Gloss overlap extensions for a semantic network algorithm: building a better semantic distance measure

Thimal Jayasooriya and Suresh Manandhar

Department of Computer Science,
The University of York,
York YO10 5DD,
United Kingdom
`thimal, suresh@cs.york.ac.uk`

Abstract

Semantic similarity or inversely, semantic distance measures are useful in a variety of circumstances, from spell checking applications to a lightweight replacement for parsing within a natural language engine. Within this work, we examine the (Jiang & Conrath 1997) algorithm; evaluated by (Budanitsky & Hirst 2000) as being the best performing, and subject the algorithm to a series of tests. We also propose a novel technique which corrects a crucial weakness of the original algorithm, and show that its application improves semantic distance measures for cases where the underlying linguistic network causes deficiencies.

Introduction

Semantic distance has been used in a variety of situations and natural language processing tasks. Word sense disambiguation (Sussna 1993) (Pedersen & Banerjee 2003), identifying discourse structure, text summarization and annotation, lexical selection and information retrieval tasks are some of the areas discussed in (Budanitsky 1999)'s work.

However, semantic distance computation need not be confined to identification of synonyms such as *midday* and *noon* or *boy* and *lad*. Is there a semantic relationship between a *tire* and a *wheel* ? Between a *doctor* and a *hospital* ? Is the relatedness between a *bus* and *driver* closer than that between a *bus* and a *conductor* ? These are some of the questions that semantic distance computation is intended to answer. Giving a quantifiable numeric value to the degree of relatedness between two words is the function of numerous semantic distance algorithms.

Given the importance of semantic similarity measurements in such a wide variety of tasks, it's no surprise that a variety of techniques have been devised over the years to measure relatedness. Budanitsky (1999) discusses three main approaches adopted by these techniques - computing path length, scaling the network and integrated approaches. (Ted Pedersen & Michellizzi 2004) classify semantic distance algorithms supported in their widely

available Wordnet::Similarity module as being path based, information content based and based on gloss similarity.

Determining the best semantic distance algorithm out of the many that have been devised is subjective. However, (Budanitsky & Hirst 2000) is among several studies that have been done on various algorithms to discover the best performing for a standard series of tests. In their work, Budanitsky and Hirst (2000) conclude that Jiang and Conrath's integration of edge counting and information content performs best for a standard series of twenty word pairs. We re-examine the algorithm, as implemented by the Wordnet::Similarity module. We also evaluate our results using a subset of the (Rubinstein & Goodenough 1965) dataset for examining correlations between synonyms. Our test data is the original dataset of 20 word pairs used by (Jiang & Conrath 1997) - augmented by the more recent (Miller & Charles 1991) study, which adds human judgement estimates for each of the word pairs.

The significant outcome of this work is an enhanced algorithm for determining semantic distance. As with all other semantic distance measurement techniques implemented by Wordnet::Similarity, Jiang and Conrath's method (hereafter referred to as *jcn*) operates on Wordnet (Fellbaum 1998), a lexical database which organizes words into relations. One of the key weaknesses of *Jiang and Conrath*'s algorithm is its dependence on the network structure of Wordnet for an accurate result. By combining a semantic network approach such as Jiang and Conrath with a network agnostic semantic measure; such as extended gloss overlaps, we were able to increase the correlation coefficient for cases where an integrated node information content and path length driven measurement had failed to identify an appropriate degree of semantic relatedness. In other words, using a gloss overlap technique allowed us to augment *jcn* relatedness scores which were lowered due to clear deficiencies in the underlying semantic network.

Experiments

Our test set of 20 word pairs - which comprise part of the (Rubinstein & Goodenough 1965) test data set - is identical to that used by Jiang and Conrath (1997) in their semantic distance experiments. We also examine the scores that result

Word pair	jcn-score	vector-score	MC-score
food-rooster	0.018	0.746	0.222
noon-string	0.002	0.474	0.020
coast-forest	0.011	0.685	0.105
boy-lad	0.0919	0.855	0.950
chord-smile	0.0194	0.326	0.045
magician-wizard	0.0157	0.439	0.875
tool-implement	0.142	0.822	0.75
gem-jewel	1	1	0.970
journey-car	0.030	0.811	0.290
midday-noon	1	1	0.82
monk-slave	0.017	0.619	0.137
brother-monk	0.019	0.794	0.705
furnace-stove	0.009	0.68	0.797
glass-magician	0.008	0.43	0.025
cemetery-woodland	0.005	0.803	0.237
lad-wizard	0.023	0.472	0.105
forest-graveyard	0.008	0.803	0.210
shore-woodland	0.012	0.537	0.157
car-automobile	1	1	1
rooster-voyage	0.001	0.685	0.002

Table 1: Semantic distance score comparison between Jiang-Conrath, Vector and Miller-Charles scores [range: completely unrelated (0.0) - synonymous (1.0)]

from the vector method (Pedersen & Banerjee 2003) and show human judgement scores from the (Miller & Charles 1991) experiments for comparison. These results are seen from Table 1. In these results, the highest possible semantic distance score has been taken in all cases. The *jcn* score has been normalized using log10.

Analysis and discussion

The results from Table 1 show that there is general agreement between the *vector* method and *jcn* on three instances of synonymy - *midday-noon*, *gem/jewel* and *car-automobile* are flagged by both as being closely related, if not actual synonyms. Somewhat surprisingly, even though the automated semantic distance algorithms flagged the *midday-noon* wordpair as being related, this result did not correlate precisely with the human evaluations conducted by Miller and Charles.

Using Pearson's correlation coefficient technique, we discovered that the Jiang and Conrath algorithm results displayed a 0.62226 correlation with the Miller and Charles results showed above - while the vector method showed a 0.600489 correlation. It is worth noting that the vector method correlation results are consistent with those reported by (Pedersen & Banerjee 2003), while the *jcn* scores are significantly lower than those reported in the original (Jiang & Conrath 1997) paper.

Budanitsky (2000) used several methods of evaluation for his results - one of them being human judgement. This evaluation technique is mirrored by Jiang and Conrath (1997).

In each case, they relied on the evaluations performed by Miller and Charles (1991). Thus, our next evaluation phase examined words which were judged as being semantically *related* by human evaluators, but weren't identified as such by the semantic distance algorithms.

It's interesting to note that the word pairs *boy-lad* and *magician-wizard* have been identified as strongly related by human assessment, but have not been similarly recognized by the semantic distance algorithms. In each case, Roget's thesaurus provides the opposing term as part of its definition or as a synonym. For example, one sense of *wizard* has *magician* as its definition and *lad* is defined as a *boy*. Also included is a word pair omitted from evaluations in previous years *furnace/stove* scored highly on human evaluation results but wasn't included in the original computational experiments performed by Resnik due to a limitation in Wordnet (Jiang & Conrath 1997).

Devising a best of breed semantic distance measurement technique

From the preceding analysis and evaluation of semantic distance algorithms, it is clear that existing semantic distance algorithms can be further improved. For the purposes of our assessment of deficiencies, we use Budanitsky's (Budanitsky 1999) classification of semantic distance algorithms.

- Network scaling type algorithms (path and edge counting and graph traversal type algorithms) are affected by node density in specific areas of Wordnet. The number of enumerated nodes available in a specific area of Wordnet have an effect on the semantic distance determination. Sparser areas of the Wordnet topology may have a shorter *hop count* and consequently score much better in edge counting type algorithms, yet still be semantically less similar than warranted by the distance measurement returned.

- Information content based algorithms - Resnik and Jiang-Conrath for example, operate on frequency information which applies to the entire corpus of data. Any addition to the Wordnet database - even if the additions are not the source or target words, but have an influence on the computation of the least common subsumer (LCS) - will result in differentiated score.

- Network scaling and integrated approaches for semantic distance calculation cannot cross verb/noun boundaries due to the *is-a* hierarchy organization of Wordnet (Ted Pedersen & Michellizzi 2004). This also precludes the possibility of semantic distance calculations being performed on other parts of speech such as adjectives and adverbs.

- On the other hand, algorithms which depend on gloss overlaps for determination of semantic similarity are prone to surprising errors. Of the three gloss overlap techniques offered by the Wordnet::Similarity modules, only the vector technique identified both *midday/noon* and *car/automobile* as being closely related - the vector pairs

Word pair	jcn-score	vector-score	MC-score
boy-lad	0.0919	0.855	0.950
magician-wizard	0.0157	0.439	0.875
furnace-stove	0.009	0.68	0.797

Table 2: Semantic distance scores where human judgement scored higher than either algorithm

technique and the Lesk algorithm (Lesk 1986) adaptation had difficulty in identifying the *midday/noon* word pair.

Given these problems in the observed results with the Rubinstein-Goodenough (1965) dataset, we went on to investigate possible enhancements for increasing the accuracy of the returned semantic distance values. Of particular concern were the word pairs shown in Table 2, with clearly erroneous scores returned by the *jcn* algorithm.

Hybrid or integrated approaches

One of the original claims made by Jiang and Conrath (1997) was that an integrated approach which incorporates both path based and information content based characteristics combines the best of both approaches and provides a degree of robustness against the weaknesses of an individual technique.

The issue with Jiang-Conrath's technique, although being one of the better performing algorithms, is its reliance on the semantic network structural properties of the linguistic resource being used - in this case, Wordnet. Jiang-Conrath's algorithm (hereafter referred to as *jcn*) uses the *link strength* metric; a combination of node information content and path length (synonymously referred to as edge based) computations. This inherently places the burden of a proper semantic distance evaluation on the quality of the semantic network. Where the classification of a given word is both proper and adequately supported by a hierarchy of related *is-a* concepts, a semantic distance measurement has a high chance of success.

However, in cases where the Wordnet evaluation and placement of a particular word or synset does not agree with conventional usage, there may be a perceived deficiency in the resulting distance metric. Even Jiang and Conrath's own evaluation observed that the *furnace/stove* word pair did not produce a good result. This was explained by the super-ordinate class of both *furnace* and *stove* being *artifact* - a considerably higher level construct.

Thus, the weaknesses described earlier become applicable to the *jcn* algorithm. Additionally, *jcn* also depends on an *invisible* construct - that of the superordinate class, or according the algorithm description, the information content of the least common subsumer (LCS). Therefore, it is our contention that the *jcn* construct can indeed be affected by network density and depth - given that the density of nodes in a specific area of the semantic network has a direct bearing on the specificity of the superordinate class for a given word pair.

We can now observe that there are three main choices for devising a new integrated approach for determining semantic distance. The *jcn* algorithm already uses path length computation and node information content. We require further integration which minimizes the effect of the semantic network - while increasing accuracy.

Pertinently, gloss overlap techniques; on which the *vector* method is based, are insensitive to the structural properties of Wordnet or any other lexical resource (Ted Pedersen & Michellizzi 2004). Additionally, they do not constrain themselves to the *is-a* hierarchy as do other semantic distance algorithm implementations, but can instead compute similarities between different parts of speech; and even consider other relations that are non-hierarchical in nature, such as *has-part* and *is-made-of* etc. Another important characteristic of glosses is that it is unlikely to change with subsequent revisions of Wordnet - a feature allows for relative longevity of semantic distance scores. However, relatedness measurements using gloss overlap are fragile and somewhat dependent on a relatively subjective definition of a given word.

Therefore, our next phase of experiments merged the properties of second order co-occurrence vectors of the glosses of word senses with an integrated approach combining node information content and path length computation characteristics.

A gloss overlap aware semantic network metric

The quality of the semantic network is one of the key weakness of the *jcn* algorithm and other semantic network reliant algorithms. Specifically considering the Jiang-Conrath implementation, it can be seen from the results in Table 3 that determining the closest superordinate class is crucial to the eventual semantic distance result.

Words	LCS	jcn-score	human
food-rooster	entity	0.018	0.22
boy-lad	male	0.0919	0.95
furnace-stove	artifact	0.009	0.79
magician-wizard	person	0.015	0.87
car-automobile	car	1	1
midday-noon	noon	1	0.82

Table 3: Least common subsumers (LCS) for *jcn* algorithm calculations

Table 3 indicates the returned results for a selected subset of the Miller-Charles dataset (1991). As the figures demonstrate, there is a very large margin of error between the *jcn* score and the Miller-Charles results where the least common subsumer has not been sufficiently specific to the common concept between word pairs. In the case of the *food-rooster* word pair, the genericity of the LCS is

unsurprising. However, in comparison, the *boy/lad* and *magician/wizard* word pairs do not really have a sufficiently specific LCS concept.

The discrepancy between the Miller Charles judgement and the Jiang-Conrath semantic distance measurement also demonstrates that relatedness measurement is extremely dependent on the position of the least common subsumer in the *is-a* Wordnet hierarchy. The more generic the LCS (and consequently, closer to the root node), the lower the relatedness score.

For examination of our proposed technique, we initially gathered the following data about a given word pair.

- The semantic distances between the word pairs - as computed by the *jcn* algorithm
- The least common subsumer (LCS) between the two words
- The path length from each of the words in the word pair and the LCS to the root node of Wordnet (can also be referred to as the depth from the root node)
- The semantic distances between the word pairs - as computed by the chosen gloss overlap technique. In our case, the second order co-occurrence vector means of individual glosses.

We also defined a *depth* factor. The depth factor denotes the number of nodes between a given word and the root node. In our case, the depth factor is used to indicate the threshold synset depth of the least common subsumer of a word pair.

Simply described, our algorithm places relative weights on the scores returned by the *jcn* means as well as on the scores returned by the gloss overlap means. Given a predefined depth factor between the LCS and the individual words in the word pair; we place a higher or lower weight on the scores returned by the gloss overlap technique. Thus, our *gloss overlap aware semantic network metric* relies more on the properties of the semantic network when the least common subsumer is closer to the examined word pairs; and conversely, relies more on the properties of gloss overlap where the least common subsumer is further away from the examined word pair.

Although simply expressed, there are several practical difficulties in the combining of disparate semantic distance algorithms. One such consideration is that of normalization of scores. Within the original *jcn* algorithm, semantically identical elements have a distance of 0; thus a relatedness score of ∞. However, Wordnet::Similarity treats the *jcn* theoretical maximum slightly differently. Here, the maximum possible value is arrived at using the following formula:

$$similarity = 1/(-log((freq(root) - 0.01)/freq(root)))$$
(1)

where freq(root) is the frequency information for the root node of the semantic network. Applying equation 1 above

to a synonymous word pair yields the maximum score of 2.959×10^7 . On the other hand, the gloss overlap technique we chose is much simpler in its range of returned scores - the vector algorithm ((Pedersen & Banerjee 2003)) with a range between 0 (semantically unrelated) and 1 (highest possible score).

We determine the relative weight to be given to each technique using the equation shown below. Given that w1 is the first word in the pair and w2 is the second; with LCS being the least common subsumer:

$$depth\Delta = (w1depth + w2depth) - 2 \times LCSdepth \quad (2)$$

For a given depth factor of 6, a depth Δ of 6 would produce the following adjusted semantic distance score:

$$AdjustedScore = (jcn\text{-}score \times 0.5) + (vector\text{-}score \times 0.5)$$
(3)

thus, for a depth factor of 6 and an examined depth Δ of 6, the adjusted semantic distance score would be equally affected by both the gloss overlap and the *jcn* technique scores; with both scores contributing equally to the final adjusted score.

$$median - score = (jcn\text{-}score \times 0.5) + (vector\text{-}score \times 0.5)$$
(4)

However, for a depth Δ of larger than 6, the gloss overlap technique would increasing contribute a higher percentage of the final adjusted score; while a depth Δ closer to 0 would give prominence to the *jcn* score. The maximum depth has been experimentally set at 20.

Assume that the median score for our depth Δ is expressed as *median-score*. For a maximum depth of 20 and a depth factor of 6, we can divide depth into two discrete areas: all depth Δ larger than the depth factor, and all depth Δ values which are smaller than the depth factor.

The Δ values smaller than the depth factor have a greater influence from the *jcn* algorithm score.

$$Adjusted - Score = median\text{-}score + ((jcn\text{-}score \times 0.5) \times (0.5 \times (depth\text{-}difference \div depth\text{-}range))) - ((vector\text{-}score \times 0.5) \times (0.5 * depth\text{-}difference \div depth\text{-}range))$$

In the above equation, assuming that the depth Δ is determined to be 4, the depth range value would be (*depth-factor* - 0) = 6, the depth difference is (*depth-factor - depth Δ*) = 2.

$$AdjustedScore = median\text{-}score - ((jcn\text{-}score \times 0.5) \times (0.5 \times depth\text{-}difference \div depth\text{-}range)) + ((vector\text{-}score \times 0.5) \times (0.5 times depth\text{-}difference \div depth\text{-}range))$$

The above equation is only applied when the Δ value is larger than the predefined depth factor. Assuming that the depth Δ is determined to be 12, the depth range value would be (*max-depth - depth-factor*) = 14, the depth difference is (*depth Δ - depth-factor*) = 6.

Word pair (synset depth)	LCS (synset depth)	jcn-score	vector-score	depth Δ	adjusted score
food-rooster (4, 13)	entity (2)	0.018	0.746	13	0.473
noon-string (9, 7)	DNC (1)	0.002	0.474	14	0.305
coast-forest (9, 5)	DNC (1)	0.011	0.685	12	0.420
boy-lad (7, 7)	male (5)	0.0919	0.855	4	0.537
chord-smile (7, 9)	abstraction (2)	0.019	0.326	12	0.205
magician-wizard (7, 6)	person (4)	0.015	0.439	5	0.245
tool-implement (8, 6)	implement (8)	0.142	0.822	-2	0.652
gem-jewel (7, 8)	jewel (8)	1	1	-1	1
journey-car (7, 11)	DNC (1)	0.030	0.811	16	0.567
midday-noon (9, 11)	noon (9)	1	1	2	1
monk-slave (7, 5)	person (4)	0.017	0.619	4	0.368
brother-monk (8, 7)	person (4)	0.019	0.794	7	0.420
furnace-stove (7, 8)	artifact (4)	0.009	0.68	7	0.357
glass-magician (6, 7)	entity (2)	0.008	0.43	9	0.242
cemetery-woodland (8, 5)	entity (2)	0.005	0.803	9	0.447
lad-wizard (7, 6)	person (4)	0.023	0.472	5	0.266
forest-graveyard (5, 8)	DNC (1)	0.008	0.803	11	0.476
shore-woodland (6, 5)	object (3)	0.0128	0.537	5	0.297
car-automobile (11, 11)	car (11)	1	1	0	1
rooster-voyage (13, 8)	DNC (1)	0.001	0.685	19	0.361

Table 4: Experimental data for gloss overlap extension to *jcn*

In Table 4, we present the results of applying our improved semantic distance measure to the Miller-Charles dataset. In each case, we display the depth of the word (as represented in Wordnet) within parentheses. The least common subsumer (LCS) is also displayed. The notation DNC denotes either a root node reference being returned - or that the LCS did not return a valid node due to lack of information.

The figures shown in Table 4 show encouraging results for the combination of gloss overlap with the Jiang-Conrath metric. Given a depth factor of 6; the correlation figures climbed from an overall 0.6222 to 0.6822. Also, of particular relevance to our study, we discovered that the extended gloss overlap combination served to correct some clearly erroneous LCS selections made by the *jcn* algorithm; and pushed the relatedness scores of such cases higher. This finding essentially validates our case for incorporating a semantic network agnostic measure into the Jiang and Conrath semantic distance scores.

Consider the examples cited earlier: the *boy/lad* word pair saw an improved score of 0.537 as opposed to the earlier 0.0919; the *magician/wizard* word pair score climbed to 0.245 from 0.015 and the *tool/implement* word pair saw an increase from 0.142 in the *jcn* result to a score of 0.652. The previously cited *furnace/stove* word pair saw a similar rise from 0.009 to 0.357.

Despite the improvement in correlation, not all results showed an improvement in their individual scores. One reason for this discrepancy is fundamental to our algorithm - the semantic distance scores of word pairs with an extremely generic LCS are biased towards the gloss overlap technique score. This is both productive and useful in the cases where the underlying semantic network has failed to produce a sufficiently specific LCS (for example, in the cases of *furnace/stove*). However, genuinely unrelated word pairs (such as *noon/string* or *rooster/voyage*) should have a high depth Δ - the most common concept between unrelated word pairs should be in fact, be a generic word such as *entity*.

Our algorithm biases unrelated word pair results towards a gloss overlap score - and in some cases, a clearly erroneous gloss overlap semantic distance score has skewed the overall semantic distance score. A good example of this situation is found in the *food/rooster* word pair. Given that the gloss overlap technique operates on a scale of 0.0 to 1.0, it seems highly improbable that a gloss overlap score of 0.746 is accurate for the *food/rooster* word pair. Although the influence of the (accurately) low *jcn* score lowers the gloss overlap score, it still retains nearly half the value and thus produces an incorrect result. A similar situation exists for *glass/magician* and *noon/string* word pairs.

Conclusion and further work

In this paper, we've attempted to duplicate the experiments performed by Jiang and Conrath (1997) in developing their algorithm for determining semantic distance. Our experiments focused on the implementation offered by the Wordnet::Similarity Perl modules. We provide experimental results based on the widely cited (Miller & Charles 1991) data set. We show a comparison between the vector technique (Pedersen & Banerjee 2003), the jiang and conrath technique (Jiang & Conrath 1997) and the Miller Charles

study of human relatedness assessments (Miller & Charles 1991).

Next, we investigated the properties of the Jiang and Conrath approach for measuring semantic similarity - a measure which combines lexical taxonomy structure with corpus statistical information. Contrary to the assertion made by the authors; we established that the *jcn* algorithm was indeed affected by semantic network density and depth - such that the determination of the most appropriate least common subsumer proved to be of crucial importance in the final assessment of semantic similarity. In cases where the least common subsumer proved to be farther away from the words examined, the relatedness scores were low. This is exactly as expected for unrelated or distantly related word pairs. However, it is sometimes the case that closely related words are programmatically determined as having a LCS that is further away than optimal due to a deficiency in Wordnet. In the original Jiang and Conrath work (1997), they cite one example word pair to demonstrate this phenomenon - *furnace-stove*. In our experiments, we uncovered several other word pairs which are similarly impeded.

We went on to propose a means of incorporating gloss overlap techniques (using techniques proposed and implemented by Lesk, Pedersen and Banerjee among others) into existing *jcn* style integrated semantic network measurements - a means of diminishing the negative effect of a sparse semantic network. Experimental results indicate that our technique performs optimally where related components are separated due to a sparse semantic network. In summary, our algorithm returns a *jcn* biased score where the least common subsumer is smaller than the *depth factor*, optimally set at 6. For depth Δ values larger than 6 (closer to the root node, distant from the examined words), our algorithm returns a score which is biased towards the extended gloss overlap technique.

Our results demonstrate that the combination of gloss overlap with an integrated approach such as Jiang and Conrath's algorithm has positive outcomes for the co-occurrence vector based gloss overlap semantic distance scores. In 13 of the 20 word pairs examined, our combined score improved the gloss overlap semantic distance measure. In 5 of the 20 cases examined, our combined score bettered the Jiang and Conrath based distance measure. The Jiang and Conrath improvements are particularly pertinent - the intended effect of our algorithm was to offset the lower *jcn* relevance scores in the word pairs which had an insufficiently specific least common subsumer.

A particularly interesting offshoot of these experiments was found in manually selecting the best scores from the different senses for a word. Within the experimental results displayed above, we always chose the *highest* scores for a given word pair. Choosing the highest score is useful where the word pair is semantically proximal, but it has a negative effect on the scores of unrelated word pairs. Essentially, the concept of the *best score* differs from word to word

- semantically unrelated word pairs such as *noon/string* would be better served by picking the *lowest* semantic distance score instead of the highest. Manually selecting the *best* scores for all the word pairs further increased correlation from the reported 0.682 to an impressive 0.724 - an increase of slightly more than 0.10 over the original Jiang and Conrath scores.

A semantic distance measurement technique such as one proposed within our work has a number of uses - some of which were discussed earlier (see page 1). In the context of our own work, we demonstrate the utility of this technique by its use in a natural language engine built for devices in a smart home environment. Further work on the technique will focus on refinement; so as to reduce the number of processor intensive calculations and thus, more suited for a resource constrained environment.

References

Budanitsky, A., and Hirst, G. 2000. Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures.

Budanitsky, A. 1999. Lexical semantic relatedness and its application in natural language processing. Technical report, University of Toronto.

Fellbaum, C. 1998. *Wordnet - an electronic lexical database*. MIT Press.

Jiang, J., and Conrath, D. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of international conference on research in computational linguistics*.

Lesk, M. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from a ice cream cone. In *Proceedings of SIGDOC, '86*.

Miller, G., and Charles, W. 1991. Contextual correlates of semantic similarity. In *Language and cognitive processes*, volume 6.

Pedersen, T., and Banerjee, S. 2003. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the 18th International joint conference on Artificial Intelligence*, pp. 805–810.

Rubinstein, H., and Goodenough, J. B. 1965. Contextual correlates in synonymy. In *Communications of the ACM, 8(10):627633*.

Sussna, M. 1993. Word sense disambiguation for free-text indexing using a massive semantic network. In *Proceedings of the Second International Conference on Information and Knowledge Management (CIKM-93)*, pp. 67 – 64.

Ted Pedersen, S. P., and Michellizzi, J. 2004. WORD-NET::SIMILARITY, measuring the relatedness of concepts. In *AAAI 2004*.