

# Using WordNet-based Context Vectors to Estimate the Semantic Relatedness of Concepts

**Siddharth Patwardhan**

School of Computing  
University of Utah  
Salt Lake City, UT, 84112, USA  
sidd@cs.utah.edu

**Ted Pedersen**

Department of Computer Science  
University of Minnesota, Duluth  
Duluth, MN, 55812, USA  
tpederse@d.umn.edu

## Abstract

In this paper, we introduce a WordNet-based measure of semantic relatedness by combining the structure and content of WordNet with co-occurrence information derived from raw text. We use the co-occurrence information along with the WordNet definitions to build *gloss vectors* corresponding to each concept in WordNet. Numeric scores of relatedness are assigned to a pair of concepts by measuring the cosine of the angle between their respective gloss vectors. We show that this measure compares favorably to other measures with respect to human judgments of semantic relatedness, and that it performs well when used in a word sense disambiguation algorithm that relies on semantic relatedness. This measure is flexible in that it can make comparisons between any two concepts without regard to their part of speech. In addition, it can be adapted to different domains, since any plain text corpus can be used to derive the co-occurrence information.

## 1 Introduction

Humans are able to quickly judge the relative semantic relatedness of pairs of concepts. For example, most would agree that *feather* is more related to *bird* than it is to *tree*.

This ability to assess the semantic relatedness among concepts is important for Natural Language Understanding. Consider the following sentence: *He swung the bat, hitting the ball into the stands*. A reader likely uses domain knowledge of sports along with the realization that the baseball

senses of *hitting*, *bat*, *ball* and *stands* are all semantically related, in order to determine that the event being described is a baseball game.

Consequently, a number of techniques have been proposed over the years, that attempt to automatically compute the semantic relatedness of concepts to correspond closely with human judgments (Resnik, 1995; Jiang and Conrath, 1997; Lin, 1998; Leacock and Chodorow, 1998). It has also been shown that these techniques prove useful for tasks such as word sense disambiguation (Patwardhan et al., 2003), real-word spelling correction (Budanitsky and Hirst, 2001) and information extraction (Stevenson and Greenwood, 2005), among others.

In this paper we introduce a WordNet-based measure of semantic relatedness inspired by Harris' *Distributional Hypothesis* (Harris, 1985). The distributional hypothesis suggests that words that are similar in meaning tend to occur in similar linguistic contexts. Additionally, numerous studies (Carnine et al., 1984; Miller and Charles, 1991; McDonald and Ramscar, 2001) have shown that context plays a vital role in defining the meanings of words. (Landauer and Dumais, 1997) describe a context vector-based method that simulates learning of word meanings from raw text. (Schütze, 1998) has also shown that vectors built from the contexts of words are useful representations of word meanings.

Our *Gloss Vector* measure of semantic relatedness is based on second order co-occurrence vectors (Schütze, 1998) in combination with the structure and content of WordNet (Fellbaum, 1998), a semantic network of concepts. This measure captures semantic information for concepts from contextual information drawn from corpora of text.

We show that this measure compares favorably

to other measures with respect to human judgments of semantic relatedness, and that it performs well when used in a word sense disambiguation algorithm that relies on semantic relatedness. This measure is flexible in that it can make comparisons between any two concepts without regard to their part of speech. In addition, it is adaptable since any corpora can be used to derive the word vectors.

This paper is organized as follows. We start with a description of second order context vectors in general, and then define the Gloss Vector measure in particular. We present an extensive evaluation of the measure, both with respect to human relatedness judgments and also relative to its performance when used in a word sense disambiguation algorithm based on semantic relatedness. The paper concludes with an analysis of our results, and some discussion of related and future work.

## 2 Second Order Context Vectors

Context vectors are widely used in Information Retrieval and Natural Language Processing. Most often they represent first order co-occurrences, which are simply words that occur near each other in a corpus of text. For example, *police* and *car* are likely first order co-occurrences since they commonly occur together. A first order context vector for a given word would simply indicate all the first order co-occurrences of that word as found in a corpus.

However, our Gloss Vector measure is based on second order co-occurrences (Schütze, 1998). For example, if *car* and *mechanic* are first order co-occurrences, then *mechanic* and *police* would be second order co-occurrences since they are both first order co-occurrences of *car*.

Schütze's method starts by creating a *Word Space*, which is a co-occurrence matrix where each row can be viewed as a first order context vector. Each cell in this matrix represents the frequency with which two words occur near one another in a corpus of text. The Word Space is usually quite large and sparse, since there are many words in the corpus and most of them don't occur near each other. In order to reduce the dimensionality and the amount of noise, non-content stop words such as *the*, *for*, *a*, etc. are excluded from being rows or columns in the Word Space.

Given a Word Space, a context can then be represented by second order co-occurrences (context

vector). This is done by finding the resultant of the first order context vectors corresponding to each of the words in that context. If a word in a context does not have a first order context vector created for it, or if it is a stop word, then it is excluded from the resultant.

For example, suppose we have the following context:

*The paintings were displayed in the art gallery.*

The second order context vector would be the resultant of the first order context vectors for *painting*, *display*, *art*, and *gallery*. The words *were*, *in*, and *the* are excluded from the resultant since we consider them as stop words in this example. Figure 1 shows how the second order context vector might be visualized in a 2-dimensional space.

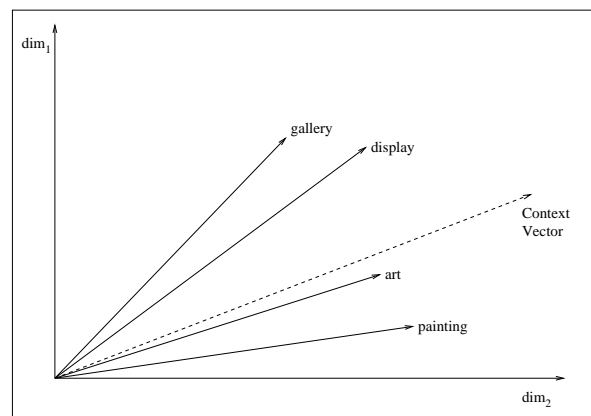


Figure 1: Creating a context vector from word vectors

Intuitively, the orientation of each second order context vector is an indicator of the domains or topics (such as *biology* or *baseball*) that the context is associated with. Two context vectors that lie close together indicate a considerable contextual overlap, which suggests that they are pertaining to the same meaning of the target word.

## 3 Gloss Vectors in Semantic Relatedness

In this research, we create a Gloss Vector for each concept (or word sense) represented in a dictionary. While we use WordNet as our dictionary, the method can apply to other lexical resources.

### 3.1 Creating Vectors from WordNet Glosses

A Gloss Vector is a second order context vector formed by treating the dictionary definition of a

concept as a context, and finding the resultant of the first order context vectors of the words in the definition.

In particular, we define a Word Space by creating first order context vectors for every word  $w$  that is not a stop word and that occurs above a minimum frequency in our corpus. The specific steps are as follows:

1. Initialize the first order context vector to a zero vector  $\vec{w}$ .
2. Find every occurrence of  $w$  in the given corpus.
3. For each occurrence of  $w$ , increment those dimensions of  $\vec{w}$  that correspond to the words from the Word Space and are present within a given number of positions around  $w$  in the corpus.

The first order context vector  $\vec{w}$ , therefore, encodes the co-occurrence information of word  $w$ .

For example, consider the gloss of *lamp* – *an artificial source of visible illumination*. The Gloss Vector for *lamp* would be formed by adding the first order context vectors of *artificial*, *source*, *visible* and *illumination*.

In these experiments, we use WordNet as the corpus of text for deriving first order context vectors. We take the glosses for all of the concepts in WordNet and view that as a large corpus of text. This corpus consists of approximately 1.4 million words, and results in a Word Space of approximately 20,000 dimensions, once low frequency and stop words are removed. We chose the WordNet glosses as a corpus because we felt the glosses were likely to contain content rich terms that would distinguish between the various concepts more distinctly than would text drawn from a more generic corpus. However, in our future work we will experiment with other corpora as the source of first order context vectors, and other dictionaries as the source of glosses.

The first order context vectors as well as the Gloss Vectors usually have a very large number of dimensions (usually tens of thousands) and it is very difficult to visualize this space. Figure 2 attempts to illustrate these vectors in two dimensions. *Tennis* and *food* are the dimensions of this 2-dimensional space. We see that the first order context vector for *serve* is approximately halfway between *tennis* and *food*, since the word *serve*

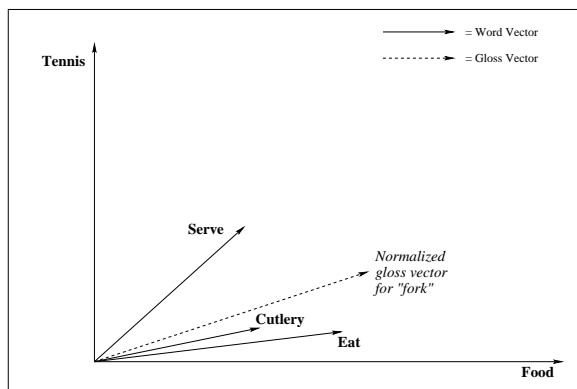


Figure 2: *First Order Context Vectors and a Gloss Vector*

could mean to “serve the ball” in the context of tennis or could mean “to serve food” in another context.

The first order context vectors for *eat* and *cutlery* are very close to *food*, since they do not have a sense that is related to tennis. The gloss for the word *fork*, “cutlery used to serve and eat food”, contains the words *cutlery*, *serve*, *eat* and *food*. The Gloss Vector for *fork* is formed by adding the first order context vectors of *cutlery*, *serve*, *eat* and *food*. Thus, *fork* has a Gloss Vector which is heavily weighted towards *food*. *Food*, therefore, is in the same semantic space as and is related to the concept of *fork*.

Similarly, we expect that in a high dimensional space, the Gloss Vector of *fork* would be heavily weighted towards all concepts that are semantically related to the concept of *fork*. Additionally, the previous demonstration involved a small gloss for representing *fork*. Using augmented glosses, we achieve better representations of concepts to build Gloss Vectors upon.

### 3.2 Augmenting Glosses Using WordNet Relations

The formulation of the Gloss Vector measure described above is independent of the dictionary used and is independent of the corpus used. However, dictionary glosses tend to be rather short, and it is possible that even closely related concepts will be defined using different sets of words. Our belief is that two synonyms that are used in different glosses will tend to have similar Word Vectors (because their co-occurrence behavior should be similar). However, the brevity of dictionary glosses may still make it difficult to measure relatedness

despite this characteristic.

(Banerjee and Pedersen, 2003) experience a similar issue when measuring semantic relatedness by counting the number of matching words between the glosses of two different concepts. They expand the glosses of concepts in WordNet with the glosses of concepts that are directly linked by a WordNet relation. We adopt the same technique here, and use the relations in WordNet to augment glosses for the Gloss Vector measure. We take the gloss of a concept  $c$ , and concatenate to it the glosses of all the concepts to which it is directly related according to WordNet. The Gloss Vector is then created from this big concatenated gloss.

## 4 Other Measures of Relatedness

Below we briefly describe five alternative measures of semantic relatedness, and then go on to include them as points of comparison in our experimental evaluation of the Gloss Vector measure. All of these measures depend in some way upon WordNet. Four of them limit their measurements to nouns located in the WordNet *is-a* hierarchy.

Each of these measures takes two WordNet concepts (i.e., word senses or synsets)  $c_1$  and  $c_2$  as input and return a numeric score that quantifies their degree of relatedness.

(Leacock and Chodorow, 1998) finds the path length between  $c_1$  and  $c_2$  in the *is-a* hierarchy of WordNet. The path length is then scaled by the depth of the hierarchy (D) in which they reside to obtain the relatedness of the two concepts.

(Resnik, 1995) introduced a measure that is based on *information content*, which is a value that indicates the specificity of a concept. These values are derived from corpora, and are used to augment the concepts in the WordNet *is-a* hierarchy. The measure of relatedness between two concepts is the information content of the most specific concept that both concepts have in common (i.e., their lowest common subsumer in the *is-a* hierarchy).

(Jiang and Conrath, 1997) extends Resnik’s measure to combine the information contents of  $c_1$ ,  $c_2$  and their lowest common subsumer.

(Lin, 1998) also extends Resnik’s measure, by taking the ratio of the shared information content to that of the individual concepts.

(Banerjee and Pedersen, 2003) introduce Extended Gloss Overlaps, which is a measure that determines the relatedness of concepts proportional to the extent of overlap of their WordNet glosses. This simple definition is extended to take advantage of the complex network of relations in WordNet, and allows the glosses of concepts to include the glosses of synsets to which they are directly related in WordNet.

## 5 Evaluation

As was done by (Budanitsky and Hirst, 2001), we evaluated the measures of relatedness in two ways. First, they were compared against human judgments of relatedness. Second, they were used in an application that would benefit from the measures. The effectiveness of the particular application was an indirect indicator of the accuracy of the relatedness measure used.

### 5.1 Comparison with Human Judgment

One obvious metric for evaluating a measure of semantic relatedness is its correspondence with the human perception of relatedness. Since semantic relatedness is subjective and depends on the human view of the world, it only makes sense to compare the measures with human judgment. This was done by (Budanitsky and Hirst, 2001) in their comparison of five measures of semantic relatedness. We follow a similar approach in evaluating the Gloss Vector measure.

We use a set of 30 word pairs from a study carried out by (Miller and Charles, 1991). These word pairs are a subset of 65 word pairs used by (Rubenstein and Goodenough, 1965), in a similar study almost 25 years earlier. In this study, human subjects assigned relatedness scores to the selected word pairs. The word pairs selected for this study ranged from highly related pairs to unrelated pairs. We use these human judgments for our evaluation.

Each of the word pairs have been scored by humans on a scale of 0 to 5, where 5 is the most related. The mean of the scores of each pair from all subjects is considered as the “human relatedness score” for that pair. The pairs are then ranked with respect to their scores. The most related pair is the first on the list and the least related pair is at the end of the list. We then have each of the measures of relatedness score the word pairs and a another ranking of the word pairs is created corresponding to each of the measures.

Table 1: Correlation to human perception

Relatedness Measures	M & C	R & G
Gloss Vector	0.91	0.90
Extended Gloss Overlaps	0.81	0.83
Jiang & Conrath	0.73	0.75
Resnik	0.72	0.72
Lin	0.70	0.72
Leacock & Chodorow	0.74	0.77

Spearman’s Correlation Coefficient (Spearman, 1904) is used to assess the closeness of two rankings. If the two rankings are exactly the same, the Spearman’s correlation coefficient between these two rankings is 1. A completely reversed ranking gets a value of  $-1$ . The value is 0 when there is no relation between the rankings.

We determine the correlation coefficient of the ranking of each measure with that of the human relatedness. We use the relatedness scores from both the human studies – the Miller and Charles study as well as the Rubenstein and Goodenough research. Table 1 summarizes the results of our experiment. We observe that the Gloss Vector has the highest correlation with humans in both cases.

Note that in our experiments with the Gloss Vector measure, we have used not only the gloss of the concept but augmented that with the gloss of all the concepts directly related to it according to WordNet. We observed a significant drop in performance when we used just the glosses of the concept alone, showing that the expansion is necessary. In addition, the frequency cutoffs used to construct the Word Space played a critical role. The best setting of the frequency cutoffs removed both low and high frequency words, which eliminates two different sources of noise. Very low frequency words do not occur enough to draw distinctions among different glosses, whereas high frequency words occur in many glosses, and again do not provide useful information to distinguish among glosses.

## 5.2 Application-based Evaluation

An application-oriented comparison of five measures of semantic relatedness was presented in (Budanitsky and Hirst, 2001). In that study they evaluate five WordNet-based measures of semantic relatedness with respect to their performance in context sensitive spelling correction.

We present the results of an application-oriented

Table 2: WSD on SENSEVAL-2 (nouns)

Measure	Nouns
Jiang & Conrath	0.45
Extended Gloss Overlaps	0.44
Gloss Vector	0.41
Lin	0.36
Resnik	0.30
Leacock & Chodorow	0.30

evaluation of the measures of semantic relatedness. Each of the seven measures of semantic relatedness was used in a word sense disambiguation algorithm described by Banerjee and Pedersen.

Word sense disambiguation is the task of determining the meaning (from multiple possibilities) of a word in a given context. For example, in the sentence *The ex-cons broke into the bank on Elm street*, the word *bank* has the “financial institution” sense as opposed to the “edge of a river” sense.

Banerjee and Pedersen attempt to perform this task measuring the relatedness of the senses of the target word to those of the words in its context. The sense of the target word that is most related to its context is selected as the implied sense of the target word.

The experimental data used for this evaluation is the SENSEVAL-2 test data. This consists of 4,328 instances (or contexts) that include a single ambiguous target word. Each instance consists of approximately 2-3 sentences and one occurrence of a target word. 1,754 of the instances include nouns as target words, while 1,806 are verbs and 768 are adjectives. We use the noun data to compare all six of the measures, since four of the measures are limited to nouns as input. The accuracy of disambiguation when performed using each of the measures for nouns is shown in Table 2.

## 6 Gloss Vector Tuning

As discussed in earlier sections, the Gloss Vector measure builds a word space consisting of first order context vectors corresponding to every word in a corpus. Gloss vectors are the resultant of a number of first order context vectors. All of these vectors encode semantic information about the word or the gloss that the vectors represent.

We note that the quality of the words used as the dimensions of these vectors plays a big part in getting accurate relatedness scores. We find that

words corresponding to very specific concepts and are highly indicative of a few topics, make good dimensions. Words that are very general in nature and that appear all over the place add noise to the vectors.

In an earlier section we discussed using stop words and frequency cutoffs to keep only the high “information content” words. In addition to those, we also experimented with a *term frequency · inverse document frequency* cutoff.

*Term frequency* and *inverse document frequency* are commonly used metrics in information retrieval. For a given word, term frequency (*tf*) is the number of times a word appears in the corpus. The *document frequency* is number of documents in which the word occurs. Inverse document frequency (*idf*) is then computed as

$$idf = \log \frac{\text{Number of Documents}}{\text{Document Frequency}} \quad (1)$$

The  $tf \cdot idf$  value is an indicator of the specificity of a word. The higher the  $tf \cdot idf$  value, the lower the specificity.

Figure 3 shows a plot of  $tf \cdot idf$  cutoff on the x-axis against the correlation of the Gloss Vector measure with human judgments on the y-axis.

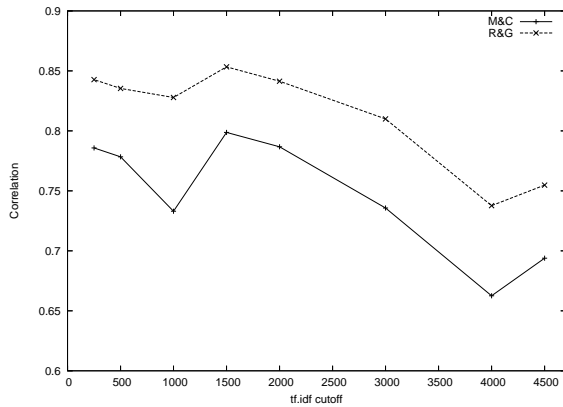


Figure 3: Plot of  $tf \cdot idf$  cutoff vs. correlation

The  $tf \cdot idf$  values ranged from 0 to about 4200. Note that we get lower correlation as the cutoff is raised.

## 7 Analysis

We observe from the experimental results that the Gloss Vector measure corresponds the most with human judgment of relatedness (with a correlation of almost 0.9). We believe this is probably because the Gloss Vector measure most closely imitates the representation of concepts in the human

mind. (Miller and Charles, 1991) suggest that the cognitive representation of a word is an abstraction derived from its contexts (encountered by the person). Their study also suggested the semantic similarity of two words depends on the overlap between their contextual representations. The Gloss Vector measure uses the contexts of the words and creates a vector representation of these. The overlap between these vector representations is used to compute the semantic similarity of concepts.

(Landauer and Dumais, 1997) additionally perform *singular value decomposition* (SVD) on their context vector representation of words and they show that reducing the number of dimensions of the vectors using SVD more accurately simulates learning in humans. We plan to try SVD on the Gloss Vector measure in future work.

In the application-oriented evaluation, the Gloss Vector measure performed relatively well (about 41% accuracy). However, unlike the human study, it did not outperform all the other measures. We think there are two possible explanations for this. First, the word pairs used in the human relatedness study are all nouns, and it is possible that the Gloss Vector measure performs better on nouns than on other parts of speech. In the application-oriented evaluation the measure had to make judgments for all parts of speech. Second, the application itself affects the performance of the measure. The Word Sense Disambiguation algorithm starts by selecting a context of 5 words from around the target word. These context words contain words from all parts of speech. Since the Jiang-Conrath measure assigns relatedness scores only to noun concepts, its behavior would differ from that of the Vector measure which would accept all words and would be affected by the noise introduced from unrelated concepts. Thus the context selection factors into the accuracy obtained. However, for evaluating the measure as being suitable for use in real applications, the Gloss Vector measure proves highly accurate.

The Gloss Vector measure can draw conclusions about any two concepts, irrespective of part-of-speech. The only other measure that can make this same claim is the Extended Gloss Overlaps measure. We would argue that Gloss Vectors present certain advantages over it. The Extended Gloss Overlap measure looks for exact string overlaps to measure relatedness. This “exactness” works against the measure, in that it misses po-

tential matches that intuitively would contribute to the score (For example, *silverware* with *spoon*). The Gloss Vector measure is more robust than the Extended Gloss Overlap measure, in that exact matches are not required to identify relatedness. The Gloss Vector measure attempts to overcome this “exactness” by using vectors that capture the contextual representation of all words. So even though *silverware* and *spoon* do not overlap, their contextual representations would overlap to some extent.

## 8 Related Work

(Wilks et al., 1990) describe a word sense disambiguation algorithm that also uses vectors to determine the intended sense of an ambiguous word. In their approach, they use dictionary definitions from LDOCE (Procter, 1978). The words in these definitions are used to build a co-occurrence matrix, which is very similar to our technique of using the WordNet glosses for our Word Space. They augment their dictionary definitions with similar words, which are determined using the co-occurrence matrix. Each concept in LDOCE is then represented by an aggregate vector created by adding the co-occurrence counts for each of the words in the augmented definition of the concept.

The next step in their algorithm is to form a context vector. The context of the ambiguous word is first augmented using the co-occurrence matrix, just like the definitions. The aggregate of the word vectors of the words in the augmented context forms the context vector. To disambiguate the target word, the context vector is compared to the vectors corresponding to each meaning of the target word in LDOCE, and that meaning is selected whose vector is mathematically closest to that of the context.

Our approach differs from theirs in two primary respects. First, rather than creating an aggregate vector for the context we compare the vector of each meaning of the ambiguous word with the vectors of each of the meanings of the words in the context. This adds another level of indirection in the comparison and attempts to use only the relevant meanings of the context words. Secondly, we use the structure of WordNet to augment the short glosses with other related glosses.

(Niwa and Nitta, 1994) compare dictionary based vectors with co-occurrence based vectors, where the vector of a word is the probability that

an origin word occurs in the context (in a given window of words) of the word. These two representations are evaluated by applying them to real world applications and quantifying the results. Both measures are first applied to *word sense disambiguation* and then to the *learning of positives or negatives*, where it is required to determine whether a word has a positive or negative meaning. It was observed that the co-occurrence based idea works better for the *word sense disambiguation* and the dictionary based approach gives better results for the *learning of positives or negatives*. From this, the conclusion is that the dictionary based vectors contain some different semantic information about the words and warrants further investigation. It is also observed that for the dictionary based vectors, the network of words is almost independent of the dictionary that is used, i.e. any dictionary gives us almost the same network.

(Inkpen and Hirst, 2003) also use gloss-based context vectors in their work on the disambiguation of near-synonyms – words whose senses are almost indistinguishable. They disambiguate near-synonyms in text using various *indicators*, one of which is context-vector-based. Context Vectors are created for the context of the target word and also for the glosses of each sense of the target word. Each gloss is considered as a bag of words, where each word has a corresponding Word Vector. These vectors for the words in a gloss are averaged to get a Context Vector corresponding to the gloss. The distance between the vector corresponding to the text and that corresponding to the gloss was measured (as the cosine of the angle between the vectors). The nearness of the vectors is used as an indicator to pick the correct sense of the target word.

## 9 Conclusion

We introduced a new measure of semantic relatedness based on the idea of creating a Gloss Vector that combines dictionary content with corpus based data. We find that this measure correlates extremely well with the results of these human studies, and this is indeed encouraging. We believe that this is due to the fact that the context vector may be closer to the semantic representation of concepts in humans. This measure can be tailored to particular domains depending on the corpus used to derive the co-occurrence matrices, and

makes no restrictions on the parts of speech of the concept pairs to be compared.

We also demonstrated that the Vector measure is performs relatively well in an application-oriented setup and can be conveniently deployed in a real world application. It can be easily tweaked and modified to work in a restricted domain, such as bio-informatics or medicine, by selecting a specialized corpus to build the vectors.

## 10 Acknowledgments

This research was partially supported by a National Science Foundation Faculty Early CAREER Development Award (#0092784).

All of the experiments in this paper were carried out with the WordNet::Similarity package, which is freely available for download from <http://search.cpan.org/dist/WordNet-Similarity>.

## References

- S. Banerjee and T. Pedersen. 2003. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence (IJCAI-03)*, Acapulco, Mexico, August.
- A. Budanitsky and G. Hirst. 2001. 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, June.
- D. Carnine, E. J. Kameenui, and G. Coyle. 1984. Utilization of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly*, 19:188–204.
- C. Fellbaum, editor. 1998. *WordNet: An electronic lexical database*. MIT Press.
- Z. Harris. 1985. Distributional structure. In J. J. Katz, editor, *The Philosophy of Linguistics*, pages 26–47. Oxford University Press, New York.
- D. Inkpen and G. Hirst. 2003. Automatic sense disambiguation of the near-synonyms in a dictionary entry. In *Proceedings of the 4th Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2003)*, pages 258–267, Mexico City, February.
- J. Jiang and D. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings on International Conference on Research in Computational Linguistics*, Taiwan.
- T. K. Landauer and S. T. Dumais. 1997. A solution to plato’s problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104:211–240.
- C. Leacock and M. Chodorow. 1998. Combining local context and WordNet similarity for word sense identification. In C. Fellbaum, editor, *WordNet: An electronic lexical database*, pages 265–283. MIT Press.
- D. Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of International Conference on Machine Learning*, Madison, Wisconsin, August.
- S. McDonald and M. Ramscar. 2001. Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, Edinburgh, Scotland.
- G.A. Miller and W.G. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28.
- Y. Niwa and Y. Nitta. 1994. Co-occurrence vectors from corpora versus distance vectors from dictionaries. In *Proceedings of the Fifteenth International Conference on Computational Linguistics*, pages 304–309, Kyoto, Japan.
- S. Patwardhan, S. Banerjee, and T. Pedersen. 2003. Using measures of semantic relatedness for word sense disambiguation. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-03)*, Mexico City, Mexico, February.
- P. Procter, editor. 1978. *Longman Dictionary of Contemporary English*. Longman Group Ltd., Essex, UK.
- P. Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, August.
- H. Rubenstein and J.B. Goodenough. 1965. Contextual correlates of synonymy. *Communications of the ACM*, 8:627–633, October.
- H. Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.
- C. Spearman. 1904. Proof and measurement of association between two things. *American Journal of Psychology*, 15:72–101.
- M. Stevenson and M. Greenwood. 2005. A semantic approach to ie pattern induction. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 379–386, Ann Arbor, Michigan, June.
- Y. Wilks, D. Fass, C. Guo, J. McDonald, T. Plate, and B. Slator. 1990. Providing machine tractable dictionary tools. *Machine Translation*, 5:99–154.