A Semantic Relatedness Measure Based on Combined Encyclopedic, Ontological and Collocational Knowledge

Yannis Haralambous

Institut Télécom – Télécom Bretagne
Département Informatique
UMR CNRS 3192 Lab-STICC
Technopôle Brest Iroise
CS 83818, 29238 Brest Cedex 3, France
`yannis.haralambous@telecom-bretagne.eu`

Vitaly Klyuev

University of Aizu
Aizu-Wakamatsu
Fukushima-ken 965-8580, Japan
`vkluev@u-aizu.ac.jp`

Abstract

We describe a new semantic relatedness measure combining the Wikipedia-based Explicit Semantic Analysis measure, the WordNet path measure and the mixed collocation index. Our measure achieves the currently highest results on the WS-353 test: a Spearman ρ coefficient of 0.79 (vs. 0.75 in (Gabrilovich and Markovitch, 2007)) when applying the measure directly, and a value of 0.87 (vs. 0.78 in (Agirre et al., 2009)) when using the prediction of a polynomial SVM classifier trained on our measure.

In the appendix we discuss the adaptation of ESA to 2011 Wikipedia data, as well as various unsuccessful attempts to enhance ESA by filtering at word, sentence, and section level.

1 Introduction

1.1 Semantic Relatedness and Corpora

Semantic relatedness describes the degree to which concepts are associated via any kind of semantic relationship (Scriver, 2006). Its evaluation is a fundamental NLP problem, with applications in word-sense disambiguation, text classification, information retrieval, automatic summarization and many other fields. In recent decades, a great variety of relatedness measures have been defined, based on corpora such as Wikipedia, Wiktionary, WordNet, etc.

Wikipedia is one of the most successful collaborative projects of all time. By a constantly growing number of additions, corrections and verifications, its contents grows in both quantity and quality, and is considered by many linguists as the corpus they had always dreamed of (Medelyan et al., 2008).

By measuring the normalized tfidf values of words in a page, we can consider the page to be a weighted vector in the space of words. Inverting the matrix of these vectors we obtain weighted vectors of words in the space of pages. As every page deals with a single topic, we consider these vectors as being concept vectors. The ESA (Explicit Semantic Analysis) measure between two words is obtained by taking the cosine of their concept vectors (Gabrilovich and Markovitch, 2007).

Unlike Wikipedia, WordNet (Miller, 1995), a semi-formal lexical ontology (Huang et al., 2010), has a fine and carefully-crafted ontological structure: word senses are represented by sets of synonyms ("synsets"), and there is a graph structure on synsets based on hypernymic relations. Several WordNet-based semantic relatedness measures have been defined, based on distances in the hypernymic graph, and often combined with word distribution in sense-tagged corpora.

1.2 Evaluation of Results, WS-353 Test

(Finkelstein et al., 2001) introduce WS-353, a semantic relatedness test set consisting of 353 word pairs[1] and a gold standard defined as the mean value of evaluations by up to 17 human judges. Although this test suite contains some quite controversial word pairs,[2] it has been widely used in literature and has become the de facto standard for semantic relatedness measure evaluation.

Technically, the final result of the test is the Spearman ρ rank correlation coefficient (Spearman, 1904) between the relatedness ranking of pairs by human judges and that by the tested algorithm. So, in fact, it is not the value obtained for each pair that counts, but only the ranks.

1.3 Our Approach

By closely examining word pairs that failed to be ranked correctly by ESA, we came to the conclusion that the WS-353 word pairs belong (non-exclusively) to four classes, corresponding to different kinds of semantic relatedness and requiring different kinds of knowledge:

1. *encyclopedic*: see Section 2;

2. *ontological*: see Section 3;

3. *collocational*: see Section 4;

4. *pragmatic*: see Section 6.

In this paper, we define a new semantic relatedness measure by combining knowledge related to these four classes.

---

[1] Actually 352 pairs, since "money / cash" appears twice.
[2] For example: "Arafat / terror" (0.765), "Arafat / peace" (0.673), "Jerusalem / Israel" (0.846), "Jerusalem / Palestinian" (0.765), etc.

2 Encyclopedic Knowledge

This class contains pairs that are best sorted by ESA. We note that (Agirre et al., 2009) qualify ESA as a distributional approach. Indeed, technically two words are semantically related in ESA if they appear together frequently in Wikipedia pages. But since pages are descriptions of topics (= concepts), words are ESA-close when they appear frequently in common concept descriptions, and therefore in common semantic domains. Hence, ESA is semantically richer than a merely distributional approach.

ESA is the first and most important component of our combined relatedness measure. By adapting our implementation of the ESA algorithm to 2011 Wikipedia data (see App. A), we obtain a Spearman $\rho = 0.7394$. In the following sections we describe the components added to ESA in order to optimize its performance even further.

3 Ontological knowledge

To get a better insight into the shortcomings of ESA on WS-353, we calculate Spearman ρ for the WS-353 set minus a single pair, for every pair. In Fig. 1 one can see the top 40 "most problematic" pairs: those whose removal increases ρ the most. By taking a closer look at them we can get hints for further improvements of the measure.
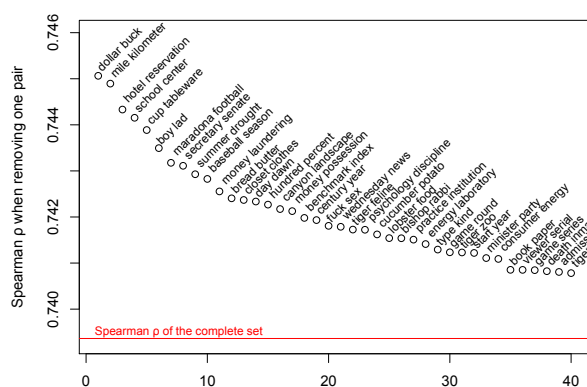

Figure 1: Spearman ρ when removing a single pair from WS-353.

First of all, we see pairs having a relation that is ontological in nature: "tiger / feline" (hyponym), "mile / kilometer" (coordinate terms, or "classmates" (Kuroda et al., 2010b)), "dollar / buck" (synonyms), etc. These relations are strong enough to justify the presence of the pairs in the test set, but do not necessarily imply high frequency of terms in common Wikipedia pages.

A good place for information of an ontological nature is WordNet. There have been several WordNet-based measures defined in the literature. When applying them[3] to the WS-353 test set we get the following ρ:

WNP (Path-based)	0.2873
WUP (Wu and Palmer, 1994)	0.1356
RES (Resnik, 1995)	0.2112
JCN (Jiang and Conrath, 1997)	**0.3172**
LCH (Leacock and Chodorow, 1998)	0.1437
HSO (Hirst and St-Onge, 1998)	0.1598
LIN (Lin, 1998)	0.1987
LESK (Banerjee and Pedersen, 2002)	0.1304

Despite the fact that JCN (which combines WordNet-graph calculations and word frequencies from a corpus[4]) rates best when used alone, the measure which we are going to use is WNP, which gives the best results when combined with ESA (see below). This measure is based exclusively on the shortest-path distance in WordNet and hence is purely ontological. For example, the WNP-measure of "wood / forest" is 1 (synonyms), "bird / cock" is 0.5 (hypernym), "century / year" is 0.33, "bishop / rabbi" is 0.25, etc.

We found that this measure provides bad results in its lower range (since the path length between distant nodes strongly depends on the density of WordNet for each knowledge domain). To understand the behavior of ESA and WNP measures in their low ranges, we progressively remove pairs from WS-353 *in order of increasing relatedness.*
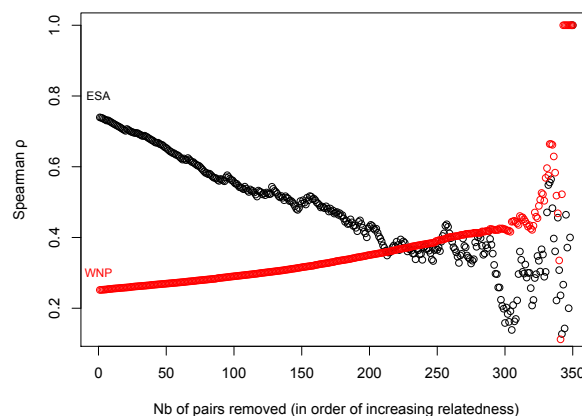

Figure 2: The effect on Spearman ρ of the progressive removal of pairs in order of increasing relatedness, for ESA and WNP measures.

As we can see in Fig. 2, removing pairs in the small-value range of the measure *strongly decreases* ESA (which, after half of the pairs are removed, becomes chaotic), while the same operation steadily *increases* WNP. In other words, small-value pairs are crucial positive contributors for ESA, but rather negative contributors for WNP. For this reason, we use only the upper range of WNP, and ignore its results for low-valued pairs. To achieve a smooth "fade-out" of WNP's lower range we multiply it by a sigmoid logistic function. We

---

[3]In fact, these measures apply to synsets rather than to words. To avoid going through a sense-disambiguation process, we take the optimistic approach of using for each pair

of words, the pair of senses which are the most closely related. Hence, if $\hat{\mu}$ is a synset-measure, s, s' are synsets and w, w' words, we define the induced word-measure μ as $\mu(w, w') := \max_{s \ni w, s' \ni w'} \hat{\mu}(s, s')$.

[4]For the distributional part of Jiang & Conrath, Resnik and Lin, we use the Wikipedia 2011 corpus.

hence define a new measure

$$\mu_{\text{EW}}(w_1, w_2) = \mu_{\text{ESA}}(w_1, w_2) \\ \cdot (1 + \lambda \sigma_{m,s}(\mu_{\text{WNP}}(w_1, w_2))), \quad (1)$$

where λ weights WNP with respect to ESA, m is the sigmoid inflection point (= a soft boundary of WNP's lower range), s is the steepness of the sigmoid (small s makes the central part of the sigmoid closer to a vertical line), and "EW" stands for "ESA and WordNet."

Calculations give the following optimal result:

$\lambda = 4.665, m = 0.26, s = 0.05$	$\rho = \mathbf{0.7779}$

which surpasses the (Gabrilovich and Markovitch, 2007) ESA result of 0.75 by 5.2%. The parameter values have been obtained by gradient descent. In the next section we will further enhance this result by taking collocations into account.

4 Collocational Knowledge

Returning to Fig. 1, we see that many "problematic" pairs are in fact collocations: "baseball / season," "money / laundering", "hundred / percent," etc. We claim that the collocational nature of these word pairs has motivated their inclusion in WS-353. To show this, we calculated the collocation index (defined as $\frac{2\#(w_1 w_2)}{\#(w_1) + \#(w_2)}$) of all WS-353 pairs[5].
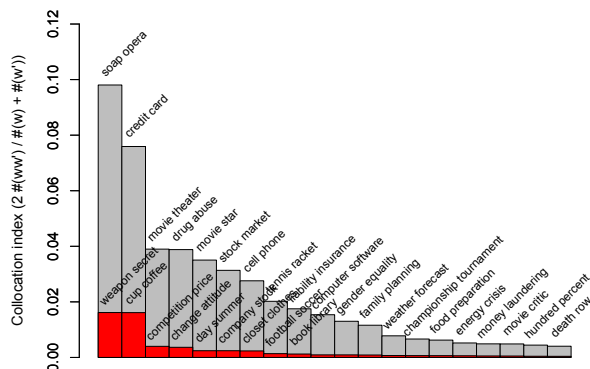
Figure 3: Top twenty direct (in gray) and inverse (in red) collocation indices for WS-353.

The primary goal of WS-353 is to evaluate relatedness measures, and these are symmetric by definition (we always have $\mu(w_1, w_2) = \mu(w_2, w_1)$). If the word pairs were chosen on strictly semantic criteria, and if collocations were purely accidental, then we would have a roughly equal number of pairs (w_1, w_2) where $w_1 w_2$ is a collocation and pairs where $w_2 w_1$ is a collocation.

Fig. 3 shows that this is not the case: for the word pairs concerned, WS-353 developers have almost systematically chosen to write the words in the order in which they form a collocation.

---

[5]We obtained WS-353 pair and word frequencies from the 53.45 billion-word GoogleBooks corpus (Michel et al., 2011). We considered only books published after 1970.

But neither ESA nor WNP recognize collocations: the former because of the bag-of-words principle underlying tfidf, and the latter only in the case where the collocational pair is a concept on its own. Indeed, most of the collocations in Fig. 3 are WordNet concepts (the exceptions being: "gender / equality," "food / preparation," "secret / weapon," "energy / crisis," etc.) but knowledge of that fact is not sufficient for ranking, since there is no mention in WordNet of the strength of the collocational relation.

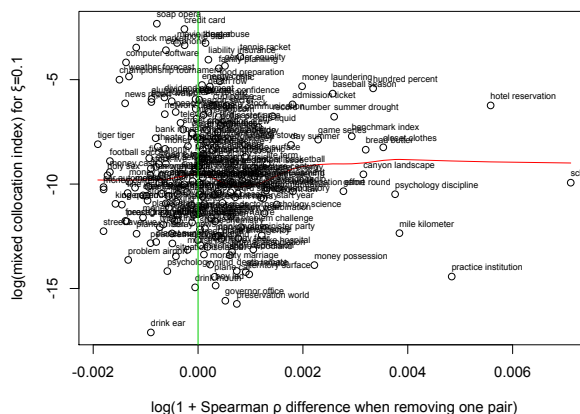We use the collocation index to further enhance our EW relatedness measure.

Figure 4: Collocation index vs. Spearman stability of EW. The red line is LOWESS polynomial regression (Cleveland, 1981).

Note that this index is *not* a measure (for example, the collocation index of "tiger / tiger" is not 1) and cannot be used directly as such.

How do collocational pairs contribute to the WS-353 Spearman ρ value? In Fig. 4 one can compare collocation index and Spearman stability (that is, the effect on ρ of the removal of a single word pair). Pairs located on the green vertical line are those whose removal does not affect Spearman ρ. Those on the right increase ρ when removed. We observe that most collocations are on the right; in other words, they are negative contributors. The most problematic ones are collocations which are not individual WordNet concepts (typical examples: "school / center," "hotel / reservation," "canyon / landscape," etc.).

On the other hand, on the left side we find collocations that contribute positively to ρ: in many cases these have a strong ontological relation ("tiger / tiger," "street / avenue," "football / soccer," etc.) which is probably the main reason for their positive contribution. The LOWESS polynomial regression line is quasi-horizontal, so we cannot infer whether or not collocation index is correlated with ρ.

An auxiliary question is whether collocation index values (at least in the high range) are correlated with the actual values of the WS-353 gold standard. Fig. 5 compares these two quantities. As we can see, LOWESS polynomial regression is almost steadily monotonically

increasing, which shows that, although not a measure per se, (high-range) collocation index could be useful for relatedness measurement.
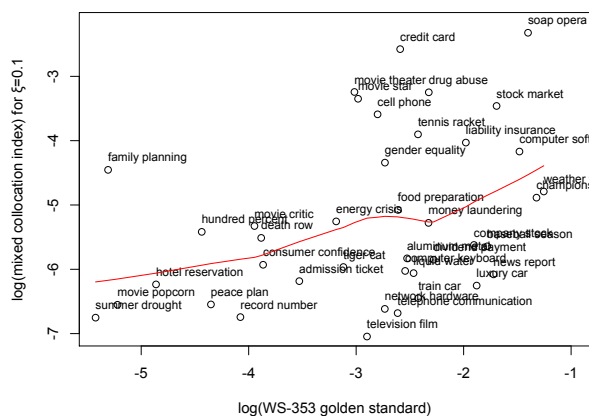

Figure 5: Collocation index vs. WS-353 gold standard. The red line is LOWESS polynomial regression.

We combine the previously defined EW measure with collocation index, by defining measure EWC (= "ESA + WordNet + collocations") as follows:

$$\mu_{\text{EWC}}(w_1, w_2) = \mu_{\text{ESA}}(w_1, w_2)$$
$$\cdot (1 + \lambda \sigma_{m,s}(\mu_{\text{WNP}}(w_1, w_2)))$$
$$\cdot (1 + \lambda' \sigma_{m',s'}(C_\xi(w_1, w_2))),$$

where λ, m, s are as in (1), λ', m' and s' are similar, and the *mixed collocation index* C_ξ is defined as follows:

$$C_\xi(w_1, w_2) = \frac{2\#(w_1 w_2)}{\#(w_1) + \#(w_2)} + \xi \frac{2\#(w_2 w_1)}{\#(w_1) + \#(w_2)}$$

where $\#(.)$ is the frequency in the corpus.

Calculations give the following optimal result:

$\lambda = 5.16, \mu = 0.25, \lambda' = 48.7,$ $\mu' = 0.19, s = s' = 0.05, \xi = 0.55$	$\rho = \mathbf{0.7874}$

which is 1.2% higher than EW and, to the best of our knowledge, currently the highest result for WS-353 by a direct measure (not using a support vector machine). The parameter values have been obtained by gradient descent.

We can interpret this result as follows: *the EWC measure works best when the lower fourth of WordNet measure and the lower fifth of collocation index values are ignored, and when inverse collocations count half as much as direct ones.*

5 Supervised Approach Using an SVM

(Agirre et al., 2009, p. 25) train an SVM on pairs of WS-353 pairs; this allows them to get an insight on performance increase obtained by combining various measures. By combining knowledge from a Web corpus and from WordNet, they obtain a highest value of Spearman $\rho = 0.78$. We calculated predictions of (4th degree polynomial) SVMs based on our EW and EWC

measures, and obtained the following results, using 10-fold cross-validation:

Measure	Result
EW (ESA + WNP)	$\rho = 0.7996$
EWC (ESA + WNP + collocations)	$\rho = \mathbf{0.8654}$

We observe that even without collocations we already get a better value than (Agirre et al., 2009), and also that the collocation component increases this value significantly, hence validating our choice of using collocational knowledge to enhance semantic relatedness measurement.

6 Pragmatic Knowledge

This class contains pairs not captured by the previous methods. The typical example is "hotel / reservation": its ESA value is very low, there is no ontological relation, and the collocation index is quite low as well. To capture the relatedness of such a pair, we need specific knowledge domain ontologies, providing relations such as "A is part of a functional process of system B" (in this case: "a 'reservation' is part of the process of renting a room in a 'hotel' "). We leave this as an open task for future development.

7 Conclusion

By combining two pre-existing semantic relatedness measures and by adding a component based on frequency of collocations, we have obtained a new measure that surpasses the one given in (Agirre et al., 2009) by 11% (when comparing results obtained by SVMs). We conjecture that this measure can further be enhanced by using pragmatic knowledge taken, for example, from specialized domain ontologies.

Appendices

A Adapting ESA to 2011 Wikipedia

The original (and unreleased) C++ ESA implementation (Gabrilovich and Markovitch, 2007) is based on 2005 Wikipedia data (2.2 GB) and achieves a Spearman $\rho = 0.75$. A later implementation in Python and Java (Çallı, 2010), based on the same corpus, achieves $\rho = 0.74$. We implemented ESA in Perl and similarly obtained $\rho = 0.7404$ when based on 2005 data. The same algorithms applied to 2011 data (31 GB), produced a disappointing $\rho = 0.7047$. Indeed, between 2005 and 2011, Wikipedia has evolved as follows:

	2005	2011
#concepts	866,881	4,178,454
#terms/concept	96.1971	97.4243

where by "concepts" we mean Wikipedia pages in the main namespace, and by "terms," distinct stemmed words.

Following advice by Gabrilovich (personal communication), we increased the generality of concepts by filtering Wikipedia pages by two criteria: minimum number of terms, and minimum number of in- and outgoing links. The original values were: 100 terms and 5 links; by requiring a minimum of 200 terms and 14

links, we have attained the 2005 ρ value (more precisely: $\rho = \mathbf{0.7394}$). Fig. 6 displays ρ as a function of our two criteria.
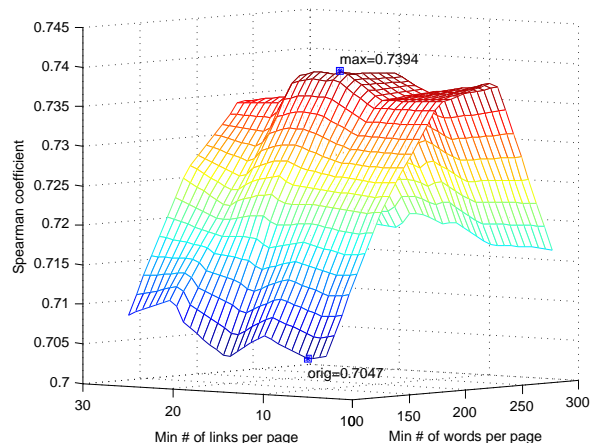

Figure 6: Adapting ESA to 2011 Wikipedia data by increasing the minimum number of distinct (stemmed) words and of in- and outgoing links per page.

In the following table, the column 2011 displays the results with original ESA setting, 2011* the ones with modified settings, $\overline{\mathrm{df}}$ is mean document frequency of terms and term density is $\frac{\overline{\mathrm{df}}}{\#\mathrm{concepts}}$:

	2005	2011	2011*
#concepts	132,689	311,209	155,767
#terms/concept	165	279	414
#terms	187,971	503,368	408,299
$\overline{\mathrm{df}}$	116.3307	173.7199	159.0395
term density	0.00088	0.00056	0.00102

As we see, terms are less densely distributed in the 2011 corpus, since the increase of their mean document frequency, though important, is overruled by an even more important increase in the number of concepts. By more efficiently pruning concepts and leaving $\overline{\mathrm{df}}$ relatively stable, we manage to increase term density anew and hence, enhance performance.

B Experiments

(Giraud-Carrier and Dunham, 2010) emphasize the importance of sharing negative results. Responding to their call, here are some of our failed attempts at increasing ESA performance on the 2005 corpus. Note that the standard ESA value we challenge is $\rho = 0.7404$.

B.1 At the Word Level: Lemmatization and POS Filtering

ESA removes stop words and words with fewer than three letters before applying the Porter stemmer thrice. Instead of stemming, we lemmatized and then applied two strategies: keeping only nouns and proper names (Penn tags NN, NNP, and plurals), or also verbs and adjectives (tags starting with NN, NNP, VB, and JJ). Here are the results obtained:

Penn tags NN, NNS, NNP, NNPS	$\rho = 0.7194$
Penn tags NN*, NNP*, VB*, JJ*	$\rho = 0.7178$

The performance loss is due to lemmatization, proving once again that while Porter stemming may seem a brutal technique, it works better than anything else. Note that, surprisingly, when adding verbs and adjectives we get a (slightly) *smaller* ρ.

B.2 Filtering at the Sentence Level

We attempted to triple the weight of sentences containing either the page title, or one of the (non stop-)words of the page title, or one of the anchors pointing to the page. This operation affected 1,399,165 sentences. Here are the results obtained:

Tripling weight of selected sentences	$\rho = 0.7293$

B.3 Filtering at the Section Level

The idea is to avoid "historical sections" in pages describing current notions or objects. Historical sections are detected by a higher frequency of past-tense verbs, unless of course the whole page is of a historical nature, and hence using primarily the past tense. Let $\pi = \frac{\#\text{ past-tense verbs}}{\#\text{ verbs}}$ for each Wikipedia page. We pruned sections of $\pi \geq 0.8$ when the page had $\pi < 0.8$. We also pruned sections named "History," "External links," "References," "See also," "Further reading," and "Bibliography." This affected 111,028 sections out of 470,948. Here are the results obtained:

Pruning of "historical" and other sections	$\rho = 0.6608$

C Implementation Details

Implementation of ESA was done from scratch in Lex and Perl. To access WordNet v3, we used the Perl module `WordNet::Similarity` (Pedersen et al., 2004). SVM calculations as well as 2D figures were done in R, and the 3D figure in Matlab. For lemmatizing and POS-tagging, we used TreeTagger (Schmid, 1994). Our code is publicly available at `http://omega2.enstb.org/yannis/similarity.php`.

D Acknowledgments

References

Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Paşca, and Aitor Soroa. 2009. A study on similarity and relatedness using distributional and WordNet-based approaches. In *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*, pages 19–27.

Eneko Agirre, Montse Cuadros, German Rigau, and Aitor Soroa. 2010. Exploring knowledge bases for similarity. In *Proceedings of IJCAI*, pages 373–377.

Satanjeev Banerjee and Ted Pedersen. 2002. An adapted Lesk algorithm for word sense disambiguation using WordNet. *Springer Lecture Notes in Computer Science*, 2276:136–145.

Alexander Budanitsky and Graeme Hirst. 2006. Evaluating WordNet-based measures of lexical semantic relatedness. *Comput. Linguist.*, 32:13–47, March.

Çağatay Çallı. 2010. Improving search result clustering by integrating semantic information from Wikipedia. Master's thesis, Middle East Technical University, Ankara.

W.S. Cleveland. 1981. LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, 35:54.

Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. 2001. Placing search in context: the concept revisited. In *Tenth International World Wide Web Conference (WWW10), Hong Kong*, pages 406–414.

Evgeniy Gabrilovich and Shaul Markovitch. 2007. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In *IJCAI'07: Proceedings of the 20th International Joint Conference on Artificial Intelligence*.

Evgeniy Gabrilovich and Shaul Markovitch. 2009. Wikipedia-based semantic interpretation for natural language processing. *Journal of Artificial Intelligence Research*, 34(1):443–498.

Christophe Giraud-Carrier and Margaret H. Dunham. 2010. On the importance of sharing negative results. *SIGKDD explorations*, 12(2):3–4.

Graeme Hirst and David St-Onge. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. In Christiane Fellbaum, editor, *WordNet: An electronic lexical database*, pages 305–332. The MIT Press.

Cha-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, and Laurent Prévot, editors. 2010. *Ontology and the lexicon*. Studies in Natural Language Processing. Cambridge.

Jay J. Jiang and David W. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings on International Conference on Research in Computational Linguistics, Taiwan 1997*.

Kow Kuroda, Francis Bond, and Kentaro Torisawa. 2010a. Why Wikipedia needs to make friends with WordNet. In *The 5th International Conference of the Global WordNet Association (GWC-2010)*.

Kow Kuroda, Jun'ichi Kazama, and Kentaro Torisawa. 2010b. A look inside the distributionally similar terms. In *Proceedings of the Second Workshop on NLP Challenges in the Information Explosion Era (NLPIX 2010)*, pages 40–49.

Claudia Leacock and Martin Chodorow. 1998. Combining local context and WordNet similarity for word sense identification. In Christiane Fellbaum, editor, *WordNet: An electronic lexical database*, pages 265–283. The MIT Press.

Dekang Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of 15th International Conference On Machine Learning, Madison WI, 1998*.

Olena Medelyan, Catherine Legg, David Milne, and Ian H. Witten. 2008. Mining meaning from Wikipedia. Technical report, Department of Computer Science, University of Waikato, New Zealand.

Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.

George A. Miller. 1995. WordNet: a lexical database for English. *Commun. ACM*, 38:39–41, November.

Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. 2004. WordNet::Similarity - Measuring the relatedness of concepts. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04)*, pages 1024–1025.

Simone Paolo Ponzetto and Michael Strube. 2006. Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*, pages 192–199.

Simone Paolo Ponzetto and Michael Strube. 2007. Knowledge derived from Wikipedia for computing semantic relatedness. *Journal of Artificial Intelligence Research*, 30:181–212.

R Development Core Team, 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Kira Radinsky, Eugene Agichtein, Evgeniy Gabrilovich, and Shaul Markovitch. 2011. A word at a time: Computing word relatedness using temporal semantic analysis. In *WWW 2011*.

Daniel Ramage, Anna N. Rafferty, and Christopher D. Manning. 2009. Random walks for text semantic similarity. In *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*, TextGraphs-4, pages 23–31, Stroudsburg, PA, USA. ACL.

Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montréal*, pages 448–453.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

Aaron D. Scriver. 2006. Semantic distance in WordNet: A simplified and improved measure of semantic relatedness. Master's thesis, University of Waterloo.

Charles Spearman. 1904. The proof and measurement of association between two things. *Amer. J. Psychol.*, 15:72–101.

Zhi-Biao Wu and Martha Palmer. 1994. Verb semantics and lexical selection. In *Proceedings of the 32nd Annual Meetings of the Association for Computational Linguistics*, pages 133–138.

Torsten Zesch, Iryna Gurevych, and Max Mühlhäuser. 2007. Analyzing and accessing Wikipedia as a lexical semantic resource. Preprint of the Technische Universität Darmstadt.

Torsten Zesch, Christof Müller, and Iryna Gurevych. 2008. Extracting lexical semantic knowledge from Wikipedia and Wiktionary. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*.