Semantic Annotation Workflow using Bio-Ontologies
Emmanuel Castanier, Clement Jonquet, Soumia Melzi, Pierre Larmande, Manuel Ruiz, Patrick Valduriez

To cite this version:

HAL Id: lirmm-01052546
https://hal-lirmm.ccsd.cnrs.fr/lirmm-01052546
Submitted on 28 Jul 2014
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