Semantic Annotation Workflow using Bio-Ontologies

Institut de biologie computationnelle

Castanier E.[1], Jonquet C.[2], Melzi S.[2], Larmande P.[3], Ruiz M.[4], Valduriez P.[1]

1 – Zenith, INRIA-LIRMM
2 – SMILE, UM2-LIRMM
3 – DIADE, IRD
4 – AGAP, CIRAD

Montpellier, France

Introduction

Biologists have adopted ontologies:
- To provide canonical representation of scientific knowledge
- To annotate experimental data to enable interpretation, comparison, and discovery across databases
- To facilitate knowledge-based applications for decision-support, natural language processing, and data integration

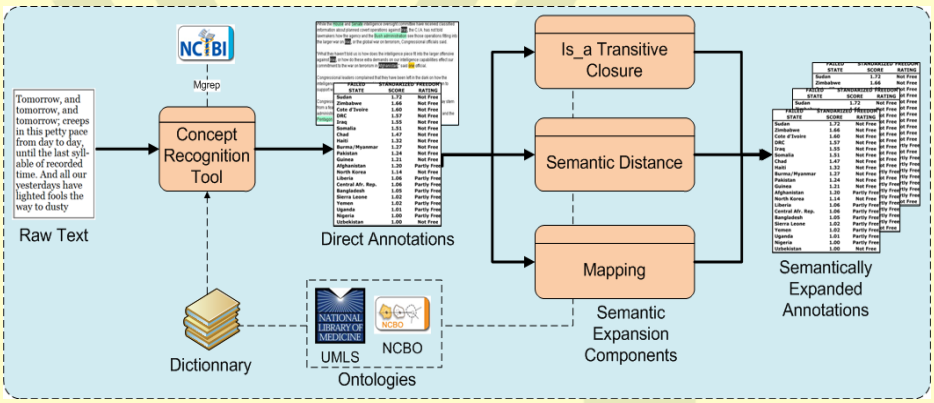
But **off-the-shelf solutions for the biologist** to use ontologies are rare (versions, format, availability, license, overlap, etc.)

The challenge

Automatically process complex biological resources text content and generate annotations :
- Large-scale – to scale up to many resources and ontologies
- Automatic – to keep precision and accuracy
- Easy to use and to access – web service approach
- Customizable – to fit very specific needs
- Smart – to leverage the knowledge contained in ontologies

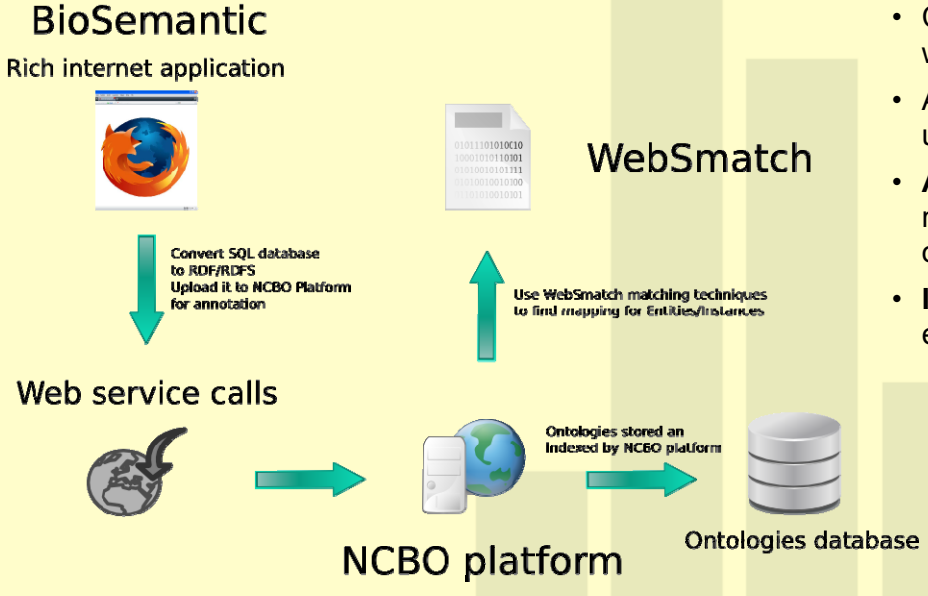
There have been **success stories to reproduce**: GO annotations, PubMed indexing, etc.

NCBO Annotator: Ontology-based annotation workflow



- First, direct annotations are created by **recognizing concepts** in raw text.
- Second, annotations are **semantically expanded** using knowledge of the ontologies.
- Third, all annotations are **aggregated and scored** according to the context in which they have been created.

Customized IBC Annotator for database schemas



- Convert **SQL database schemas to RDF/RDFS** with BioSemantic
- Annotate with NCBO Annotator and WebSmatch using customized NCBO services
- **Annotator relies on WebSmatch** to create mappings between elements of schemas and ontological concepts
- **Indexing IBC related data** with the workflow to enhance semantic search and mining of data

R. Coletta, E. Castanier, P. Valduriez, C. Frisch, D-H. Ngo, Z. Bellahsene: Public data integration with WebSmatch. *Workshop on Open Data*. 2012, pp. 5-12.

C. Jonquet, N. H. Shah, M. A. Musen. **The Open Biomedical Annotator**, In *AMIA Symposium on Translational BioInformatics*. 2009. pp. 56-60.

C. Jonquet, P. LePendu, S. Falconer, A. Coulet, N. F. Noy, M. A. Musen, N. H. Shah. **NCBO Resource Index: Ontology-Based Search and Mining of Biomedical Resources**, *Web Semantics*. 2011. Vol. 9 (3), pp. 316-324.

J. Wollbrett, P. Larmande, F. De Lamotte, M. Ruiz. **Clever generation of rich SPARQL queries from annotated relational schema: application to Semantic Web Service creation for biological databases.** *BMC Bioinformatics*. 2013; 14:126.

In collaboration with SIFR project
www.lirmm.fr/sifr
THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

cirad LA RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT
IRD Institut de recherche pour le développement
LIRMM Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier
Inria informatics mathematics