An Intrinsic Information Content Metric for Semantic Similarity in WordNet

Nuno Seco[1] and **Tony Veale**[1] and **Jer Hayes**[1]

Abstract. Information Content (IC) is an important dimension of word knowledge when assessing the similarity of two terms or word senses. The conventional way of measuring the IC of word senses is to combine knowledge of their hierarchical structure from an ontology like WordNet with statistics on their actual usage in text as derived from a large corpus (e.g., [16]). In this paper we present a wholly intrinsic measure of IC that relies on hierarchical structure alone. We report that this measure is consequently easier to calculate, yet when used as the basis of a similarity mechanism it yields judgments that correlate more closely with human assessments than other, extrinsic measures of IC that additionally employ corpus analysis. We report a resulting correlation value of 0.84 between human and machine similarity judgments on the dataset of Miller and Charles [12], which is suggestively close to the upper-bound of 0.88 postulated by Resnik in [17].

1 Introduction

Semantic similarity (SS) has for a long time been a subject of intense scholarship in the fields of Artificial Intelligence, Psychology and Cognitive Science. Computational models trying to imitate this human ability date back to Quillian [14] and the spreading activation algorithm.

Nowadays, these computational models of similarity are being included in many software applications with the intent of making these seem more intelligent or even creative (see [3]). The use of SS has also found its way into the Bio-Informatics domain. Recently, Lord [9] studied the effect of using SS strategies when querying DNA and protein sequence databases. As these databases have acquired enormous amounts of textual annotations, Lord suggests the use of a SS search tool, capable of exploiting these annotations, has the potential to be a valuable addition to the armory of the bio-researcher.

A dominant trend in Natural Language Processing (NLP) has been to gather statistical data from corpora and to reason about some particular task in the light of such data, consider the use of Hidden Markov Models or Bayesian Networks for Part of Speech Tagging (e.g. [13], [19]). Some NLP systems use a hybrid approach where both statistics and a hand-crafted lexical Knowledge Base, such as WordNet, is used (e.g. [15]). SS has been no exception to this trend. Despite this movement, we feel that these knowledge bases have not yet been fully exploited, and that there is still much reasoning potential to be discovered. Hence, we present a novel metric of IC that is completely derived from WordNet without the need for external resources from which statistical data is gathered. Experimentation will show that this new metric delivers better results when we substitute our IC values with the corpus derived ones in previously established formulations of SS. These formulations, that make use of IC values, are generally known as Information Theoretic formulas, thus our main focus throughout the paper shall be on these. Nevertheless, when analyzing our results we consider alternative approaches in order to exhaustively evaluate our metric.

This paper is organized in the following manner; in section 2, we will provide a brief overview of some of the approaches that we believe are increasingly relevant to our research and that base themselves on the notion of IC and thus, are considered to be Information Theoretic approaches. The following section describes our method of deriving IC values for existing concepts in WordNet [11] along with the assumptions made and its formal definition. Section 4 presents the experimental setup and a discussion of the results obtained evaluating our metric. Comments regarding experimentation and their results will conclude this paper along with some ideas for future work.

2 Information Theoretic Approaches

Previous information theoretic approaches ([5], [16] and [8]) obtain the needed IC values by statistically analyzing corpora. They associate probabilities to each concept in the taxonomy based on word occurrences in a given corpus. These probabilities are cumulative as we go up the taxonomy from specific concepts to more abstract concepts. This means that every occurrence of a noun in the corpus is also counted as an occurrence of each taxonomic class containing it. The IC value is then obtained by considering the negative log likelihood:

$$ic_{res}(c) = -log\, p(c) \tag{1}$$

where c is some concept in WordNet and $p(c)$ is the probability of encountering c in a given corpus. It should be noted that this method ensures that IC is monotonically decreasing as we move from the leaves of the taxonomy to its roots. Philip Resnik [16] was the first to consider the use of this formula, that stems from the work of Shannon [18], for the purpose of SS judgments. The basic intuition behind the use of the negative likelihood is that the more probable a concept is of appearing then the less information it conveys, in other words, infrequent words are more informative then frequent ones. Knowing the IC values for every concept we may then calculate the SS between two given concepts. According to Resnik, SS depends on the amount of information two concepts have in common, this shared information is given by the Most Specific Common Abstraction (MSCA) that subsumes both concepts. In order to find a quantitive value of shared information we must first discover the MSCA, if one does not exist then the two concepts are maximally dissimilar, otherwise the

[1] Department of Computer Science, University College Dublin, Ireland email: {nuno.seco, tony.veale, jer.hayes}@ucd.ie

shared information is equal to the IC value of the MSCA. Formally, semantic similarity is defined as:

$$sim_{res}(c_1, c_2) = \max_{c \in S(c_1,c_2)} ic_{res}(c) \quad (2)$$

where $S(c_1, c_2)$ are the set of concepts that subsume c_1 and c_2.

Another information theoretic similarity metric that used the same notion of IC was that of Lin [8]. His definition of similarity states:

> ”The similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are.”

Formally the above definition may be expressed by:

$$sim_{lin}(c_1, c_2) = \frac{2 \times sim_{res}(c_1, c_2)}{(ic_{res}(c_1) + ic_{res}(c_2))} \quad (3)$$

Jiang and Conrath [5] also continued on in the information theoretic vein and suggested a new measure of semantic distance (if we consider the opposite[2] of the distance we obtain a measure of similarity) that combined the edge-based counting method with IC serving as a decision factor. Their model takes into consideration several other factors such as local density, node depth and link type, but for the purpose of this paper we will only consider the case[3] where node depth is ignored and link type and local density both have a weight of 1. In this special case, the similarity metric is:

$$dist_{jcn}(c_1, c_2) = (ic_{res}(c_1) + ic_{res}(c_2)) - 2 \times sim_{res}(c_1, c_2) \quad (4)$$

Both Lin’s and Jiang’s formulation correct a problem existent with Resnik’s similarity metric; if one were to calculate $sim_{res}(c_1, c_1)$ one would not obtain the maximal similarity value of 1, but instead the value given by $ic_{res}(c_1)$[4]. This problem is corrected in both subsequent formulations, yielding that $sim_{lin}(c_1, c_1) = 1$ and $dist_{jcn}(c_1, c_1) = 0$.

3 Information Content in WordNet

As was made clear in the previous section, IC is obtained through statistical analysis of copora, from where probabilities of concepts occurring are inferred. Statistical analysis has been receiving must attention and has proved to be very valuable in several Natural Language Processing tasks [10]. We feel that WordNet can also be used as a statistical resource with no need for external ones. Moreover, we argue that the WordNet taxonomy may be innovatively exploited to produce the IC values needed for SS calculations.

Our method of obtaining IC values rests on the assumption that the taxonomic structure of WordNet is organized in a meaningful and structured way, where concepts with many hyponyms convey less information than concepts that are leaves. We argue that the more hyponyms a concept has the less information it expresses, otherwise there would be no need to further differentiate it. Likewise, concepts that are leaf nodes are the most specified in the taxonomy so the information they express is maximal. In other words we express the IC value of a WordNet concept as a function of the hyponyms it has. Formally we have:

$$ic_{wn}(c) = \frac{log(\frac{hypo(c)+1}{max_{wn}})}{log(\frac{1}{max_{wn}})} = 1 - \frac{log(hypo(c) + 1)}{log(max_{wn})} \quad (5)$$

where the function $hypo$ returns the number of hyponyms of a given concept and max_{wn} is a constant that is set to the maximum number of concepts that exist in the taxonomy[5]. The denominator, which is equivalent to the value of the most informative concept, serves as a normalizing factor in that it assures that IC values are in $[0, .., 1]$. The above formulation guarantees that the information content decreases monotonically as we transverse from the leaf nodes to the root nodes as can be observed in figure 1. Moreover, the information content of the imaginary top node of WordNet would yield an information content value of 0.

As result of multiple inheritance in some of WordNet’s concepts, caution must be taken so that each distinct hyponym is considered only once. Consider again the situation in figure 1, the concept *artifact* is an immediate hyponym of *whole* and *object*. Since *whole* is also a hyponym of *object* we must not consider the hyponyms of *artifact* twice when calculating the number of hyponyms of *object*.

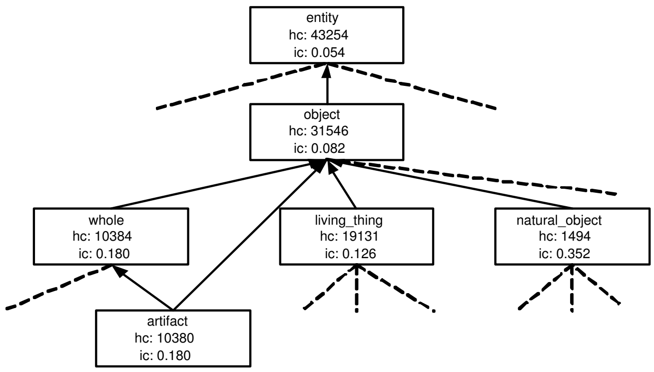

Figure 1. An example of multiple inheritance in the upper taxonomy of WordNet. **ic** and **hc** stand for Information Content and Hyponym Count respectively.

Obviously, this metric gives the same score to all leaf nodes in the taxonomy regardless of their overall depth. As a consequence of this concepts such as *blue_sky* and *mountain_rose* both yield a maximum information content value of 1 despite one being at a two link depth and the other at a nine link depth in the taxonomy, which is in accordance with our initial assumption. However, some counter examples do exist that disagree with the assumption; take the concept *anything* which is a leaf node thus yielding maximum IC. Qualitatively analyzing the amount of information conveyed by this concept may lead us to question the score given by our metric which indeed seems to over exaggerate. But yet another perspective may lead us to ask: ”Why weren’t any nodes considered as hyponyms of *anything*?” Whatever the answer may be, we must recognize that certain commitments had to be made by the designers of WordNet and that

[2] Note that we avoid using the word *inverse* which may be misleading. If one were to simply mathematically inverse the distance this would alter the magnitude of the resulting correlation coefficient. Suppose c_1 and c_2 represent the same concept hence have a semantic distance of 0, consider also that between c_3 and c_4 there is a distance of 1. If one were to consider the mathematical inverse function this would profoundly alter the magnitude of comparison. In the distance scenario we have a difference of 1 between the two pairs $\{c_1, c_2\}$ and $\{c_3, c_4\}$; in the similarity scenario we obtain a difference of infinity between the two.

[3] Which is also the most widely observed configuration in the literature.

[4] Note that the MSCA that subsumes c_1 and c_1 is c_1.

[5] There are 79689 noun concepts in WordNet 2.0.

these may not always match our present needs. Irrespective of this fact, some NLP tasks like Information Retrieval where SS is essential, we will find that words like *anything, nothing , something, ...* which yield exaggerated IC scores are frequently stored in *stop word lists* and are ignored, which will somewhat attenuate these apparent contradictions.

4 Empirical Studies

In order to evaluate our IC metric we decided to use the three formulations of SS presented in section 2 and substituted Resnik's IC metric with the one presented in equation 5. In accordance with previous research, we evaluated the results by correlating our similarity scores with that of human judgments provided by Miller and Charles [12]. In their study, 38 undergraduate subjects were given 30 pairs of nouns and were asked to rate similarity of meaning for each pair on a scale from 0 (no similarity) to 4 (perfect synonymy). The average rating for each pair represents a good estimate of how similar the two words are.

In order to make fair comparisons we decided to use a independent software package that would calculate similarity values using previously established strategies while allowing the use of WordNet 2.0. One freely available package is that of Siddharth Patwardhan and Ted Pederson[6]; which implement semantic relatedness measures described by Leacock Chodorow [7], Jiang Conrath [5], Resnik [16], Lin [8], Hirst St Onge [4], Wu Palmer [21], the adapted gloss overlap measure by Banerjee and Pedersen [1]. Despite our focus being on SS, a special case of Semantic Relatedness, we decided to also evaluate how all of these algorithms would judge the similarity of the 30 pairs of words using WordNet 2.0. In addition to these we also used Latent Semantic Analysis [6] to perform similarity judgments by means of a web interface available at the LSA website[7].

Table 5 presents the similarity obtained with the chosen algorithms and their correlation factor with human judgments. Each of the capital letters heading each column represents a different semantic relatedness algorithm. The columns are organized in following manner:

- A — The data gathered by Miller and Charles Regarding human Judgments.
- B — The results obtained using the independent implementation of the Leacock Chodorow measure.
- C — The results obtained using the independent implementation of the simple edge-counts measure.
- D — The results obtained using the independent implementation of the Hirst St. Onge measure.
- E — The results obtained using the independent implementation of the Jiang Conrath measure.
- F — The results obtained using the independent implementation of the adapted gloss overlap measure.
- G — The results obtained using the independent implementation of the Lin measure.
- H — The results obtained using the independent implementation of the Resnik measure.
- I — The results obtained using the independent implementation of the Wu Palmer measure.
- J — The results obtained using the independent implementation of the LSA measure.
- K — The results obtained using our implementation of the Resnik measure.
- L — The results obtained using our implementation of the Lin measure.
- M — The results obtained using our implementation of the Jiang Conrath measure.

It should be noted that in two of the configurations, namely E and G, two word pairs were not considered in the correlation calculation. This is due to the fact that SemCor, a small portion of the Brown Corpus, was used in obtaining the concept frequencies to calculate the IC values. SemCor is a relatively small sized corpus which contains about 25% of the existing nouns in WordNet. The word *crane* (nor none of its hyponyms) that appear twice in the Miller dataset does not appear in the corpus, thus no information content value may derived for the word. Due to this fact we decided to ignore the entries that would need these values in their assessment and calculated correlation without considering them.

One last observation regarding our implementations must be made before we discuss the results. Resnik's and Lin's measure yield results in $[0, .., 1]$ where 1 is maximum similarity and 0 corresponds to no similarity whatsoever. However, Jiang and Conrath's measure is a measure of semantic distance, in order to maintain the coherency of our implementations we decided to apply a linear transformation on every distance value in order to obtain a similarity value[8]. Yet this transformation will only yield similarity values instead of distance, so normalization was also required in order to constrain the output to values between 0 ... 1. The resulting formulation is:

$$sim_{jcn}(c_1, c_2) = 1 - \left(\frac{ic_{wn}(c_1) + ic_{wn}(c_2) - 2 \times sim_{res'}(c_1, c_2)}{2}\right) \quad (6)$$

Note that $sim_{res'}$ corresponds to Resnik's similarity function but now accommodating our IC values.

4.1 Discussion of Results

Observing table 5 we see that the algorithms performed fairly well. Established algorithms for which there are published results regarding the Miller compilation appear to be the same. The results obtained using our IC values in the information theoretic formulas (K, L and M) seem to have outperformed their homologues (H, G and E), which suggests that the initial assumption concerning the taxonomic structure of WordNet is correct. It should be noted that the maximum value obtained, using Jiang and Conrath's formulation, is very close to what Resnik [17] proposed as a computational upper bound. Reproducing the experiment performed by Jiang and Conrath where they removed the pair *furnace — stove* from their evaluation claiming that MCSA for the pair is not reasonable[9], we obtain a correlation value of 0,87.

5 Conclusion and Future Work

Obviously, the use of such a small dataset does not allow us to be conclusive regarding the true correlation between computational approaches of SS and human judgments of similarity. Nevertheless, when our IC metric is applied in previously established semantic

[6] This software can be downloaded at http://www.d.umn.edu/ tpederse/.
[7] The web interface can be accessed at http://lsa.colorado.edu/.

[8] This transformation will not change the magnitude of the resulting correlation coefficient, although its sign may change from negative to positive [5].
[9] We agree with their claim, in that a more informative subsumer should have been chosen, but we also think that algorithms dealing with manually constructed knowledge bases must be able to deal with these situations as they are inescapable. Fortunately, some research has emerged that looks for these inconsistencies allowing a restructure of the taxonomy ([20], [2]).

similarity formulations, we find a very motivating quislingism. One major advantage of this approach is that it does not rely on corpora analysis, thus we avoid the sparse data problem which was evident in these experiments when judging pairs that contained the word *crane*.

The experiments conducted in this paper only regard nouns that exist in WordNet. We suggest that other classes, that can also be organized in a hierarchal manner, should also be used. For example, if we take the word pair *journey - voyage*, only considering the noun hierarchy we obtain a similarity value of 0,878, but considering the verb hierarchy we obtain a slightly higher value of 0.923 which would strengthen the correlation with human judgments. Although the opposite may also occur, considering the verb taxonomy may weaken correlation with human judgments, we feel that this aspect deserves further attention.

Future research regarding the Information Content metric will make use of taxonomies other than WordNet, such as the Gene Ontology. This will allow us to conclude if our metric generalizes and can be used with other hierarchal knowledge bases.

REFERENCES

[1] Satanjeev Banerjee and Ted Pedersen, 'Extended gloss overlaps as a measure of semantic relatedness', in *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, pp. 805–810, Acapulco, Mexico, (August 2003).

[2] Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider, 'Sweetening ontologies with dolce', in *Proceedings of the European Workshop on Knowledge Acquisition, Modeling, and Management*, (2002).

[3] Paulo Gomes, Nuno Seco, Francisco C. Pereira, Paulo Paiva, Paulo Carreiro, José L. Ferreira, and Carlos Bento, 'The importance of retrieval in creative design analogies', in *Procceedings of the International Joint Conference on Artificial Intelligence IJCAI'03 Workshop: "3rd Workshop on Creative Systems"*, (2003).

[4] Graeme Hirst and David St-Onge, 'Lexical chains as representations of context for the detection and correction of malapropisms', in *WordNet: An Electronic Lexical Database*, ed., Christiane Fellbaum, chapter 13, 305–332, MIT Press, (1998).

[5] J. Jiang and D. Conrath, 'Semantic similarity based on corpus statistics and lexical taxonomy', in *Proceedings of the International Conference on Research in Computational Linguistics*, (1998).

[6] T. K. Landauer, P. W. Foltz, and D. Laham, 'Introduction to latent semantic analysis', *Discourse Processes*, 259–284, (1998).

[7] C. Leacock and M. Chodorow, 'Combining local context and wordnet similarity for word sense identification', in *WordNet: An Electronic Lexical Database*, ed., Christiane Fellbaum, 265–283, MIT Press, (1998).

[8] Dekang Lin, 'An information-theoretic definition of similarity', in *Proceedings of the 15th International Conf. on Machine Learning*, pp. 296–304. Morgan Kaufmann, San Francisco, CA, (1998).

[9] P.W. Lord, R.D. Stevens, A. Brass, and C.A. Goble, 'Semantic similarity measures as tools for exploring the gene ontology', in *Procceedings of the 8th Pacific Symposium on Biocomputing*, pp. 601–612, (2003).

[10] Christopher D. Manning and Hinrich Schütze, *Foundations of statistical natural language processing*, MIT Press, 1999.

[11] George Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller, 'Introduction to wordnet: an on-line lexical database.', *International Journal of Lexicography*, **3**(4), 235 – 244, (1990).

[12] George Miller and W.G. Charles, 'Contextual correlates of semantic similarity', *Language and Cognitive Processes*, **6**, 1–28, (1991).

[13] Leonid Peshkin, Avi Pfefer, and Virginia Savova, 'Bayesian nets in syntactic categorization of novel words', in *Procceedings of the NAACL-HLT'2003*, (2003).

[14] M. Quillian, *Semantic Memory*, 227–270, MIT Press, 1968.

[15] Ganesh Ramakrishnan, B. P. Prithviraj, A. Deepa, Pushpak Bhattacharya, and Soumen Chakrabarti, 'Soft Word Sense Disambiguation', in *Proceedings of the Second International WordNet Conference*, pp. 291–298.

[16] Philip Resnik, 'Using information content to evaluate semantic similarity in a taxonomy', in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp. 448–453, (1995).

[17] Philip Resnik, 'Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language', *Journal of Artificial Intelligence Research*, **11**, 95–130, (1999).

[18] C.E. Shannon, 'A mathematical theory of communication', *Bell System Technical Journal*, **27**, 379–423 and 623–656, (July and October 1948).

[19] Scott M. Thede and Mary P. Harper, 'A second-order hidden markov model for part-of-speech tagging', in *Proceedings of the 37th conference on Association for Computational Linguistics*, pp. 175–182. Association for Computational Linguistics, (1999).

[20] Tony Veale, 'The analogical thesaurus: An emerging application at the juncture of lexical metaphor and information retrieval', in *Proceedings of the International Conference on Innovative Applications of Artificial Intelligence*, (2003).

[21] Z. Wu and M. Palmer, 'Verb semantics and lexical selection', in *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94), Las Cruces, NM, 133-138*, (1994).

Algorithm		A	B	C	D	E	F	G	H	I	J	K	L	M
car	automobile	3,92	3,47	1,00	16,00	0,00	9577,00	1,00	6,11	0,89	0,60	0,68	1,00	1,00
gem	jewel	3,84	3,47	1,00	16,00	0,00	2297,00	1,00	10,52	0,86	0,21	1,00	1,00	1,00
journey	voyage	3,84	2,77	0,50	4,00	4,95	192,00	0,69	5,82	0,92	0,43	0,66	0,84	0,88
boy	lad	3,76	2,77	0,50	5,00	3,41	154,00	0,82	7,57	0,80	0,43	0,76	0,86	0,88
coast	shore	3,70	2,77	0,50	4,00	0,62	336,00	0,97	8,93	0,91	0,40	0,78	0,98	0,99
asylum	madhouse	3,61	2,77	0,50	4,00	0,41	104,00	0,98	11,50	0,82	0,12	0,94	0,97	0,97
magician	wizard	3,50	3,47	1,00	16,00	0,00	976,00	1,00	11,91	0,80	0,29	0,80	1,00	1,00
midday	noon	3,42	3,47	1,00	16,00	0,00	152,00	1,00	10,40	0,88	0,59	1,00	1,00	1,00
furnace	stove	3,11	1,39	0,13	5,00	18,13	202,00	0,220	2,56	0,46	0,28	0,18	0,23	0,39
food	fruit	3,08	1,39	0,13	0,00	11,65	128,00	0,13	0,86	0,22	0,39	0,05	0,13	0,63
bird	cock	3,05	2,77	0,50	6,00	3,76	200,00	0,80	7,74	0,94	0,38	0,40	0,60	0,73
bird	crane	2,97	2,08	0,25	5,00	*	102,00	*	7,74	0,84	0,31	0,40	0,60	0,73
tool	implement	2,95	2,77	0,50	4,00	1,23	542,00	0,92	7,10	0,91	0,13	0,42	0,93	0,97
brother	monk	2,82	2,77	0,50	4,00	14,90	503,00	0,25	10,99	0,92	0,03	0,18	0,22	0,33
crane	implement	1,68	1,86	0,20	3,00	*	51,00	*	3,74	0,67	-0,05	0,24	0,37	0,59
lad	brother	1,66	1,86	0,20	3,00	12,47	28,00	0,29	2,54	0,60	0,24	0,18	0,20	0,28
journey	car	1,16	0,83	0,07	0,00	11,93	158,00	0,00	0,00	0,00	0,10	0,00	0,00	0,00
monk	oracle	1,10	1,39	0,13	0,00	17,42	35,00	0,23	2,54	0,46	0,06	0,18	0,22	0,34
cemetery	woodland	0,95	1,16	0,10	0,00	19,75	21,00	0,08	0,86	0,18	-0,01	0,05	0,06	0,19
food	rooster	0,89	0,83	0,07	0,000	15,19	38,00	0,10	0,86	0,13	0,03	0,05	0,08	0,40
coast	hill	0,87	1,86	0,20	4,00	5,37	123,00	0,71	6,57	0,67	0,05	0,50	0,63	0,71
forest	graveyard	0,84	1,16	0,10	0,00	18,70	25,00	0,08	0,86	0,18	-0,01	0,05	0,06	0,19
shore	woodland	0,63	1,67	0,17	2,00	17,00	78,00	0,14	1,37	0,44	0,14	0,08	0,11	0,30
monk	slave	0,55	1,86	0,20	3,00	15,52	73,00	0,25	2,54	0,60	-0,02	0,18	0,23	0,39
coast	forest	0,42	1,52	0,14	0,00	17,60	89,00	0,13	1,37	0,40	0,14	0,08	0,10	0,29
lad	wizard	0,42	1,86	0,20	3,00	13,60	13,00	0,27	2,54	0,60	0,20	0,18	0,21	0,32
chord	smile	0,13	1,07	0,09	0,00	14,86	31,00	0,27	2,80	0,44	0,05	0,25	0,28	0,35
glass	magician	0,11	1,39	0,13	0,00	18,07	57,00	0,13	2,50	0,36	0,14	0,18	0,20	0,31
noon	string	0,08	0,98	0,08	0,00	18,32	16,00	0,00	0,00	0,00	0,09	0,00	0,00	0,00
rooster	voyage	0,08	0,47	0,05	0,00	21,61	16,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00
Correlation		1,00	0,82	0,77	0,68	-0,81	0,37	0,80	0,77	0,74	0,72	0,77	0,81	0,84

Table 1. Results obtained evaluating correlation with human judgments using several algorithms and WordNet 2.0.