

Selecting an Ontology for Biomedical Text Mining

He Tan, Patrick Lambrix

Department of Computer and Information Science
Linköpings universitet, Sweden

Abstract

Text mining for biomedicine requires a significant amount of domain knowledge. Much of this information is contained in biomedical ontologies. Developers of text mining applications often look for appropriate ontologies that can be integrated into their systems, rather than develop new ontologies from scratch. However, there is often a lack of documentation of the qualities of the ontologies. A number of methodologies for evaluating ontologies have been developed, but it is difficult for users by using these methods to select an ontology. In this paper, we propose a framework for selecting the most appropriate ontology for a particular text mining application. The framework comprises three components, each of which considers different aspects of requirements of text mining applications on ontologies. We also present an experiment based on the framework choosing an ontology for a gene normalization system.

1 Introduction

With the explosive growth of the volume of published biomedical research, it is challenging to keep up to date with the underlying knowledge available in the form of free text. The necessity of understanding actions of individual biological components in system context rather than in isolation, extends the coverage of literature far beyond the capabilities of individual scientists. Text mining is an emerging field that attempts to deal with these challenges (Ananiadou and McHought, 2006; Cohen and Hersh, 2005; Spasic et al., 2005). Text mining requires a significant amount of domain knowledge.

A large number of biomedical ontologies already exist, and hold much of the information. Some of the ontologies have been designed for modeling domain knowledge, e.g. FMA (Rosse and Mejino, 2003) and GO (Ashburner et al., 2000), others are developed for potential applications, e.g. MeSH (Lowe and Barnett, 1994) for indexing the medical literature. Whatever purposes the ontologies were built for, they are used to support text mining for tasks such as access to text, natural language processing, and information integration. Developers of text mining applications often look for appropriate ontologies that can be integrated into their systems, rather than develop new ontologies from scratch. The choice of ontology may, however, have a major impact on the performance of the text mining system, including the quality of the results.

Selecting an appropriate ontology relies on evaluation and comparison of the available ontologies. Unfortunately, there is often a lack of documentation of the qualities of the ontologies. A number of methodologies for evaluating ontologies have been proposed, mainly for evaluating one ontology at a time. However, it is difficult for users by using these methods to make a decision on ontology selection. The various evaluation methods can be classified into three main categories: 1) Those that evaluate ontologies against a set of criteria defined by human (e.g. (Lozano and Gómez, 2004) suggests 160 criteria). Most of the evaluation criteria are from the point of view of ontology building. They are not applicable for the selection of ontologies for a particular application. 2) Those that include methods to gather statistics about the nature of ontologies (e.g. (Gangemi et al., 2006) proposes 32 measures for depth, width, fan-out, etc). The problem

for this kind of methods is that it is not clear how to use these statistics for ontology selection among alternatives. 3) Those that are application-dependent evaluations. For example, ontologies are evaluated against a corpus in (Brewster et al., 2004) regarding the domain of an IE system, and (Porzel and Malaka, 2004) evaluates the quality of an ontology regarding a relation tagging task by using gold standard data tagged by human. While evaluating ontologies for the particular application is a relatively straightforward method, evaluations may be sensitive to the test dataset, and it may also be expensive to perform evaluations for many ontologies.

In this paper we propose a framework for selecting the most appropriate ontology for a particular text mining application. The framework comprises three components for selection, and uses and extends some of existing ontology evaluation and validation methods regarding the requirements of text mining in the area. After a brief overview of the related work, we introduce our framework in section 3. In section 4 we show how to use the framework in the setting of selecting an ontology for the gene normalization system described in (Tan, 2008).

2 Related Work

Not much work has been done yet for ontology selection in the biomedical area. Biportal (Biportal, 2.0) is an ontology library, currently mainly for browsing and navigating biomedical ontologies. The system also allows users of ontologies to submit information about their applications and comments on the content of ontologies, and stores mappings between concepts of ontologies in the library. This information may help users to select ontologies from the repository.

Some work has been done for ontology selection for the Semantic Web. In (Sabou et al., 2006), authors indicate the challenges for ontology evaluation posed by ontology selection for the Semantic Web, such as the need for automation and good performance. Two examples of ontology selection for the Semantic Web are OntoSelect (Buitelaar et al., 2004) and AKTiveRank (Alani and Brewster, 2005). Both are mainly based on the second category of ontology evaluation methods. OntoSelect is an ontology library that gives a functionality for selecting

ontologies for a given knowledge markup task. The selection relies on measuring the number of concepts and properties, and popularity of ontologies. The ontology selection algorithm in AKTiveRank combines the measures of coverage of an ontology given search terms, and locations of the terms and semantic similarity between the terms in the structure of the ontology.

OntoMetric is a hierarchical framework proposed in (Lozano and Gómez, 2004) for general ontology selection. The tool offers a default hierarchy of criteria to evaluate ontologies from the point of view of building ontologies, and also allows users to adapt it for their needs. The selection is based on a weighted addition of value of each criteria.

3 The Framework

In principle biomedical ontologies provide formal representations of domain knowledge for text mining in the area, but they are used for different purposes such as providing a model for storing, searching and querying a repository of text; providing domain knowledge for natural language processing; providing a framework for information integration; or several of the above purposes.

Based on a literature study and experience in building systems, in this section we suggest criteria for selecting an ontology for a given biomedical text mining application. The criteria are organized in three components of a framework.

Component 1

In the first component the criteria for ontology selection are given in two dimensions: content and supporting technologies. The combined requirements from the two dimensions lead to a list of candidate ontologies.

Content What are the requirements on the content of ontology given an application? We propose three characteristics to be considered,

- *Type* What is the kind of the ontology required in the scenario? Ontologies can be distinguished into three basic types (Stenzhorn et al., 2005): *top ontologies*, e.g. BFO (Smith, 2004), contain only a restricted set of highly general classes, such as Function and Object, which are not tied to any particular domain of interest;

top-domain ontologies, e.g. BioTop (Stenzhorn et al., 2005), contain all classes that are essentially needed to describe a certain domain, such as Organism, Cell and Tissue in the case of biology; and *domain ontologies*, e.g. GO (Ashburner et al., 2000), contain classes that comprehensively describe a certain domain of interest. For example, for natural language processing tasks such as entity recognition, a top-domain ontology may be sufficient.

- *Scope* What are the main subjects that need to be covered in a top-domain or domain ontology required by the application? For example, gene and disease are the domains that concern (Hristovski et al., 2005).
- *Representation* What kind of information needs to be present in the ontology? From a knowledge representation point of view, ontologies can have the following components: *concepts*, which represent sets or classes of entities in a domain; *relations* between concepts, *instances*, which represent the actual entities; and *axioms*, which represent facts that are always true in the topic area of the ontology (Lambrix et al., 2007). Which components should be present in the ontology, depends to some extent on the purpose of the ontology in the application. For example, if an ontology is used as resource for NLP in the application, components *concepts* and *instances* both may be necessary, but not *relations* and *axioms*.

Supporting technologies Three questions are considered in this dimension:

1. What technologies are needed to support the use of the ontology in the scenario? Supporting technologies may include ontology representation languages, ontology browsers, tools for ontology alignment, reasoning services, and ontology-driven NLP tools.
2. Are tools available to provide the supporting technologies for the ontology? For example, ontology alignment systems, e.g. (Lambrix and Tan, 2008) are available for biomedical ontologies in OWL and OBO format.

3. What is the cost to develop new tools to support the use of an ontology if there is no existing tools? Does the cost meet the expectations of the application?

Component 2

In this level the criteria for selection focus on detailed content of candidate ontologies which are considered in two aspects: verification and evaluation.

Verification Considering the maturity level of current biomedical ontologies, verification of taxonomic knowledge in candidate ontologies is valuable for selection. Reasoners, such as Racer, Pellet, and FaCT++, can check consistency, incompleteness and redundancy in ontologies.

Evaluation First, we list a set of characteristics of ontologies that are of interest for text mining applications,

- *Coverage* Does an ontology cover the concepts or/and relations concerned in the application? Do their definitions meet the expected scope?
- *Taxonomic Knowledge* The two relations *is-a* and *part-of* have a central role in almost all ontologies (Smith et al., 2005). Do the locations of essential concepts in the *is-a* and *part-of* hierarchies meet the expectation?
- *Instance* The instantiation of concepts and the number of instances for each concept could effect greatly the performance of many text mining tasks such as entity recognition. Do they satisfy the requirements?

These characteristics may be evaluated,

- against domain experts or references if they exist. For example, if an ontology involves integration of literature with data from biological databases, schemas of databases can be the references.
- among candidate ontologies. There is no gold standard in this situation and therefore the candidate ontologies are compared directly to each other. For instance, aligning the ontologies can provide information about their similarities and differences regarding their coverage.

Component 3

The results from component 1 and 2 may not lead to a direct and confident decision. Also, in many cases there exist gold standards or benchmarks which we can use to evaluate our application. Therefore, the purpose of component 3 is to evaluate the use of ontologies in a real system. This component could provide a relatively straightforward method for evaluating the quality of ontologies for use in the application.

The field of biomedical text mining has expended considerable effort in building evaluation resources. A number of challenges, such as TREC Genomics track, BioCreative, BioNLP/JNLPBA 2004 and LLL05, have been organized in the community. They contribute to the creation of shared gold standard datasets, prepared by domain experts and also suggest evaluation measures. Comparison of the system individually bundled with candidate ontologies can be performed by using these gold standard datasets and measures.

4 Experiment

In this section we present an experiment in which we use our framework to select a biomedical ontology for the gene normalization task.

4.1 Scenario

The purpose of gene normalization (GN) is to link genes and proteins mentioned in the literature to their entries in structured databases of biological data. It has a substantial impact on tasks such as retrieval of relevant documents, identification of interesting genes and proteins, and relationships between them. The task is challenging even for scientists, since there is no community wide agreement on how a particular gene and gene product should be named. Heavy use of short forms (acronyms or abbreviations) in biomedical literature makes the problem worse. In (Tan, 2008) we developed an ontology-centred system for gene normalization. It relies on information about gene candidates, contexts of the symbol and external knowledge sources. Information about gene candidates is extracted from gene databases. Ontologies are used for semantic interpretation of contexts of gene symbols and identification of their relevance to gene candidate informa-

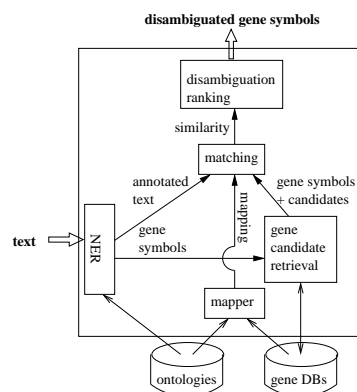


Figure 1: The components and data flows of the system

tion. Normalization is based on matching contexts of the symbol to relevant information about gene candidates.

Figure 1 illustrates the components and data flows of the system. The system receives a raw text as the input, and yields database entries of genes appearing in the text: 1) *Mapper* receives the ontology and schemas of gene databases, and outputs a list of concepts from the ontology and their mappings to categories of information stored in gene databases. 2) *Named Entity Recognizer (NER)* recognizes relevant biomedical entities in text by linking them to the concepts of the ontology which have been mapped to categories of gene information in *mapper*. 3) *Gene candidate retrieval* retrieves gene candidates for each gene symbol. The categories of information about each gene candidate are collected. 4) *Matching* includes various algorithms that match contexts of a gene symbol to information about its gene candidates according to the mappings obtained in *mapper*, and returns similarity values between them. 5) *Normalization ranking* ranks gene candidates based on results of *Matching* for each gene symbol appearing in a text.

4.2 Selecting the ontology

In this application the components *NER* and *Mapper* count on the deployment of ontologies. The ontology-driven *NER* component involves associating text with the correct concepts in the ontology by means of associating mentions in the text with instances in the ontology. The ontology provides the formal representation of domain knowledge for the *NER*. The component *Mapper* involves mapping the

ontology to database schemas. The ontology provides the model to link information extracted from literature and data in biological databases.

Component 1

Following the criteria of the component 1, we define the requirements on the ontology.

Type Based on studies of categories of information stored in gene databases such as EntrezGene (Maggloot et al., 2005), we decide that top-domain ontologies are suitable for our application.

Scope The ontology should represent domain knowledge about genes and gene products. The ontology must contain concepts covering the categories of gene information, e.g. location, functions, and biological process.

Representation In order to support the task of *NER*, the ontology must at least have concepts and instances.

Supporting Technologies In the system the supporting technologies include an ontology-driven NLP tool that support the task of *NER*, and ontology alignment algorithms, that can be used to map the ontology to categories of information in databases.

We look for suitable ontologies by searching OBO (OBO, 2009) and Bioportal, and reviewing literature on biomedical ontologies. There are a few ontologies covering genes and gene products such as GO, MeSH, the Unified Medical Language System (UMLS) knowledge sources (Lingberg et al., 1993) and ontologies from the GENIA project (Kim et al., 2003). Only two of these meet all the above requirements, the UMLS knowledge sources and the GENIA ontologies. The UMLS Semantic Network and GENIA entity and event ontologies can be seen as top-domain ontologies, of which the concepts are on the level of the requirement. The UMLS Metathesaurus is primarily concerned with names used in the domain. It contains a huge number of biomedical entities and events (called concepts in the Metathesaurus) and their various names from more than 100 biomedical vocabularies. All of the concepts are assigned to at least one semantic type in the Semantic Network. The MetaMap program (Aronson, 2001) is available to map text to the concepts and semantic type. The GENIA project collects a set of MEDLINE articles concerning transcription factors in human blood cells. Every biomedical entity and event

appearing in the articles are identified with a concept in the GENIA ontology. Several NLP tools, e.g. LingPipe (LingPipe, 3.7.0), support statistical name entity recognition by using the GENIA corpus as training data.

Component 2

Verification We checked the consistency and redundancy in the UMLS Semantic Network and GENIA ontologies, respectively, by translating them into OWL format and then sending to the reasoner, Racer. Both of them are consistent and have minimal representations.

Evaluation We perform two evaluations. The first evaluation is to investigate the coverage of concepts of the two ontologies against categories of EntrezGene. The coverage determines the extent of information that can be used for gene normalization. In the second one we compare biomedical entities and events belonging to concepts of the two ontologies, since they influence the performance of *NER*.

- *Evaluation 1* Manually, we identify a list of categories of EntrezGene that are used for gene normalization. Evaluation is performed by aligning concepts of the two ontologies to the categories. A first observation is that UMLS covers more topics than GENIA and therefore may give better results for *NER*. The topics of the GENIA corpus can be represented by a query using human, blood cell, transcription factor. To be able to compare the coverage of UMLS and GENIA on an overlapping piece of a domain we align the two ontologies based on the GENIA topics. The evaluation is based on an instance-based strategy proposed in (Tan et al., 2006). The algorithm consists of the following steps: 1) *Instance collection*: We retrieve a list of human genes from the database using human, blood cell, and transcription factor. All entities appearing in the categories of information are collected as instances. 2) *Classification*: For each ontology, each instance is classified to concepts by a NLP tool. The MetaMap program is used for the UMLS Semantic Network, and the LingPipe entity recognition program is used for the GENIA ontology. An instance can be classified to more than one concept. 3) *Calculation*: The similarity between a concept from an ontology (A) and a category from the EntrezGene (B_j) is determined by the extent to which the instances of the category

EntrezGene	UMLS (value)	GENIA (value)
Name/Syn.	Gene or Genome (0,66)	protein_molecule (0,83)
Chromosome	Cell Component (1,0)	protein_molecule (1,0)
Map Location	Nucleotide Sequence (1,0)	DNA_domain_or_region (0,5) protein_molecule (0,4)
Protein	Gene or Genome (0,42) Amino Acid, Peptide, or Protein (0,25)	protein_molecule (0,60)
Pathway	Molecular Function (0,29) Cell Function (0,24)	other_name (0,76)
GOA_function	Molecular Function (0,75)	other_name (0,85)
GOA_component	Cell Component (0,96)	cell_component (0,40) other_name (0,15)
GOA_process	Cell Function (0,49) Molecular Function (0,16)	other_name (0,78)

Table 1: Comparison: EntrezGene, UMLS and GENIA

support the concept. The similarity value is calculated as: $sim(A, B_j) = \frac{\sum_{i=1}^n S(a_i, B_j)}{\sum_{k=1}^m \sum_{i=1}^n S(a_i, B_k)}$, where a_i are the instances belonging to A , n is the number of instances belonging to A , m is the number of categories, and S is a function calculating the degree to which an instance of an ontology concept supports a category from EntrezGene. S is defined as:

$$S(a_i, B_j) = \begin{cases} 0 & \text{if } a_i \text{ does not associate with } B_j \\ \frac{1}{p} & \text{otherwise; } p \text{ is the number of categories } B_k \\ & \text{that } a_i \text{ associates with} \end{cases}$$

Table 1 shows the alignment results. The left column gives the categories of gene information we use for normalization. If the similarity value between a category and the UMLS semantic type or GENIA concept is higher than 0.6, they are considered as corresponding to the category. If no similarity value is higher than 0.6, we list the two semantic types or concepts with the highest values. Three of eight alignments fall into this situation for UMLS semantic types, and two for GENIA concepts. We also note that the GENIA concept *other_name* appears 4 times, but the meaning of this concept is not well-defined. Most other categories are aligned to *protein_molecule*, although the categories are very different. In this evaluation, it is more likely that the UMLS semantic network is more appropriate than the GENIA ontology for our system.

- *Evaluation 2* 35,515 entities of interest in the biological domain are identified in the GENIA entity corpus, 7,089 entities of which have been found in the UMLS Metathesaurus. Since we could not obtain all the UMLS Metathesaurus concepts for each semantic type, we decided to compare the UMLS semantic types and GENIA concepts that appear in the first evaluation, based on these 7,089 shared instances. The comparison is based on the same instance-based alignment strategy used in the first

evaluation. Tables 2 and 3 show the alignment results. In table 2 the value illustrates the extent to which instances of GENIA concepts support the UMLS semantic types. We list the three concepts with the highest values in the table, if they exist. The concepts in italics are event concepts, the others are entities. All the UMLS semantic types cover instances from more than one GENIA concept belonging to a subtree. For example, Gene or Genome could be aligned to the three GENIA concepts *DNA_family_or_group*, *DNA_domain_or_region* and *RNA_molecule*, which are leaves of the subtree *Nucleotide_acid_compound*. The granularity of the part of the GENIA ontology concerning the application is finer than the corresponding part of the UMLS semantic network. Table 3 shows the distribution of the instances of GENIA concepts when they support UMLS semantic types. Instances of *DNA_domain_or_region* and *cell_component* mainly support Gene or Genome and Cell Component, respectively, which is consistent with our observation in table 1. Another observation from this table is that *protein_molecule* covers both Gene or Genome and Amino Acid, Peptide, or Protein. The result also shows that *other_name* is not well defined. However, this evaluation does not give a conclusive answer to which ontology supports the task of *NER* better.

Component 3

Since component 1 and 2 could not lead to a direct and confident decision, we evaluate the use of the two ontologies in our system. We use a training dataset from the BioCreative II gene normalization task (Morgan et al., 2008) in the evaluation. The BioCreative datasets contain Medline abstracts along with the EntrezGene identifiers corresponding to the human genes and direct gene products appearing in the abstracts. The training dataset contain 281 Medline abstracts and 640 human genes. The MetaMap program is the NLP tool that uses the UMLS Semantic Network, and the LingPipe entity recognition program employs GENIA ontology.

Table 4 illustrates the quality of normalization in the two systems. The quality is measured in terms of precision, recall and f-measure. Recall (Re.) is defined as the number of correctly disambiguated gene symbols divided by the number of the gene symbols to be disambiguated. Precision is measured in

GENIA	UMLS	value
DNA_family_or_group	Gene or Genome	0.63
DNA_domain_or_region	Gene or Genome	0.54
RNA_molecule	Gene or Genome	0.44
DNA_N/A	Nucleotide Sequence	0.23
DNA_substructure	Nucleotide Sequence	0.23
DNA_domain_or_region	Nucleotide Sequence	0.17
cell_component	Cell Component	0.76
RNA_N/A	Cell Component	0.50
DNA_molecule	Cell Component	0.46
protein_family_or_group	Amino Acid, Peptide, or Protein	0.37
protein_subunit	Amino Acid, Peptide, or Protein	0.33
amino_acid_monomer	Amino Acid, Peptide, or Protein	0.33
protein_domain_or_region	Amino Acid Sequence	0.42
Cellular_physiological_process	Cell Function	0.26
Cell_communication	Cell Function	0.24
Cell_differentiation	Cell Function	0.19
Protein_amino_acid_deacetylation	Molecular Function	1.0
Protein_amino_acid_dephosphorylation	Molecular Function	0.80
Protein_ubiquitination	Molecular Function	0.50

Table 2: Comparison: UMLS and GENIA (1)

GENIA	UMLS	value
protein_molecule	Gene or Genome	0.30
protein_molecule	Amino Acid, Peptide, or Protein	0.29
protein_molecule	Biologically Active Substance	0.10
DNA_domain_or_region	Gene or Genome	0.54
DNA_domain_or_region	Biologically Active Substance	0.07
DNA_domain_or_region	Nucleotide Sequence	0.06
cell_component	Cell Component	0.76
cell_component	Biomedical or Dental Material	0.03
cell_component	Amino Acid, Peptide, or Protein	0.02
other_name	Disease or Syndrome	0.15
other_name	Cell Function	0.09
other_name	Neoplastic Process	0.08

Table 3: Comparison: UMLS and GENIA (2)

two ways. Pre.-1 is defined as the number of gene symbols correctly and uniquely identified to the real gene, divided by the total number of genes proposed in the result regarding the whole data set. Pre.-2 is computed as the average of the precision of disambiguation for each gene symbol. F-measure is the evenly weighted harmonic mean of precision and recall. Pre.-1 and Pre.-2 are used in the computation of Fm.-1 and Fm.-2, respectively. For both datasets the quality of normalization from the system bundled with the UMLS is better than the one with the GENIA ontology.

Result

Overall, the UMLS knowledge source can be considered as the ontology that is most appropriate for our gene normalization system. The ontology covers the subject, genes and gene products, well. The meaning of the concepts is defined well enough for the use in the application. The granularity of the part of the ontology meets the need of the task. The system bundled with the UMLS and its supporting technologies produced better results in a gold standard dataset than the other one.

KB	Dataset	Pre.-1	Pre.-2	Re.	Fm.-1	Fm.-2
GENIA	dataset 1	0.45	0.65	0.78	0.57	0.71
	dataset 2	0.50	0.63	0.72	0.59	0.67
UMLS	dataset 1	0.48	0.69	0.82	0.61	0.75
	dataset 2	0.52	0.67	0.78	0.62	0.72

Table 4: Quality of normalization

Although we have not run additional experiments to confirm that we have indeed made the right selection for our gene normalization system, the tests do corroborate our results. The chosen ontology leads to a better result for both datasets that we used. Therefore, each dataset can be seen as a confirmation of the framework where we only used the other dataset.

5 Conclusions

In this paper we proposed a framework for selecting an appropriate ontology for a particular biomedical text mining application. The framework deals with ontology selection in three components, each of which considers different aspects of requirements of text mining applications on ontologies. Then we present an experiment in which we select a biomedical ontology for a gene normalization system, using the framework. Within the framework, evaluation results lead us to a relatively concrete choice of an ontology for our system.

In the future we want to evaluate our framework with more applications and ontologies. Further, currently there is no service to support ontology selection for biomedical text mining. Therefore, an important track for future work is to build or extend an existing portal with information about the ontologies and their use in text mining applications that is structured according to our framework. The information in such a portal will constitute valuable data and experiences regarding ontology selection that will be useful for future applications.

Acknowledgments

We thank Bo Serenius for comments on the project, Qiang Liu for generating part of the data and David Byers for proof-reading the paper. We also acknowledge the financial support of the Center for Industrial Information Technology.

References

- Alani H, Brewster C. (2005) Ontology Ranking based on the Analysis of Concept Structures. *Proceedings of the 3th International Conference on Knowledge Capture*, 51-58.
- Alias-i. 2008. LingPipe 3.7.0. <http://alias-i.com/lingpipe>
- Ananiadou S, McNaught J. (editors) (2006) *Text Mining for Biology and Biomedicine*. Artech House.
- Aronson AR. (2001) Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proceedings of the AMIA Symposium*, 17-21.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1):25-9.
- Bioportal. <http://bioportal.bioontology.org/>
- Brewster C, Alani H, Dasmahapatra S, Wilks Y. (2004) Data Driven Ontology Evaluation. *International Conference on Language Resources and Evaluation*.
- Buitelaar P, Eigner T, Declerck T. (2004) OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection. *Proceedings of the Demo Session at the International Semantic Web Conference*.
- Cohen AM, Hersh WR. (2005) A survey of current work in biomedical text mining. *Briefings in Bioinformatics*, 6:57-71.
- Gangemi A, Catenacci C, Ciaramita M, Lehmann J. (2006) Modelling Ontology Evaluation and Validation. *Proceedings of the 2006 European Semantic Web Conference*.
- Hristovski D, Peterlin B, Mitchell JA, Humphrey SM. (2005) Using literature-based discovery to identify disease candidate genes. *International Journal of Medical Informatics*, 74(2-4):289-298.
- Kim J-D, Tomoko O, Yuka T, Tsujii J. (2003). GENIA corpus - a semantically annotated corpus for biotextmining. *Bioinformatics*, 19(suppl. 1):180-182.
- Lambrix P, Tan H, Jakoniene V, Strömbäck L. (2007) Biological Ontologies. *Chapter 4 in Baker, Cheung (eds), Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, 85-99, Springer.
- Lambrix P, Tan H. (2008) Ontology alignment and merging. *Chapter 6 in Burger, Davidson, Baldock, (eds), Anatomy Ontologies for Bioinformatics: Principles and Practice*, 133-150, Springer.
- Lindberg DA, Humphreys BL, McCray AT. (1993) The Unified Medical Language System. *Methods of Information in Medicine*, 32:281-91.
- Lowe HJ, Barnett GO. (1994) Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *Journal of the American Medical Association*, 271(14):1103-1108.
- Lozano-Tello A, Gómez-Pérez A. (2004) ONTOMETRIC: A Method to Choose the Appropriate Ontology. *Journal of Database Management. Special Issue on Ontological Analysis, Evaluation, and Engineering of Business Systems Analysis Methods*, 15(2):1-18.
- Maglott D, Ostell J, Pruitt KD, Tatusova T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research*. 33:D54-8.
- Morgan AA, Lu Z, Wang X, Cohen AM, Fluck J, Ruch P, Divoli A, Fundel K, Leaman R, Hakenberg J, Sun C, Liu HH, Torres R, Krauthammer M, Lau WW, Liu H, Hsu CN, Schuemie M, Cohen KB, Hirschman L. (2008) Overview of BioCreative II gene normalization. *Genome Biology*, 9(Suppl 2):S3.
- OBO. <http://obofoundry.org/>
- Porzel R, Malaka R. (2004) A task-based approach for ontology evaluation. *ECAI 2004 Workshop Ontology Learning and Population*.
- Rosse C, Mejino JL Jr. (2003) A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of Biomedical Informatics*, 36(6):478-500.
- Sabou M, Lopez V, Motta E, Uren V. (2006) Ontology Selection: Ontology Evaluation on the Real Semantic Web. *Proceedings of the 4th International EON Workshop, Evaluation of Ontologies for the Web*.
- Smith B. (2004) Beyond concepts: ontology as reality representation. *Proceedings of Formal Ontology and Information Systems*, 73-78.
- Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. (2005) Relations in biomedical ontologies. *Genome Biology*, 6(5):R46.
- Spasic I, Ananiadou S, McNaught J, Kumar A. (2005) Text mining and ontologies in biomedicine: making sense of raw text. *Briefings in Bioinformatics*, 6(3):239-251.
- Stenzhorn H, Schulz S, Beißwanger E. (2007) Towards a Top-Domain Ontology for Linking Biomedical Ontologies *Proceedings of the 12th World Congress on Health (Medical) Informatics*.
- Tan H. (2008) Knowledge-based Gene Symbol Disambiguation. *CIKM: Proceedings of the 2nd international workshop on Data and text mining in bioinformatics*, 73-76.
- Tan H, Jakoniene V, Lambrix P, Aberg J, Shahmehri N. (2006) Alignment of Biomedical Ontologies using Life Science Literature. *Proceedings of the International Workshop on Knowledge Discovery in Life Science Literature*, 1-17.