Linguistic Dependencies as a Basis for the Extraction of Semantic Relations

Thierry Declerck
DFKI GmbH, Language Technology Lab
Stuhlsatzenhausweg 3
66123 Saarbruecken, Germany
thierry.declerck@dfki.de

Mihaela Vela
DFKI GmbH, Language Technology Lab
Stuhlsatzenhausweg 3
66123 Saarbruecken, Germany
mihaela.vela@dfki.de

ABSTRACT

In this paper we present an on-going investigation on how linguistic dependency analysis can help in the automated extraction of shallow semantic relations from a biomedical text corpus. We consider then such semantic relations as a possible starting point for the text-based (supervised) semi-automatic ontology creation and also for literature-based knowledge discovery in biomedicine.

Keywords

linguistic dependencies, semantic relation extraction

1. INTRODUCTION

The Semantic Web has marked a new stage in advanced automated textual analysis, ontologies becoming a key instrument in the development of applications requiring semantic resources, like for example information extraction (IE), knowledge acquisition (KA) and text-based knowledge discovery (KD).

But it remains the problem that the construction of (domain specific) ontologies is itself a time consuming task, which requires many human resources. Therefore there are investigations and projects dealing with (supervised) automated ontology extraction/learning from various sources. Ontology learning/extraction is generally defined as a set of methods and techniques used for building ontologies from scratch, enriching, or adapting an existing ontology in a semi-automatic fashion using several resources. A detailed overview is given in [4]. [8] distinguishes further different ontology learning approaches, focusing on the type of input used for learning: ontology learning from text, from dictionary, from knowledge base, from semi-structured schemata and from relational schemata.

In the investigation described in this paper, we are dealing with a text-based approach to ontology extraction, whereas we situate our work at a lower level: we are trying to extract automatically from linguistically annotated text shallow semantic relations that can be used for both ontology extraction and for supporting scientific discovery in the biomedical domain.

So our work is not dealing with the semantic indexing of text, using semantic resources in the biomedical domain like described in [9]. But [9] or the results of the MuchMore project (see [2]), for example, will be helping us in evaluating our semantic relations extraction tools against documents already indexed/annotated with UMLS relations[1].

2. NATURAL LANGUAGE PROCESSING IN THE BIOMEDICAL DOMAIN

Biomedicine, defined as the branch of medical science that applies biological and physiological principles to clinical practice is a rapid evolving and constantly growing field, which is documented among others in a large variety of journal, conferences and workshop papers. This scientific literature can be searched for in MEDLINE, the US National Library of Medicine's (NLM's) online database[2], which contains around 11 million references to journal articles in the health sciences from over 7300 different publications from 1965 until today. It is clear that this huge amount of knowledge is practically unmanageable by traditional paper-based methods when it comes to store, retrieve or access it for performing for example text-based scientific discovery.

In this context, automated natural language processing (NLP) techniques, combined with high-level domain specific semantic resources, can facilitate the handling of the biomedical literature for knowledge extraction, knowledge discovery and extension of the already existing biomedical knowledge.

Already in the year 2002 a workshop in Tokyo has been dedicated to the relation between Natural Language Processing

[1]UMLS stands for "Unified Medical Language System" Within UMLS relevant medical concepts are defined in a "Metathesaurus" and a "Semantic Network" defines relations between the concepts. See http://www.nlm.nih.gov/research/umls/about_umls.html for more details.

[2]see http://www.ncbi.nlm.nih.gov.

and Ontology Building in Biology[3]. In this workshop [5] presented some challenges and possible solutions for NLP in accessing the biological literature. The detected challenges are mainly dealing with the great amount of biological discoveries that are connected to relevant information and relations in large ontologies and data bases. So one of the challenge relates to keeping ontologies up to date and data bases consistent, current and correct, also on the basis of large set of documents processed by NLP. The access to relevant information across many sources and the discovery of new relations from already known information is also a challenge for NLP in accessing the biological literature.

In order to support the first challenge described above, we are looking for a way to extract generic semantic relations from linguistically annotated text, and to map where possible those relations to UMLS labels for relations, and so to see if our approach can be instrumental in supporting ontology extraction from text in the biomedical domain. The very large semantic resources already available in this domain is given us the optimal background for judging the degree of coverage and accuracy of linguistic methods for extracting domain specific semantic resources from scratch.

In the following sections we present first an approach based on (conceptual) textual analysis for scientific discovery in biomedicine. Then we present in section 2.2 an approach that makes real use of NLP, within a limited range of linguistic phenemena, with the goal of extending already existing domain specific semamtic resources. In section 3 we present our own actual work. Our approach goes into more complex syntactic analysis, and more especially the so-called dependency analysis and formulate first findings and hypothesis on the kind of semantic relations that can be extracted from dependencies, and which can be used for ontology building, whereas we do not consider for the time being the issue of extending already existing semantic resources.

2.1 Text-Based Scientific Discovery for Biomedicine: The DAD System

Based on work by Swanson on scientific discovery (see [11]) [12] introduces a literature-based scientific discovery tool called the DAD-system. This system deals with concepts, taken from the UMLS Metathesaurus rather than with he words as such as the basis for the scientific discovery task.

So in a first phase the system has to provide for domain specific semantic annotation of the textual documents. And as we will see later, the system doesn't really provide for syntactic processing as the basis on which semantic annotation/extraction can be performed. The concept-based approach has at least two advantages: 1) words irrelevant for biomedicine and/or with a limited semantic content such as determiners and prepositions do not enter the system, and 2) UMLS supports the identification of compound terms such as "blood pressure", even without previous syntactic analysis, since UMLS concepts contains labels consisting in those compound words.

The DAD-system uses the following semantic and language resources: PubMed, MetaMap and UMLS. PubMed is a service of the National Library of Medicine, which includes over 15 million citations for biomedical articles back to the 1950's. PubMed links mainly to MEDLINE but also includes links to many sites providing full text articles and other related resources. The second resource is MetaMap, a text to concept mapping program, developed by the National Library of Medicine (see [12]). MetaMap is used within the described system for mapping raw text to UMLS Metathesaurus concepts.

As the reader can see, DAD doesn't seem to use natural language computational lexicons, supporting morpho-syntactic analysis, but maps directly strings in text to semantic resources. This approach is giving good results, since English is a poorly inflected language, and the semantic resources are in fact including the relevant terms in English. Things might change when one looks at other languages, with a rich morphology and having not the same coverage of terms in the available semantic resources. This is the motivation beyond the approach we propose below (even if for the sake of simplicity in this paper, we only describe English examples), which has been applied to both German and English texts.

2.2 Adjectival Modifications as the Basis for Ontology Extension

[1] describes an NLP-based approach that supports the extention of an already existing biomedical terminology on disorder and procedures. The aim of this approach is to find within MEDLINE noun phrases terms that can be added to the UMLS Metathesaurus. The two conditions for inserting a new term into the Metathesaurus are: 1) similarly modified terms exist in the terminology, for a given semantic category, and 2) a demodified term created from this phrase can be found in the terminology. A demodified term is a noun phrase whose modifier[4] has been removed, wheras [1] means by a similarly modified term the following: if the modifier removed to create the demodified term also modifies existing terms in the terminology, for a given semantic category.

The corpora for this experiment contains three million simple noun phrases[5] found in the MEDLINE citations. These phrases were submitted to a syntactic analysis tool (described in [10]), which performed the identification of the head noun and modifiers within the nominal phrases. For example, the noun phrase *catastrophic cervical spinal cord injuries* was analyzed as:

```
[[mod([catastrophic, adj]),
  mod([cervical, adj]),
  mod([spinal, adj]),
  mod([cord, adj]),
  head([injuries, noun])]].
```

At the same time the UMLS terms for disorder and procedures were also submitted to the syntactic analysis tool.

[3]http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/WS.html.

[4]Modifier is here to be understood as an adjective in a nominal phrase.

[5]Simple noun phrase are considered noun phrases excluding prepositional modification or any other complex structure.

¿From the three million MEDLINE noun phrases the ones encountered in the UMLS Metathesaurus were filtered out. The remaining noun phrases were once again sent to the syntactic analyser in order to select for phrases consisting only of one or more modifiers and the head noun. In the next stage of this approach both the remaining MEDLINE noun phrases and the disorder and procedure terms in the Metathesaurus were demodified in order to create foreach category a list of allowable adjectival modifiers. For the noun phrase *accidental arterial perforations* the demodified terms created were *accidental perforations*, *arterial perforations* and *perforations*. As stated above, the first requierement for MEDLINE candidate terms is that an allowable modifier modifies a term in the terminology. Therefore, the MEDLINE modifiers are mapped to the Metathesaurus disorder and procedure terms in order to test which allowable MEDLINE modifier already modifies a term in the Metathesaurus. For the MEDLINE candidate phrase *accidental arterial perforations*, both *accidental* and *arterial* are modifiers in Metathesaurus terms, fulfilling the first requirement.

The second requirement for a MEDLINE candidate term in this study is that a demodified term created from this phrase already exists in the Metathesaurus. In order to check this requirement, demodified MEDLINE terms had been mapped to disorder and procedure concepts in the Metathesaurus. For the candidate phrase *accidental arterial perforations*, both terms *arterial perforations* and *perforations* were found in UMLS. This way, the two conditions for the extention of the Metathesaurus with *accidental arterial perforations* are fulfilled and the phrase can be added to the existing terminology

In comparison with [12], this approach uses linguistic analysis when it comes to identify the head and modifiers (and the corresponding part of speech of both) of a noun phrase. ¿From this point of view [1]'s approach and the one introduced in section 3 are similar. Still, [1] deals only with modification phenomena within simple noun phrases (noun phrases exluding prepositional modification or any other complex feature), wheras the approach proposed in chapter 3 deals with a more complex set of linguistic phenomena. And our actual aim is to extract semantic relations that can support ontology extraction from text, and not to extend already existing semantic resources.

3. SEMANTIC RELATION EXTRACTION ON THE BASE OF ANNOTATION OF LINGUISTIC DEPENDENCIES

This section describes the steps that have to be run through in order to extract specific semantic relations from linguistically annotated text, which are here automatically provided by the system SCHUG. This modular system provides for a pipe-line architecture including part-of-speech tagging, morphological inflection and decomposition, phrase and dependency structures, such as head-complement, head-modifier and grammatical functions (see [3]).

One kind of dependency structure is internal to nominal phrases and describes for example the modification relation between adjectives and the main noun of a nominal phrase, and another kind of dependency structure is the one existing between a nominal phrase and the predicate of the sentence, whereas the nominal phrase can be for example the subject or the direct object of the predicate. The later type of dependency structure is known as "grammatical functions" of linguistic constituents. The rules for semantic extraction rest upon the dependency structures provided by the system at the end of the processing chain. The output of the

```
<LING_INFO BOS="12" EOS="18" STRING="Rheumatoid
arthritis is an immunologically mediated
inflammation of joints of unknown aetiology
and often leads to disability.">
 <CLAUSE id="1" BOC="12" EOC="14" MARKER="and"
    POLARITY="positive">
   <PRED AGR="[]" FRAG="13"
    SUBCAT_STEM="be">is</PRED>
   <PREDICATIVE_NP FRAG="14">an immunologically
   mediated inflammation of joints of unknown
   aetiology</PREDICATIVE_NP>
   <SUBJ FRAG="12">Rheumatoid arthritis</SUBJ>
 </CLAUSE
 <CLAUSE id="2" BOC="16" EOC="18" MARKER="S"
    POLARITY="positive">
   <PP_OBJ FRAG="18">to disability</PP_OBJ>
   <PRED AGR="[]" FRAG="17"
    SUBCAT_STEM="lead">leads</PRED>
   <PREDICATIVE_ADVP
    FRAG="16">often</PREDICATIVE_ADVP>
   <SUBJ FRAG="12">Rheumatoid arthritis</SUBJ>
 </CLAUSE>
</LING_INFO>
```

Figure 1: The output of the linguistic dependency analysis applied to the sentence: *Rheumatoid arthritis is an immunologically mediated inflammation of joints of unknown aetiology and often leads to disability.*

whole linguistic processing chain of SCHUG is shown below in figure 1. We show here, for reason of place, only the top-level annotation, that annotates the dependency structure between Predicate (the verb) and the Subject (and Predicative NP, Prepositional Object etc.).

On the top of the various types of dependency structures detected and annotated by SCHUG, a set of heuristics for extracting semantic relations can be described and applied. How this works is explained in the next sections. Based on the dependency structure and morpho-syntactic information provided by SCHUG, a number of semantic extraction rules have been defined. Those rules apply bottom up, which means that the relation extraction starts within the phrases and extends then to the relations between linguistic fragments.

The first type of relations is extracted from within nominal phrases (NPs) and prepositional phrases (PPs), more precisely the relations that can arise from the phrasal modifier (NP_MOD) and the nominal head (NP_HEAD). Another type of semantic relation is infered from the coordination between NPs which themselves appear within NPs and PPs. The third type of relation that can be extracted is an inter-fragmental one that emerges from the relation between the verb and its arguments.

3.1 The Relation between Modifiers and Nominal Heads

In order to determine the type of semantic relation that can be extracted from the structure modifier-nominal head, some components of the structure had to be viewed from a lexical semantic point of view. We consider here only the modifiers (adjectives and adverbs) included in NPs, and apply to then various language specific classification schemes (see [6] for adjectives and [7] for adverbs]).

In a NP like "synovial inflammation", the adjective "synovial" is being classified as a referential adjective, introducing a conceptual-part-of relation, but the adjective "wide" in the NP "a wide campaign" would be classified as a dimensional adjective, introducing a dimensional relation. We used 24 different classes of adjectives that introduce different relations, and the relation names, where possible, have been adapted to the UMLS relation naming, for the purpose of future extensive comparisons of our work with the UMLS resources and documents indexed/annotated with those resources.

For the time being we identify seven lingusitic phenomena on which the heuristics for semantic relation extraction can apply. Those heuristics (or mapping rules) are marked with the string "rel_1" to "rel_7" in the following listing.

3.1.1 *Phrases with one Pre-modifier (rel_1)*

This rule applies on NPs and PPs in which exactly one premodifier occurs. The generalized rule can be written as follows:

```
NP[NP_SPEC? NP_MOD NP_HEAD]:
 if NP_MOD(introduces some_rel)==>
    NP[NP_MODn NP_HEAD]
    RELATION_INTRODUCED_BY_NP_MOD [NP_HEAD]
```

Depending on the class of the modifier, a specific semantic relation is introduced. The presence of the determiner (NP_SPEC) in the NP is optional, but the appearance of exactly one modifier (NP_MOD) and of the head (NP_HEAD) is obligatory.

In our example above, the corresponding relation extracted is the fact that "synovial inflammation" is a conceptual part of "inflammation" because "synovial" has been classified as an adjective, which introduces the conceptual-part-of relation.

3.1.2 *Phrases with more than one Pre-modifier (rel_2)*

A rule that applies on the phrases that consist of more than one pre-modifier. We consider phrases that consist of modifiers that are not separated by any punctuation sign or conjunction:

```
n = number of modifiers in a phrase
j = the jth modifier in the list of  modifiers
j = 2...n
NP[NP_SPEC? NP_MOD_0...NP_MOD_n NP_HEAD]:
 if NP_MODj(introduces some_rel)==>
 NP[NP_MODj NP_HEAD]
 RELATION_INTRODUCED_BY_NP_MODj [NP_HEAD]
  foreach j==>
  NP[NP_MODj NP_MOD_{j-1}...NP_MODn NP_HEAD]
  RELATION_INTRODUCED_BY_NP_MODj
  NP[NP_MODj-1...NP_MEDn NP_HEAD]
```

The rule can be explained as follows. Each modifier (NP_MODj) in a nominal phrase, depending on its semantic class, introduces a specific relation between itself and the head (NP_HEAD). Furthermore, each modifier (NP_MODj) that is not a direct neighbor of the head-noun modifies the foregoing "subphrase." Out of the NP "chronic synovial inflammation", we can now extract following semantic relations: 1) "synovial inflammation" is a conceptual part of "inflammation". And 2) "chronic synovial inflammation" is a conceptual part of "synovial inflammation".

3.1.3 *Phrases with more than one Pre-modifier, connected by punctuation signs (rel_3) or/and conjunctions (rel_4)*

Here, all modifiers introduce the same relation with the head-noun. The corresponding rule can be formulated as follows.

```
n = number of modifiers in a phrase
j = the jth modifier in the list of  modifiers
j = 2...n
NP[NP_SPEC? NP_MOD0...NP_MODn NP_HEAD]:
 if NP_MODj(introduces some_rel)==>
 NP[NP_MODj NP_HEAD]
 RELATION_INTRODUCED_BY_NP_MODj [NP_HEAD]
```

¿From a phrase like "severe, destructive and premature arthritis", the following semantic relations are extracted: "severe arthritis", "destructive arthritis" and "premature arthritis", which all denote a property of "arthritis". The use of punctuation information for extracting types of relation is for sure language dependent.

3.2 Coordination between the Components of NPs and PPs

The extraction of semantic relation from the coordinated components of NPs and PPs, is actually an extension of the extraction rule mentioned above. As the modification rules, it uses the dependency structure and the morpho-syntactic information provided by SCHUG, but this time apply this information to the coordinated parts of a coordination.

3.2.1 *Coordination between Nominal Heads in Phrases (rel_5)*

Nominal phrases might have not only several modifiers but also more than one nominal head put into relation with a coordinating word. The rule that performs the semantic extraction for those cases:

```
m = number of embedded phrases in a PHRASE
j = the jth embedded phrase in a NP or PP
PHRASE = NP or PP
PHRASE[NP_SPEC? [PHRASE]0...[PHRASE]m]==>
 all the nominal heads in the embedded
 phrases are associated with each other
```

The interpretation of this rule is: the heads from each embedded phrase relate to each other by the associated-with relation. Since a phrase can in this case be analyzed by more than one rule, for these examples first the modification rules apply, and after this stage is completed the coordination between intra-fragmental phrases applies. ¿From the example "chronic synovial inflammation and joint destruction" the following relations are extracted: "synovial inflammation" is a conceptual part of "inflammation" and "chronic synovial inflammation" is a conceptual part of "synovial inflammation", as described above. The coordination rule associates than "chronic synovial inflammation" with "joint destruction".

At this level we do not consider the possible reading of the coordinated phrase where "chronic" might also be modifying "joint destruction". In order to ensure the maximal precision in our semantic relation extraction algorithm, we do not try to solve issues of ambiguitities resulting from scopus properties. The chunk parser of SCHUG is here considering that both head-nouns in the coordinated structure are modified only by adjectives that are within the syntactic scope of the corresponding NP[6]

3.3 Relations between Linguistic Fragments

3.3.1 The ISA Relation (rel_6)

Another rule proposed for semantic relation extraction is that resulted from the relation between subject and the second argument of the predicate, when the predicate is the verb "be" for English, or one of its synonyms listed in the Roget thesaurus. This rule applies inter-fragmental and take into account not only morphology but also syntactic information. We assume that if the arguments of the verb "be" are identified, the relation between the arguments is an isa relation

```
PHRASE[GF=SUBJ] VG[STEM=BE] PHRASE[GF=OBJ]==>
PHRASE[GF=SUBJ] ISA_RELATION PHRASE[GF=OBJ]
```

In the example "Rheumatoid arthritis is an immunologically mediated inflammation of joints of unknown aetiology and often leads to disability", SCHUG identifies "rheumatoid arthritis" as the subject of the predicate "is", and the linguistic fragment "an immunologically mediated inflammation of joints of unknown aetiology" as the predicative NP. The semantic relation extraction rule proposes for this type of constructions an isa relation between subject and the predicative NP.

As pointed out by an anonymous reviewer of this paper, the use of "is" is highly ambiguous in text, and there is a real danger that our heuristic might overgenerate "ISA" semantic relations in a significant way, thus reducing precision in a dramatic way. Our actual strategy to avoid this, is to restrict the application of this heuristic in to sentences that realise the syntactic structure "NP PRED NP", where the second NP is being introduced by an indefinite article.

[6]This precision is resulting from a question/comment by an anonymous reviewer.

3.3.2 The CAUSE Relation (rel_7)

In the second part of the sentence, we detect a cause relation between the subject "Rheumatoid arthritis" (annotated on the base of an ellipsis analysis, since the subject of the verb "leads" is not realized in the text), and the head noun "disability" of the NP included in the prepositional phrase "to disability". ¿From a subcategorisation lexicon for English, the CELEX lexicon, we know that a PP introduced by the preposition "to" is belonging to the so-called subcat-frame of the verb "lead". This is giving us the triggering syntactic property that allows to apply the "rel_7" extraction rule.

In this example, the SCHUG parser is applying an ellipsis analysis for determining the subject of the second clause in the sentence. This analysis step is responding to the general assumption that every sentence in English has a subject, being realised or not in syntactic terms.

4. THE IMPLEMENTATION OF THE MAPPING RULES

All the heuristics described above have been implemented in a Perl module that has been added to the processing chain of SCHUG. This module delivers the extracted semantic relations in the form of a graph. Graphs resulting from various documents can then be merged and so propose a unified structure for the semantic relations extracted from a larger set of linguistically annotated documents. The merging or the various graphs is at the same time the starting point of a starting investigation on our approach for the topic of literature-based scientific discovery, since the merged graphs can show certain relations that are not explicitely mentioned in the documents that are being considered.

5. EVALUATION

At the actual stage of development we can only provide for a small scale evaluation, which consists in establishing a small test corpus (35 sentences, 5 for each linguistic phenomenon). selected randomly out of a corpus of linguistically annotated biomedical text (from the MuchMore corpus). In this small test corpus, semantic relations have been annotated by hand (involving just one person), and the results of the SCHUG semantic relation extraction module has been compared with the manual annotation. A real scale evaluation still has to be performed, but in the figure we present in Table 1, we already get indication on the possible performances of the system and where we should improve the approach. Failures are partly due to the performance of the syntactic analysis of SCHUG. So in the case of the test sentences covering the "rel_4" case for English, SCHUG was delivering incomplete parses and the rule for semantic relation extraction didn't apply at all.

6. CONCLUSIONS AND FUTURE WORK

In this paper we presented on going work on extracting shallow semantic extraction from linguistically annotated text.

Table 1: Evaluation results for the chosen test suite

Phenomenon	Recall	Precision
rel_1	100%	97%
rel_2	90%	90%
rel_3	50%	100%
rel_4	-	-
rel_5	62%	56%
rel_6	42%	100%
rel_7	16%	25%

The first step is a generic one, whereas a second step will be dedicated in mapping the extracted generic semantic relations to available domain specific semantic resources and so to specify the kind of relations we are extracting. In the actual work, we map the semantic relations to named relation in the UMLS context. Relations described in UMLS will also allow for a better evaluation study of the quality of the extraction.

On the base of a first (very limited) evaluation, it seems to be that we can claim that linguistic dependencies might really offer appropriate means for extracting shallow semantic relation for helping in building ontologies from scratch and for supporting literature-based scientific discovery. This assumption also relies on the fact that the system was confronted only to well-written texts, displaying clear communication intentions of their authors. So we are aware that the approach described can not be applied to arbitrary text in an open domain.

Concerning the scientific discovery perspective, our approach can have the advantage that the discovery process is not only based on word annotated with semantic classes, but on semantic relations extracted from the text. But this has to be verified in furhter investigations.

We still have to provide for a full scale evaluation and for a classification of the failures of the system. We also need a mechanism that maps the shallow (generic) semantic relations extracted to domain specific ones in a more principled way. We will also apply the approach to other domains, like the financial domain, where the amount of already available high-level semantic resources is much smaller as in the case of biomedicine, and where such an approach can be thus much more benefical.

7. ACKNOWLEDGEMENT

This research was partially pursued within the Esperonto project[7]. Actual work is done in the context of the LIRICS project[8], where we investigate at the standartdisation level the relation between syntactic argument structures and semantic relations in general. We would also like to thank the anonymous reviewers for their very valuable comments on our submission to this workshop.

[7]Esperonto was a project of the Information Society Technologies (IST) Program for Research, Technology Development & Demonstration under the 5th Framework Program of the European Commission, with the number IST-2001-34373, see www.esperonto.net.

[8]LIRICS is a project in the eContent programme of the European Commission, contract Nr. EDC-22236.

8. REFERENCES

[1] O. Bodenreider, T. Rindflesch, and A. Burgun. Unsupervised, corpus-based method for extending a biomedical terminology. In *Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain*, pages 53–60, Philadelphia, USA, July 2002. Association for Computational Linguistics.

[2] P. Buitelaar and H. Uszkoreit. Muchmore: Concept-based cross-lingual information retrieval in the medical domain. *Kuenstliche Intelligenz*, 2004.

[3] T. Declerck. A set of tools for integrating linguistic and non-linguistic information. In *Proceedings of SAAKM (ECAI Workshop)*, 2002.

[4] A. Gomez-Perez and D. Manzano-Macho. Ontoweb deliverable 1.5: A survey of ontology learning methods and techniques. Technical report, Facultad of Informatica (Univerisdad Politecnica de Madrid), 2003.

[5] L. Hirschmann. Accessing the biological literature: Chalenges for natural language processing. Workshop on Natural Language Processing and Ontology Building in Biology, February 2002.

[6] S. Lee. *Untersuchungen zur Valenz des Adjektivs in der deutschen Gegenwartssprache*. Lang, 1994.

[7] A. Lobeck. *Discovering Grammar: An Introduction to English Sentence Structure*. Oxford University Press, 2000.

[8] A. Maedche and S. Staab. Ontology learning for the semantic web. In *IEEE Intelligent Systems*, volume 16. Special edition for the semantic web edition, 2001.

[9] K. Mark, P. Daumke, S. Schulz, and U. Hahn. Cross-language mesh indexing using morpho-semantic normalization. In *Proceedings of the 2003 Annual Symposium of the American Medical Informatics Association*, pages 425–429, 2003.

[10] T. C. Rindflesch, J. V. Rajan, and L. Hunter. Extracting molecular binding relationships from biomedical text. In *Proceedings of the ANLP-NAACL*, pages 188–195, 2000.

[11] D. R. Swanson. Fish oil, raynaud's syndrome, and undiscovered public knowledge. In *Perspectives in Biology and Medicine*, volume 30, pages 7–18, 1986.

[12] M. Weeber, H. Klein, A. R. Aronson, J. G. Mork, L. T. de Jong-van den Berg, and R. Vos. Text-based discovery in biomedicine: the architecture of the dad-system. In *Proceedings of the AMIA Annual Fall Symposium*, pages 903–907, Philadelphia, USA, 2000.