Chapter 4

BIOLOGICAL ONTOLOGIES

Patrick Lambrix, He Tan, Vaida Jakoniene and Lena Strömbäck
Department of Computer and Information Science, Linköpings Universitet, Sweden

Abstract: Biological ontologies define the basic terms and relations in biological domains and are being used among others, as community reference, as the basis for interoperability between systems, and for search, integration and exchange of biological data. In this chapter we present examples of biological ontologies and ontology-based knowledge, show how biological ontologies are used and discuss some important issues in ontology engineering.

Key words: ontologies, ontology alignment, ontology-based search, ontology development, Gene Ontology (GO), Medical Subject Headings (MeSH), Open Biomedical Ontologies (OBO).

1. INTRODUCTION

Intuitively[4], ontologies can be seen as defining the basic terms and relations of a domain of interest, as well as the rules for combining these terms and relations [24]. Many ontologies have already been developed and are used in several areas, including bioinformatics and systems biology [27,15,17]. They are considered to be an important technology for the Semantic Web (e.g. [18,30,23]). They are used for communication between people and organizations by providing a common terminology over a domain. They provide the basis for interoperability between systems. They can be used for making the content in information sources explicit and serve as an index to a repository of information. Further, they can be used as a basis for integration of information sources and as a query model for

---

[4] For discussions of different definitions of ontologies we refer to [10,9].

information sources. They also support clearly separating domain knowledge from application-based knowledge as well as validation of data sources. The benefits of using ontologies include reuse, sharing and portability of knowledge across platforms, and improved maintainability, documentation, maintenance, and reliability (e.g. [36]). Overall, ontologies lead to a better understanding of a field and to more effective and efficient handling of information in that field. As an example, in Figure 4-1 we see two small pieces from two ontologies, Adult Mouse Anatomy (MA) and Medical Subject Headings (MeSH), representing knowledge about the nose. The *i* symbols in MA denote is-a relationships while the *p* symbols denote part-of relationships. The – symbols in MeSH can denote is-a and part-of relationships. The terms in bold face in MA and MeSH, respectively, that are connected with a dashed line denote equivalent terms.

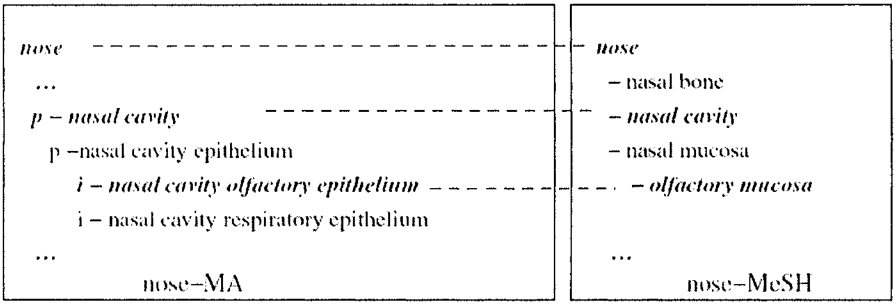


Figure 4-1. Example ontologies.

Although ontologies have been around for a while, it is only during the last decade that the creation and use of biological ontologies have emerged as important topics. The work on biological ontologies is now recognized as essential in some of the grand challenges of genomics research [6] and there is much international research cooperation for the development of biological ontologies (e.g. Open Biomedical Ontologies (OBO)) and the use of biological ontologies for the Semantic Web (e.g. the EU Network of Excellence REWERSE). The number of researchers working on methods and tools for supporting ontology engineering is constantly growing and more and more researchers and companies use ontologies in their daily work.

The use of biological ontologies has grown drastically since database builders concerned with developing systems for different (model) organisms joined to create the Gene Ontology (GO) Consortium in 1998 [7]. The goal of GO was and still is to produce a structured, precisely defined, common and dynamic controlled vocabulary that describes the roles of genes and proteins in all organisms. Another milestone was the start of Open Biomedical Ontologies as an umbrella Web address for ontologies for use within the genomics and proteomics domains [25]. The member ontologies are required to be open, to be written in a common syntax, to be orthogonal to each other, to share a unique identifier space and to include textual definitions. Many biological ontologies are already available via OBO. The field has also matured enough to start talking about standards. An example of this is the organization of the first conference on Standards and Ontologies for Functional Genomics (SOFG) in 2002 and the development of the SOFG resource on ontologies [35]. Further, in systems biology ontologies are used more and more, for instance, in the definition of standards for representation and exchange of molecular interaction data.

In this chapter we give an overview of the area of biological ontologies. First, as a background, we introduce a characterization of ontologies based on the kind of information they can represent (section 2). In section 3 we present OBO as well as some types of biological ontologies. We show how biological ontologies are used (section 4) and discuss some important issues in ontology engineering (section 5). In addition to the biological ontologies other ontology-related knowledge is available and can be used for search, integration and analysis of data. Section 7 presents this knowledge.

2. CHARACTERIZATION OF ONTOLOGIES

Ontologies differ regarding the kind of information they can represent. From a knowledge representation point of view ontologies can have the following components (e.g. [36,17]). *Concepts* represent sets or classes of entities in a domain. For instance, in Figure 4-1 nasal cavity represents all the things that are nasal cavities. The concepts may be organized in taxonomies, often based on the is-a relation or the part-of relation. *Instances* represent the actual entities. They are, however, often not represented in ontologies. Further, there are many types of *relations*. For instance, one type is the group of taxonomic relations such as the specialization relationships (e.g. nasal cavity olfactory epithelium is-a nasal cavity epithelium) and the partitive relationships (e.g. nasal cavity part-of nose). Finally, *axioms* represent facts that are always true in the topic area of the ontology. These can be such things as domain restrictions (e.g. the origin of a protein is always of the type gene coding origin type), cardinality restrictions (e.g. each protein has at least one source), or disjointness restrictions (e.g. a helix can never be a sheet and vice versa).

Ontologies can be classified according to the components and the information regarding the components they contain. A simple type of ontology is the *controlled vocabulary*. These are essentially lists of concepts. When these concepts are organized in an is-a hierarchy, we obtain a *taxonomy*. A slightly more complex kind of ontology is the *thesaurus*. In this case the concepts are organized in a graph. The arcs in the graph represent a fixed set of relations, such as synonym, narrower term, broader term, similar term. The data models allow for defining a hierarchy of classes (concepts), attributes (properties of the entities belonging to the classes, functional relations), relations and a limited form of axioms. The *knowledge bases* are often based on a logic. They can contain all types of components and provide reasoning services such as checking the consistency of the ontology.

An ontology and its components can be represented in a spectrum of representation formalisms ranging from very informal to strictly formal [15]. In general, the more formal the used representation language, the less ambiguity there is in the ontology. Formal languages are also more likely to implement correct functionality. Furthermore, the chance for interoperation is higher. In the informal languages the ontology content is hard-wired in the application. This is not the case for the formal languages as they have a well-defined semantics. However, building ontologies using formal languages is not an easy task.

In practice, biological ontologies have often started out as controlled vocabularies. This allowed the ontology builders, which were domain experts, but not necessarily experts in knowledge representation, to focus on the gathering of knowledge and the agreeing upon definitions. More advanced representation and functionality was a secondary requirement and was left as future work. However, some of the biological ontologies have reached a high level of maturity and stability regarding the ontology engineering process and their developers have now started investigating how the usefulness of the ontologies can be augmented using more advanced representation formalisms and added functionality. Moreover, some recent efforts, such as FungalWeb Ontology [1], have started out immediately as knowledge bases.

3. EXAMPLES OF BIOLOGICAL ONTOLOGIES

There are many biological ontologies. They differ in the type of biological knowledge they describe, their intended use, the level of abstraction and the knowledge representation language. There are ontologies focusing on things such as protein functions, organism development, anatomy and pathways. Most biological ontologies are controlled vocabularies, taxonomies or thesauri, but there are also ontologies that are knowledge bases and use OWL (Web Ontology Language, a language building on the Resource Description Framework (RDF) and RDF Schema) as their representation language. With respect to the abstraction level the ontologies may range from high level ontologies that define general biological knowledge to ontologies that describe selected aspects. For instance, some general biological knowledge is covered in the TAMBIS ontology [8] (e.g. protein and nucleic acid are biomolecules, and motif is-component-of protein). The GO molecular function ontology defines the whole space of possible biological functions (e.g. signal transducer activity and the more specific function receptor activity).

In this section we describe one of the important efforts in the area, OBO, and present a selection of ontologies that appear often in current research.

3.1 Open Biomedical Ontologies

Many biological ontologies are available via OBO, an umbrella web address that provides ontologies for shared use across different biomedical domains. In June 2006, 58 ontologies were available via the website. Some were under development and a few were deprecated and replaced by newer ontologies. Many of the OBO ontologies are stored in the SourceForge CVS (Concurrent Versions System) repository, which allows the ontologies to be updated daily while keeping a record of all changes.

```
[Term]
id: MA:0000281
name: nose
is_a: MA:0000017 ! sensory organ
is_a: MA:0000581 ! head organ
relationship: part_of MA:0000327 ! respiratory system
relationship: part_of MA:0002445 ! olfactory system
relationship: part_of MA:0002473 ! face
```

Figure 4-2. Example entry from Adult Mouse Anatomy (OBO).

The allowed representation formats for ontologies in OBO are the OBO syntax, extensions of this or OWL. The OBO flat file format is the most common file format in the OBO collection and aims to achieve human readability, ease of parsing, extensibility and minimal redundancy in the ontology files. Figure 4-2 shows an entry in OBO syntax. It represents the term nose (name) and has as identifier *MA:0000281* (id). The nose is a sensory organ (which has identifier *MA:0000017*) and a head organ (which has identifier *MA:0000581*). Further, the nose is part of the respiratory system (*MA:0000327*), the olfactory system (*MA:0002445*), and the face (*MA:0002473*). Other information, such as definition, synonyms, and comments may also be described. The same information in OWL is presented in Figure 4-3. For a complete description of the OBO syntax, we refer to http://geneontology.org/GO.format.shtml#oboflat, and for a description of OWL we refer to http://www.w3.org/2004/OWL/ and another chapter in this book.

```
<owl:Class rdf:ID="MA:0000281">
  <rdfs:label xml:lang="en">nose</rdfs:label>
  <rdfs:subClassOf rdf:resource="#MA:0000017"/>
  <rdfs:subClassOf rdf:resource="#MA:0000581"/>
  <rdfs:subClassOf> <owl:Restriction>
     <owl:onProperty> <owl:ObjectProperty rdf:about="#part_of"/>
        </owl:onProperty>
        <owl:someValuesFrom rdf:resource="#MA:0000327"/>
      </owl:Restriction> </rdfs:subClassOf>
  <rdfs:subClassOf> <owl:Restriction>
     <owl:onProperty> <owl:ObjectProperty rdf:about="#part_of"/>
        </owl:onProperty>
        <owl:someValuesFrom rdf:resource="#MA:0002445"/>
      </owl:Restriction> </rdfs:subClassOf>
  <rdfs:subClassOf> <owl:Restriction>
     <owl:onProperty> <owl:ObjectProperty rdf:about="#part_of"/>
        </owl:onProperty>
        <owl:someValuesFrom rdf:resource="# MA:0002473"/>
      </owl:Restriction> </rdfs:subClassOf>
</owl:Class>
```

Figure 4-3. Example OWL entry.

Editing of OBO flat files is often performed using the OBO-Edit tool (previously called DAG-Edit). The Ontology Lookup Service [5] provides a user-friendly single entry point for (June 2006, circa 40) ontologies in the OBO format. There are some ontology development tools, such as Protégé, that support OWL-based ontologies.

3.2 Frequently Used Ontologies

The *GO* Consortium is a joint project with the goal to produce a structured, precisely defined, common and dynamic controlled vocabulary that describes the roles of genes and proteins in all organisms. Currently, there are three independent ontologies publicly available: biological process (ca 11000 terms), molecular function (ca 8000 terms) and cellular component (ca 1800 terms) (June 2006). The GO ontologies are a de facto standard and many biological data sources are today annotated with GO terms. The terms in GO are arranged as nodes in a directed acyclic graph, where multiple inheritance is allowed. The GO ontologies are available via OBO. They are still being further developed and efforts are made to improve the quality of the ontologies (e.g. [16]).

Medical Subject Headings (MeSH) [22] is a controlled vocabulary produced by the American National Library of Medicine and used for indexing, cataloging, and searching for biomedical and health-related information and documents. It organizes terms in a hierarchical structure and it includes different categories, including anatomy, organisms, and diseases, most of which are available via OBO. The version available via the Ontology Lookup Service contains circa 15000 terms (version December 2005). MeSH uses 'is-a' to represent both the is-a relation and the part-of relation.

An area where many ontologies have been developed is *anatomy*. OBO lists 18 different anatomy ontologies (June 2006) and MeSH which has an anatomy category. The ontologies cover different organisms (*C. elegans*, *Drosophila,* Medaka fish, Zebrafish, Human, Mosquito, Mouse, Fungi, *Dictyostelium discoideum*, *Arabidopsis*, Cereal, Maize and Plant), cell types and enzyme sources. Some of the plant related ontologies are deprecated (e.g. *Arabidopsis* anatomy and Cereal anatomy) and have been replaced by the Plant anatomy ontology. SOFG has focused on integration of human and mouse anatomy ontologies and several are available via their web site. The number of terms in these ontologies differs a lot. For instance, Fungal anatomy and *Dictyostelium* anatomy contain less than 100 terms, while Mouse anatomy and development contains over 13000 terms. The anatomy ontologies are often taxonomies.

In *systems biology* ontologies are currently being developed in conjunction with the development of standards for the representation of molecular interaction data. These standards (see e.g. overviews in [39,37,38]) aim to provide the ability to supply information on molecular pathways in a format that supports efficient exchange and integration. This is seen as an important prerequisite for advances in the area. For instance, the Systems Biology Ontology, connected to the Systems Biology Markup Language (SBML) [32,12], defines terms used in quantitative biochemistry in four controlled vocabularies: roles of reaction participants, quantitative parameters, rate laws, and simulation frameworks. The Protein-protein interaction ontology, connected to the Proteomics Standards Initiative - Molecular Interaction [11,28], defines terms related to protein-protein interactions such as interaction detection methods, experimental roles and biological roles. The Systems Biology Ontology and the Protein-protein interaction ontology are available via OBO. The Biological Pathway Exchange (BioPAX) [3] standard aims to provide an OWL-based data exchange format for pathway data and is developed as an ontology.

4. USE OF BIOLOGICAL ONTOLOGIES

We have already mentioned advantages of using ontologies in the introduction. Regarding biological ontologies the main focus has been on data source annotation, ontology-based search, data integration, data exchange and the use of ontologies as a community reference (e.g. [36,33,31]).[5]

Many biological data sources use ontologies for annotation of their data entries and many tools exist to support annotating data sources or to predict annotations for data entries (e.g. BLAST2GO, GOFigure, GOtcha). The annotations are used in several ways. Search engines can take advantage of the annotations as they give extra information. Further, several kinds of systems use GO annotations to compute a semantic similarity measure between entries in data sources (e.g. FuSSiMeG). Entries annotated with similar sets of GO terms are considered likely to be similar themselves [21]. Such a similarity measure can be used for data integration and grouping of data entries [14]. There are also many tools that use GO annotations to interpret gene expression analysis on multiple genes (e.g. EASE, FatiGO, FuncAssociate, GOstat, Onto-Compare). For instance, given a list of genes from a microarray experiment, systems calculate over- or under-representation statistics for each GO term related to the genes in the experiment. This provides a description of significant features of the genes in the list. Ontologies and annotations are also used in text mining. For instance, Genes2Diseases uses occurrence counts of GO and MeSH terms in research literature as well as data sources to connect genes to genetically inherited diseases.

Ontologies are also used in different steps in ontology-based search. An ontology can be used as an index to the information in the information sources. A user can browse the ontology and use the terms in the ontology as query terms. For instance, TAIR Keyword Browser (Fly), GOFish (Yeast, Fly, Mouse, Worm) and MGI GO Browser (Mouse) use GO to browse databases. MeSH is used to index PubMed, an archive for biomedical and life sciences journal literature, and GOPubMed connects GO to PubMed. A module of Whatizit marks all GO terms in a document and links them to their entries in GO. An ontology may also be used for query refining and expansion by moving up and down in the hierarchy of concepts. For instance, when a user searches in a database for 'immune response' and gets

---

[5] In this section we exemplify the uses of biological ontologies using a number of systems and tools. The list of systems and tools is not intended to be complete. For the sake of brevity, we also do not provide references to each of the tools, but the tools without reference are available from the GO Web page under 'Tools'.

only very few results, the user may decide to query with a more general term to find more answers. The ontology can be used to find these more general terms, in this case, for instance 'defense response'.

5. ISSUES IN ONTOLOGY ENGINEERING

The ontology engineering process contains different phases such as development and maintenance. Although there exist many tools that support these phases, such as ontology development tools, ontology integration tools, ontology evaluation tools, ontology-based annotation tools, ontology learning tools and ontology storage and querying tools [27], not all phases are well understood yet and several issues need further investigation. In this section we briefly discuss such issues that currently receive attention.

5.1 Ontology Development Best Practice Principles

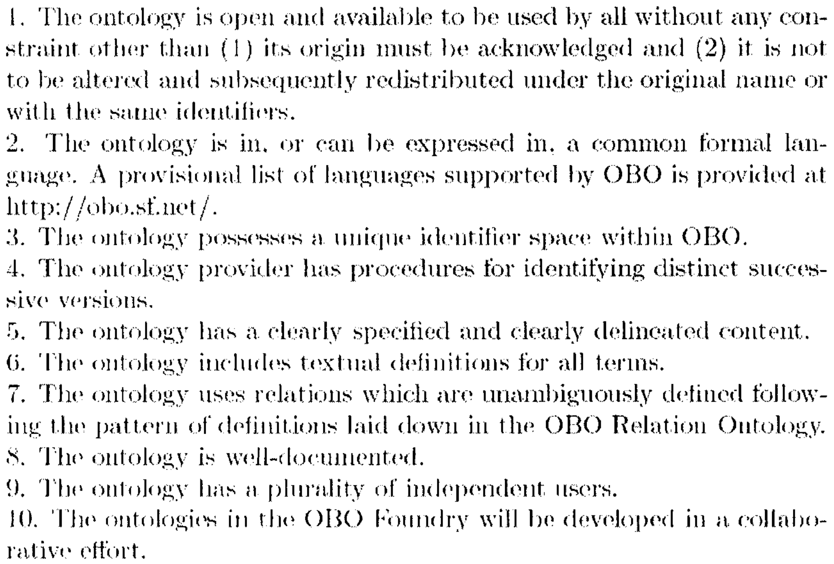
Although OBO ontologies are required to be open, to use a common syntax, to be orthogonal to each other, to share a unique identifier space and to include textual definitions, there are still a number of problems regarding interoperability between the ontologies and the quality and formal rigor in the ontologies. For instance, not all OBO ontologies use the is-a and part-of relations in the same way. For this reason the OBO Foundry was created [26]. Ontology developers joining the OBO Foundry commit to a set of best practice principles for ontology development (Figure 4-4). Most of the OBO criteria are included in the OBO Foundry best practice principles.[6]

The criteria are connected to the main goals and intended uses of ontologies. For instance, criterion 1 requires that the ontologies are open and available, criterion 9 that there are many users, and criterion 10 that the ontologies are developed collaboratively. These are basic requirements if the ontologies are to become community references. Further, by not allowing changes without changing names (criterion 1), using unique identifiers (criterion 3), using textual definitions to reduce potential ambiguity (criterion 6) and using relations from the OBO Relation Ontology [34] (criterion 7), clear and unambiguous definitions of the terms in the ontologies are promoted. This leads to improved understanding and usefulness of the ontologies as well as improved interoperability between the applications using these ontologies. Interoperability and integration are also supported by criterion 2 (use of common formal languages) and criterion 5

---

[6] OBO Foundry criteria 1, 2, 3, 6 are also OBO criteria. In addition, OBO requires that the ontologies in OBO are orthogonal to each other.

(terms outside the scope of the ontologies should not be defined within the ontology, but the ontology should rather refer to their definitions in other ontologies). Criteria 2 and 7 also support reasoning. Criterion 4 (versioning) addresses an important, and currently not so well supported, aspect of ontology engineering. Finally, criteria 6 and 8 (documenting the content, use and development process of the ontologies) are particularly important for human users.



1. The ontology is open and available to be used by all without any constraint other than (1) its origin must be acknowledged and (2) it is not to be altered and subsequently redistributed under the original name or with the same identifiers.
2. The ontology is in, or can be expressed in, a common formal language. A provisional list of languages supported by OBO is provided at http://obo.sf.net/.
3. The ontology possesses a unique identifier space within OBO.
4. The ontology provider has procedures for identifying distinct successive versions.
5. The ontology has a clearly specified and clearly delineated content.
6. The ontology includes textual definitions for all terms.
7. The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the OBO Relation Ontology.
8. The ontology is well-documented.
9. The ontology has a plurality of independent users.
10. The ontologies in the OBO Foundry will be developed in a collaborative effort.

Figure 4-4. OBO Foundry criteria (from [26]).

5.2 Ontology Instantiation

Vast amounts of biological data, e.g. research articles, are available on the Web. However, the knowledge in these Web documents is not readily available for analysis and use in applications. Therefore, in ontology instantiation (also called ontology population) specific knowledge is extracted from these documents based on the knowledge available in ontologies. The ontologies define the kind of information that is extracted. The instantiated ontology becomes a knowledge base. Manually instantiating ontologies is a time-consuming and error-prone task. Research has started on developing tools to support (semi-)automatic instantiation, but, currently, few tools exist. The most promising approaches use information extraction techniques [2,4] for retrieving the knowledge, (see also Chapter 13). Another step is then required to detect redundancy and check consistency of the knowledge base.

5.3 Ontology Alignment

Many of the currently developed ontologies, such as the OBO and SOFG anatomy ontologies, contain overlapping information. As an example, in Figure 5-1 the terms in bold face in MA and MeSH, respectively, denote equivalent terms. Often we would want to be able to use multiple ontologies. For instance, companies may want to use community standard ontologies and use them together with company-specific ontologies. Applications may need to use ontologies from different areas or from different views on one area. Ontology builders may want to use already existing ontologies as the basis for the creation of new ontologies by extending the existing ontologies or by combining knowledge from different smaller ontologies. In each of these cases it is important to know the relationships between the terms in the different ontologies. Furthermore, different data sources in the same domain may have annotated their data with different but similar ontologies. Knowledge of the inter-ontology relationships would lead to improvements in search, integration and analysis of biomedical data. We say that we align two ontologies when we define the relations between terms in the different ontologies. We merge two ontologies when we, based on the alignment relations between the ontologies, create a new ontology containing the knowledge included in the source ontologies. It has been realized that this is a major issue and some organizations have started to deal with it. For instance, SOFG developed the SOFG Anatomy Entry List which defines cross species anatomical terms relevant to functional genomics and which can be used as an entry point to anatomical ontologies.

There exist a number of ontology alignment systems that support the user to find inter-ontology relationships. Some of these systems are also ontology merge systems. These systems implement strategies based on linguistic matching, structure-based strategies, constraint-based approaches, instance-based strategies, strategies that use auxiliary information (such as thesauri or domain knowledge) or a combination of these. Some systems are automatic, but most systems are semi-automatic, requiring a human expert to validate the results of the system. For an overview of ontology alignment systems we refer to [19,20] and http://www.ontologymatching.org/.

6. ONTOLOGICAL KNOWLEDGE

In addition to the ontologies there is also other publicly available ontological knowledge that can be used for data search, integration and analysis [13,14]. This knowledge includes ontology alignments (i.e. inter-ontology relationships), ontological annotations of data sources, and mappings between data values and ontological terms.

Ontology alignments. As mentioned before, knowing inter-ontology relationships is a major issue and some organizations have started to address it. As a result of these efforts, a number of alignments have been generated. We already mentioned the SOFG Anatomy Entry List. Further, there are alignments between GO and other ontologies such as the Enzyme Nomenclature and MetaCyc. These are available from the GO Consortium web pages. Also the Unified Medical Language System (UMLS) [40] may be seen as a collection of alignments. In the near future we expect an increase of such knowledge as many ontology alignment tools are currently being developed to support the identification of such alignments.

Annotations. Many data sources annotate their data entries with ontological terms. For instance, terms from the GO molecular function ontology are used to describe gene and protein functions. Annotations can be stored as separate mapping rules, included in an ontology or stored in a data source entry. For instance, different data source annotations by GO terms can be found on the GO Consortium web pages.

Mappings between data values and ontological terms. In a similar way as whole data entries in data sources are related to ontological terms, the allowed values for certain data properties can be indexed based on ontology terms. For instance, keywords describing data entries in UniProt, a data source of protein sequences and related data, are mapped to terms in GO ontologies.

7. SUMMARY

In this chapter we presented important efforts and issues related to biological ontologies. We presented OBO as well as ontologies that are often used in current research. We found that many of these biological ontologies are controlled vocabularies, taxonomies or thesauri. Additionally, we discussed the use of biological ontologies in data source annotation and search. We also discussed some important issues in ontology development, ontology instantiation and ontology alignment. Finally, we drew attention to publicly available ontological knowledge that can be used for data search, integration and analysis.

ACKNOWLEDGMENTS

We thank Emma Lundman for discussions. We acknowledge the financial support of the Center for Industrial Information Technology, the Swedish Research Council, the Swedish national graduate school in computer science, and the EU Network of Excellence REWERSE (Sixth Framework Programme project 506779).

REFERENCES

[1] Baker C.J.O., Shaban-Nejad A., Su X., Haarslev V., and Butler G. Semantic Web Infrastructure for Fungal Enzyme Biotechnologists, *Journal of Web Semantics, Special issue on Semantic Web for the Life Sciences* 4(3), 2006.

[2] BioCreAtIvE, Critical Assessment for Information Extraction in Biology; http://biocreative.sourceforge.net/.

[3] BioPAX, Biological Pathway Exchange; http://www.biopax.org/.

[4] Blaschke C., Hirschman L., and Valencia A. Information extraction in molecular biology, *Briefings in Bioinformatics* 3(2):154-165, 2002.

[5] Coté R., Jones P., Apweiler R., and Hermjakob H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries, *BMC Bioinformatics* 7:97, 2006.

[6] Collins F., Green E., Guttmacher A., and Guyer M. A vision for the future of genomics research, *Nature* 422:835-847, 2003.

[7] GO, The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology, *Nature Genetics* 25(1):25-29, 2000; http://www.geneontology.org/.

[8] Goble C., Stevens R., Ng G., Bechhofer S., Paton N., Baker P., Peim M., and Brass A. Transparent access to multiple bioinformatics information sources, *IBM Systems Journal* 40(2):532-551, 2001.

[9] Gómez-Pérez A. Ontological Engineering: A state of the Art, *Expert Update* 2(3):33-43, 1999.

[10] Guarino N. and Giaretta P. Ontologies and Knowledge Bases: Towards a Terminological Clarification, in: *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, Mars, ed., IOS Press, 25-32, 1995.

[11] Hermjakob H., Montecchi-Palazzi L., Bader G., Wojcik J., Salwinski L., Ceol A., et al. The HUPO PSI's Molecular Interaction format - a community standard for the representation of protein interaction data, *Nature Biotechnology* 22(2):177-183, 2004.

[12] Hucka M., Finney A., Sauro H., Bolouri H., Doyle J., Kitano H., and the rest of the SBML Forum. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, *Bioinformatics* 19(4):524-531, 2003.

[13] Jakoniene V. and Lambrix P. Ontology-based integration for bioinformatics, in: *Proceedings of the VLDB Workshop on Ontologies-based techniques for DataBases and Information Systems*, 55-58, 2005.

[14] Jakoniene V., Rundqvist D., and Lambrix P. A method for similarity-based grouping of biological data, in: *Proceedings of the 3rd International Workshop on Data Integration in the Life Sciences*, LNBI 4075, 136-151, 2006.

[15] Jasper R. and Uschold M. A Framework for Understanding and Classifying Ontology Applications, in: *Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends,* 1999.

[16] Köhler J., Munn K., Rüegg A., Skusa A, and Smith B. Quality control for terms and definitions in ontologies and taxonomies, *BMC Bioinformatics* 7:212, 2006.

[17] Lambrix, P. Ontologies in Bioinformatics and Systems Biology, in: *Artificial Intelligence Methods and Tools for Systems Biology,* Dubitzky and Azuaje, eds., Springer, chapter 8, 129-146, 2004.

[18] Lambrix, P. Towards a Semantic Web for Bioinformatics using Ontology-based Annotation, in: *Proceedings of the 14th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises,* 3-7. Invited talk, 2005.

[19] Lambrix P. and Tan H. SAMBO - A System for Aligning and Merging Biomedical Ontologies, *Journal of Web Semantics, Special issue on Semantic Web for the Life Sciences* 4(3), 2006a.

[20] Lambrix, P. and Tan H. Ontology Alignment and Merging, in: *Anatomy Ontologies for Bioinformatics: Principles and Practice,* Burger, Davidson and Baldock, eds., Springer. To appear, 2006b.

[21] Lord P., Stevens R., Brass A., and Goble C. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation, *Bioinformatics* 19(10):1275-1283, 2003.

[22] MeSH, Medical Subject Headings; http://www.nlm.nih.gov/mesh/.

[23] Mukherjea S. Information retrieval and knowledge discovery utilising a biomedical Semantic Web, *Briefings in Bioinformatics* 6(3):252-262, 2005.

[24] Neches R., Fikes R., Finin T., Gruber T., Senator T., and Swartout, W. Enabling technology for knowledge engineering, *AI Magazine* 12(3):26-56, 1991.

[25] OBO, Open Biomedical Ontologies; http://obo.sourceforge.net/.

[26] OBO Foundry; http://obofoundry.org/.

[27] OntoWeb, 2002, Deliverable 1.3: A survey on ontology tools; 2002, Deliverable 2.1: Successful Scenarios for Ontology-based Applications; 2002, Deliverable 2.2: Guidelines for the selection of techniques for kinds of ontology-based applications; 2004, Deliverable 1.6: A survey on ontology-based applications. E-commerce, knowledge management, multimedia, information sharing and educational applications; http://www.ontoweb.org/.

[28] Orchard S., Montecchi-Palazzi L., Hermjakob H., and Apweiler R. The Use of Common Ontologies and Controlled Vocabularies to Enable Data Exchange and Deposition for Complex Proteomic Experiments, in: *Proceedings of the Pacific Symposium on Biocomputing* 10:186-196, 2005.

[29] Protégé; http://protege.stanford.edu/.

[30] REWERSE, EU Network of Excellence on Reasoning on the Web with Rules and Semantics, Working group A2; http://rewerse.net/.

[31] Rojas I., Ratsch E., Saric J., and Wittig U. Notes on the use of ontologies in the biochemical domain, *In Silico Biology* 4:0009, 2003.

[32] SBML, Systems Biology Markup Language; http://sbml.org.

[33] Schulze-Kremer S. Ontologies for molecular biology and bioinformatics, *In Silico Biology* 2:0017, 2002.

[34] Smith B., Ceusters W., Klagges B., Köhler J., Kumar A., Lomax J., Mungall C., Neuhaus F., Rector A., and Rosse C. Relations in biomedical ontologies, *Genome Biology* 6:R46, 2005.

[35] SOFG, Standards and Ontologies for Functional Genomics; http://www.sofg.org/.

[36] Stevens R., Goble C., and Bechhofer S. Ontology-based knowledge representation for bioinformatics, *Briefings in Bioinformatics* 1(4):398-414, 2000.
[37] Strömbäck L., Hall D., and Lambrix P. A review of standards for data exchange within systems biology, *Proteomics*. Invited contribution. To appear, 2006a.
[38] Strömbäck L., Jakoniene V., Tan H., and Lambrix P. Representing, storing and accessing molecular interaction data: a review of models and tools, *Briefings in Bioinformatics*. Invited contribution. To appear, 2006b.
[39] Strömbäck L., and Lambrix P. Representations of molecular pathways: An evaluation of SBML, PSI MI and BioPAX, *Bioinformatics* 21(24):4401-4407, 2005.
[40] UMLS, Unified Medical Language System; http://www.nlm.nih.gov/pubs/factsheets/umls.html.