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

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Semantic Annotation Workflow using Bio-Ontologies

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

BioSemantic
Rich internet application

WebSmatch

• 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