Semantic Web Technologies for Interpreting DNA Microarray Analyses: The MEAT System

Khaled Khelif[1], Rose Dieng-Kuntz[1], and Pascal Barbry[2]

[1] INRIA Sophia Antipolis, 2004 route des lucioles 06902 Sophia Antipolis, France
{khaled.khelif, rose.dieng}@sophia.inria.fr
[2] IPMC, 660 route des lucioles 06560, Sophia Antipolis, France
barbry@ipmc.fr

Abstract. This paper describes MEAT (Memory of Experiments for the Analysis of Transcriptomes), a project aiming at supporting biologists working on DNA microarrays. We provide methodological and software support to build an experiment memory for this domain. Our approach, based on Semantic Web Technologies, is relying on formalized ontologies and semantic annotations of scientific articles and other knowledge sources. It can probably be extended to other massive analyses of biological events (as provided by proteomics, metabolomics…).

1 Introduction

Most of our knowledge can be stored in documents published on the web, in databases or correspond to human interpretations of experimental results. This knowledge is essential for checking, validating and enriching new research work. But the large amount of data, either from sources that are internal or external to users' organisations make the efficient detection, storage and use of this(ese) knowledge(s) huge tasks. This is especially true for investigators manipulating huge amounts of biological data, such as those working with DNA microarray[1] experiments, where several hundreds of experimental conditions can be easily analysed against 100,000 probes, and are to be linked to thousands of scientific reports. Under these situations, biologists need helping tools suitable for the interpretation and/or validation of their experiments, which would ultimately facilitate the planning of further experiments.

The MEAT project presented in this paper aims at proposing solutions to address these needs.

1.1 Context

The technology DNA Microarray has been developed after the full sequencing of many genomes in order to get about gene functions under many different biological contexts. Typical microarray experiments can assess thousands of genes simultaneously. Thus, they provide a huge amount of information and present difficulties for a biologist, in particular when s/he has to validate and interpret the obtained results.

---

[1] http://www.gene-chips.com/

The needs of biologists can be summarized as follows:

− *A view on related experiments*: trying to identify relations between experiments (local databases, on-line data…) and to discover new research paths and different perspective for exploration.
− *Support the validation of experimental results*: by searching in documents about the studied phenomenon information which argue, confirm or invalidate the biologist's assumptions; this requires rich annotations.
− *Support the interpretation of experimental results*: by identifying new/known relations or/and interactions between genes, cellular components and biological processes; this requires inference capabilities over the annotations.

MEAT has been developed in collaboration with biologists of the Nice Sophia Antipolis DNA Microarray platform (located at the IPMC[2] laboratory with the aim to provide an experiment memory for this domain: as such, it constitutes a specific case of corporate semantic web at the scale of a community.

1.2 Corporate Memory and Semantic Web

Currently, semantic web techniques can play a major role in knowledge management and in the building of a corporate memory.

In fact, ontologies can be used throughout knowledge representation and provide formal grounding for representing the semantics of the knowledge pieces and can guide the creation of semantic annotations constituting a set of all meta-level characterizations easing knowledge source description, evaluation, and access.

[7] proposes to materialize a corporate memory through a "corporate semantic web" by using ontologies to formalize the vocabulary shared by a community, and semantic annotations based on these ontologies to describe heterogeneous knowledge sources (textual corpora, databases,…) and facilitate their access via intranet/internet.

In our case this corporate semantic web is constituted of:

− Resources: databases of experiments, persons and articles which can come from internal sources such as specific documentation databases for each biologist or from external sources such as on-line documentation databases (e.g. Medline[3]).
− Ontologies: MeatOnto, which is a multicomponent ontology composed of 3 ontologies (cf. §3.1).
− Ontology-based semantic annotations: which describe experiments stored in the databases (results, interpretations) and knowledge extracted from scientific articles. These annotations can be generated manually (Annotation editor) or automatically (Text mining techniques).

2 Our Approach

Taking into account biologist's needs, we aimed at building an experiment memory for the DNA microarray domain.


We first made an inventory of different knowledge sources which would constitute this memory:

− MEDIANTE[4]: an information system for biochip experiments developed at IPMC. It supports the management of a microarray project from the design of the DNA probes to the storage of the results.
− Journal articles: for each biochip project, biologists constitute a textual corpus of papers concerning genes supposed a priori interesting for experiments. Of particular interest was the selection by the biologists of review articles, such as those provided by series or found on the web. Such a selection was useful, in the sense that it provided overviews of (a) specific field(s), written by a specialist of this field, and selected by another specialist (i.e. the biologist doing the microarray experiments). Our hypothesis was that such selection would decrease the number of contradictory points of view provided by wider approaches using the web scanning.
− Biologist's viewpoint: interpretation of results, connection/correlation of phenomena or experiments…

Figure 1, which summarizes the MEAT architecture, recapitulates the different stages of our approach:

1. *Construction of an ontology* which describes all kinds of resources used by biologists (experiments database, scientific papers, biomedical features); this stage produced the *MeatOnto ontology*.
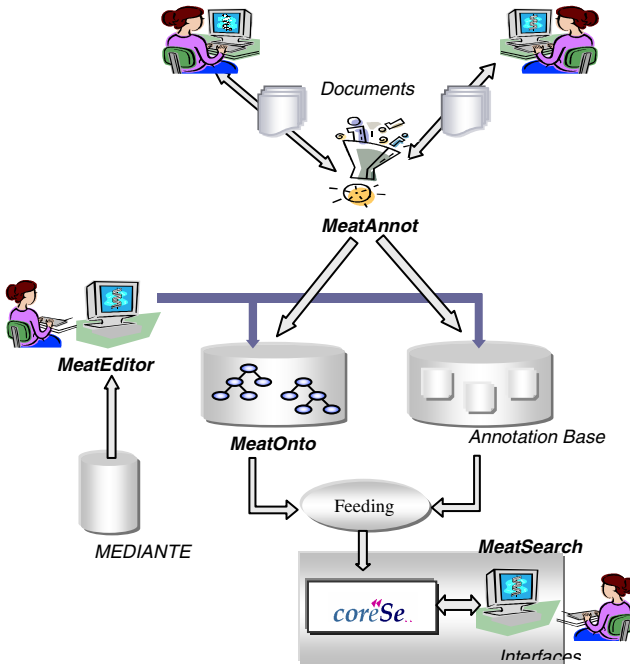
Fig. 1. The MEAT Architecure

---

2. *Structuring knowledge contained in the MEDIANTE database* fields describing experiments in order to enable biologists to carry out searches, with the aim of finding similarities/correlations between experiments; this stage takes advantage of *semantic annotations based on MeatOnto.*
3. *Automation of the semantic annotation journal articles* interesting for experiments, using the MeatOnto ontology; for this, we designed and implemented the *MeatAnnot module.*
4. *Supplying the biologists with interfaces* to add new MeatOnto-based annotations on experiments or papers; this led to the implementation of *MeatEditor.*
5. *Facilitating the validation of experimental results* by proposing bibliographical searches guided by MeatOnto and by using the semantic annotations; this resulted in the *MeatSearch module.*
6. *Facilitating the interpretation of results* thanks to advanced inferences on semantic annotations which assist the explanation of a particular behaviour; this led to additional functionalities in the *MeatSearch* module.

In the following, we will present briefly MeatOnto (lack of space) and we will focus on MeatAnnot, the evaluation study on the generated annotations and MeatSearch with examples of use.

3   Meat Components

3.1   MEAT Ontologies: MeatOnto

MeatOnto is a modular ontology composed of 3 sub-ontologies, each dedicated to the description of a precise field:

− UMLS: the goal of the Unified Medical Language System project is to help health professionals and researchers to use biomedical information from a variety of different sources [12]. It consists of a metathesaurus which collects millions of terms belonging to the nomenclatures and terminologies defined in the biomedical domain and a semantic network which consists of 135 semantic types and 54 relationships, and it represents a high-level abstraction for the metathesaurus. To describe the biological domain (drugs, cells, genes, processes…), we chose UMLS, and we considered the UMLS semantic network as a general ontology: the hierarchy of semantic types can be regarded as a hierarchy of concepts and the terms of the metathesaurus as instances of these concepts.
− MGED: it was proposed by the MGED[5] group (Microarray Gene Expression Data) to describe microarray experiments [22] in order to facilitate the sharing of biochip experiment results. This ontology is used in MEAT to describe the experiments stored in MEDIANTE.
− DocOnto: we developed this ontology to describe metadata on papers (authors, source…) and on annotations (generated_by, validated_by…). This ontology describes the structure of papers (abstract, sentence, relation…) and it links documents to UMLS concepts (has_relation, speaks_about_genes…).

The UMLS component was built automatically by a script translating the text format of the semantic network to an RDFS format and the DocOnto component was constructed progressively in order to cover our needs for the description of the knowledge contained in the scientific papers.

3.2 MeatAnnot

3.2.1 Generation of Annotations

In spite of its advantages, the creation of semantic annotations is a difficult and time-consuming process for biologists. Thus, we developed a system called MeatAnnot which, starting from texts (papers provided by biologists), allows to generate a structured annotation, based on MeatOnto, and describes the semantic content of this text.

MeatAnnot uses the NLP (Natural language processing) tools GATE[6], TreeTagger[11], RASP[2] and our own extensions dedicated to extract semantic relations and UMLS concepts. It processes texts and extracts interactions between genes and the others UMLS concepts.

So, for each sentence it detects an instance of UMLS relationship, it tries to detect the instances of UMLS concepts linked by this relationship and it generates an annotation describing this interaction (see more details in [13]).

The generation method breaks up into three steps described below:

Step1: Relation detection
In this step we used JAPE [6], a language based on regular expressions and allowing us to write information extraction grammar for texts processed by GATE. So, for each UMLS relation (such as interacts_with, expressed_in, disrupts…), an extraction grammar was manually created to extract all instances of this relation.

The example below shows a grammar which allows detection of instances of the semantic relation "Have an effect" with its different lexical forms in the text (e.g. has an effect, had effects, have a positive effect…).

```
{Token.lemme == "have"}|{SpaceToken}({Token.string == "a"}|{Token.string == "an"})?({SpaceToken})?
({Token.string == "additive"}|({Token.string == "synergistic"}|{Token.string == "inhibitory"}|{Token.string ==
"greater"}|{Token.string == "functional"}|{Token.string == "protective"}|{Token.string == "mono-
genic"}|{Token.string == "positive"})?({SpaceToken})?{Token.lemme == "effect"}
```

Step2: Term extraction
To extract terms, MeatAnnot uses the Tokeniser module of GATE and the TreeTagger. The tokeniser splits text into tokens, such as numbers, punctuation and words, and the TreeTagger assigns a grammatical category (noun, verb...) to each token.

After tokenizing and tagging texts, MeatAnnot uses an extraction window of four (four successive words are considered as a candidate term) and for each candidate term, if it exists in UMLS, MeatAnnot processes the following word, otherwise it decreases the size of the window till zero.

To interrogate UMLS, MeatAnnot uses the UMLSKS (KS: Knowledge Server). This server provides access and navigation in the UMLS metathesaurus and the UMLS semantic network. The answer received in XML format (if the term exists in UMLS) is parsed to obtain information about the term (semantic type, synonyms …)

Step3: Annotation generation

In this step MeatAnnot uses the RASP module which assigns a linguistic role to sentence words (subj, obj …) which allow to find out concepts linked by the relation. So for each relation, MeatAnnot verifies if its subjects and objects were detected as UMLS concepts and generates an annotation describing an instance of this relation.

The example below summarizes the process steps. Let us consider the sentence:

"*In vitro assays demonstrated that only p38alpha and p38beta are inhibited by csaids.*"

First: by applying the extraction grammars on this sentence, MeatAnnot detects that it contains the UMLS relation "*inhibits*".

Second: the table below describes the result of the term extraction phase:

Table 1. Term extraction results

Term	Semantic type	Synonyms
in vitro	Qualitative Concept	N/C
P38alpha	Gene or Genome	MAPK14 gene, CSBP1…
P38beta	Gene or Genome	MAPK11 gene, SAPK2…
Csaids	Pharmacologic Substance	Cytokine-Suppressant Anti-Inflammatory Drug

Third: MeatAnnot applies the RASP module on the sentence and parses the result to detect the different linguistic roles of words.

The result of RASP on this sentence is:

```
(|ncsubj| |demonstrate+ed:4_VVN| |assay+s:3_NN2| _ )
(|clause|| |demonstrate+ed:4_VVN| |inhibit+ed:11_VVN|)
(|ncsubj| |inhibit+ed:11_VVN| |p38alpaha:7_NN2| |obj|)
(|ncsubj| |inhibit+ed:11_VVN| |p38beta:9_NN2| |obj|)
(|arg_mod| |by:12_II| |inhibit+ed:11_VVN| |csaid.+s:13_NN2| |subj|)
(|conj| _ |p38alpaha:7_NN2| |p38beta:9_NN2|)
(|mod| _ |p38alpaha:7_NN2| |only:6_RR|)
(|mod| _ |p38beta:9_NN2| |only:6_RR|)
```

p38alpha and *p38beta* are detected as the objects of the relation *inhibits* and *csaids* as its subject.

MeatAnnot generates an RDF annotation for these two instances and adds it to the paper annotation.

```
<m: Pharmacologic_Substance rdf:about='csaids#'>
   <m:inhibits>
        <m:Gene_or_Genome rdf:about='p38alpha#'/>
   </m:inhibits >
   <m:inhibits >
        <m: Gene_or_Genome rdf:about='p38beta#'/>
   </m:inhibits >
</m:Pharmacologic_Substance>
```

After text processing, MeatAnnot generates an RDF annotation describing all these interactions in the article and stores it in the directory containing the annotations

of the other papers. These annotations can then be used, either in a bibliographical search or in a more complex IR (Information Retrieval) scenario such as searching interactions between genes or of genes with the other biomedical features.

3.2.2 Validation of Annotations

To validate our annotations we adopted a user-centered approach: we chose randomly a test corpus (2540 sentences) from the documents provided by biologists and we presented the suggestions proposed by MeatAnnot to biologists via an interface in order to evaluate their quality.

Since these annotations will be used in an IR context, we focused on classic IR quality measures for indexing and we adapted them to our case.

We noticed also that some suggestions were considered as correct but not useful to the biologists since they described a basic or vague knowledge. So, we introduced a new measurement called *usefulness* which measures the rate of useful suggestions.

Table 2. Measures for the quality of the annotations

	Measures
Precision	$\dfrac{\text{Nb suggestions correctly extracted}}{\text{Nb all suggestions extracted}}$
Recall	$\dfrac{\text{Nb suggestions correctly extracted}}{\text{Nb suggestions that should be extracted}}$
Usefulness	$\dfrac{\text{Nb useful suggestions extracted}}{\text{Nb suggestions correctly extracted}}$

Precision relates to the absence of noise (also called commission) in the extraction and *recall* relates to the absence of silence (also called omission).

Table 3. Quality of Meatannot suggestions

	Suggestion	Correct	Missing	Useful	Precision	Recall	Usefulness
Result	454	372	224	357	0,82	0,62	0,96

The second column describes the number of relations correctly extracted from texts. The difference with the number of suggestions proposed by MeatAnnot is due principally to the errors generated by the NLP tools (e.g. wrong grammatical category or linguistic role) and to the missing terms in UMLS (subject or object of a relation not found in UMLS). Nonetheless a good precision is obtained since 82% of the suggestions were correct.

The third column describes the number of relations not extracted by MeatAnnot, these missing suggestions are also due to the errors generated by the NLP tools and mainly to relations deduced by the biologist (when s/he reads the sentence) and which cannot be generated automatically.

Example of errors generated by the NLP tools:

"TRP gene, which belongs to the TRP-homolog group is expressed in neurons"

In this sentence where the relation "expressed_in" is detected, the RASP module suggests that "*which*" is the subject of the relation, so MeatAnnot does not generate the annotation because "*which*" is not an UMLS term and it losts the interaction between the "*TRP gene*" and "*neurons*".

Example of missing relations:

"*Upon interferon-gamma induction, after viral infection for example, a regulator of the proteasome, PA28 plays a role in antigen processing.*"

In this example, MeatAnnot extracts automatically the relation "PA28 plays_role antigen processing" but a biologist who reads this sentence can deduce, using his/her implicit knowledge, another relation which is "interferon-gamma have_effect PA28".

Finally, MeatAnnot has a good usefulness since 96% of correct suggestions are considered useful by biologists.

These results prove that MeatAnnot generates good quality annotations, an essential trait for a use in an information retrieval context.

3.3   Use of Annotations: MeatSearch

For enabling the biologists to use these annotations, we developed a tool called Meat-Search based on CORESE (Conceptual Resource Search Engine) [4][5] and composed of a set of GUI which allows users to ask queries on the annotation base.

MeatSearch translates the CORESE results to graphical or/and textual presentation (using adequate XSLT [3] style sheets) which is understandable by biologists. It also provides links to the sentence from which the annotation was extracted and to the document containing this sentence. This offers an interesting documentation on the annotations and this ability to trace the provenance is very useful for validation.

3.3.1   Use of CORESE

To formalize our ontologies and annotations, we chose RDFS [15] and RDF [14] languages, which are recommended by W3C, respectively to represent light ontologies and to describe web resources using ontology-based semantic annotations.

This choice enabled us to use the search engine CORESE which allows to:

− Navigate in the annotation bases taking into account the concept hierarchy and the relation hierarchy defined in the ontologies.
− Add rules which complete the annotation bases.
− Reason on the whole annotation base constructed from different and heterogeneous sources (papers, experiment database): it allows the biologist to deduce implicit and explicit knowledge about a gene.
− Use different levels of access (admin, public, group…) to the annotation base.
− Have different views on the annotations.

3.3.2   Examples of Use

CORESE proposes a query language for RDF very similar to SPARQL[6](currently under discussion as a w3c working draft); it enables to write queries constituted of a Boolean combination of RDF triples.

For example, the following query enables to retrieve all relations between a gene called "cav3.2" and a part of the human body:

```
select ?g ?r ?b where
?g   rdf:type   m:Gene_or_Genome
?g   =   'cav3.2'
?g   ?r   ?b
?b   rdf:type   m:Body_Part__Organ_or_Organ_Component
```

This query is generated automatically by MeatSearch and the result is formatted in a graphical representation (Figure 2) to facilitate its visualisation.
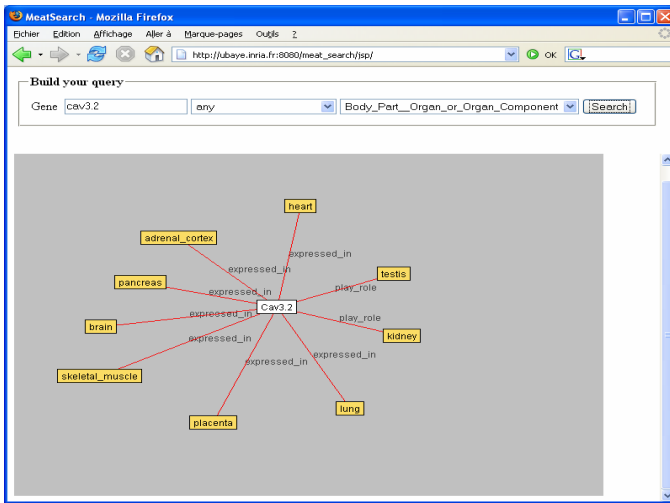

Fig 2. Result page of MeatSearch for the previous query

CORESE provides a rule language [4] which enables us to deduce new knowledge from existing one. The production rules are applied on the annotation base to complete it and to add more information in order to reduce the silence in the IR phase.

These rules are produced progressively by discussing with biologists.

An example of use of rule is:

"For each receptor which activates a molecular function, if this function plays a role in an organism function, the receptor can play the same role"

This rule is expressed as:

---

[6] http://www.w3.org/TR/rdf-sparql-query/

```
IF    ?r   rdf:type    m:Receptor
      ?r   m:activate   ?mf
      ?mf   rdf:type    m:Molecular_Function
      ?mf  m:play_role ?of
      ?of rdf:type    m:Organism_Function
THEN
      ?r    m:play_role   ?of
```

A last example concerns the addition of metadata on an annotation so as to give more information on:

− The source of the resource: the biologist who provided the paper to annotate or the biologist who made the experiments.
− The source of the annotation: generated automatically by MeatAnnot vs. added/validated by a biologist.
− The general topic of the annotation: the different biologists may have different centers of interest about the same experiment.

The annotation below describes a paper provided by a biologist named Pascal and related to the lung development.

```
<do:paper rdf:about='http://www-sop.inria.fr/acacia/meat/lungrepair.pdf'>
    <do:providedBy>Pascal</do:providedBy >
    <do:relatedTo >
        <m: Organ_or_Tissue_Function  rdf:about='lung_development #'/>
    </do:relatedTo >
    ....Annotation…
    <do:generatedBy>MeatAnnot</do:generatedBy>
    <do:validatedBy>Pascal</do:validatedBy>
    …Annotation…
</do:paper>
```

MeatSearch uses these metadata to propose different views on the annotations base related to sources (annotation source), the context (general topic of the annotation) and the method of the annotation generation (automatically vs. manually).

Queries on these metadata are very useful for browsing the annotation base, for checking its coherence and for all the validation phase.

4   Conclusions

4.1 Discussion

In this paper, we presented a method based on semantic web technologies for building an experiment memory for the biochip domain. This method can be generalized to any life science domain (chemical domain, physical domain…), since they have the same needs such as the support for the validation and the interpretation of experiments results. In fact, the modules presented are reusable and rely on standard technologies. MeatAnnot is independent of NLP tools and can use any ontology.

Our approach proposes some solutions to problems raised in the final discussion of W3C Workshop on Semantic Web for Life Sciences [23]:

- Good quality of the annotations extracted automatically: MeatAnnot annotations.
- Adequate representation of the context: our metadata on annotations.
- Possibility of reasoning on annotations: CORESE enables such reasoning.
- Semantic web browsing: we offer an automatic association of semantics to the memory resources and we provide a user-interface support.

One originality of this work consists of (a) the integration of metadata on annotations which gives new ways of reasoning and more information on the annotation base, (b) the use of several technologies (such as NLP, Ontologies, Semantic annotations, CORESE) to provide a robust Corporate Semantic Web Application.

Finally, we think that an evaluation study on the generated annotations (as we propose in this paper) is necessary since this generation phase is expensive and often irreversible.

4.2   Related Work

The method on which MeatAnnot rests can be compared with (a) work using information extraction for the biology domain [21] and (b) work on the generation of semantic annotations for the semantic web [9]. In the domain of article mining in biology, statistical techniques and machine-learning algorithms has been proposed [20] for discovering interactions among genes from article abstracts in biology in the PubMed base. Our approach relying on linguistic tools differs from machine-learning-based approaches proposed by [1] [17]. It also differs from [8] that offers generation of annotations consisting of concept instances, in order to enrich the ontology: our approach allows the generation of semantic annotations based not only on concept instances but also on relation instances.

The couple MeatAnnot/MeatSearch generates and uses ontology-based semantic annotations extracted from texts. It offers an information retrieval system based on these annotations: this system has some similarities with Textpresso system proposed by [16]. Textpresso is an ontology-based information retrieval and extraction system for biological literature. It identifies terms (instances of the ontology concepts) by matching them against regular expressions and encloses them with xml tags. Textpresso also offers user interfaces to query these annotations. But Textpresso has the following drawbacks: (i) the annotation is embedded in the text, which makes difficult its reuse by other systems (while MeatAnnot generates an RDF annotation separate from the document), and (ii) it needs thousands of regular expressions to extract relevant terms (while in MeatAnnot, linguistic analysis performs the text matching task).

Relying on NLP techniques, MeatAnnot differs from semantic annotation systems that use a machine-learning based information extraction approach. These systems (for example S-CREAM [10] and MnM [24]) rely on manually annotated documents on which the method can be trained.

Relation extraction was studied by [19] that propose the CAMELEON method/system which allows the extraction of semantic relations between terms using linguistic patterns (For example "X is Indefinite_Article Y" for hypernomy relation). This method relies on morpho-syntactic regularity in texts and needs a pre-processing phase to define specific patterns for a domain.

Our method also uses patterns (JAPE grammar) to detect relations but it relies on an advanced syntactic analysis of texts (cf. use of linguistic roles) to extract terms linked by the relation. Methods like CAMELEON could be used in our system as complement to improve the relation extraction phase (for example when the system fails to assign the correct linguistic role in a sentence).

Finally, MeatSearch can also be compared with web reasoning systems [18] applied on corporate memory, since it integrates CORESE and enables advanced information retrieval and reasoning on annotations.

4.3 Further Work

As a further work, we will try to integrate a new module to MeatAnnot which takes into account graphics and tables in the papers. We will install a local version of UMLS and manage its evolution which implies the evolution of the annotations. Finally, we will create with biologists several scenarios of use with typical queries to facilitate the navigation in the annotation base.

Acknowledgements

References

1. Blaschke C. & Valencia A., Molecular biology nomenclature thwarts information-extraction progress. IEEE Intelligent Systems & their Applications, p. 73-76. (2002).
2. Briscoe, E. and J. Carroll (2002) Robust accurate statistical annotation of general text. In *Proceedings of the Third IC LR E,* Las Palmas, Gran Canaria. 1499-1504.
3. Clark J., "XSL Transformations (XSLT) Version 1.0," W3C Recommendation, http://www.w3.org/TR/xslt (November, 1999).
4. Corby, O., Faron-Zucker, C.: Corese: A Corporate Semantic Web Engine. In Workshop on Real World RDF and Semantic Web Applications - WWW'02 Hawaii. 2002.
5. Corby, O., Dieng-Kuntz R. Faron-Zucker, C., Querying the Semantic Web with the CORESE engine. ECAI'2004, Valencia, August 2004, IOS Press, p.705-709
6. Cunningham H., Maynard D., Bontcheva K. & Tablan V. (2002). GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. ACL'02.
7. Dieng-Kuntz R., Corporate Semantic Webs, To appear in Encyclopaedia of Knowledge Management, D. Schwartz ed., Idea Publishing Group, July 2005
8. Golebiowska J. & Dieng-Kuntz R. & Corby O., & Mousseau D. Building and Exploiting Ontologies for an Automobile Project Memory. .In Proc. of K-CAP, Canada. (2001)
9. Handschuh S. , Koivunen M., Dieng R. & Staab S., eds, Proc. of KCAP'2003 Workshop on Knowledge Markup and Semantic Annotation, Sanibel, Florida, October 26. (2003)

---

10. Handschuh S., Staab S. and Ciravegna F. (2002), S-CREAM – Semi-automatic CREAtion of Metadata.The 13th International Conference on Knowledge Engineering and Management (EKAW 2002), ed Gomez-Perez, A., Springer Verlag
11. Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In Proceedings of International Conference on New Methods in Language Processing.
12. Humphreys BL, Lindberg DAB. The UMLS project: making the conceptual connection between users and the information they need. Bull Med Libr Assoc. 1993
13. Khelif K., Dieng-Kuntz R. - Ontology-Based Semantic Annotations for Biochip Domain, Proc. of EKAW 2004 Workshop on the Application of Language and Semantic Technologies to support KM Processes, U.K., 2004, http://CEUR-WS.org/Vol-121/.
14. Lassila O. and Swick R. (2001). W3C Resource Description framework (RDF) Model and Syntax Specification, http://www.w3.org/TR/REC-rdf-syntax/
15. McBride B.,"RDF Vocabulary Description Language 1.0: RDF Schema", W3C Recommendation, http://www.w3.org/TR/rdf-schema/ (February, 2004).
16. Muller, H.M., Kenny, E.E., Sternberg, P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature *PLoS Biol.*, 2, E309
17. Nédellec C., Bibliographical Information Extraction in Genomics. IEEE Intelligent Systems & their Applications, p.76-80, March/April. (2002)
18. Ohlbach H. J., Schaffert S. (eds) Workshop on Principles and Practice of Semantic Web Reasoning at The 20th ICLP, St Malo, France, 2004
19. Séguéla P. and Aussenac-Gilles N. (1999). Extraction de relations sémantiques entre termes et enrichissement de modèles du domaine. In IC'99, Paris, 79-88
20. Shatkay H., Edwards S. & Boguski M., Information Retrieval Meets Gene Analysis. IEEE Intelligent Systems & their Applications, p. 45-53. (2002)
21. Staab S., eds., Mining Information for Functional Genomics. IEEE Intelligent Systems & their Applications, p. 66-80, March-April. (2002)
22. Stoeckert, C.J., Jr and Parkinson H. (2003) The MGED ontology: a framework for describing functional genomics experiments. Comp. Funct. Genomics, 4, 127 132.
23. Summary Report - W3C Workshop on Semantic Web for Life Sciences. http://www.w3.org/2004/10/swls-workshop-report.html
24. 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, In Gomez-Perez A. and Benjamins R. eds, Proc. of the 13th International Conference on Knowledge Engineering and Knowledge Management EKAW 2002, Springer Verlag LNAI 2473, 379-391.