

The Cell Cycle Ontology: an application ontology supporting the Life Sciences

Erick Antezana^{1,2}, Mikel Egaña³, Ward Blondé^{1,2}, Robert Stevens³, Bernard De Baets⁴, Vladimir Mironov⁵, and Martin Kuiper⁵

¹ Dept. of Plant Systems Biology, VIB, Gent, Belgium

² Dept. of Molecular Genetics, Ghent University, Gent, Belgium

³ School of Computer Science, The University of Manchester, UK

⁴ Dept. of Applied Mathematics, Biometrics and Process Control, Ghent University, Gent, Belgium

⁵ Dept. of Biology, Norwegian University of Science and Technology, Norway

<http://www.CellCycleOntology.org>

Abstract

The terms and relationships provided by existing bio-ontologies only capture a small part of our biological understanding, thus the potential of applying computational analysis on such information remains limited. The Cell Cycle Ontology (CCO) is designed to capture detailed information of the cell cycle process by combining representations from several sources. CCO is an application ontology that is supplied as an integrated turnkey system for exploratory analysis, advanced querying, and automated reasoning. CCO supports four model organisms (Human, Arabidopsis, Baker's yeast and Fission yeast) with separate ontologies but also one integrated ontology. CCO holds more than 65000 concepts and more than 20 types of relationships. CCO comprises data from existing resources such as the Gene Ontology (GO), the Relations Ontology (RO), the IntAct database (MI), the NCBI taxonomy, the UniProt knowledge base as well as orthology data. An automatic pipeline builds CCO from scratch periodically: initially some existing ontologies (GO, RO, MI, in-house ones) are automatically fetched, integrated and merged, producing in turn a core cell cycle ontology. Then, organism-specific protein and gene data are added from UniProt and from the GO Annotation files, generating four organism-specific ontologies. Those four ontologies are merged and more terms are included from an ontology built automatically from the OrthoMCL execution on the cell cycle proteins. Finally, during the maintenance phase, a semantic improvement on the OWL version is carried out: ontology design patterns are included using the Ontology Pre-Processor Language. The resulting CCO is designed to provide a richer view of the cell cycle regulatory process, in particular by accommodating the intrinsic dynamics of this process. CCO is available in: OBOF, RDF, XML, OWL, GML, and DOT. A SPARQL endpoint allows building complex queries, such as "get the cell cycle related proteins in *A. thaliana* participating in the same interaction but having different locations". Visual exploration can be done via the BioPortal, the Ontology Lookup Service, the Ontology Online service, or the DIAMONDS platform.