Word Sense Disambiguation based on Semantic Density

Rada Mihalcea and Dan I. Moldovan
Department of Computer Science and Engineering
Southern Methodist University
Dallas, Texas, 75275-0122
{rada,moldovan}@seas.smu.edu

Abstract

This paper presents a Word Sense Disambiguation method based on the idea of semantic density between words. The disambiguation is done in the context of WordNet. The Internet is used as a raw corpora to provide statistical information for word associations. A metric is introduced and used to measure the semantic density and to rank all possible combinations of the senses of two words. This method provides a precision of 58% in indicating the correct sense for both words at the same time. The precision increases as we consider more choices: 70% for top two ranked and 73% for top three ranked.

1 Introduction

Word Sense Disambiguation (WSD) is an open problem in Natural Language Processing. Its solution impacts other tasks such as discourse, reference resolution, coherence, inference and others. WSD methods can be broadly classified into three types:

1. WSD that make use of the information provided by machine readable dictionaries (Cowie et al.1992), (Miller et al.1994), (Agirre and Rigau, 1995), (Li et al.1995), (McRoy, 1992);
2. WSD that use information gathered from training on a corpus that has already been semantically disambiguated (supervised training methods) (Gale, Church et al., 1992), (Ng and Lee, 1996);
3. WSD that use information gathered from raw corpora (unsupervised training methods) (Yarowsky 1995) (Resnik 1997).

There are also hybrid methods that combine several sources of knowledge such as lexicon information, heuristics, collocations and others (McRoy, 1992) (Bruce and Wiebe, 1994) (Ng and Lee, 1996) (Rigau, Asterias et al., 1997).

Statistical methods produce high accuracy results for small number of preselected words. A lack of widely available semantically tagged corpora almost excludes supervised learning methods. On the other hand, the disambiguation using unsupervised methods has the disadvantage that the senses are not well defined. To our knowledge, none of the statistical methods disambiguate adjectives or adverbs so far.

One approach to WSD is to determine the *conceptual distance* between words, that is to measure the semantic closeness of the words within a semantic network. Essentially, it is the length of the shortest path connecting the concepts (Rada et al.1989), (Rigau, Asterias et al., 1997). By measuring the conceptual distance between words, it is possible to determine the likelihood of word sense associations. For example, the method proposed in (Li et al.1995) tries to determine the possible sense of a noun associated with a verb using WordNet and a large text. Based on other occurrences of the verb or semantically related verbs in the text, the possible object is determined by measuring the semantic similarity between the noun objects.

Methods that do not need large corpora are usually based exclusively on MRD. A proposal in this sense has been made in (Agirre and Rigau, 1995); they measure the conceptual density between nouns, by using WordNet, but the method proposed in their paper cannot be applied to measuring a conceptual distance between a verb and a noun, as no direct links are provided in MRDs between the nouns and verbs hierarchies. A WordNet-based method for measuring the semantic similarity between nouns was also proposed in (Richardson et al., 1994). Their method consists of using hierarchical concept graphs constructed from WordNet data files, and a semantic similarity formula. Still, the method does not provide a link between different part-of-speech words.

2 Our approach

The approach described in this paper is based on the idea of *semantic density*. This can be measured by the number of common words that are within a semantic distance of two or more words. The closer the semantic relationship between two words the higher the semantic density between them. The way it is defined here, the semantic density works well in the case of uniform MRD. In reality there are gaps in the knowledge representations and the semantic density can provide only an estimation of the actual semantic relatedness between words.

We introduce the semantic density because it is

relatively easy to measure it on a MRD like WordNet. This is done by counting the number of concepts two words have in common. A metric is introduced in this sense which when applied to all possible combinations of the senses of two or more words it ranks them.

Another idea of this paper is to use the Internet as a raw corpora. Thus we have two sources of information: (1) the Internet for gathering statistics and (2) WordNet for measuring semantic density. As will be shown below, a ranking of words senses results from each of these two sources. The issue now is how to combine these two rankings in order to provide an overall ranking. One possibility is to use them in parallel and the other one is to use them serially. We have tried both and the serial approach provided better results. Thus, for a verb - noun pair, the WSD method consists of two Algorithms, the first one ranks the noun senses, of which we retain only the best two senses; and a second Algorithm takes the output produced by the first Algorithm and ranks the pairs of verb - noun senses. Extensions of this method to other pairs than verb - noun are discussed, and larger windows of more than two words are considered.

An essential aspect of the WSD method presented here is that we provide a raking of possible associations between words instead of a binary yes/no decision for each possible sense combination. This allows for a controllable precision as other modules may be able to distinguish later the correct sense association from such a small pool.

WordNet is a fine grain MRD and this makes it more difficult to pinpoint the correct sense combination since there are many to choose from and many are semantically close. For applications such as machine translation, fine grain disambiguation works well but for information extraction and some other applications this is an overkill, and some senses may be lumped together.

A simple sentence or question can usually be briefly described by an action and an object; for example, the main idea from the sentence *He has to investigate all the reports* can be described by the action-object pair *investigate-report*. Even the phrase may be ambiguous by having a poor context, still the results of a search or interface based on such a sentence can be improved if the possible associations between the senses of the verb and the noun are determined.

In WordNet (Miller 1990), the gloss of a verb synset provides a noun-context for that verb, i.e. the possible nouns occurring in the context of that particular verb. The glosses are used here in the same way a corpus is used.

3 Ranking the possible senses of the noun

In order to improve the precision of determining the conceptual density between a verb and a noun, the senses of the noun should be ranked, such as to indicate with a reasonable accuracy the first possible senses that it might have.

The approach we considered for this task is the use of unsupervised statistical methods on large texts. The larger the collection of texts, the bigger is the probability to provide an accurate ranking of senses. As the biggest number of texts electronically stored - and thus favoring an automatic processing - is contained on the Web, we thought of using the Internet as a source of corpora for ranking the senses of the words.

This first step of our method takes into consideration verb-noun pairs $V - N$, and it creates pairs in which the verb remains constant, i.e. V, and the noun is replaced by the words in its similarity lists. Using WordNet, a similarity list is created for each sense of the noun, and it contains: the words from the noun synset and the words from the noun hypernym synset.

Algorithm 1
Input: untagged verb - noun pair
Output: ranking of noun senses
Procedure:

1. *Form a similarity list for each noun sense.* Consider, for example, that the noun N has m senses. This means that N appears in m similarity lists,
   $(N^1, N^{1(1)}, N^{1(2)}, ..., N^{1(k_1)})$
   $(N^2, N^{2(1)}, N^{2(2)}, ..., N^{2(k_2)})$
   ...
   $(N^m, N^{m(1)}, N^{m(2)}, ..., N^{m(k_m)})$
   where N^1, N^2, ..., N^m represent the different senses of N, and $N^{i(s)}$ represents the synonym number s of the sense N^i of the noun N as defined in WordNet.

2. *Form verb - noun pairs.* The pairs that may be formed are:
   $(V - N^1, V - N^{1(1)}, V - N^{1(2)}, ..., V - N^{1(k_1)})$
   $(V - N^2, V - N^{2(1)}, V - N^{2(2)}, ..., V - N^{2(k_2)})$
   ...
   $(V - N^m, V - N^{m(1)}, V - N^{m(2)}, ..., V - N^{m(k_m)})$

3. *Search the Internet and rank senses.* A search performed on the Internet for each of these groups will indicate a ranking over the possible senses of the noun N.

   In our experiments we used (AltaVista) since it is one of the most powerful search engines currently available.

Verb	Noun	Sense of noun in SemCor	Hits provided by AltaVista for V-N where N has the sense no.:								Result
			#1	#2	#3	#4	#5	#6	#7	#8	
rescind	action	6	9	49	0	0	2	27	1	0	1
set-aside	resolution	7	5	0	75	7	0	0	17	2	2
reject	amendment	1	48								1
allow	legislator	1	172								1
ask	person	1	15628	0	1912	0					1
endorse	support	8	101	162	31	13	361	3	0	134	4
expend	fund	1	846	123							1
provide	increase	2	110189	5268	4429	1543	0				2
defeat	person	1	340	0	428						2
wait	term	2	0	27321	1	0	762				1
receive	vote	1	1271	0	0	406	62				1
revise	law	2	224	2829	648	640	37	397	0		1
expect	resignation	3	12	0	554						1
comment-on	topic	2	1801	5517							1
hold	meeting	1	205	128	8	1164	20	69	2227		3
remedy	problem	2	107	345	266						1
place	burden	4	2327	2031	12842	3271					2
award	fee	1	1	1284							2
award	compensation	1	22	126							2
protect	court	1	2574	3120	360	540	916	722	433		2

Table 1: A sample of the result we obtained in ranking the noun senses using the Internet

Using the operators provided by AltaVista, the verb-noun groups derived above can be expressed in two query-forms:

(a) ("V^* N^1*" OR "V^* $N^{i(1)}$*" OR "V^* $N^{i(2)}$*" OR ... OR "V^* $N^{i(k_i)}$*")

(b) ((V^* NEAR N^1*) OR (V^* NEAR $N^{i(1)}$*) OR (V^* NEAR $N^{i(2)}$*) OR ... OR (V^* NEAR $N^{i(k_i)}$*))

where the asterisk (*) is used as a wildcard indicating that we want to find all words containing a match for the specified pattern of letters.

Using one of these queries, we can get the number of hits for each sense i of the noun and this provides a ranking of the m senses of the noun as they relate with the verb V.

We tested this method for 80 verb-noun pairs extracted from SemCor 1.5 of the Brown corpus. [1]

Using query form (a) as an input to the search engine, we obtained an accuracy of 83% in providing a ranking over the noun senses, such as the sense indicated in SemCor was one of the first two senses in this classification. In Table 1, we present a sample of the results we obtained. The column *Result* in this table presents the ranking over the noun senses: a 1 in this column means that the sense indicated in SemCor was also indicated by our method; 2 means that the sense indicated in SemCor was in top two of the sense ranking provided by our method; similarly, 3 or 4 indicates that the sense of the noun, as specified in SemCor, was in the top three, respectively four, of this sense ranking.

We used also the query form (b), but the results we obtained have been proved to be similar; using the operator $NEAR$, a bigger number of hits is reported, but the sense ranking remains the same.

It is interesting to observe that even we are creating queries starting with a verb-noun pair, it is not guaranteed that the search on the web will identify only words linked by such a lexical relation. We based our idea on the fact that: (1) the noun directly following a verb is highly probable to be an object of the verb (as in the expression "*Verb* Noun**") and (2) for our method, we are actually interested in determining possible senses of a verb and a noun that can share a common context.

[1] These verb-noun pairs have been extracted from the file br-a01.

4 Determining the conceptual density between verbs and nouns

A measure of the relatedness between words can be a knowledge source for several decisions in the NLP applications. The conceptual density between verbs and nouns seems difficult to determine, without large corpora or a without a machine-readable dictionary having semantic links between verbs and nouns. Such semantic links can be traced however if we consider the glosses for the verbs, which are providing a possible context of a verb.

Algorithm 2
Input: untagged verb - noun pair and a ranking of noun senses (as determined by Algorithm 1)
Output: sense tagged verb - noun pair
Procedure:

1. Given a verb-noun pair $V - N$, determine all the possible senses for the verb and the noun, by using WordNet. Let us denote them by $< v_1, v_2, ..., v_k >$ and $< n_1, n_2, ..., n_l >$ respectively.
2. Using the method described in section 3, the senses of the noun are ranked. Only the first two possible senses indicated by this step will be considered.
3. For each possible pair $v_i - n_j$, the conceptual density is computed as follows:

(a) extract all the glosses from the sub-hierarchy including v_i (the rationale of the method used to determine these sub-hierarchies is explained below)

(b) Determine the nouns from these glosses. These constitute the noun-context of the verb. All these nouns are stored together with the level of the associated verb within the sub-hierarchy of v_i.

(c) Determine the nouns from the sub-hierarchy including n_i.

(d) Determine the number C_{ij} of common concepts between the nouns obtained at (b) and the nouns obtained at (c).

4. The most suitable combinations between the senses of the verb and the noun $v_i - n_j$ are the ones that provide the biggest values for C_{ij}.

In order to determine the sub-hierarchies that should be used for v_i and n_j, we used statistics provided by SemCor, a sense tagged version of the Brown corpus (Francis and Kucera, 1967) (Miller, Leacock et al., 1993), containing 250,000 words. Each word (noun, verb, adjective, adverb) is included in a synset within a hierarchy. The tops of these hierarchies denominate the class of the word. The sense in SemCor for a word W is indicated by the class C of the word W, and the sense of the word within the class C. For example, the SemCor entry:

```
<wf cmd=done pos=NN lemma=investigation wnsn=1
lexsn=1:09:00::>investigation</wf>
```

indicates:
word: investigation
part of speech: common noun
sense in WordNet: 1

A statistic measure performed on SemCor, indicates the following probabilities for the sense of a word within a class:

Parts of speech	Number of words in SemCor	within a class, the probability to have sense number:					
		0	1	2	3	4	5
noun	47,799	85%	10%	3%	1%	-	-
verb	27,637	60%	14%	5%	12%	3%	2%

Table 2: The probabilities for the sense of a word within a class

As shown in Table 2, the class of the noun indicates with a probability of 85% a correct sense 1 within that class.

Thus, for this algorithm, we consider for a noun the hierarchy including the noun (if the class of the noun n_i is C, then the method considers all the nouns from the class C).

This does not work for the verbs, as the probability to indicate a correct sense knowing the class is much smaller (only 60%). For this reason, and based on the experiments we computed, the sub-hierarchy including a verb v_i is determined as follows: (i) consider the hypernym h_i of the verb v_i and (ii) consider the hierarchy having h_i as top.

It is necessary to consider a bigger hierarchy then just the one provided by synonyms and direct hyponyms, since providing accuracy in a metric computation needs large corpora. As we replaced the corpora with the glosses, better results are achieved if more glosses are considered. Still, we do not have to enlarge too much the context, in order not to miss the correct answers.

Conceptual Density Metric

For determining the conceptual density between a noun n_i and a verb v_j, the algorithm considers:

- the list of nouns sv_k associated with the glosses of the verbs within the hierarchy determined by h_j: (sv_k,w_k), where:
  - h_j is the hypernym of v_j
  - w_k is the level in this hierarchy
- the list of nouns sn_l within the class of n_i : (sn_l)

The common words between these two lists (sv_k,w_k) and (sn_l) will produce a list of common concepts with the associated weights $cd_{ij} < w_k >$. The conceptual density between n_i and v_j is given by the formula:

$$(1) \qquad C_{ij} = \frac{\sum_{k}^{|cd_{ij}|} w_k}{log(desc_i)}$$

where:

- $|cd_{ij}|$ is the number of common concepts between the hierarchies of n_i and v_j
- w_k are the weights associated with the nouns from the noun-context of the verb v_j
- $desc_i$ is the total number of words within the hierarchy of noun n_i

As the nouns with a big hierarchy tend to indicate a big value for $|cd_{ij}|$, the weighted sum of common concepts has to be normalized in respect with the dimension of the noun hierarchy. This is estimated as the logarithm of the total number of descendants in the hierarchy (i.e. $log(desc_i)$).

We also took into consideration other metrics, like:

(2) The number of common concepts between the noun and verb hierarchies, without considering the weights.

(3) A weighted summation of the common concepts between the noun and verb hierarchies, as indicated in (1), but without a normalization in rapport with the noun hierarchy.

We considered also the metrics indicated in (Agirre and Rigau, 1995). But after running the program on several examples, the formula indicated in **(1)** provided the best results.

A possible improvement to the metric **(1)** is to consider the weights for the levels in the noun hierarchy, in addition to the levels in the verb hierarchy.

5 An example

Consider as example of a verb-noun pair the phrase *revise law*. The verb *revise* has two possible senses in WordNet 1.5:

Sense 1
revise, make revisions in
gloss: (revise a thesis, for example)

⇒ rewrite, write differently, alter by writing
gloss: ("The student rewrote his thesis")

Sense 2
re tool, revise

⇒ reorganize, shake up, organize an

The noun *law* has 7 possible senses

Sense 1
law, jurisprudence
gloss: (the collection of rules imposed by authority; "civilization presupposes respect for the law")

⇒ collection, aggregation, accumulation, assemblage
gloss: (several things grouped together)

Sense 2
law
gloss: (one of a set of rules governing a particular activity or a legal document setting forth such a rule; "there is a law against kidnapping")

⇒ rule, prescript
gloss: (prescribed guide for conduct or action)

⇒ legal document, legal instrument, official document, instrument

Sense 3
law, natural law
gloss: (a rule or body of rules of conduct inherent in human nature and essential to or binding upon human society)

⇒ concept, conception
gloss: (an abstract or general idea inferred or derived from specific instances)

Sense 4
law, law of nature
gloss: (a generalization based on recurring facts or events (in science or mathematics etc): "the laws of thermodynamics)

⇒ concept, conception
gloss: (an abstract or general idea inferred or derived from specific instances)

Sense 5
jurisprudence, law, legal philosophy
gloss: (the branch of philosophy concerned with the law)

⇒ philosophy
gloss: (the rational investigation of questions about existence and knowledge and ethics)

Sense 6
police, police force, constabulary, law
gloss: (the force of policemen and officers; "the law came looking for him")

⇒ force, personnel
gloss: (group of people willing to obey orders)

Sense 7
law, practice of law
gloss: (the learned profession that is mastered by graduate study in a law school and that is responsible for the judicial system; "he studied law at Yale")

⇒ learned profession
gloss: (one of the three professions traditionally believed to require advanced learning and high principles)

We searched on Internet, using AltaVista, for all possible pairs V-N that may be created using *revise* and the words from the similarity lists of *law*. Over the seven possible senses for this noun, the first step of our method indicated the following ranking (we indicate the number of hits between parenthesis):*law#2*(2829), *law#3*(648), *law#4*(640), *law#6*(397), *law#1*(224), *law#5*(37), *law#7*(0). Thus, only the sense *#2* and *#3* of the noun *law* are eligible to be used for the next algorithm.

For each of the two senses of the verb, we determined the noun-context, including the nouns from the glosses in the sub-hierarchy of the verb, and the associated weights.

For each of the two possible senses of the noun, we determined the nouns from the class of each sense.

In Table 3, we present: (1) the values obtained for the combinations of different senses, i.e. the number of common concepts between the verb and noun hierarchies - $|cd_{ij}|$ (columns 2-3); (2) the summations of the weights associated with each noun within the noun-context of the verb v_j (columns 4-5); (3) the total number of nouns within the hierarchy of each sense n_i, i.e. $desc_i$ (columns 6-7); (4) the conceptual density C_{ij} for each pair $n_i - v_j$, derived using the formula presented above (columns 8-9).

	$\|cd_{ij}\|$		weights		$desc_i$		C_{ij}	
	2	3	4	5	6	7	8	9
	n_2	n_3	n_2	n_3	n_2	n_3	n_2	n_3
v_1	5	4	2.06	2	975	1265	0.30	0.28
v_2	0	0	0	0	975	1265	0	0

Table 3: Values used in computing the conceptual density and the conceptual density C_{ij}

In this table:
- v_i indicates the sense number i of verb *revise*
- n_i indicates the sense number i of noun *law*

The biggest value for conceptual density is given by $v_1 - n_2$:

$$revise\#1/2 - law\#2/5 \quad C_{11} = \quad 0.30$$

This combination of verb-noun senses[2] appears in SemCor, file br-a01.

6 Tests against SemCor

We tested this method by using verb-noun pairs from SemCor. A randomly selected sample from the entire table with 80 pairs is presented in Table 4.

For each pair verb-noun, we indicate the sense of the verb (column B), the sense of the noun (column C), as they result from SemCor; the total number of possible senses for both the verb (column D)

[2]The notation $\#i/n$ means sense i out of n possible.

Verb	Noun	Results					
A		B	C	D	E	F	G
rescind	action	1	6	1	8	8	1
set-aside	resolution	2	7	2	8	16	1
reject	amendment	2	1	4	1	4	2
allow	legislator	2	1	8	1	8	2
ask	person	1	1	5	3	15	1
endorse	support	1	8	4	13	52	-
expend	fund	2	1	2	2	4	1
provide	increase	1	2	4	5	20	1
defeat	person	1	1	2	3	6	1
wait	term	3	2	4	6	24	2
receive	vote	2	1	11	5	55	1
revise	law	1	2	2	7	14	1
expect	resignation	1	3	6	3	18	1
comment-on	topic	1	2	1	2	2	1
hold	meeting	2	1	21	7	147	-
remedy	problem	1	2	2	3	6	1
place	burden	1	4	14	4	56	1
award	fee	2	1	3	2	6	1
award	compensation	2	1	3	2	6	1
protect	court	1	1	2	6	12	2

Table 4: A sample of the results obtained on verb-noun pairs extracted from SemCor

and the noun (column E). Column F illustrates the number of possible combinations between different senses of the verb and the noun, computed as total_senses_verb x total_senses_noun (product of numbers in column D and E). Then column G indicates the results obtained by using the conceptual density method in determining the possible associations between verbs and nouns.

The numbers in the last column should be interpreted as follows: 1 means the combination of senses for verb and noun, as resulted from SemCor, was indicated as the first possible combination by our method, 2, 3 or 4 means that the combination of senses for verb and noun, as resulted from SemCor, was in the top two, respectively three or four, possible combinations indicated by our method; a dash − means that the sense of the noun was not included in the first two senses determined by our sense ranking method.

Discussion of results: The last column in Table 4 indicates the ranking provided by this method, with respect to the combinations of senses specified in SemCor. In Table 5, we present a summary of the results.

	Number of pairs considered considered	Ranking provided by this method, such as the combination indicated in SemCor was in top:		
		1	2	3
Value	80	47	56	60
Percent	100%	58%	70%	73%

Table 5: Summary of results

The senses specified in SemCor for a verb-noun pair matched our first choice with a precision of 58%, matched our top two choices with 70% and our top three choices with 73%.

When evaluating these results, one should take into consideration that:

- Considering the glosses as a base for calculating the conceptual density it has the advantage of eliminating the use of a large corpus. But a disadvantage comes with the use of glosses, as they are not part-of-speech tagged, like the corpora usually are. For this reason, when determining the nouns from the verb glosses, an error rate is introduced, as some verbs (like *make, have, go, do*) have also a noun function in WordNet.
- SemCor provides a possible combination of senses for a verb-noun pair, but in some cases this combination is determined by the context. For example, for the pair *protect court* SemCor indicates a combination such that the court means *an assembly to conduct judicial business.* Instead, the method we used indicates as a first possible sense for court - *a room in which a law court sits*, and the second possible sense is the one indicated in SemCor.

Evaluation: As outlined in (Resnik and Yarowsky, 1997), it is hard to compare the WSD methods, as long as distinctions reside in the approach considered (MRD based methods, supervised or unsupervised statistical methods), in the words that are disambiguated (words a priori established, only nouns etc.). To our knowledge, there is no other MRD based algorithm that tries to disambiguate both the verb and the noun in a verb-noun pair, so we cannot make a comparison.

But we can consider the accuracy of the WSD method that assigns to each word its most frequent sense, as a base to evaluate our method. Two cases have been considered:

1. The particular case of the 80 pairs considered for the tests: assigning the most frequent sense, results in an accuracy of 60% for the noun, and 69% for the verb. Thus, an overall accuracy of 41% for a verb-noun pair, while assigning the most frequent sense for both the verb and the noun.
2. We measured on SemCor that the first (most frequent) WordNet sense occurs with a probability of 80.35% for nouns, 62.53% for verbs, 81.78% for adjectives and 84.38% for adverbs. It results then an overall accuracy of 50.24% (= 80.35% x 62.53%) in disambiguating both the verb and the noun in a verb-noun pair, while assuming the most frequent sense for each of these words. Thus, an important improvement has been achieved by our method that increases this accuracy to 58%.

Extensions: Extensions of the method proposed in this paper can be done in two different directions:

1. The method presented here can be applied in a similar way to determine the conceptual density within noun-noun pairs, or verb-verb pairs (in

these cases, the *NEAR* operator should be used for the first step of this algorithm).

2. The number of words considered at a time can be increased, from two to three, four or even more words.

7 Conclusion

In this paper, we have presented a method for WSD that is based on measuring the conceptual density between words using WordNet. The metric proposed may be further improved by considering the weights for verbs as well as for nouns. The senses of the words are ranked, and an user may select the first choice or the first few choices, depending upon the application. We have also proposed to use the Internet as a source of statistics on a raw corpora.

The method extends well to considering more than two words at a time, and also for all parts of speech covered by WordNet.

It is difficult to compare the precision obtained by this method with other methods, since we consider here the collective meaning of two or more words, while most of other methods consider one word at a time. However, an estimation can be done by extracting the square root of the accuracy for a pair of verb-noun words; and that is 76.15% for the first choice, 83.66% for the first two choices and 85.44% for the first three choices. Since the disambiguation precision for nouns is usually higher than for verbs, those numbers provide only an average.

References

Digital Equipment Corporation. AltaVista Home Page. URL:*http://www.altavista.digital.com*.

E. Agirre and G. Rigau, A Proposal for Word Sense Disambiguation using Conceptual Distance, Proceedings of the 1st International Conference on Recent Advances in Natural Language Processing, Velingrad, 1995

R. Bruce and J. Wiebe, Word Sense Disambiguation using Decomposable Models, Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94), 139-146, LasCruces, NM, June 1994.

J. Cowie, L. Guthrie and J. Guthrie, Lexical disambiguation using simulated annealing. *Proceedings of the Fifth International Conference on Computational Linguistics COLING-92*, 157-161, 1992.

S. Francis and H. Kucera, *Computational Analisys of present-day American English*, Providence, RI: Brown University Press, 1967

W. Gale, K. Church and D. Yarowsky, One Sense per Discourse, Proceedings of the DARPA Speech and Natural Language Workshop, Harriman, New York, 1992.

X. Li, S. Szpakowicz and S. Matwin. A WordNet-based algorithm for word semantic sense disambiguation. Proceedings of the 14th International Joint Conference on Artificial Intelligence IJCAI-95, Montreal, Canada, 1995.

S. McRoy, Using multiple Knowledge Sources for Word Sense Disambiguation, *Computational Linguistics*, 18(1):1-30, 1992.

G.A. Miller, WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235-312, 1990.

G.A. Miller, C. Leacock, T. Randee and R. Bunker, A Semantic Concordance. Proceedings of the 3rd DARPA Workshop on Human Language Technology, 303-308, Plainsboro, New Jersey, 1993

G.A. Miller, M. Chodorow, S. Landes, C. Leacock and R.G. Thomas, Using a semantic concordance for sense identification. *Proceedings of the ARPA Human Language Technology Workshop*, 240-243, 1994.

H.T. Ng and H.B. Lee, Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Examplar-Based Approach, Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL-96), Santa Cruz, 1996.

R. Rada, H. Mili, E. Bickell and M. Blettner. Development and Application of a Metric on Semantic Nets. *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, pp 17-30, Jan/Feb 1989.

P. Resnik, Selectional Preference and Sense Disambiguation, Proceedings of ACL Siglex Workshop on Tagging Text with Lexical Semantics, *Why, What and How?*, Washington, April 4-5, 1997.

P. Resnik and D. Yarowsky, A Perspective on Word Sense Disambiguation Methods and Their Evaluation, Proceedings of ACL Siglex Workshop on Tagging Text with Lexical Semantics, *Why, What and How?*, Washington, April 4-5, 1997.

R. Richardson, A.F. Smeaton and J. Murphy, Using WordNet as a Knowledge Base for Measuring Semantic Similarity between Words, Technical Report, Working paper CA–1294, School of Computer Applications, Dublin City University. Dublin, Ireland, 1994.

G. Rigau, J. Atserias and E. Agirre, Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation. *Computational Linguistics /9704007*, 1997.

D. Yarowsky, Unsupervised Word Sense Disambiguation rivaling Supervised Methods. Proceedings of the 33rd Association of Computational Linguistics, 1995.