical research. In a similar effort, the AgroPortal tool will be used in the www.d2kab.org (D2KAB) project primarily funded by the French National Research Agency

Keywords

SIFRm, biomedical, ontology, annotator, AgroPortal, BioPortal, French language, clinical data, indexing

Discover other articles in the same domain of application



Advancing long-term digital preservation of scientific data





It's double the fun when supercomputing meets quantum simulation





The rehabilitation robot finding work outside the hospital





Getting royalties to film content creators



Project Information

SIFRm

Grant agreement ID: 701771

Project website 🗹

DOI 10.3030/701771

Project closed

EC signature date 29 April 2016

Start date 1 September 2016 End date 15 October 2019

Last update: 5 May 2020

Funded under EXCELLENT SCIENCE - Marie Skłodowska-Curie Actions

Total cost € 264 668,40

EU contribution € 264 668,40

Coordinated by UNIVERSITE DE MONTPELLIER France **Permalink:** <u>https://cordis.europa.eu/article/id/415957-an-open-web-tool-for-annotation-of-biomedical-text-in-the-french-language</u>

European Union, 2025