The Paradoxical Role of Similarity in Creative Reasoning

Nuno Seco, Tony Veale and Jer Hayes

Department of Computer Science
University College Dublin
Dublin 4, Ireland

Abstract

In this paper we present a semantic similarity metric that wholly relies on the hierarchical structure of WordNet which makes it amenable as a means of evaluating creativity when considering creative recategorizations of concepts in an Ontology (Veale, 2004). Many creative discoveries are only acknowledged long after their conception due to changes in the evaluation criteria (Bento and Cardoso, 2004), therefore evaluation plays a critical role in creative reasoning systems. We evaluate the similarity function and report a correlation value of 0.84 between human and machine similarity judgments on the dataset of (Miller and Charles, 1991), which is suggestively close to the upper-bound of 0.88 postulated by (Resnik, 1999). We then use the similarity metric as basis for evaluating some examples of creative categorizations. An extension of the metric is also suggested as a means of assessing analogical similarity by looking for analogical cues in the taxonomy.

1. Introduction

Creativity is a vexing phenomenon to pin down formally (Wiggins, 2003), which is perhaps why we tend to think of it in largely metaphoric terms. For example, creativity is often conceived as a form of mental agility that allows gifted individuals to make astonishing mental leaps from one concept to another (Hutton, 1982). Alternately, it is popularly conceived as a form of lateral thinking that allows those who use it to insightfully cut sideways through the hierarchical rigidity of conventional categories (de Bono, 1994). Common to most of these metaphors is the idea that creativity involves recategorization, the ability to meaningfully move a concept from one category to another in a way that unlocks hidden value, perhaps by revealing a new and useful functional property of the concept. For example, psychometric tests such as the Torrance test of creative thinking (Torrance, 1990) try to measure this ability with tasks that, e.g., ask a subject to list as many unusual and interesting uses of old tin cans as possible.

The ad-hoc nature of creativity is such that most ontologies do not and can not provide the kinds of lateral linkages between concepts to allow this kind of inventive recategorization. Instead, ontologies tend to concentrate their representational energies on the hierarchical structures that, from the lateral thinking perspective, are as much a hindrance as an inducement to creativity. This is certainly true of WordNet (Miller et al., 1990), whose **isa** hierarchy is the most richly developed part of its lexical ontology, but it is also true of language independent ontologies like Cyc (Lenat and Guha, 1990), which are rich in non-hierarchical relations but not of the kind that capture deep similarity between superficially different concepts. It is connections like these that most readily fuel the recategorization process.

Withal, (Veale, 2004) has suggested several ways of detecting these lateral linkages in WordNet by exploiting existing polysemies. Polysemy is a form of lexical ambiguity in which a word has multiple related meanings. The form of polysemy that interests us most from a creativity perspective is function-transforming polysemy, which reflects at the lexical level the way concepts can be extended to fulfill new purposes. For instance, English has a variety of words that denote both animals and the meat derived from them (e.g., chicken, lamb, cod), and this polysemy reflects the transformation potential of animals to be used as meat.

(Veale, 2004) further points out that if one can identify all such instances of function-transforming polysemy in WordNet, we can generalize from these a collection of pathways that allow a system to hypothesize creative uses for other concepts that are not so entrenched via polysemy. For example, WordNet defines several senses of *knife*, one as an *edge tool* used for cutting and one as a *weapon* used for injuring. Each sense describes structurally similar objects (sharp flat objects with handles) with a common behavior (cutting) that differ primarily in function (i.e., slicing vs. stabbing). This polysemy suggests a generalization that captures the functional potential of any other *edge tool*, such as *scissors* and *shears*, to also be used as a *weapon*.

Some recategorizations will exhibit more creativity than others, largely because they represent more of a mental leap within the ontology. We can measure this distance using any of a variety of taxonomic metrics, and thus rank the creative outputs of our system. For instance, it is more creative to reuse a *coffee can* as a *percussion instrument* than as a *chamberpot*, since like *tin can* the latter is already taxonomized in WordNet as a *container*. Any similarity metric (called σ, say) that measures the relative distance to the Most Specific Common Abstraction (MSCA) will thus attribute greater similarity to *coffee can* and *chamberpot* than to *coffee can* and *tympan*. This reckoning suggests that the creative distance in a recategorization of a concept c_1 from α to φ may be given by $1 - \sigma(\alpha, \varphi)$.

Of course, distance is not the only component of creativity, as any recategorization must also possess some utility to make it worthwhile (e.g., there is a greater distance still between *tin cans* and *fish gills*, but the former cannot be sensibly reused as the latter). In other words, a creative product must be unfamiliar enough to be innovative but familiar enough to be judged relative to what we know already works. This is the paradox at the heart of ontological creativity: to be creative a recategorization must involve a significant mental leap in function but not in form, yet typically (e.g., in WordNet), both of these qualities are ontologically expressed in the same way, via taxonomic structure. This suggests that the similarity σ must be simultaneously maximized (to preserve structural compatibility) and minimized (to yield a creative leap).

Fortunately, polysemy offers a way to resolve this paradox (Veale, 2004). If a creative leap from α to φ is facilitated by a polysemous link between β and χ where β is a hyponym of α and χ is a hyponym of φ, the sensibility of the recategorization of c_1 can be measured as $\sigma(c_1, \beta)$ while the creativity of the leap can be measured as $1 - (\alpha, \varphi)$. The value of a creative product will be a function of both distance and sensibility, as the former without the latter is unusable, and the latter without the former is banal. The harmonic mean is one way of balancing this dependency on both measures:

$$value(c_1, \varphi) = \frac{2 \times \sigma(c_1, \beta) \times (1 - \sigma(\alpha, \varphi))}{1 + \sigma(c_1, \beta) - \sigma(\alpha, \varphi)} \quad (1)$$

Considering the example of an *ax* being categorized as a *weapon* would lead to the following instantiation:

- c_1 = *ax*
- α = *edge tool*
- β = *knife* (the edge tool sense)
- χ = *knife* (the weapon sense)
- φ = *weapon*

It is precisely the issue of Semantic Similarity (SS) that this paper will address. We present a wholly intrinsic measure of similarity 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 that additionally employ corpus analysis. Given the hierarchical nature of our metric we argue that it is an ideal candidate for the role of σ presented in equation 1.

This paper is organized in the following manner; in section 2. we 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 Information Content (IC) (Resnik, 1995) which is the cornerstone of our metric. These approaches are usually dubbed Information Theoretic, a terminology that we will also employ in the present paper. The following section describes our method of deriving IC values for existing concepts in WordNet (Miller et al., 1990) 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 against human ratings of similarity. When analyzing our results we also consider alternative approaches (i.e. non-information theoretic) in order to exhaustively evaluate our metric. In section 5. we suggest how this similarity metric may be used for evaluating creative recategorizations, possible extensions that may facilitate the assessment of analogical similarity according to the WordNet ontology are given in section 6. Comments regarding our similarity metric will conclude this paper.

2. Information Theoretic Approaches

A recent 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. Some NLP systems use a hybrid approach where both statistics and a hand-crafted lexical Knowledge Base, such as WordNet, is used. 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.

Previous information theoretic approaches ((Jiang and Conrath, 1998), (Resnik, 1995) and (Lin, 1998)) 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:

$$ic_{res}(c) = -log\, p(c) \quad (2)$$

where c is some concept in WordNet and $p(c)$ is its probability according to its frequency in a 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. (Resnik, 1995) was the first to consider the use of this formula, that stems from the work of (Shannon, 1948), 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 allows us to 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 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 shared information is equal to the IC value of the MSCA. Formally, semantic similarity is defined as:

$$sim_{res}(c_1, c_2) = \max_{c \in S(c_1, c_2)} ic_{res}(c) \quad (3)$$

where $S(c_1, c_2)$ is 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, 1998). 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:

$$sim_{lin}(c_1, c_2) = \frac{2 \times sim_{res}(c_1, c_2)}{(ic_{res}(c_1) + ic_{res}(c_2))} \qquad (4)$$

(Jiang and Conrath, 1998) also continued on in the information theoretic vein and suggested a new measure of semantic distance (if we consider the opposite[1] 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[2] where node depth is ignored and link type and local density both have a weight of 1. In this special case, the distance metric is:

$$dist_{jcn}(c_1, c_2) = (ic_{res}(c_1)+ic_{res}(c_2))-2\times sim_{res}(c_1, c_2) \qquad (5)$$

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, but instead the value given by $ic_{res}(c_1)$[3]. This problem is corrected in both subsequent formulations, yielding that $sim_{lin}(c_1, c_1)$ is maximal and $dist_{jcn}(c_1, c_1)$ is minimal.

3. Information Content in WordNet

As was made clear in the previous section, IC is obtained through statistical analysis of corpora, from where probabilities of concepts occurring are inferred. Statistical analysis has been receiving much attention and has proved to be very valuable in several NLP tasks (Manning and Schütze, 1999). 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:

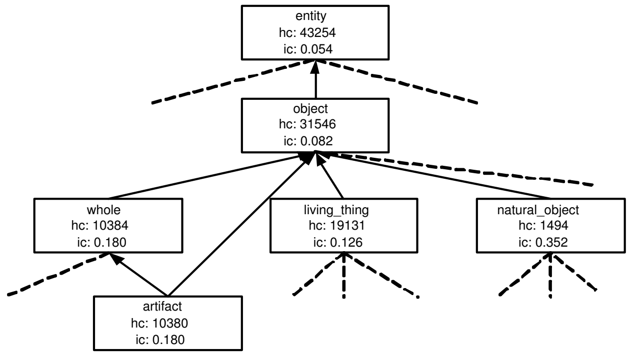

Figure 1: An example of multiple inheritance in the upper taxonomy of WordNet. **ic** and **hc** stand for Information Content and Hyponym Count respectively.

$$ic_{wn}(c) = \frac{log(\frac{hypo(c)+1}{max_{wn}})}{-log(max_{wn})} \qquad (6)$$

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[4]. The denominator, which is equivalent to the value of the most informative concept, serves as normalizing factor in that it assures that IC values are in $[0, .., 1]$. The above formulation guarantees that IC decreases monotonically as we transverse from the leaf nodes to the root nodes as can be observed in figure 1. Moreover, the IC 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*.

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 IC 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

[1]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 w_1 and w_2 represent the same concept hence have a semantic distance of 0, consider also that between w_3 and w_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; in the similarity scenario we obtain a difference of infinity between the two.

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

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

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

that these may not always match our present needs. Irrespective of this fact, in 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 6. In accordance with previous research, we evaluated the results by correlating our similarity scores with that of human judgments provided by (Miller and Charles, 1991). 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[5]; which implement semantic relatedness measures described by (Leacock and Chodorow, 1998), (Jiang and Conrath, 1998), (Resnik, 1995), (Lin, 1998), (Hirst and St-Onge, 1998), (Wu and Palmer, 1994) and the adapted gloss overlap measure by (Banerjee and Pedersen, 2003). 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 (Landauer et al., 1998) to perform similarity judgments by means of a web interface available at the LSA website[6].

Table 4.1. presents the similarity values 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 IC 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. Using Resnik's and Lin's formulas yields 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[7]. Yet this transformation will only yield similarity values instead of distance, so normalization factor was also required in order to constrain the output to values to $[0, .., 1]$. The resulting formulation is:

$$sim_{jcn}(c_1, c_2) = 1 - \left(\frac{ic_{wn}(c_1) + ic_{wn}(c_2) - 2 \times sim_{res'}(c_1, c_2)}{2}\right) \quad (7)$$

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

4.1. Discussion of Results

Observing table 4.1. 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

[5]This software can be downloaded at http://www.d.umn.edu/ tpederse/.

[6]The web interface can be accessed at http://lsa.colorado.edu/.

[7]This transformation will not change the magnitude of the resulting correlation coefficient, although its sign may change from negative to positive (Jiang and Conrath, 1998).

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, 1999) 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[8], we obtain a correlation value of 0,87.

5. Similarity in Creative Recategorization

Considering the high correlation value obtained with configuration M and the hierarchical nature of the metric we believe that it is an ideal candidate to fulfill the role of σ presented in equation 1. As a starting point for the validation of the above hypothesis, we conducted an exploratory experiment in which we generate new recategorizations and then assess their creative value by substituting σ in equation 1 with the SS metric used in configuration M. The recategorizations are generated by a process dubbed **Category Broadening** (Veale, 2004).

As an example of this process imagine we want to broaden the WordNet category *weapon*. The members of this category can be enumerated by recursively visiting every hyponym of the category, which will include *knife*, *gun*, *artillery*, *pike*, etc. But by traversing polysemy links as well as **isa** relations, additional prospective members can be reached and admitted on the basis of their functional potential. Thus, the polysemy of *knife* causes not only *dagger* and *bayonet* but *steak knife* and *scalpel* to be visited. Stretching category boundaries even further, we may generalize that all *edge tools* maybe considered *weapons*, thereby allowing *scissors*, *ax*, *razor* and all other sharp-edged tools to be recognized as having weapon-like potential.

At the heart of the broadnening process is the use of polysemy links. Since WordNet does not contain these links explicitly a patchwork of polysemy detectors are needed. As such we implemented the polysemy detectors presented in (Mihalcea and Moldovan, 2001) and (Veale, 2004) to find the needed facilitating links. The new domain pointers of WordNet 2.0 were also used; basically we consider that if two senses of the same word belong to same domain then they are polysemous. We then applied the broadening process described above to the WordNet 2.0 noun hierarchy and divided the generated recategorizations into 3 groups according to their creative value:

- High — the creative value of the recategorization is in $[0.66, 1]$.
- Medium — the creative value of the recategorization is in $[0.33, 0.66[$.
- Low — the creative value of the recategorization is in $[0, 0.33[$.

Some examples from each of these groups are given in table 2.

6. Analogical Similarity

Analogy is regarded as an important creative reasoning mechanism, as such we feel that extending our metric to deal with analogical similarity is very appealing. Obviously, a simple taxonomic metric will not be able to capture some of the deep similarities of an analogical insight, but taxonomic cues do exist that may shed some light on a potential analogy. As suggested by (Veale, submitted manuscript), WordNet defines *seed* as hyponym of *reproductive structure* and *egg* as a hyponym of *reproductive cell*. Reproduction is thus the unifying theme of the analogy $\{$seed-plant; egg-bird$\}$. The strict taxonomic similarity between *seed* and *egg* is very low yielding a value of 0.37, as their lowest common WordNet hypernym is the root node *entity*. However, if *reproductive structure* and *reproductive cell* are treated as equivalent by considering the average of their IC values as the IC value of a hypothetical analogical pivot we obtain a value of 0.88. We feel this value indicates the analogical similarities between *egg* and *seed*.

7. 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 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*.

Future work will consist of a more thorough evaluation of our metric regarding both its literal facet and also its potential to evaluate creative recategorizations. Another aspect that will also deserve our future attention is the application of our metric to other taxonomic knowledge bases (e.g. Gene Ontology), allowing us to conclude if our intuition about IC is generalizable to other taxonomic resources.

8. References

Banerjee, Satanjeev and Ted Pedersen, 2003. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*. Acapulco, Mexico.

Bento, Carlos and Amilcar Cardoso, 2004. Studying creativity in ai. *Fachbereich Kunstliche Intelligenz der Gesellschaft fur Informatik*:45–46.

de Bono, E., 1994. *Parallel Thinking*. London: Viking Press.

Gangemi, Aldo, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider, 2002. Sweetening ontologies with dolce. In *Proceeding of the European Workshop on Knowledge Acquisition, Modeling, and Management*.

[8]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 ((Veale, 2003), (Gangemi et al., 2002)).

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.

High	Medium	Low
dog collar isa *tie*	*cigar band* isa *necklace*	*dancing* isa *performance*
plane ticket isa *leave of absence*	*smoking room* isa *hiding place*	*coat* isa *plumage*
priest doctor isa *sorcerer*	*scissors* isa *weapon*	*outdoorsman* isa *worker*

Table 2: Some examples of creative recategorizations grouped by their creative value.

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

Hutton, J., 1982. *Aristotle's Poetics*. New York: Norton.

Jiang, J. and D. Conrath, 1998. Semantic similarity based on corpus statistics and lexical taxonomy.

Landauer, T. K., P. W. Foltz, and D. Laham, 1998. Introduction to latent semantic analysis. *Discourse Processes*:259–284.

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

Lenat, Douglas B. and R. V. Guha, 1990. *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*. Reading, Massachusetts: Addison-Wesley.

Lin, Dekang, 1998. An information-theoretic definition of similarity. In *Proc. 15th International Conf. on Machine Learning*. Morgan Kaufmann, San Francisco, CA.

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

Mihalcea, Rada and Dan Moldovan, 2001. Ez.wordnet: Principles for automatic generation of a coarse grained wordnet. In *Proceedings of Flairs 2001*.

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

Miller, George and W.G. Charles, 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6:1–28.

Resnik, Philip, 1995. Using information content to evaluate semantic similarity in a taxonomy. In *IJCAI*.

Resnik, Philip, 1999. 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.

Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656.

Torrance, E. P., 1990. *The Torrance Tests of Creative Thinking*. Illinois: Bensonville.

Veale, Tony, 2003. The analogical thesaurus: An emerging application at the juncture of lexical metaphor and information retrieval. In *Proceedings International Conference on Innovative Applications of Artificial Intelligence*.

Veale, Tony, 2004. Pathways to creativity in lexical ontologies. In *Proceedings of the 2nd Global WordNet Conference*.

Wiggins, Geraint, 2003. Categorizing creative systems. In *Proceedings of 3rd Workshop on Creative Systems*.

Wu, Z. and M. Palmer, 1994. Verb semantics and lexical selection.