Using the Intension of Classes and Properties definition in Ontologies for Word Sense Disambiguation

Khaled Khelif[1], Fabien Gandon[1], Olivier Corby[1], Rose Dieng-Kuntz[1]

[1] INRIA Sophia Antipolis Méditerranée - 2004 route des Lucioles 06902, BP93 Sophia Antipolis - France
{khaled.khelif, fabien.gandon, olivier.corby, rose.dieng}@sophia.inria.fr

Abstract. We present an ontology-driven word sense disambiguation process. The main idea consists of using the context of the ambiguous word to decide which class can be assigned to it. The disambiguation relies on similarities between classes assigned to the ambiguous word, classes assigned to terms close to it in the text, and on the type of properties that could occur between them. The computation of the similarity uses domain ontologies to provide semantic distances based on definitions in intension. We tested our approach in the extraction of annotations from biomedical texts.

Keywords: word sense disambiguation, ontology, semantic distances

1   Introduction

In [11], we proposed an approach based on semantic web technologies to generate and use ontology-based semantic annotations. The resulting system, MeatAnnot, relies on UMLS [14] and NLMs MetaMap system [17], which maps candidate terms to semantic types in the UMLS semantic network. Experiments showed that about 12% of MetaMap results are ambiguous. In fact, if a candidate term maps to more than one class with an equally high confidence score, this system cannot determine which class is the correct mapping. Here is an example of ambiguity:

"The <u>blood pressure</u> lowering effect and tolerability of the angiotensin-converting enzyme inhibitor enalapril combined with a very low dose of hydrochlorothiazide (HCTZ) were compared with the selective betareceptor blocker atenolol in patients with mild-to-moderate hypertension."

The expression "blood pressure" of this sentence is mapped to three classes, each time with an equally high confidence score: blood pressure $\Rightarrow$ Organism Function
blood pressure $\Rightarrow$ Diagnostic Procedure
blood pressure $\Rightarrow$ Laboratory or Test Result

This example shows the problem of ambiguity in UMLS. This problem is mainly due to word ambiguity in English and to the fact that the context of the word is not taken into account during the mapping.

The approach we describe in this paper uses the intension of the definitions in the ontology (in our case the definition of the UMLS semantic network) to disambiguate terms extracted from texts. To do this, we propose methods for computing semantic distances between two classes of the ontology, which take into account the hierarchy of classes, the hierarchy of properties and the signature of properties (domain and range). Our hypothesis is that a combination of distances over these networks can improve the computation the disambiguation.

2   Conceptual distances: ontologies as metric spaces

The idea of evaluating conceptual relatedness from semantic networks representation dates back to the early works on simulating the humans' semantic memory [7] [5]. Relatedness of two concepts can take many forms for instance, functional complementarity (e.g. nail and hammer) or functional similarity (e.g. hammer and screwdriver). The latter example belongs to the family of semantic similarities where the relatedness of concepts is based on the definitional features they share (e.g. both the hammer and the screwdriver are hand tools). The natural structure supporting semantic similarities reasoning is the taxonomy of classes where is-a links group classes according to the characteristic they share (e.g. hammer, screwdriver, saw, plane, pliers, etc. are subclasses of hand tool). When applied to a semantic network using only is-a links, the relatedness calculated by a spreading algorithm gives a form of semantic distance. The graph of the class hierarchy and the distance calculation algorithm turn the ontology into a metric space. Rada et al. [8] defined their conceptual distance between two classes C_A and C_B as the minimum number of edges separating them. However they also showed that this distance exhibited counter intuitive behaviours around zero. Starting from this introduction we can identify two main trends in defining a semantic distance over a class hierarchy: (1) the approaches that include additional external information in the distance, e.g. statistics on the use of a concept; (2) the approaches trying to rely solely on the structure of the hierarchy to tune the behaviour of the distances.

For the first approaches, relying on additional external information, we can quote Resnik [9] whose work was influenced by information theory. The author tried to apply this technique directly to words and encountered counter intuitive behaviours due to polysemia. An improvement in [6] integrates the value of the information content of the compared concepts to the calculation. A comparison of different distances on using WordNet, is proposed in [4]. But these techniques require statistical analysis of the corpus to evaluate the probabilities and thus require finding a relevant corpus to effectively approximate the probabilities by frequencies.

The second trend essentially explores the different ways of combining the depths of the concepts and their deepest common super concept in the hierarchy. The simplest one is the one of Rada et al. [8] presented before. An alternative in [10] is based on the ratios between depths. In the domain of Conceptual Graphs, a use for such a distance is to propose a non binary projection, i.e. a similarity $S : C^2 \rightarrow [0, 1]$ where 1 is the perfect match and 0 the absolute mismatch, and this can be used to propose approximate search algorithms intelligently relaxing typing constraints [1].

In the past we used several kinds of these distances to provide approximate search algorithms [1] or protocols for distributed annotation and query management [2]. We also extended them for instance to provide ultrametrics for clustering algorithms [3].

Most of the distances belonging to the first approaches (i.e. relying on additional external information) capture somehow the extensional use of classes i.e. how they are effectively used in a corpus. They require a pre-processing to evaluate frequencies approximating the probabilities needed in their formulas.

Most of the distances belonging to the second approaches (i.e. relying on the ontology) limit their use of the metric space to the hierarchy of classes i.e. only the graph of direct subsumption links is used in defining the metric space. In this article we propose to go beyond and selectively include other relations in the metric space.

3 Signature distance principle

We extend the RDFS model with a symmetric property subsuming `rdfs:domain` and `rdfs:range` properties used in specifying the signature of a property. In the following, let H_C be the hierarchy of classes and H_P be the hierarchy of properties.

Definition 1: the property *cos:signature* is such that:

 rdfs:domain(P, C) $\Rightarrow$ *cos:signature(P, C)*   *// rdfs:domain < cos:signature*
 rdfs:range(P, C) $\Rightarrow$ *cos:signature(P, C)*   *// rdfs:range < cos:signature*
 cos:signature(X,Y) $\Leftrightarrow$ *cos:signature(Y,X)*   *// symmetric property*

Definition 2: a signature path $P_s(C_x,C_y)$ between two classes $C_x, C_y \in H_C^2$ is a path from C_x to C_y exclusively composed of arcs typed as *cos:signature* and infered from the signatures in H_R. We note: $P_s(C_x,C_y) :=< C_x$, *signature*, C_1, *signature*, C_2, *signature*, ... *signature*, C_n, *signature*, $C_y>$

Definition 3: the signature distance $d_s(C_x,C_y)$ between two classes $C_x, C_y \in H_C^2$ is defined by $d_s(C_x,C_x):=0$ and $d_s(C_x,C_y):= min_{\{Pi \ \in \ \{ \ Ps(Cx,Cy) \ \}\}} \ length(P_i)$ with *length(<C_x, signature, X_1, signature, X_2, ... signature, X_n, signature, C_y >) := n*

Intuitively, with this distance, two classes are close if there exists (in intension) a possibility to use them in a concise annotation graph. For instance the concepts *document* and *country* are close in an ontology with the following declarations:

 rdfs:domain(author, document) $\Rightarrow$ *cos:signature(document, author)*
 rdfs:range(author, person) $\Rightarrow$ *cos:signature(author, person)*
 rdfs:domain(nationality, person) $\Rightarrow$ *cos:signature(person, nationality)*
 rdfs:range(nationality, country) $\Rightarrow$ *cos:signature(nationality, country)*
         leading to $d_S(document, country)=3$

Note that RDFS signatures are supposed to be used as derivation rules not as type checking rules but that we made the choice here to consider that they give a hint of the intended use of the relation.

4   Merging signature and hierarchies graphs

The definitions and examples in the previous section illustrate the core principle of the distance but they discard the hierarchy paths. We now extend the previous base definitions to obtain a complete distance definition used in our experiment. Mixing signature links and subClassOf/subPropertyOf links can introduce noise thus we parameterize the distance using weights: w_{sig} for signatures, $w_{subclass}$ for class subsumption links and $w_{subprop}$ for property subsumption links.

Definition 4: the property *cos:subtype-aware-signature* is such that:

$cos:signature(X,Y)$ $\Rightarrow cos:subtype-aware-signature(X,Y,w_{sig})$
$rdfs:subClassOf\ (C_x,C_y)$ $\Rightarrow cos:subtype-aware-signature(C_x,C_y,w_{subclass})$
$rdfs:subPropertyOf(P_x,P_y)$ $\Rightarrow cos:subtype-aware-signature(P_x,P_y,w_{subprop})$
$cos:subtype-aware-signature\ (X,Y,w)$ $\Leftrightarrow cos:subtype-aware-signature\ (Y,X,w)$

Definition 5: a subtype-aware signature path $P_{ST}(C_x,C_y)$ between two classes $C_x,C_y \in H_C^2$ is a path from C_x to C_y composed exclusively of *subtype-aware-signature* links inferred from the declaration in H_R and H_C : $P_{ST}(C_x,C_y) := <$ *subtype-aware-signature(C_x, X_1, w_0), subtype-aware-signature(X_1, X_2, w_1), ..., subtype-aware-signature(X_n, C_y, w_n)* $>$

Definition 6: the subtype-aware signature distance $d_{ST}(C_x,C_y)$ between two classes $C_x,C_y \in H_C^2$ is the length of the shortest subtype-aware signature paths between them: $d_{ST}(C_x,C_x):=0$ and $d_{ST}(C_x,C_y):= min_{\{Pi\ \in\ \{\ PS(Cx,Cy)\ \}\}}\ length(P_i)$ with *length($<$subtype-aware-signature(C_x, X_1, w_0), subtype-aware-signature(X_1, X_2, w_1), ..., subtype-aware-signature(X_n, C_y, w_n) $>$) :=Σw_n*

This is the distance we use in our disambiguation process. To do so the Corese engine extends the SPARQL language in order to offer the possibility for computing paths in RDF(S) graphs [20]. This extension also allows us to specify constraints on the types of the properties that can be used by a path and we apply it to extract paths using the subtype-aware-signature property.

For example, the query below enables us to find all paths between two classes 'Body_Part_Organ_or_Organ_Component' and 'Amino_Acid_Peptide_or_Protein', to order them by length and thus identify the shortest one.

```
prefix umls: <http://umlsinfo.nlm.nih.gov/umls#>
select list display xml ?path pathLength(?path) as ?length  where
{    umls: Body_Part_Organ_or_Organ_Component
     direct::<umls:subtype-aware-signature>::?path
     umls: Amino_Acid_Peptide_or_Protein . }
order by pathLength(?path)
```

Figure 2 shows a result to this query with the different steps of the path computation. The edges are RDFS properties between classes and properties of the ontology.
So we can see that '*Body_Part_Organ_or_Organ_Component*' is a subsumption of '*Fully_Formed_Anatomical_Structure*', which belongs to the signature of the relation 'produces'. This relation has as range 'Organic_Chemical' which is a super-class of 'Amino_Acid_Peptide_or_Protein'. The length of this path is 4.

The semantic similarity *semSim(C1, C2)* (see Definition 7) between theses classes is then computed by (i) weighting the edges of the path (using weights defined in Definition 5), and (ii) using a normalisation formula. For example, if w_{sig} =0.2, $w_{subclass}$ =0.4 and $w_{subprop}$ =0.4, *semSim(Body_Part_Organ_or_Organ_Component, Amino_Acid_Peptide_or_Protein) = 0.66*.
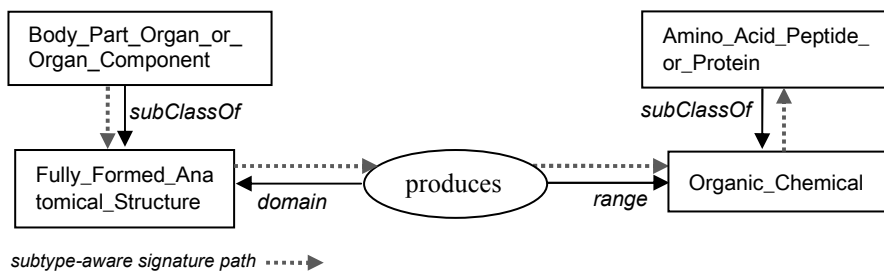

Fig. 2. The path found between 'Body_Part_Organ_or_Organ_Component' and 'Amino_Acid_Peptide_or_Protein' in the UMLS semantic network

5 The WSD Approach

In this section we present the word sense disambiguation approach that we based on the distance we just defined. The main idea of this method is to use the context of the ambiguous word to decide to which class we can assign it. This context consists of the set of terms which co-occur with the ambiguous word in the same paragraph. So, if MeatAnnot assigns several classes to a candidate term, the disambiguation module tries to find the right class computing similarities between classes assigned to the ambiguous word, those assigned to its neighbours in the paragraph, and on the type of properties that could occur between them. The computation of similarity is based on the inverse of the semantic distances described previously.

This approach was tested on a standard collection containing the most ambiguous word in UMLS. To help developers test their disambiguation algorithms, the NLM proposed a WSD test collection [12] of 50 highly ambiguous words in UMLS. This collection consists of 5000 Medline abstracts containing these ambiguities, 100 abstract per ambiguous word. The main goal of the WSD test collection is to establish a standard for evaluating disambiguation methods by comparing their results to human suggestions; the 5000 instances have been disambiguated by human raters.

The WSD algorithm takes as input: (i) the ambiguous term, (ii) semantic classes assigned to it, (iii) classes found in the sentence (and/or the paragraph) containing the ambiguity, and (iv) the UMLS semantic network. For each ambiguous term, the disambiguator builds a VSTn vector describing the context of this term. The VSTn vector contains the classes assigned to terms detected as neighbours of the ambiguous one. Then, it computes a similarity between each STa in VSTa (vector of classes assigned to the ambiguous term) and the VSTn, this similarity consists of the average

of the similarities between STa and STn elements. Finally, the class which STa has the highest similarity is proposed as the best class for the ambiguous term.

Definition 7: Similarities are computed inversing the semantic distances described above; we call it semSim. For each $STa \in VSTa$, $Sim(STa, VSTn) = Avg_{STn \in VSTn}(semSim(STa, STn))$ $BestST(VSTa, V STn) = Max_{STa \in VSTa}(Sim(STa, V STn))$ where $semSim(STa, STn) = 2/(2+dist(STa, STn))$ with *dist* being one of the distances defined previously.

Let us give a processing example with the sentence:

'I also tracked lipid profiles, HBA1C, blood pressure, body mass index, hostility, and nicotine use'.

Results of the MeatAnnot term extractor on this sentence:

```
<Sentence>
    <Concept type="Laboratory_Procedure" term="lipid profile"/>
    <Concept type="Gene_or_Genome" term="HBA1C"/>
    <Concept type="Organism_Function" term="blood pressure"/>
    <Concept type="Diagnostic_Procedure" term="blood pressure"/>
    <Concept type="Laboratory_or_Test_Result" term="blood pressure"/>
    <Concept type="Diagnostic_Procedure" term="body mass index"/>
    <Concept type="Mental_Process" term="hostility"/>
    <Concept type="Organic_Chemical" term="nicotine"/>
</Sentence>
```

We can notice that the term 'blood pressure' is ambiguous since there are three classes assigned to it. The vectors for the ambiguous term and its context are:

VSTa = {"Organism_Function", "Diagnostic_Procedure", "Laboratory_or_Test_Result"}
VSTn = {"Laboratory_Procedure", "Gene_or_Genome", "Diagnostic_Procedure", "Mental_Process", "Organic_Chemical"}

Table 1. Similarities between the STs affected to 'blood pressure' and those affected to its neighbours in the sentence

	Organism_Function	Diagnostic_Procedure	Laboratory_or_Test_Result
Laboratory_Procedure	0.291	0.862	0.167
Gene_or_Genome	0.166	0.167	0.3
Diagnostic_Procedure	0.29	1	0.169
Mental_Process	0.98	0.28	0.168
Average	0.431	**0.549**	0.201

The following table summarizes the obtained disambiguation results using the distance which relies only on the subsumption hierarchy of the ontology. In this example MeatAnnot assigns 'Diagnostic_Procedure' as the class best matching the meaning of the term 'blood pressure' w.r.t the studied context in this sentence.

6 Evaluation

To evaluate our algorithm, we tested it on the UMLS Word Sense Disambiguation test collection. We launched MeatAnnot on the entire collection (5000 abstracts) to detect UMLS terms and their classes and to disambiguate terms provided with the WSD collection. In this experiment, in a first experiment, we decided to use the semantic distance described in Definition 6 and to vary the different weights, starting from a distance which favours the use of the subsumption path between classes to a distance which favours the path using the signature of properties. In a second experiment, we selected weights which maximise this distance and we compared its results to three other distances: (i) uses only the class subsumption hierarchy, (ii) uses only the signature of relations, and (iii) uses both features. The metric used in this evaluation is Precision (the percentage of correct mappings), which computes the rate of correctly resolved ambiguities.

Table 2 shows the results of the first experiment in which we varied the weights of our semantic distance as defined in definition 4: w_{sig} for signature links, $w_{subclass}$ for class subsumption links and $w_{subprop}$ for property subsumption links. The variation of the weights was chosen to represent the progressive transformation of the distance from a distance purely based on subsumption paths to a distance purely based on signature paths. The values are the average precision scores in retrieving the right class for the terms.

Table 2. Summary of some precision scores for nine weight sets

weight sets	1	2	3	**4**	5	6	7	8	9
$w_{subclass}$	0	0.1	0.2	**0.2**	0.2	0.3	0.4	0.5	1
$w_{subprop}$	0	0.1	0.2	**0.4**	0.6	0.3	0.4	0.5	0
w_{sig}	1	0.8	0.6	**0.4**	0.2	0.3	0.2	0	0
Average	63.4	63.73	66.26	**83.62**	67.63	67.11	64.99	54.21	51.56

Table 3 shows the results of a second experiment in which we compared the previous semantic distance with the optimal weights obtained in the first experiment to three other distances: a classical ontology based attenuated distance [1] based on classes subsumption only, a distance using only signature links (definition 1) and a distance mixing non weighted signature and subsumption links. The values are the precision scores in retrieving the right class for the term given in the left side column. The last line is the average precision over these terms.

The first experiment seems to indicate that it is indeed interesting to mix two aspects of the intentional definitions of types: their subsumption neighbourhood and their potential properties. More precisely, what table 2 seems to indicate is that there is an interesting optimum in combining subsumption distances and signature distances (here corresponding to the weight set number 4 where $w_{subclass}$= 0.2, $w_{subprop}$= 0.4 and w_{sig}=0.4) and that this optimum could perform better than classical distances. Because the improvement is sometimes quite small and because the terms have very different behaviours (Table 3) we intend to study how the sub-domains of the ontology

influence the performances and how the design rationale of the ontology may be responsible for these differences.

Table 3. Summary of some precision scores (10 from the 30 tested words) comparing different distances studied

	Subsumption	Simple signature	Non weighted signature and subsumption	Optimised
adjustment	65.59	29.03	61.29	65.59
ganglion	92	90	93	93
extraction	62.06	52.87	65.51	65.51
japanese	64.55	92.4	91.13	92.4
pressure	97.91	10.04	37.5	86.45
surgery	61.11	53.3	55.55	60
depression	86.2	86.2	82.75	96.55
lead	64.28	40.81	60.2	60.2
radiation	94.11	29.41	56.86	92.15
sensitivity	98	81	70	74
Average on 30 terms	**80.43**	**63.39**	**72.77**	**82.64**

7   Related Work in Word Sense Disambiguation

In addition to related work on semantic distances cited in section 2, we notice that [13] describes an experiment using NLM's WSD test collection and compares four versions of the Journal Descriptor Indexing methodology to a baseline frequency methodology. JDI approach is a method of indexing based on NLM's practice of maintaining a subject index to journal titles using journal descriptors (JDs), which are terms corresponding to biomedical specialties. For each word JDI builds a set of weighted ST vectors describing the semantic context in which it occurs. For 45 ambiguities studied, the overall average of precision is 0.783.

In [15], for each ambiguous word, the authors use a collection of MEDLINE abstracts containing this word and rely on words related to it by UMLS conceptual relationships to determine his adequate sense.

In [21], the authors proposed an approach for WSD in medical documents using related UMLS terms found in the same context. They rely on the context of the ambiguous word in order to compute a score for each sense candidate and then they choose the one having the highest score. This score consists of the number of terms (in the document) which are related, in UMLS, with the different senses of the ambiguous word. In comparison with our method, this approach is different in two main aspects: (i) it does not use the hierarchies of concepts and properties in UMLS, and, (ii) it ignores terms which co-occur with the ambiguous word in the same context but do not have a direct link with him in UMLS.

In [18], the authors propose a method that assigns a word the sense that is most related to the senses of its neighbours and that relies on finding paths in a concept network (which is more flexible than an ontology and does not contain relations). The proposed method in [19] is different from ours since it relies on instances of relations (i.e. extension, whereas we use the intension) and it does not take advantage from a formal model (the hierarchy of relations).

For a review on general disambiguation methods, see [16].

8   Conclusion

We have described an experiment using NLM's WSD test collection to compare four versions of semantic distances in order to disambiguate term mapping to the UMLS semantic network. This experiment has shown that the use of the ontology definition can improve significantly the precision of WSD algorithms. Over the 22 ambiguous words used in this experiment, the highest average precision was 82.64%. This method is generic and domain-independent since it relies only on the structure of the ontology and the chosen distances. Future work falls into three categories:

- Improving the disambiguation algorithm by optimizing the path finding in the distance computation and by proposing other distances taking into account existing annotations (i.e. combination of instances and ontology structure).
- Studying the use of the approach in other applications: testing the WSD algorithm in different domains using different ontologies and different term extractors.
- Reintroducing in our distance the computation of the depth to attenuate subsumption lengths as classically done in standard distances [1].

References

1. Corby O., Dieng-Kuntz R., Faron-Zucker C., Gandon F., Searching the Semantic Web: Approximate Query Processing Based on Ontologies, IEEE Intelligent Systems, January/February (Vol. 21, No. 1), pp. 20-27, ISSN: 1541-1672, 2006
2. Gandon F., Distributed Artificial Intelligence and Knowledge Management: ontologies and multi-agent systems for a corporate semantic web, PhD Thesis., INRIA, 2002
3. Gandon F., Corby O., Giboin A., Gronnier N., Guigard C., Graph-based inferences in a Semantic Web Server for the Cartography of Competencies in a Telecom Valley, ISWC, Lecture Notes in Computer Science LNCS 3729, Galway, 2005
4. Budanitsky A., Hirst G., Semantic distance in WordNet: An Experimental, Application-oriented Evaluation of five Measures, In Workshop on WordNet and Other Lexical Resources, Second meeting of the North American Chapter of the Association for Computational Linguistics. Pittsburgh, PA, 2001
5. Collins A., Loftus E., A Spreading Activation Theory of Semantic Processing. Psychological Review, vol. 82, pp. 407-428, 1975

6. Jiang J., Conrath D., Semantic Similarity based on Corpus Statistics and Lexical Taxonomy. In Proc. of International Conference on Research in Computational Linguistics, 1997
7. Quillian M., Semantic Memory, in M. Minsky (ed.), Semantic Information Processing, pp 227-270, MIT Press; Readings in Cognitive Science, section 2.1
8. Rada R., Mili H., Bicknell E., Blettner M., Development and Application of a Metric on Semantic Nets. In IEEE Transaction on Systems, Man, and Cybernetics, pp. 17-30, 1989
9. Resnik P., Semantic Similarity in a Taxonomy: An Information-Based Measure and its Applications to Problems of Ambiguity in Natural Language. In Journal of Artificial Intelligence Research, vol 11, pp. 95-130, 1995
10. Wu Z., Palmer M.. Verb Semantics and Lexical Selection. In Proc. of the 32nd Annual Meeting of the Association for Computational Linguistics, Las Cruces, New Mexico, 1994
11. Khelif, K., Dieng-Kuntz, R., and Barbry, P., An ontology-based approach to support text mining and information retrieval in the biological domain. In Journal of Universal Computer Science (JUCS), Vol. 13, No. 12, pp. 1881-1907, 2007
12. Weeber, M., Mork, J., and Aronson, A., Developing a test collection for biomedical word sense disambiguation, 2001
13. Humphrey, S. M., Rogers, W. J., Kilicoglu, H., Demner-Fushman, D., and Rindflesch, T. C., Word sense disambiguation by selecting the best semantic type based on journal descriptor indexing: Preliminary experiment. J. Am. Soc. Inf. Sci. Technol., 96–113, 2006
14. Humphreys, B. and Lindberg, D., The umls project: making the conceptual connection between users and the information they need. In Bull Med Libr Assoc, page 170177, 1993
15. Liu, H., Johnson, S., and Friedman, C., Automatic resolution of ambiguous terms based on machine learning and conceptual relations in the umls. J Am Med Inform Assoc, , 2002
16. Edmonds, P. and Kilgarriff, A., Introduction to the special issue on evaluating word sense disambiguation systems. Nat. Lang. Eng., 8(4):279–291, 2002
17. Aronson A.R., Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc. AMIA Symp.*, 17–21, 2001
18. Pedersen T., Banerjee S., and Patwardhan S. Maximizing Semantic Relatedness to Perform Word Sense Disambiguation, University of Minnesota, Research Report UMSI 2005/25.
19. Hassell J., Aleman-Meza B., Arpinar I. B. Ontology-driven automatic entity disambiguation in unstructured text. In ISWC'06, pp. 44–57, 2006
20. Corby O. Web, Graphs & Semantics. Proc. of the 16th International Conference on Conceptual Structures (ICCS'2008), Toulouse, France , 2008
21. Widdows D., Peters S., Cederberg S., Chan C., Steffen D., Buitelaar P.: Unsupervised monolingual and bilingual word-sense disambiguation of medical documents using UMLS. In Natural Language Processing in Biomedicine, ACL 2003 Workshop. Sapporo, 2003