'Watskeburt': heuristic support for hypothesis construction from literature

Marco Roos[a], Sophia Katrenko, Edgar Meij, Willem van Hage, Frans Verster, Scott Marshall, Pieter Adriaans

Adaptive Information Disclosure, Institute of Informatics, Faculty of Science, University of Amsterdam, The Netherlands.
(a) roos@science.uva.nl

Introduction

Our objective is to provide automated support for hypothesis construction from literature

We start with a seed of knowledge, a 'proto-ontology', that we want to extend, for instance information from a review about histones and disease.
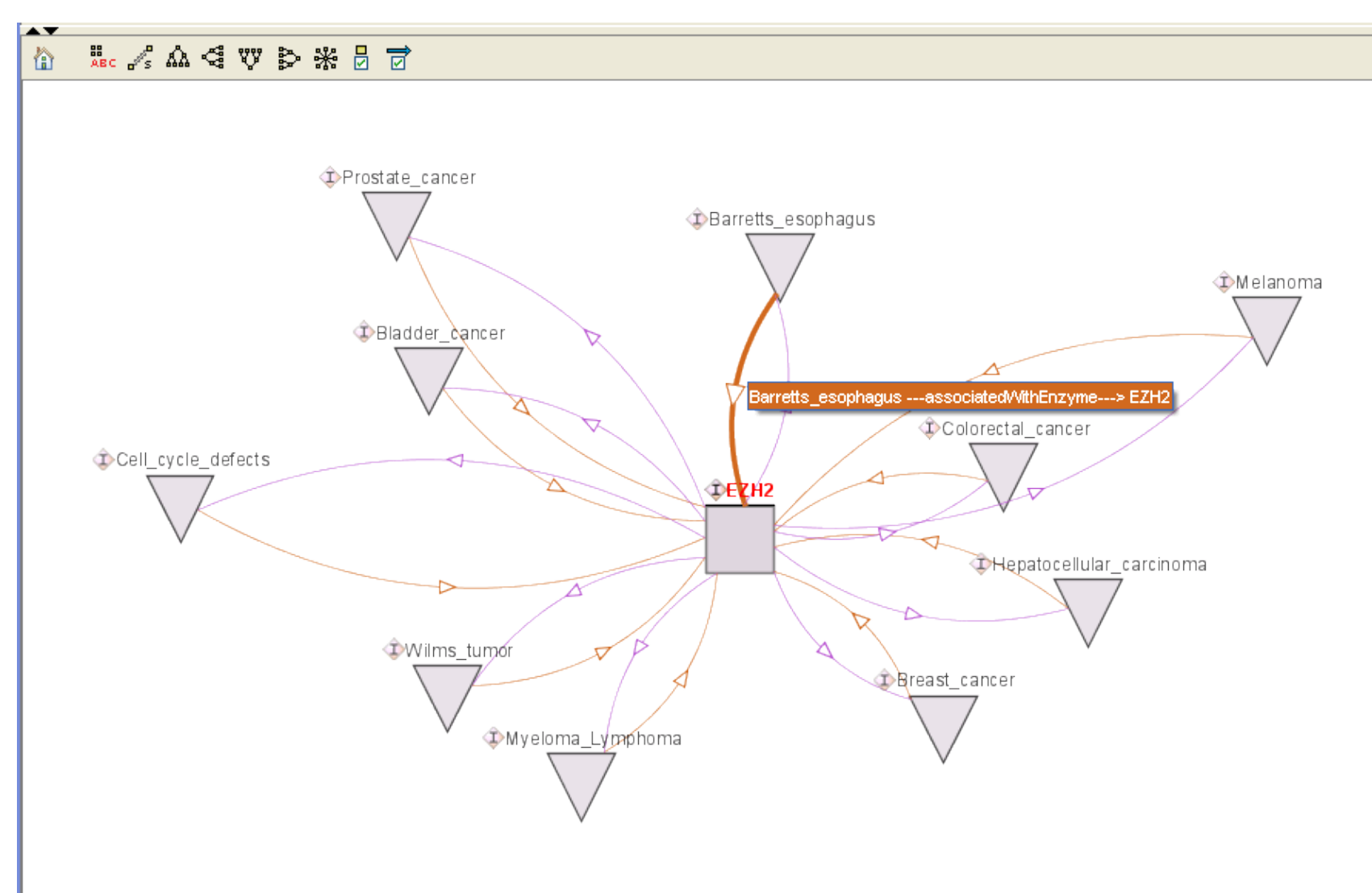

Example proto-ontology

This proto-ontology represents reported relationships between Enhancer of Zeste and various diseases. As an example we extend this ontology with hypothetically related genomics terms discovered in MEDLINE abstracts.

>Snapshot of Jambalaya plugin for Protégé/OWL

'Watskeburt' workflow

We implemented 'Watskeburt' using the workflow enactor tool Taverna*, services it supports, and services from our AIDA toolbox. One version imports the proto-ontology's terms from an excel file, retrieves relevant papers, and then identifies genomics entities based on a pre-learned genomics model.
*http://taverna.sourceforge.net
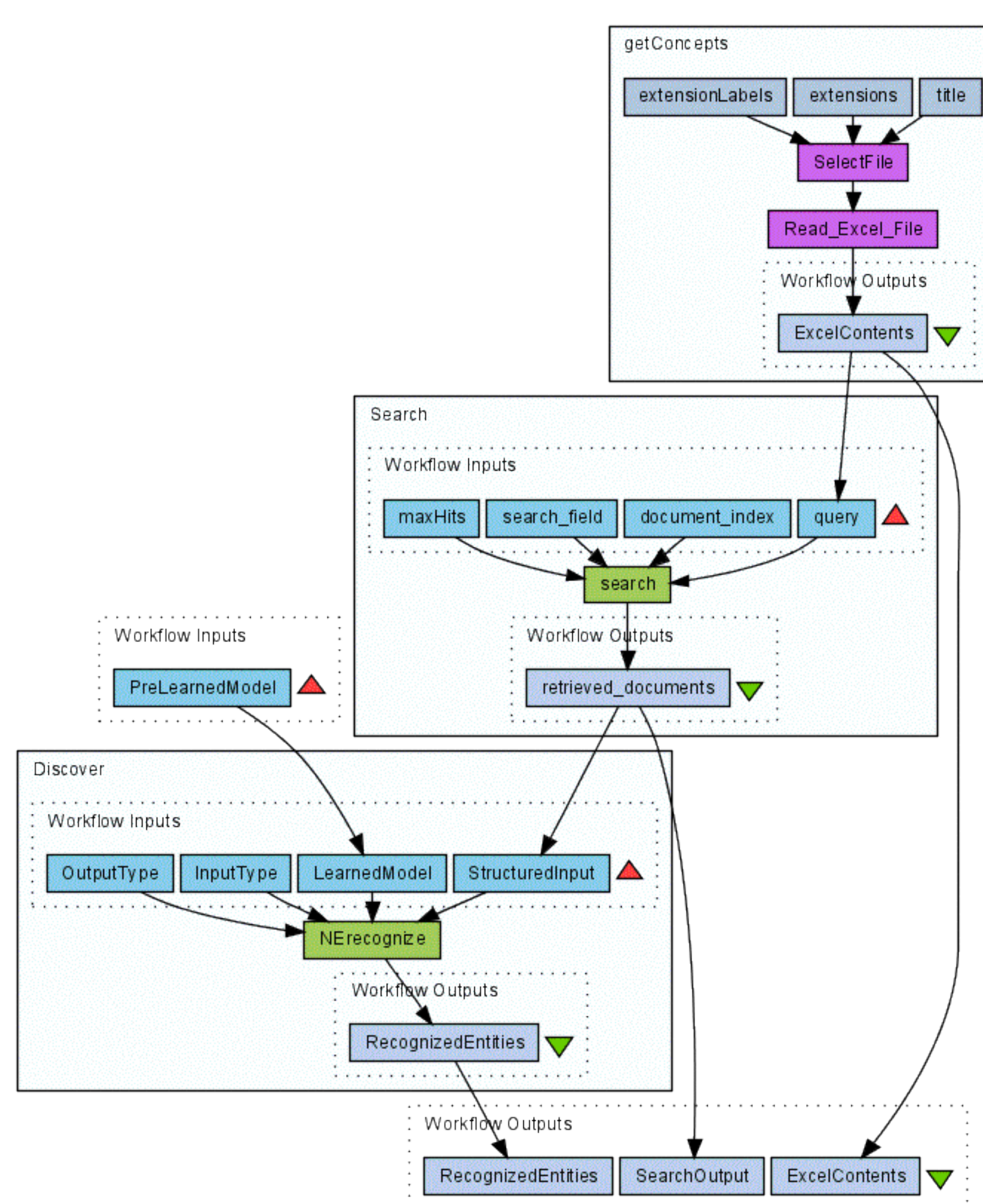
The results are tagged genomics entities that were re-imported into the ontology tool Protégé/OWL.
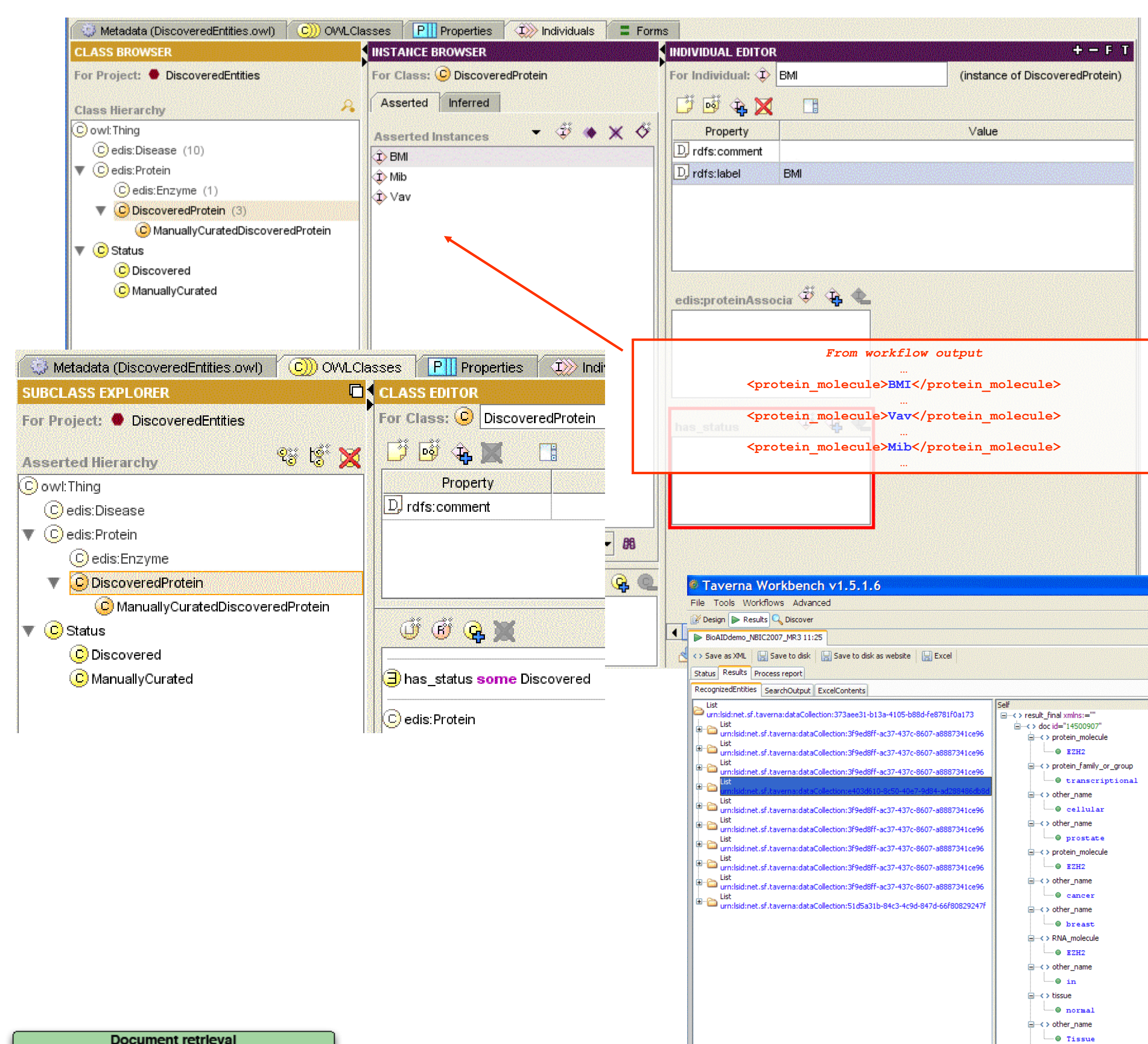
The AIDA toolbox

An alternative to building special purpose bioinformatics tools is to use *web services* within a general computing environment such as the workflow tool *Taverna*.

'Watskeburt' applies this 'e-science' approach for text mining. Its components come from our 'AIDA* toolbox', which contains services that can be flexibly combined for various applications, including text mining.

*Adaptive Information Disclosure Application
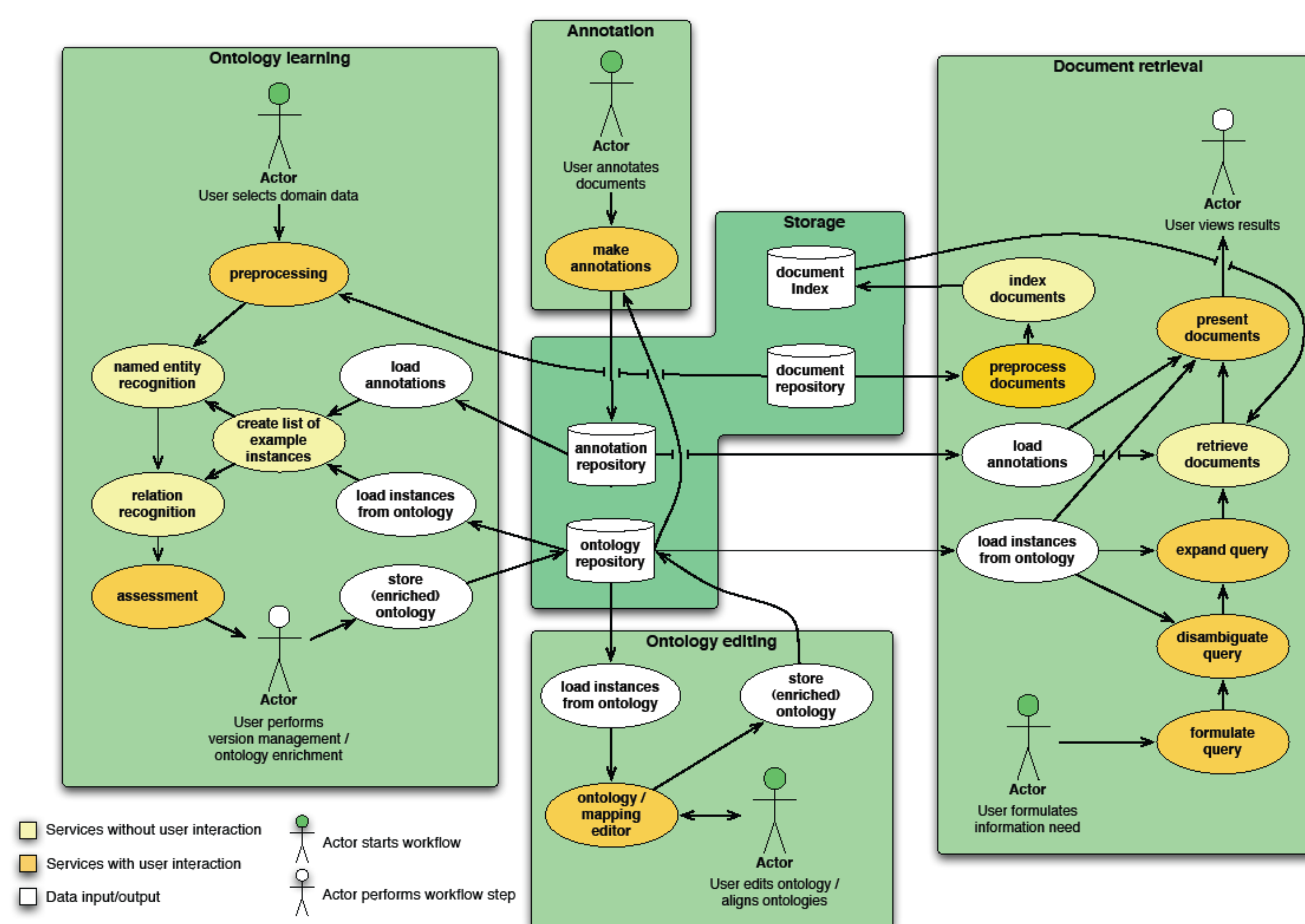
Discussion

'Watskeburt' supports automated hypothesis construction by enriching an ontology with hypothetically related elements discovered in literature.
We can explore various text-mining strategies by varying the workflow, for instance to discover proteins that putatively link two separate topics of interest.

Future work includes optimization of the services and 'personalisation' by machine learning services that allow to recognize concepts from ones own ontology.