An Evaluation of Annotation Tools for Biomedical Texts

Kele T. Belloze¹, Daniel Igor S. B. Monteiro², Túlio F. Lima², Floriano P. Silva-Jr¹, Maria Cláudia Cavalcanti²

¹Laboratório de Bioquímica de Proteínas e Peptídeos – Instituto Oswaldo Cruz Avenida Brasil 4365 – 21.040-360 – Rio de Janeiro – RJ – Brazil

²Departamento de Ciência da Computação Instituto Militar de Engenharia (IME) – Rio de Janeiro, RJ – Brazil

Abstract. Biomedical texts are a rich information source that cannot be ignored. There are several text annotation tools that may be used to extract useful information from these texts. However, the multi-domain characteristic of these texts, and the diversity of ontologies available in this area, demands a careful analysis before choosing an annotation tool. This work presents an evaluation of the existing annotation tools, with focus on biomedical texts. Initially, based on a set of required characteristics, a tool selection was conducted. AutôMeta and Gate tools were selected for a more detailed evaluation. They were quantitatively and qualitatively evaluated. Results of such evaluation are discussed and bring to light the best/worst of each tool.

1. Introdução

The constant growth of data and publications in the Biomedical area has been pushing the creation and reuse of domain ontologies in that area, not only for structured data annotation, but also for text indexation and annotation. Particularly, text bases are a rich information extraction source, since many biomedical findings are available only in textual format. PubMed¹ is one of the most popular digital biomedical citation reference (more than 21 million texts). Each text citation is associated (indexed) using MeSH² thesaurus. However, in order to facilitate the extraction of information from texts, a more automated and detailed indexation is required.

Biomedical area texts are typically multi-domain, and require different ontologies for their annotation. The Open Biological and Biomedical Ontologies (OBO) Foundry [Smith et al. 2007] and the NCBO BioPortal [Noy et al. 2009] provide together more than 300 ontologies. The motivation of this work is to provide support for annotation with multiple ontologies. For instance, a paper about drug targets usually refers to proteins, diseases, organisms, pharmacogenomics, etc. Each of these terms can be annotated by different domain ontologies such as: GO (Gene Ontology) [The Gene Ontology Consortium 2000], for gene and protein annotations, NCBITaxon³ (NCBI organismal classification), for organisms, and PHARE (The PHArmacogenomic

¹ http://www.ncbi.nlm.nih.gov/pubmed/

² http://www.nlm.nih.gov/mesh/

³ http://bioportal.bioontology.org/ontologies/1132/

RElationships $Ontology)^4$, for pharmacogenomics techniques, such as the knockout technique. Based on these annotations, it is possible to establish useful correlations. For instance, a text may describe that the application of the knockout technique over a certain gene G of an organism O, led to its death. Thus, if annotated with the mentioned ontologies, an additional annotation extracted from this text would inferred: gene G is essential for organism O.

There are already a variety of (semi) automatic tools for text annotation, i.e., which provide support for the association of text expressions to ontology terms. The main goal of this work was to identify and compare such tools, with focus on texts and ontologies of the biomedical area. Initially, a set of tools has been selected. After, relevant requirements for biomedical text annotation, such as the support for manual and automatic annotation, and the flexibility for loading ontologies were observed. Only two of the selected tools attended these requirements. These two tools were then analyzed with respect to their annotation results, in terms of quantity and quality. An additional contribution of this work is to provide guidelines for annotation tool analysis.

The remainder of this work is organized as follows: section 2 introduces semantic annotation basic concepts and illustrates it in the biomedical scenario. Section 3 describes and analyzes semantic annotation tools. Section 4 reports the realized experiment, results and difficulties. Finally, conclusions and future works are presented in Section 5.

2. Semantic Annotation

Semantic annotation is an approach to achieve the concepts of the Semantic Web, whose information organization provides a means, in which the logical connection of terms establishes interoperability between systems [Shadbolt et al. 2006]. It proposes to annotate a document using semantic information from domain ontologies. Popov et al. (2003a) define semantic annotation as a "specific schema for generation and use of metadata, enabling new methods of information access". According to Ding et al. (2006), the semantic annotation should be explicit, formal and unambiguous, so that is publicly accessible, understood and identifiable, respectively.

More specifically, we emphasize that semantic annotation is an association between relevant expressions or terms of a document or from metadata, and concepts and instances described in the ontology. Figure 1 illustrates the associations between terms in a piece of text and terms of ontologies and taxonomies, and how these associations can enrich the text with the knowledge embedded in the ontology. Annotations can be inserted in the same document file or saved separately. They contribute to the information retrieval mechanisms that are able to interpret them.

The multi-domain characteristic of biomedical articles makes it difficult to obtain a well-annotated text with a single ontology. The ontologies of this area are built focused in only one domain. Therefore, for an article to be well-annotated, the use of multiple ontologies or taxonomies is needed. However, as mentioned previously, there are many available ontologies. Hence, a prior analysis of which ontologies are compliant with the domains of the articles is needed. In Figure 1 we can see that in a

⁴ http://bioportal.bioontology.org/ontologies/1550/

small text fragment it was necessary to use a thesaurus, a taxonomy and an ontology for the annotation.

3. Technologies to Support the Semantic Annotation of Texts

There are tools that provide support for the semantic annotation of documents (or texts) available on the Web. These may include different characteristics such as access to ontologies, intuitive graphical user interface, editors and repositories for ontologies storage, etc.



Figure 1. Associations between term_article and classe_ontology for ontologybased semantic annotation.

Regarding the kind of annotation, they are classified as manual and automatic. In the manual annotation, the user performs the whole process of marking the document, selecting the parts to be annotated and describing the annotation associated to a term of an ontology. In automatic annotation, the tool performs the annotation without user intervention, through the use of techniques such as natural language processing (NLP), machine learning and information extraction among others, to associate text expressions to ontology terms. There are tools that provide support for both manual and automatic annotation, and are considered to be hybrid. Another important characteristic is how the annotation is saved. It can be intrusive, which means the annotation is saved in the document, or non-intrusive, which means the annotation is stored in another file and do not modify the original document.

Other characteristics refer to the types of software platforms (desktop and Web), types of documents to be annotated (txt, pdf, etc.) and the use of ontologies for the annotation (which format and whether a user-choice ontology may be used). For this work, several tools have been analyzed and tested in accordance to these characteristics and are presented in the following section.

3.1. Semantic Annotation Tools

The selected tools are presented below, and Table 1 summarizes the characteristics previously described, observed in each of these tools.

Annotea [Kahan et al. 2001] is a project of the World Wide Web Consortium (W3C). The annotations of this tools refer to comments, notes, explanations or general comments on Web documents. It is part of the efforts of the Semantic Web and uses an annotation scheme based on Resource Description Framework (RDF). It stores the metadata of the annotations locally or on annotation servers.

 $Annozilla^5$ has the same characteristics as Annotea, but works as a plugin for Mozilla Firefox browser. It stores the annotations as RDF on a server. It also highlights the annotation of the documents, which remains when it is reloaded.

AutôMeta (Automatic Metadata annotation tool) [Fontes 2011] allows the annotation of one or more documents using an ontology previously selected. The annotations generated by the tool are stored using RDFa standard (Resource Description Framework in attributes)⁶.

GATE (General Architecture for Text Engineering) [Cunningham et al. 2002] is a tool for natural language processing applications. It integrates a development environment which includes plugins and other components that allow both the annotation or information extraction.

GoNTogle [Bikakis et al. 2010] is a tool for annotation and search. It also provides search facilities using a combination of semantic search and keyword. The annotations are saved as an instance in the ontology server and added to a list on annotations editor.

KIM [Popov et al. 2003b] is a web based platform for semantic search and annotation of data and documents. It has its own ontologies which includes general interest entities. The access to the features at KIM server is done through a Web interface (KIM Web UI), which allows traditional methods of searching by keyword or semantic search (entities, patterns).

Knowtator [Ogren 2006] is a plugin of Protégé, and allows an increase in ontologies to adapt to the user application. The annotation is done with the ontologies present in Protégé, from the region of the text selected to be annotated and the specification of the ontology to be used.

Melita [Ciravegna et al. 2002] is a tool that has its own ontologies, allowing the users to add their results to the used ontology, increasing it in each satisfactory annotation.

MnM [Vargas-Vera et al. 2002] is a tool that allows annotation on Web pages. It uses a learning algorithm on the annotations to posteriorly calculate the precision and recall of the annotations in the corpus. It integrates a Web browser with an ontology editor and provides APIs (Application Programming Interface) for connection between ontologies servers and information extraction tools.

⁵ http://annozilla.mozdev.org/

⁶ http://www.w3.org/TR/xhtml-rdfa-primer/

ONTEA [Laclavik et al. 2006] uses its own ontologies which is only related to addresses, names and e-mails.

RDFaCE (RDFa Content Editor) [Khalili and Auer 2011] is a plugin for TinyMCE Javascript WYSIWYG Editor that allows the intrusive annotation in RDFa standard. Instead of ontologies, uses APIs that suggest resources for the annotation. These resources provide appropriate URIs for objects, properties and namespaces.

RDFa Editor [Duma 2011] presents itself as a promising tool that uses RDFa as standard for the annotations. It allows arbitrary ontologies.

*Yawas*⁷ is a plugin developed for Mozilla Firefox and Google Chrome browser. The annotations are Web pages highlights, without using any semantic resource.

 Table 1. Characteristics of tools. Kind of annotation (A=automatic, H=hybrid, M=manual),

 Saved annotation (I=intrusive, NI=non-intrusive), Platform (D=desktop, W=web)

Tool	Kind of annotation	Saved annotation	Format of input documents	Format of ontologies	Arbitrary ontology	Platform
Annotea	М	NI	Web documents	-	No	W
Annozilla	М	NI	Web documents	-	No	W
Autômeta	Н	Ι	ТХТ	N-Triple, RDF, OWL, XML	Yes	D
GATE	Н	NI	PDF, TXT, HTML, DOC, ODT	RDF, OWL	Yes	D
GoNTogle	Н	NI	PDF, RTF, TXT, DOC, ODT	OWL	Yes	D
KIM	Α	NI	HTML	RDF, OWL	No	W
Knowtator	М	NI	PDF, TXT, HTML, DOC, ODT	RDF, OWL, XML	Yes	D
Melita	М	NI	PDF, TXT, HTML, DOC, ODT	OWL	No	D
MnM	М	NI	HTML, TXT	DAML + OIL, RDF	Yes	W
Ontea	А	NI	PDF, TXT, DOC, e-mails, e- mail attachments in HTML	OWL	No	D
RDFaCE	М	Ι	PDF, TXT, HTML, DOC, ODT	-	No	D
RDFa Editor	A	NI	PDF, TXT, HTML, DOC, ODT	RDF, OWL, XML	Yes	D
Yawas	М	Ι	Web pages	-	No	W

3.2. Tools Analysis

This analysis aims to identify which tools attend the required characteristics for the semantic annotation of biomedical documents. Such characteristics include the kind of

⁷ http://www.keeness.net/yawas/index.htm/

annotation and the possibility of using an arbitrary ontology for annotation. We aimed at selecting automatic annotation tools that also provide support for manual annotation. On one hand, the automation was needed due to the large volume of texts and also to the difficulty and high cost to keep specialists responsible for the manual annotation task. On the other hand, the manual annotation would be used for extra annotations, according to the user needs. Regarding the use of arbitrary ontologies, as stated before, the user needs different domain ontologies for the annotation of a biomedical text, and therefore, the tools should be flexible enough to load the selected ontologies.

In this preliminary evaluation, various tools have been dismissed for not meeting the required characteristics previously mentioned. They are: Annotea, Annozila, KIM, Knowtator, Melita, Ontea, RDFa Editor, RDFaCE and Yawas.

Despite the fact that the KIM tool was dismissed it provides a friendly interface and has an excellent support through a mailing list. However, to make annotations using an arbitrary ontology, the KIM uses GATE platform processing resources via command line. Due to that, the GATE tool was used directly.

RDFa Editor and MnM tools were discarded for technical problems. Although RDFa Editor attends the required characteristics, it is still under development and was not available for download and testing until the moment of this article closure. MnM tool, although well documented and capable of performing automatic annotation based on ontologies, presented technical limitations that precluded its use. The extraction system information necessary for the integration with the tool is no longer available on the developer portal, and it seems the project was discontinued.

GoNTogle tool presented failures in the connection with the Protégé server, this problem was not solved until the closure of this work, despite recommendations suggested by the developers. This tool could not be effectively tested, but as it has all the characteristics desired for the work, this problem has not definitively ruled out the use of this tool in future trials.

The AutôMeta and GATE tools were the only ones that did not presented failures and could participate on a further experiment, for the annotation of articles in the biomedical scenario. The experiment and its results are described in the next section.

Additionally, when conducting these tests, other problems related to the use of such tools were found. AutôMeta presented a problem using its graphical interface, however it could be used through command line. GATE failed when loading large ontologies, probably caused by problems with memory management. For the storage of non-intrusive annotation, each document must be saved individually, which can become a limitation for tests involving the annotation of a very large set of texts.

Overall, we have observed that many tools are described in order to make semantic annotation of documents, but actually they highlight the text, without reference to semantic content through ontologies or taxonomies. Another consideration relates to the lack of documentation. Many of them have only the article which describes the tool or a basic explanation on a Web page. Finally, there is also the problem of project discontinuation.

4. Experiment and Results

4.1. Scenario

AutôMeta and GATE were selected for the semantic (ontology-based) annotation experiment. As the biomedical field is very large, we restrict ourselves to the subdomain of studies on targets for drug development to combat neglected tropical diseases, which is the focus of some researchers at the Proteins and Peptides Biochemistry Laboratory at Fiocruz. To build the corpus of this experiment 108 full papers in the biomedical area were selected, extracted by searching in PubMed for keywords related to this subdomain. The files were obtained in HTML format.

PHARE ontology, which is available at NCBO BioPortal, was selected. It was chosen because it belongs to the same domain of the corpus and also for its OWL format and small size (it includes 228 classes and 83 properties, and a taxonomy depth of 5 classes), which are required to use the GATE tool. This ontology describes concepts and functions that represent the interests of pharmacogenomics relationships.

4.2. Analysis and Results

In order to analyze the output files of the executions of the AutôMeta and GATE tools, it was necessary to develop different scripts for each tool. These results were then quantitatively analyzed. The number of distinct annotated classes and terms in each article were calculated.

Figures 2 and 3 illustrate graphs comparing the number of distinct annotated terms and the number of distinct annotated classes, respectively. It is worth noting that GATE annotated more terms than AutôMeta, and that it used a larger amount of classes of the ontology. Interestingly, despite these quantitative differences, the annotation profiles are similar, with both tools featuring high peak annotation for the same documents. These high peaks indicate an adaptability of the ontology to the subdomain of the articles.

Both tools behaved similarly when the least number of terms and classes were annotated, which could be noted in articles 2;7;27;81;98 and 101. In this case, while AutôMeta annotated 33 terms and 5 classes, GATE annotated 18 terms and 31 classes. The most annotated article using AutôMeta tool was the article 77, with 165 terms. This article was the fourth most annotated in GATE (169 terms). In the case of GATE tool, article 85 was the most annotated one, with 200 terms. The same article, using AutôMeta, was the sixth most annotated (151 terms). These numbers show a slightly different behavior with respect to the most annotated articles. We also note from Figure 2, that lines get closer (almost intersect each other) when there is a low number of terms annotated in the article.

Observing the output files (annotated articles) with respect to the number classes used for annotation, for each tool, we could note that for the same part of the text, the GATE tool was capable to recognize much more classes related to the domain ontology. The average number of classes used was 27 for AutôMeta, while for GATE it was 143. The additional annotation obtained using the GATE tool was due to the fact that it is in essence a natural language processing tool, and maybe, also to its ability to use synonymy information.





Figure 2. Comparative analysis between number of distinct terms.

Figure 3. Comparative analysis between the number of distinct classes.

Unfortunately, it was not possible to evaluate all annotated articles in a qualitative manner. However, the article 85 was chosen to a qualitative analysis for its representativeness with respect to the number of annotated terms.

As AutôMeta annotates using RDFa standard, this facilitated the extraction of information as well-formed RDF triples (subject, predicate, object), as shown in Table 2. Moreover, Autômeta is able to make inferences, and produce additional annotations, not only for hierarchical relations, but also for domain specific relations. However, its focus is not on the use of natural language processing resources, which explains its low performance with respect to the number of annotated terms.

Subject	Predicate	Object	
phare:Intramuscular	Label	Intramuscularly	
phare:Organism	Label	living system	
phare:Drug	Label	content	
phare:Infection	Label	http://www.stanford.edu/~coulet/phare.owl#Symptom	

Table 2. Examples of RDF triples resulting from the Autômeta annotation

Regarding GATE, although it does not use the RDFa standard, it generates its output in its own format, which facilitates information extraction (via script). A small part of the result extracted from the GATE annotation in article 85 can be found in Table 3. It is worth noting that GATE was able to annotate based on synonymy information (owl metadata). This is the case of "DrugSentivity", "DiseaseExacerbation" and "DrugDose" classes, which were used to annotate synonymous terms, "Tolerance", "Drug" and "G", respectively.

Class	Term
http://www.stanford.edu/~coulet/phare.owl#DrugSensitivity	Tolerance
http://www.stanford.edu/~coulet/phare.owl#Drug	Drug
http://www.stanford.edu/~coulet/phare.owl#DiseaseExacerbation	Growth
http://www.stanford.edu/~coulet/phare.owl#DrugDose	G

Table 3. Information	derived	from the	GATE	annotation
-----------------------------	---------	----------	------	------------

Analyzing the results of each tool annotation for article 85, we identified annotated terms that were not relevant, i.e., wrong annotations (when the used class is not related to the annotated term) and superfluous annotations (not really important to the focus of the article). Therefore, it was possible to calculate the annotation *precision*, which is the rate between the number of relevant annotations and the number of annotations (relevant/total). Both tools showed a low performance, 56% for AutôMeta and 53% for GATE. The *recall* was not calculated because it depends on what is relevant for the users, i.e., a manual annotation performed by a domain specialist would be necessary.

Although the annotations with PHARE ontology were significant, other important terms for the subdomain of the articles would also be relevant. However, PHARE did not cover all these subdomains, such as names of diseases, organisms, genes and proteins. This emphasizes, one more time, the need for annotation with multiple ontologies, so that these subdomains could also be covered. This study did not include the annotation with multiple ontologies due to problems found, as reported in the next section.

In summary, the main advantage of AutôMeta tool is that it uses RDFa standard, and that it supports the load of large ontologies such as the Molecule Role and NCI Thesaurus. On the other hand, GATE is a very solid and mature tool. Its main advantage is that it uses natural language processing resources. Both are user-friendly, but GATE is a bit more complicated at first. With respect to the inference ability, both AutôMeta and GATE include it. However, the focus of AutôMeta is on an intensive exploration of the ontology inference potential. Finally, both tools have a good documentation and are free.

4.3. Difficulties

One of the objectives of this study was to highlight the importance of semantic annotation with multiple ontologies. For this purpose, we designed an experiment where, in addition to the PHARE ontology, other ontologies were planned to be used: the Molecule Role⁸ ontology and NCBITaxon taxonomy. The first refers to an ontology

⁸ http://bioportal.bioontology.org/ontologies/1029/

used to annotate names of proteins and protein families, and the second refers to a taxonomic classification of living organisms. However, it was not possible to conduct such experiment because we could not load these other ontologies into the GATE tool. This was due to their large size (Molecule Role has 9,217 classes and 41.8Mb in the XML format and NCBITaxon has 513,248 classes and 243Mb in the OWL format). Therefore, we can conclude that the GATE tool is not prepared to handle large ontologies, a typical feature of biomedical ontologies. Although the AutôMeta tool presented long-term executions, it was able to make annotations with all the chosen ontologies.

A possible solution to this problem was envisaged. Dividing these ontologies into modules, taking only the parts of interest to the user, would reduce their size and facilitate their reuse. There are tools available on the web for the modularization and/or module extraction of ontologies [Garcia et al., 2012]. For such tools it is also required a large memory capacity to load them and generate the corresponding modules. Nevertheless, these tools are not prepared to deal with biomedical ontologies because they generate modules based on the names of the classes. Typically, in most biomedical ontologies, class names correspond to numeric identifiers of ontologies, and not to the corresponding terms. For example, in the Molecule Role ontology, the class name for the "enzyme" term (label) has value "IMR0000207" (enzyme identifier). Therefore, because of time restrictions, such alternative was left for future work.

5. Conclusions

This work surveys semantic annotation tools in the light of the biomedical scenario. Among the characteristics analyzed, the focus was on investigating their ability for automatic and manual annotation, their flexibility with respect to loading arbitrary ontologies, and their compliance to input/output standards.

Among a larger set of tools, AutôMeta and GATE were identified as the most adequate tools to attend the biomedical domain requirements, as both of them are able to load arbitrary ontologies and provide support for manual and automatic annotation. An experiment was then conducted to further evaluate these two tools. According to its results each tool has benefits and drawbacks. The AutôMeta tool is able to generate annotations using the RDFa standard, and to support the load of large ontologies. On the other hand, it shows a low performance with respect to the use of natural language processing resources, which is the main feature of the GATE tool. Both are userfriendly, provide inference ability, have a good documentation and are free.

From the experiment, we identified how important, in the biomedical scenario, it is to support annotation with multiple ontologies. Therefore, as future work we intend to modify the AutôMeta tool to use multiple ontologies, and new experiments will be carried out. The choice of AutôMeta is mainly due to its compliance to the RDFa standard, which facilitates the structuring of semantic data about each text and the consequent use of these data. Moreover, further improvements to AutôMeta include using additional natural language processing resources.

Acknowledgements

The authors would like to thank CNPq(309307/2009-0; 486157/2011-3), Fundação Ricardo Franco (PBIP), FAPERJ(E-26/111.147/2011; E-26/102.521/2010) for partially funding their research projects.

References

- Bikakis, N., Giannopoulos, G., Dalamagas, T. and Sellis, T. (2010) "Integrating Keywords and Semantics on Document Annotation and Search". Proc. of the Int. Conf. on the move to meaningful internet systems: Part II, Hersonissos, Crete, Greece.
- Ciravegna, F., Dingli, A., Petrelli, D. and Wilks, Y. (2002) "Timely and Non-Intrusive Active Document Annotation Via Adaptive Information Extraction". In Semantic Authoring, Annotation and Knowledge Markup (SAAKM02), ECAI.
- Cunningham, H., Maynard, D., Bontcheva, K. and Tablan, V. (2002) "GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications". Proceedings of the ACL.
- Ding, Y., Embley, D.W. and Liddle, S.W. (2006) "Automatic Creation and Simplified Querying of Semantic Web Content: An Approach Based on Information-Extraction Ontologies". Proceedings of the First Asian Semantic Web Conference (ASWC'06). Berlin Heidelberg: Springer, pp. 400-414.
- Duma, M. (2011) "RDFa Editor for Ontological Annotation". Proceedings of the Student Research Workshop associated with RANLP 2011, pages 54–59, Hissar, Bulgaria.
- Fontes, C.A. (2011) "Explorando Inferência em um Sistema de Anotação Semântica", Master's thesis, Dept. of Computer Science, Military Inst. of Engineering, Rio de Janeiro, Brazil.
- Garcia, A.C., Tiveron, L., Justel, C., Cavalcanti, M.C. (2012) "Applying Partitioning Algorithms to Modularize Large Ontologies". In: Proc. of ONTOBRAS.
- Kahan, J., Koivunen, M., Prud'Hommeaux, E. and Swick, R.R. (2001) "Annotea: An Open RDF Infrastructure for Shared Web Annotations", Proceedings. of the WWW10 International Conference, Hong Kong.
- Khalili, A. and Auer, S. (2011) "The RDFa Content Editor From WYSIWYG to WYSIWYM". http://svn.aksw.org/papers/2011/ISWC_RDFaEditor/public.pdf, September.
- Laclavik, M., Seleng, M., Gatial, E., Balogh, Z. and Hluchy L. (2006) "Ontology based Text Annotation – OnTeA". Proceedings of 16-th European-Japanese Conference on Information Modelling and Knowledge Bases (EJC'2006), pp. 280-284, Trojanovice, Czech Republic.
- Noy, N. F., Shah, N.H., Whetzel, P.L. et al. (2009) "BioPortal: Ontologies and Integrated Data Resources at the Click of a Mouse". Nucleic Acids Res. Jul 1;37(Web Server issue):W170-3. PMID: 19483092.

- Ogren, P.V. (2006) "Knowtator: a Plug-in for Creating Training and Evaluation Data Sets for Biomedical Natural Language Systems". Proceedings of the 9th International Protégé Conference 73–76.
- Popov, B., Kiryakov, A., Ognyanoff, D., et al. (2003a) "Towards Semantic Web Information Extraction". Human Language Technologies Workshop at the 2nd International Semantic Web Conference (ISWC2003), Florida, USA, 20 October.
- Popov, B., Kiryakov, A., Kirilov, A., Manov, D., Ognyanoff, D. and Goranov, M. (2003b) "KIM - Semantic Annotation Platform". In 2nd International Semantic Web Conference, (Florida, USA, 2003), 834-849.
- Shadbolt, N., Hall, W. and Berners-Lee, T. (2006) "The Semantic Web Revisited". IEEE Intell. Syst., 21(3): 96-101.
- Smith, B., Ashburner, M., Rosse, C., et al. (2007) "The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration". Nature Biotechnology 25, 1251 - 1255.
- The Gene Ontology Consortium. (2000) "Gene ontology: Tool for the Unification of Biology". Nat. Genet., 25(1):25-9.
- Vargas-Vera, M., Motta, E., Domingue, J., Lanzoni, M., Stutt, A. and Ciravegna, F. (2002) "MnM: Ontology Driven Semi-automatic and Automatic Support for Semantic Markup". Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, p.379-391.