From Gene Ontology to Cell Ontology: interoperability among biological knowledge bases

ID - 228

Calabrese Antonio[1], Ammaturo Laura[2], Mele Francesco[1], Talamo Oliviero[1], Tino Angela[1], Tortiglione Claudia[1]

[1]Consiglio Nazionale delle Ricerche, Istituto di Cibernetica, Pozzuoli

[2]Università degli Studi Federico II, Napoli

Motivation

Recently in bioinformatics specific methodologies have been borrowed from the field of ontologies to build representation schemata for biological knowledge bases. It is therefore necessary to propose adequate taxonomies of concepts and representations of meronimic relations and it is also necessary to assure that every new representation is interoperable with the existing ones in the field.

Gene Ontology (GO) is a controlled vocabulary that describes the roles of genes and proteins in all organisms. Analyzing Gene Ontology we detected various representation anomalies beside those already put in evidence in literature. In GO the is-a relationship is frequently misused to represent an unambiguous part-of relationship.

A rigorous ontology represents these relationships as non-hierarchical relationships among classes (the chromatophore membrane isnt a kind of chromatophore; it is a chromatophore component). Moreover the relations part-of are used to represent both essential parts and not essential parts: the relation part-of (membrane, cell) means that every cell has a membrane and part-of (flagellum, cell) means that a flagellum is a part of some cells. Another anomaly of part-of relations in GO, originates from the fact that every type of meronomic relation is represented in the same way, and this produce the typical incongruity when transitive rules are applied. Therefore we constructed a new ontology that not presents these anomalies and that proposes a proper representation for cellular biology knowledge bases.

Methods

The proposed ontology, that we called Cell Ontology (CO), includes all the taxonomic relations already existing in GO, and presents new explicit spatial and temporal relations to represent existing relations in the domain of cell biology. Starting from this new sets of relations, formalized as ontology classes, we also defined rules for meronomic reasoning. We implemented these rules axiomatically, therefore they are also applicable to other knowledge domains.

The peculiar features of Cell Ontology (CO) are the following: 1. the main classes of CO (Molecular_Function, Biological_Process and Cellular_Component) are represented as explicit classes (in GO they were implemented as implicit metaclasses); 2. all meronomic relations, most of which are not present in GO, are explicitly defined (component-totality relations (part-of), functional relations, material-object relations (made_of), phases relations). They are all included in a single class of relations; 3. explicit temporal qualitative relations are introduced; 4. explicit qualitative and quantitative spatial relations; 5. redundancies detected in GO (i.e. classes that are subclasses of different classes) are not present in CO.

Once built the new ontology, we evaluated the possibility of using data present both in Gene Ontology and in Cell Ontology. We not considered Data Migration from GO to CO in that it would oblige GO users to change querying syntax and data insertion modality. We introduced a specific interoperability between the two ontologies, allowing to use equivalent access operations, also with different query languages, when the same information, represented by means of different metadata, is searched.

In this way the GO user is allowed not only to continue to use its ontology but also to access to the information entered in the new ontology. Obviously also CO user may use information contained both in GO and in CO.

Moreover we realized an interface allowing different user typologies to access knowledge in both ontologies. The different kind of user considered are the following: (a) old GO users, (b) CO users and (c) new users of both ontologies. Such an interface was constructed introducing specific bridge rules between various representations and typical user access languages (a), (b) and (c). Such bridge rules were implemented taking advantage of a Flora2 property, the specific language of the Frame Logic used, allowing to separately define in modules the various representations (a kind of contexts defined by local facts and rules). The modules exchange data through XSB Prolog/Flora2 shared variables as in traditional Logic Programming languages.

Results

We detected some new representation anomalies in Gene Ontology beside the ones already reported in

literature. We constructed a new ontology, named Cell Ontology, that not exhibit these representation anomalies and introduces new concepts to represent molecular biology knowledge.
We also built an user interface that allow to share (provide and harvest) metadata records (interoperability) among different biological ontologies (in the specific case between Gene Ontology and Cell Ontology). The software architecture of this interface allows interoperability between data schemata of different ontologies and was built using some libraries of XSB Prolog and Flora2: Interprolog and Java2Flora.
Email: a.calabrese@cib.na.cnr.it