

# Web-Based Measure of Semantic Relatedness

Jorge Gracia and Eduardo Mena

IIS Department, Univ. of Zaragoza, María de Luna 1, 50018 Zaragoza, Spain

{jgracia, emena}@unizar.es

<http://sid.cps.unizar.es>

**Abstract.** Semantic relatedness measures quantify the degree in which some words or concepts are related, considering not only similarity but any possible semantic relationship among them. Relatedness computation is of great interest in different areas, such as Natural Language Processing, Information Retrieval, or the Semantic Web. Different methods have been proposed in the past; however, current relatedness measures lack some desirable properties for a new generation of Semantic Web applications: maximum coverage, domain independence, and universality.

In this paper, we explore the use of a semantic relatedness measure between words, that uses the Web as knowledge source. This measure exploits the information about frequencies of use provided by existing search engines. Furthermore, taking this measure as basis, we define a new semantic relatedness measure among ontology terms. The proposed measure fulfils the above mentioned desirable properties to be used on the Semantic Web. We have tested extensively this semantic measure to show that it correlates well with human judgment, and helps solving some particular tasks, as word sense disambiguation or ontology matching.

**Keywords:** semantic web, relatedness, relationship discovering.

## 1 Introduction

Many applications, in Natural Language Processing (NLP) and other fields, benefit from calculating measures to determine numerically how semantically related two words are. Semantic measures can also be defined between lexically expressed word senses, or between whole texts. Three main kind of measures are considered in the literature about this topic: *semantic similarity*, *semantic relatedness* and *semantic distance*. Unfortunately they have not been interpreted always in the same way by different authors. We adopt here the interpretation given in [3].

1. *Semantic similarity*: It is usually defined by considering lexical relations of *synonymy* (e.g.  $\langle car, automobile \rangle$ ) and *hyponymy* (the meaning of a word is encompassed by the another more general term, as in  $\langle car, vehicle \rangle$ ).
2. *Semantic relatedness*: It covers any kind of lexical or functional association, so it is a more general concept than *semantic similarity*. Dissimilar entities may still be related by many possible relationships, such as *meronymy* (or “part of” relation, as in  $\langle finger, hand \rangle$ ), *antonymy* (opposite meanings, as  $\langle hot, cold \rangle$ ), or any kind of functional relationship or frequent association (for example,  $\langle penguin, Antarctica \rangle$ , that are not linked by any lexical relation).

3. *Semantic distance*: It is the inverse of *semantic relatedness*. The more two terms are semantically related, the more semantically close they are.

In this paper we start discussing the interest of using relatedness measures on the Semantic Web, identifying some desirable characteristics that they must accomplish, mainly: domain independence, universality, and maximum coverage. The set of current relatedness measures between words that fulfil these previous requirements is rather limited. Moreover, we have not found measures that explore relatedness between word senses (expressed as *ontological terms*), instead of just between plain words, and accomplish these properties fully.

Next, we choose a well-founded existent semantic distance [5], that uses the Web as knowledge source, to compute a web-based relatedness measure between plain words.

The main contribution of this paper is that, based on the latter, we define a relatedness measure among ontological terms, by exploring their semantic descriptions. This measure is independent of its final purpose, it does not depend on particular lexical resources, it does not need pre-processing tasks, and can operate with any ontology. An evaluation of the web-based word relatedness measure is also presented, by comparison with human judgment, and exploring the effect of using different search engines to access the Web data. Finally, we evaluate the suitability of the proposed relatedness measure among ontological terms when applied to two particular scenarios: ontology matching and word sense disambiguation. Our results show a good behaviour of the measure in all these experiments.

The rest of this paper is organized as follows. Section 2 summarizes the desirable features for a semantic relatedness measure to be used on the Semantic Web. In Section 3 we discuss some related work. Our proposal of semantic relatedness measure appears in Section 4. Experimental results can be found in Section 5, and conclusions and future work appear in Section 6.

## 2 Applicability and Desirable Features

The *Semantic Web* is conceived as an extension of the current Web, “in which information is given well-defined meaning, better enabling computers and people to work in cooperation” [1]. Semantic measures play an important role in many Semantic Web applications. We start mentioning *word sense disambiguation*, which is a fundamental problem not only in NLP but in the Semantic Web as well. Disambiguation techniques try to pick up the most suitable sense of an ambiguous word according to the context. For example, the word *plant* could mean<sup>1</sup> “building for carrying on industrial labor” or “a living organism lacking the power of locomotion”. It is expected that, in a text about car manufacturing, it is used in the first sense, while the second interpretation may be the right one in a web page about vegetal life. Disambiguation methods compare the senses of ambiguous words with words in the context, measuring how related they

---

<sup>1</sup> According to WordNet 2.0 definitions.

are. Many traditional methods have used similarity measures in this task [3], but relatedness is more convenient, because the context that activates the right meaning of an ambiguous word can be related to it by any kind of relationship (not only by similarity).

Another example of applicability is *ontology matching*, the task of determining relations that hold among terms of two different ontologies [8]. An ontology is a specification of a conceptualization [11], which facilitates interoperability among heterogeneous systems by specifying their data semantics. Most ontologies describe classes, instances, and relations (properties) among them (we will refer to these elements as *ontological terms* in the rest of the paper). Similarity measures give good results in discovering equivalences and hierarchical based relationships among ontology terms. Nevertheless, other relationships may remain hidden by using only similarity measures. We consider that relatedness measures can complement the use of similarity to improve ontology matching tasks. For example, it is expected that two entities related by a non-taxonomical relationship (e.g.  $\langle \textit{penguin}, \textit{Antarctica} \rangle$ ) show a low similarity, whereas a high degree of relatedness could reveal that they are linked in some other way.

Relatedness measures can be used in many other applications, such as analysis of structure of texts, annotation, information retrieval, automatic indexing, or spelling correction [3], as well as entity recognition, Semantic Web mining and semantic ranking [13]. According to [20], relationships are in the core of the Semantic Web. Therefore it is of great importance to provide well founded and useful ways to measure relatedness degree; not only among words but, more interestingly, among concepts, expressed as terms in ontologies.

**Desirable features for a relatedness measure.** We summarize some characteristics that, in our opinion, are desirable for a semantic measure to be used in current Semantic Web applications.

1. *Domain independence.* Nowadays, an increasing amount of online ontologies and semantic data is available on the Web, enabling a new generation of semantic applications [16]. If we want to develop that kind of domain independent applications, we have to deal with this increasing heterogeneity, without establishing in advance the ontologies to be accessed.
2. *Universality.* Semantic measures, in the highly dynamic context of the Web, must be flexible and general enough to be used independently of their final purpose, and without relying on specific lexical resources or knowledge representation languages.
3. *Maximum coverage.* We consider that, in the context of web applications with no predefined domain, *maximum coverage* of possible interpretations of the words must be warranted. If we are limited to a particular knowledge source, such as WordNet<sup>2</sup>, or a certain set of ontologies, we are constraining the scope of our applications.

---

<sup>2</sup> <http://wordnet.princeton.edu/>

This latter issue has motivated us to focus on the Web as possible source of knowledge. The Web is an information resource with virtually unlimited potential, where millions of people contribute with billions of web pages. As only a minority of users are domain experts, we assume that the Web is not a high-quality corpus. But, due to its immense size, it is likely that extremes cancel and the average web content still hides meaningful implicit semantics. As a matter of fact, the use of the Web as a corpus is gaining a growing popularity [14].

It is also desirable, for a semantic measure, the ability to establish comparisons not only among *plain words*, but also among *senses of words* which, in a Semantic Web context, will be expressed as *ontological terms*.

Finally, absence of pre-processing tasks are desirable, as well as a portable, flexible and scalable implementation, in order to deal with the highly dynamic and heterogeneous environment of the Web.

### 3 Related Work

Many semantic measures have been proposed in the past to compute degrees of relatedness among words, texts or concepts. Among all possible classifications, we organize the rest of this section according to the source of knowledge utilized.

#### 3.1 Measures Based on Thesauri and Other Lexical Resources

Most of traditional methods to compute semantic measures exploit particular lexical resources: corpus, dictionaries, or well structured taxonomies such as WordNet. Some of them explore path lengths among nodes in taxonomies. Others exploit glosses (textual descriptions of concepts) in dictionaries, while a last group rely on annotated corpora to compute information content [3,22]. These methods result in a limited coverage: for example, as reported in [21], WordNet 2.1 does not include many named entities (e.g. *The Rolling Stones*) or many specialized terms (e.g. *exocytosis*).

Latent Semantic Analysis [6] is a statistical technique that leverages word cooccurrence from large unlabeled corpus of texts. Although it can be successfully used for many purposes, its coverage is still restricted to the corpus used as input, and it needs costly pre-processing tasks.

#### 3.2 Measures Based on Wikipedia

Some recent research efforts has focused on using Wikipedia<sup>3</sup> to improve coverage with respect to traditional thesauri-based methods. Nowadays Wikipedia is rapidly growing in size, and it is not difficult to find new terms and named entities on it. In [21], some classic measures are adapted to use Wikipedia instead of WordNet as knowledge source, showing promising results.

A further step in using Wikipedia is found in [9]. They propose a method to represent the meaning of texts or words as weighted vectors of Wikipedia-based

---

<sup>3</sup> <http://en.wikipedia.org>

concepts, using machine learning techniques. According to their results, they provide even better correlation with human judgment than [21].

Although they have clear benefits, Wikipedia is still not comparable with the whole Web in the task of discovering and evaluation of implicit relationships. For example, at the time of writing this paper<sup>4</sup>, the terms “stomach disease” and “aspirin” do not appear together in any Wikipedia page, but can be found together in 2360 web pages (according to Yahoo!), so their implicit relationship could be inferred by accessing the Web.

### 3.3 Measures Based on the Web

In order to guarantee the maximum coverage, we have focused on methods that exploit the Web<sup>5</sup> as source of knowledge. In [2] they propose a similarity measure that combines various similarity scores based on page counts, with another one based on lexico-syntactic patterns extracted from text snippets. Other web-based semantic similarity measures can be found in [4] and [19]. They also explore text snippets, but without adding page counts as they do in [2]. We agree with all these works in using web content to compute semantic measures. However they are designed to capture similarity instead of the more general relatedness.

OntoNL semantic relatedness measure for OWL ontologies [13] explores semantic relationships by computing path-based conceptual distances, as well as exploring commonalties of two concepts. A remarkable advantage of this method is that it considers relatedness between ontological concepts (instead of between words, as almost all the other methods do). However, it depends on the particular syntax of OWL (thus not fulfilling our universality requirement), and some of their parameters have to be experimentally determined for a given domain ontology. Moreover, it only compares terms belonging to the same ontology.

The Cilibrasi and Vitányi’s Normalized Google Distance [5] (NGD) uses the relative frequency whereupon two terms appear on the Web within the same documents. NGD is well-founded on information distance and Kolmogorov complexity theories, and it does not preclude any kind of relationship between compared words. It fulfils our desirable requirements mentioned in Section 2. However, it does not perform direct comparisons between ontological terms. In Section 4 we show how we use NGD to construct the relatedness measure among ontological terms that we were looking for.

## 4 Web-Based Semantic Relatedness

In this section we propose a transformation of the Normalized Google Distance [5] into a word relatedness measure, generalizing it to any web search engine. Then we describe a method to compute a web-based relatedness degree among ontology terms, by taking into account their ontological context. Finally, we consider a mixed way of relatedness measure, between ontology terms and plain words.

<sup>4</sup> March 2008.

<sup>5</sup> At least, the Web reachable by current web search engines.

## 4.1 Semantic Relatedness among Words

As it was mentioned in Section 3, a semantic distance based on Google page counts is defined in [5]. Actually, as the authors indicate, the discussion about the Google Distance is independent of the particular search engine we use to access the Web. Different search engines use different indexes and retrieval methods, thus providing different results in page counts. This motivated us to try other existent web search engines, in order to compare their behaviour (and to choose the most suitable one eventually).

First of all, we generalized the Cilibrasi and Vitányi’s Normalized Google Distance  $NGD(x, y)$  between search terms  $x$  and  $y$ , by calling *Normalized Web Distance*  $NWD(x, y)$ , the same NGD formula they define in [5], but using any web search engine as source of frequencies.

The smaller the value of  $NWD$ , the greater the semantic relation between words, e.g.  $NWD_{google}(red, blue) = 0.25$ ,  $NWD_{google}(blue, October) = 0.48$ . Although most of  $NWD$  values fall between 0 and 1, it actually ranges from 0 to  $\infty$ . Nevertheless, to obtain a proper relatedness measure, we need a bounded value (in the range  $[0, 1]$ ) that increases inversely to distance. This can be obtained with the following transformation, which defines our proposed *web-based semantic relatedness* measure between two search terms  $x$  and  $y$  as (see [10]):

$$relWeb(x, y) = e^{-2NWD(x, y)} \quad (1)$$

We have considered the following web search engines in our experiments: Google, Yahoo!, Live Search, Altavista, Exalead, Ask, and Clusty<sup>6</sup>. Exalead-based measures show the best correlation with human judgment (see Section 5), closely followed by Yahoo! and Altavista.

Up until this point we have merely “translated” the Cilibrasi and Vitányi’s Google Distance into a relatedness measure between words. The main contribution of this paper starts from the following, where we use it to compute relatedness among ontology terms, and where we evaluate extensively its use with different search engines and for different purposes.

## 4.2 Semantic Relatedness among Ontology Terms

In the following we explain our proposed method to compute relatedness among word senses (expressed as terms from any ontology), instead of only among “plain” words. Initially, previously presented Equation 1 can be applied to *any* pair of search terms indexed by a search engine. Nevertheless, we are interested in providing a measure of how much semantically related a pair of senses (expressed as ontological terms) are. The idea is to exploit some elements of the available semantic description of the ontological terms, and to perform “elementary” comparisons among the words that describe them, by using Equation 1.

<sup>6</sup> <http://www.google.com>, <http://www.yahoo.com>, <http://www.msn.com>, <http://www.altavista.com>, <http://www.exalead.com>, <http://www.ask.com>, and <http://www.clusty.com>, respectively.

In [10] we proposed a first approach to compute a Google-based semantic relatedness measure between ontological terms. We devised a description of an ontological term by means of combining, with logical operators (ANDs, ORs), synonyms and hyperonyms of the ontological term. A search term was then constructed, to be used as input in a search engine, in order to compute semantic relatedness with respect to other terms.

However, this approach was hampered by a technical limitation of most popular search engines: they do not always follow Boolean logic, not giving the expected results in number of hits<sup>7</sup>. Due to this limitation, we have reconsidered the heuristics used to compute semantic relatedness between ontology terms, proposing the general scheme that we detail in the following paragraphs.

**Ontological context.** Given an ontological term  $t$ , we define its *ontological context*, denoted by  $OC(t)$ , as the minimum set of other ontological terms (belonging to its semantic description) which locate the term in the ontology and characterize its meaning. Such an informal definition is better understood with a simple example: in WordNet ontology, the class “Java” (in the sense of “an Indonesian island”), is well characterized, as well as distinguished from other senses, by considering its direct hypernym “Island”. Another example: in an ontology about trips, the property “star” could be well distinguished by specifying the domain it belongs to: “Hotel”.

To define the set of elements (other ontological terms) that can be found in the ontological context of the considered term  $t$ , we follow this strategy:

1.  $t$  is a class  $\Rightarrow OC(t)$  is the set of its direct hypernyms
2.  $t$  is a property  $\Rightarrow OC(t)$  is the set of its domain classes
3.  $t$  is an instance  $\Rightarrow OC(t)$  is the class it belongs to

We do not establish in advance the source where *ontological context* is obtained from. It could be extracted from two different ontologies, or from a run-time integration of many of them (as we do in [23]), for example. In the following we consider any two ontologies as source of *ontological context*, without considering whether they integrate information from other sources (other ontologies, or even lexical resources) or not.

**Relatedness computation.** Let us suppose that  $a$  and  $b$  are two ontological terms (classes, properties or instances) belonging to ontologies  $O_a$  and  $O_b$ , respectively. We consider that, in the search space of the Web, each sense represented by  $a$  and  $b$  can be characterized by taking into account two levels of their semantic description: *Level 0*) the term label and its synonyms, and *Level 1*) its *ontological context*, as it was defined in previous paragraphs. Let us call  $Syn(a) = \{syn_{a_1}, syn_{a_2}, \dots\}$  and  $Syn(b) = \{syn_{b_1}, syn_{b_2}, \dots\}$  the set of synonyms (equivalent labels, including the term label) of ontological terms  $a$  and

<sup>7</sup> For example, they fail the distributive property: the query “*driver AND (car OR train)*” gives 12,700,000 hits in Google (at the time of writing this paper, on March 2008), while the query “*(driver AND car) OR (driver AND train)*” returns 1,140,000.



$b$ , respectively, and  $OC(a) = \{oc_{a_1}, oc_{a_2}, \dots\}$  and  $OC(b) = \{oc_{b_1}, oc_{b_2}, \dots\}$  the terms of their ontological context. Notice that  $|Syn(x)| \geq 1$  and  $|OC(x)| \geq 0$ . Then, comparisons are performed as follows:

$$rel_0(a, b) = \frac{\sum_{i,j} relWeb(syn_{a_i}, syn_{b_j})}{|Syn(a)| \cdot |Syn(b)} \quad \begin{matrix} i = 1..|Syn(a)| \\ j = 1..|Syn(b)| \end{matrix} \quad (2)$$

Equation 2 provides a measure of how related both terms are at *Level 0*, by averaging the different degree in which different synonyms of the compared terms appear together on the Web. A sum is performed (instead of a maximum, for example) to let all synonyms take part in comparisons.

$$rel_1(a, b) = \frac{\sum_{i,j} rel_0(oc_{a_i}, oc_{b_j})}{|OC(a)| \cdot |OC(b)} \quad \begin{matrix} i = 1..|OC(a)| \\ j = 1..|OC(b)| \end{matrix} \quad (3)$$

Equation 3 averages the relatedness at *Level 0* among ontological context terms, to measure how related  $a$  and  $b$  are at *Level 1*. Notice that it cannot be computed if one of the terms lacks in semantic description (thus not having ontological context:  $|OC(x)| = 0$ ). If that is the case, we still can proceed as we will see in Section 4.3.

Therefore Equations 2 and 3 provide the degree of semantic relatedness between the two corresponding levels that characterize terms  $a$  and  $b$ <sup>8</sup>. We weight these values to provide a final relatedness between two ontological terms as follows:

$$rel(a, b) = w_0 \cdot rel_0(a, b) + w_1 \cdot rel_1(a, b) \quad (4)$$

where  $w_0 \geq 0, w_1 \geq 0$  and  $w_0 + w_1 = 1$ .

Figure 1 illustrates, with a very simple example, the result of computing  $rel(a, b)$  by combining elementary word relatedness calculations. In the example  $Syn(a) = \{“astronomy”, “uranology”\}$ ,  $Syn(b) = \{“star”\}$ ,  $Syn(oc_a) = \{“physics”\}$  and  $Syn(oc_b) = \{“celestial body”\}$ . Elementary computations for *Level 0* obtain values for  $relWeb(“astronomy”, “star”)$  and  $relWeb(“uranology”, “star”)$  (0.257 and 0.158 respectively). Their combination leads to  $rel_0(a, b) = 0.207$ . Similarly,  $rel_1(a, b) = 0.247$ , and a final  $rel(a, b) = 0.227$  is obtained (with  $w_0 = w_1 = 0.5$ ).

**Additional remarks.** There are other possible ways to define  $rel(a, b)$ . For example one could consider an arbitrary number of higher levels in the semantic characterization of a term. However, we have restricted its number following this intuition: The higher a word is in the hierarchy that characterize an ontology term, the lesser *information content* it has [17], so it is less significant to characterize the ontology term. For example “Java” (in the sense of “an Indonesian

<sup>8</sup> Notice that Equation 2 is computable in polynomial time on  $|Syn(a)|, |Syn(b)|$  and Equation 3 is computable in polynomial time on  $|OC(a)|, |OC(b)|, |Syn(oc_{a_i})|, |Syn(oc_{b_j})|$ . Therefore by controlling the values  $|Syn|$  and  $|OC|$  we handle the performance of the computation.



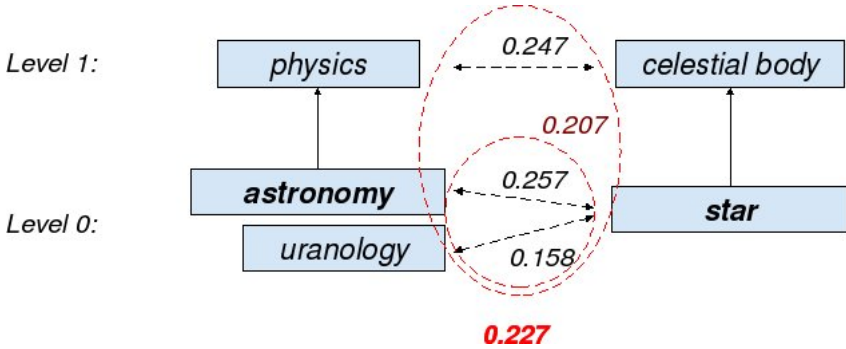


Fig. 1. Example of relatedness computation between ontology terms

island”) is closer characterized by the word “island” (direct hypernym) than by “physical object” or “thing” (higher in the hierarchy, thus with less *information content*). Therefore we consider two levels of semantic description. Additional levels could introduce many low significant words in comparisons, also increasing the computation time, as we have found out empirically.

Nevertheless, other variations of the method may be explored in the future, such as weighting differently the synonyms in Equation 2, or considering alternative definitions of  $OC(t)$ .

### 4.3 Semantic Relatedness among Ontology Terms and Words

Finally, there are possible scenarios where a *mixed* relatedness measure, between an ontology term  $t$  and a plain word  $w$ , is required. In that case, we use the following to compute *Levels* 0 and 1:

$$rel_0(t, w) = \frac{\sum_i relWeb(syn_{t_i}, w)}{|Syn(t)|} \quad i = 1..|Syn(t)| \tag{5}$$

$$rel_1(t, w) = \frac{\sum_i rel_0(oc_{t_i}, w)}{|OC(t)|} \quad i = 1..|OC(t)| \tag{6}$$

combining their results in this way:

$$rel(t, w) = w_0 \cdot rel_0(t, w) + w_1 \cdot rel_1(t, w) \tag{7}$$

where  $w_0 \geq 0, w_1 \geq 0$  and  $w_0 + w_1 = 1$ .

These previous equations provide a numerical value that indicates the relatedness degree between a sense, described as a term in an ontology, and a word. It can be useful for different purposes, such as the disambiguation experiment we describe in the following section.

Notice that the measures proposed in this section accomplish the good properties we were looking for: they use the Web as source of knowledge, they can be applied to any ontology, and do not rely on particular lexical resources or ontology languages.

## 5 Experimental Evaluation

In this section, we discuss the experiments that we have carried out to test our proposed relatedness measure. Firstly, we discuss the lack of gold standards to evaluate relatedness measures, presenting our own experiment for a proper evaluation of word relatedness degree, with respect to human judgment. We used it to evaluate how Cilibrasi and Vitányi’s based relatedness behaves, in comparison with some other well-established semantic measures. Secondly, the potential of the relatedness measure among ontology terms that we propose in Sections 4.2 and 4.3, is shown in the context of two particular tasks: disambiguation and ontology matching.

### 5.1 Correlation with Relatedness Human Judgment

Shortage in gold standards to evaluate semantic relatedness measures is a well known problem [3]. The small amount of available data is still inadequate, and it is mainly oriented to evaluate similarity, not relatedness. Obtaining large-enough set of pairs and their correspondent human judgments, with a solid methodology, is still a pending task for relatedness evaluation. Meanwhile, the two more utilized benchmark are: Miller and Charles’s data set [15], and WordSim353<sup>9</sup>. The first one is a set of 30 pairs of nouns and their similarity degree according to human opinion. The second is a larger set of 353 word pairs, where subjects were asked for a relationship degree slightly broader than similarity. However, it is still not a relatedness evaluation, and its methodology has been largely discussed [12].

Wikipedia-based methods [21,9] include an evaluation with some of these benchmarks. To let a direct comparison with one of these methods, we computed Equation 1 with Miller and Charles’s similarity data sets. Our result, a linear correlation coefficient with human opinion of 0.54, is slightly higher than the 0.46 obtained by WikiRelate [21]. However, this is only partially illustrative, as the experiment only considers *similarity*, instead of the more general *relatedness*.

As these benchmarks are not good enough to evaluate relatedness, instead of only similarity, we were motivated to create a new test data set, focused on relatedness evaluation, as we describe in the following paragraphs.

The experiment we have carried out is similar to Miller and Charles’s one. We selected 30 pairs of English nouns, where some kind of relationship are present in most of them: similarity (such as  $\langle person, soul \rangle$ ), meronymy (e.g.  $\langle hour, minute \rangle$ ), frequent association (e.g.  $\langle penguin, Antarctica \rangle$ ), and others. There are, however, other weakly related pairs (e.g.  $\langle transfusion, guitar \rangle$ ). A group of 30 university graduated people (from seven different nationalities, and high skilled in English language) were asked to assess the semantic relatedness between those pairs, in a scale from 0.0 to 4.0 (from no relatedness at all, to identical or strongly related words). In our survey we asked for “how much related the words are”, taking into account any possible relationship that could

<sup>9</sup> <http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353/wordsim353.html>

connect their meanings, therefore not considering only similarity but the more general concept of relatedness. Table 1 shows the selected pairs of words, as well as the average ratings for semantic relatedness according to human judgment. Word pairs are shown in the table in the same order as they were shown to human evaluators.

**Table 1.** Group of selected noun pairs and average relatedness according to humans. Score ranges from 0 to 4.

<i>Word<sub>1</sub></i>	<i>Word<sub>2</sub></i>	<i>Score</i>	<i>Word<sub>1</sub></i>	<i>Word<sub>2</sub></i>	<i>Score</i>
car	driver	3.14	person	soul	2.84
transfusion	guitar	0.05	theorem	wife	0.34
professional	actor	2.12	mathematics	theorem	3.30
person	person	4.00	atom	bomb	2.63
city	river	1.85	pencil	paper	2.90
theft	house	1.99	power	healing	1.25
cloud	computer	0.32	hour	minute	3.38
river	lake	3.19	blood	transfusion	3.28
blood	keyboard	0.12	xenon	soul	0.07
dog	friend	2.51	nanometer	feeling	0.11
ten	twelve	3.01	penguin	Antarctica	2.96
citizen	city	3.24	yes	no	3.00
sea	salt	2.87	computer	calculator	2.81
keyboard	computer	3.25	car	wheel	3.02
letter	message	3.16	pen	lamp	0.65

We computed Equation 1 to obtain semantic relatedness values for many different search engines. Results are summarized in Table 2, where comparisons with other WordNet-based measures [22] are provided. WordNet-based measures were computed by using WordNet::Similarity software<sup>10</sup>. We show in Table 2 the measures that obtained the best results<sup>11</sup>. We consider that a linear correlation coefficient it is not appropriate in this evaluation, because some measures produce noticeable non-linear results (such as adapted Lesk measure), thus difficulting comparisons. For this reason we use the Spearman correlation coefficient, which compares corrected ranks of assessments rather than absolute values.

These results show a high correlation between web-based measures and human judgment, thus confirming the validity of using the Cilibrasi and Vitányi's distance as basis to compute relatedness among words. Most of search engines analyzed provide higher correlation than the compared WordNet-based traditional methods, with the remarkable exception of Live Search and Clusty. A detailed discussion of the reasons for the differences between search engines is out of the scope of this paper.

<sup>10</sup> See <http://talisker.d.umn.edu/cgi-bin/similarity/similarity.cgi>, where additional information about the used measures is available.

<sup>11</sup> Differences with some results we obtained in [10] are due to the use of a later version of WordNet::Similarity tool, based on WordNet3.0 instead of WordNet2.0.

**Table 2.** Spearman correlation coefficients with human judgment. It ranges from -1 (total disagreement) to +1 (total agreement).

Web-based Measure	Value	WordNet-based Measure	Value
Exalead	<b>0.78</b>	Vector	0.62
Yahoo!	0.74	Resnik	0.56
Altavista	0.74	Adapted Lesk	0.56
Ask	0.72	Wu & Palmer	0.47
Google	0.71	Hirst & St-Onge	0.46
Live Search	0.44	Lin	0.46
Clusty	0.41	Leacock & Chodorow	0.41

## 5.2 Application to Word Sense Disambiguation

In [10] we proposed a multiontology disambiguation method to discover the intended meaning of a set of user keywords. The algorithm relies on a semantic relatedness measure to establish comparisons among word senses, obtaining a relevance degree for each keyword sense, to pick up the most probable one.

In this experiment we used our web-based relatedness among ontology terms and words, computed with Equation 7, within this disambiguation algorithm<sup>12</sup>. We chose Yahoo! as search engine in this experiment due its balance between good correlation with human judgment (see Table 2) and fast time response (for example, four times faster than Exalead). Other choices could be suitable also, such as Google, the only one faster than Yahoo! in our experiments, but however with a correlation with humans slightly worse than Yahoo!.

We illustrate the process by disambiguating the same list of highly ambiguous nouns used in the experiment described in [7] (*glass, earth, plant*). For each word, we have randomly selected ten sentences from SemCor2.0 corpus<sup>13</sup> where it appears with well defined WordNet senses. The target of the experiment is to disambiguate each selected word within each of their ten sentences, thus running 30 different disambiguation processes. We use the other words in the sentence as context to disambiguate, selecting a window of two significant words around the ambiguous one. Then we apply our disambiguation algorithm, to obtain the sense we consider most relevant, and compare it with the SemCor annotation. In Table 3 we show the averaged ratio of successful disambiguations for each keyword. We compare it with two baselines: one corresponding to random selection of senses, and other corresponding to selection of senses with highest frequency of use.

This example illustrate the usefulness of the web-based relatedness measure when applied to disambiguation tasks: even dealing with highly ambiguous words (e.g. *earth* has 6 possible senses in this experiment) we have a high probability of picking up the most suitable one with our method (almost 40% greater than a random selection).

<sup>12</sup> We limited to values  $|Syn(x)| = 4$  and  $|OC(x)| = 3$ , using also  $w_0 = w_1 = 0.5$ .

<sup>13</sup> <http://www.cs.unt.edu/~rada/downloads.html>

**Table 3.** Disambiguation results

Ambiguous word	Precision	Random baseline	Max. freq. baseline
glass	30%	14%	30%
earth	60%	14%	60%
plant	80%	25%	40%
AVERAGE	<b>57%</b>	18%	43%

### 5.3 Application to Ontology Matching

In this experiment we explore the behaviour of our web-based relatedness measure among ontology terms, in the context of an ontology matching experiment. We used Yahoo! as search engine, for the same reasons explained in Section 5.2.

In [18], it is described an ontology matching experiment between NALT and AGROVOC ontologies<sup>14</sup>. The method they apply derives semantic mappings by dynamically selecting, exploiting, and combining multiple and heterogeneous online ontologies. Each mapping is defined by  $\langle a, b, r \rangle$ , where  $a$  and  $b$  are the mapped terms from NALT and AGROVOC ontologies, respectively, and  $r$  is the kind of relationship. To assess the quality of the inferred mappings, an extensive manual evaluation was performed, obtaining a precision value of 70%.

We have reused the same experimental data of this ontology matching experiment, but for a slightly different purpose. The goal of our test is to compare human assessment in mappings evaluation, with an assessment based on using our relatedness measure. We expect that, in general, a valid mapping has a greater relatedness degree, between the involved terms, than an invalid one.

We randomly selected a set of 160 human assessed mappings, equally divided between invalid<sup>15</sup> and valid ones. Then, we computed semantic relatedness measures between each pair of the mapped ontology terms, by applying Equation 4.

Using a threshold of 0.19 to assess the validity of mappings according to relatedness values<sup>16</sup>, we obtain that 79% of relatedness-based assessments are correct according to human based judgment. This result shows that our measure highly correlates with human assessment of mappings, even when specific domain data are involved (both ontologies belong to agriculture domain).

## 6 Conclusions and Future Work

In this paper we have studied the state of the art of semantic relatedness measures, and their great interest for many applications. We have also identified some desirables features for a semantic relatedness measure to be used on the

<sup>14</sup> They are ontologies belonging to agriculture domain.

See <http://www.few.vu.nl/~wrvhage/oei2006/>

<sup>15</sup> From the set tagged as “invalid due to incorrect anchoring” in [18].

<sup>16</sup> This optimal threshold has been empirically inferred, and can be reused later in larger experiments.

Semantic Web: domain independence, maximum coverage and universality. In our study we have chosen a well-founded semantic distance (the Cilibrasi and Vitányi's one) to compute the relatedness between two plain words by using the Web as knowledge source. We have shown experimentally that it correlates well with respect to human judgment (even better than some other preexisting measures), exploring also the effect of using different search engines.

Due to the interest of computing relatedness between word senses (expressed as ontology terms) instead on only between plain words, we have taken this word relatedness as basis, to define a new measure with the following characteristics:

1. It computes relatedness among terms from different ontologies (even on different domains), by exploiting their semantic descriptions.
2. It does not depend on specific lexical resources or knowledge representation languages. As it uses the Web as source of data, it maximizes the coverage of possible interpretations.
3. It is general enough to be applied for many different purposes (such as word sense disambiguation, ontology matching, or others).

We have shown that this relatedness measure behaves well when applied to disambiguate a set of polysemous words. It also reproduces well the evaluation of ontology mappings according to human opinion.

As future work, we plan the use of larger data sets in our experiments. Moreover, we devise new ontology matching experiments, in order to exploit the full potential of this measure in relationship discovering tasks.

**Acknowledgments.** We thank Marta Sabou and Mathieu d'Aquin for providing us with the ontology matching data set we have used in our experiments. This work is supported by the CICYT project TIN2007-68091-C02-02.

## References

1. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American* (May 2001)
2. Bollegala, D., Matsuo, Y., Ishizuka, M.: Measuring semantic similarity between words using web search engines. In: *Proc. of WWW 2007, Banff, Canada* (2007)
3. Budanitsky, A., Hirst, G.: Evaluating wordnet-based measures of semantic distance. *Computational Linguistics* 32(1), 13–47 (2006)
4. Chen, H.-H., Lin, M.-S., Wei, Y.-C.: Novel association measures using web search with double checking. In: *Proceedings of the COLING/ACL 2006, Morristown, NJ, USA. Association for Computational Linguistics* (2006)
5. Cilibrasi, R.L., Vitányi, P.M.: The Google similarity distance. *IEEE Transactions on Knowledge and Data Engineering* 19(3), 370–383 (2007)
6. Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Hashman, R.: Indexing by Latent Semantic Indexing. *Journal of the American Society for Inf. Science* (1990)
7. Duca, A.: Sketchnet: Knowledge-based word sense disambiguation. In: *Proc. of EUROLAN07 Doctoral Consortium* (2007)
8. Euzenat, J., Shvaiko, P.: *Ontology matching*. Springer, Heidelberg (2007)

9. Gabrilovich, E., Markovitch, S.: Computing semantic relatedness using wikipedia-based explicit semantic analysis. In: Proceedings of The Twentieth International Joint Conference for Artificial Intelligence, Hyderabad, India (2007)
10. Gracia, J., Trillo, R., Espinoza, M., Mena, E.: Querying the web: A multiontology disambiguation method. In: Sixth International Conference on Web Engineering (ICWE 2006), Palo Alto (California, USA). ACM, New York (2006)
11. Gruber, T.R.: Towards principles for the design of ontologies used for knowledge sharing. In: Formal Ontology. Kluwer, Dordrecht (1993)
12. Jarmasz, M., Szapkowicz, S.: Roget's thesaurus and semantic similarity. In: Proceedings of the RANLP-2003, pp. 212–219 (2003)
13. Karanastasi, A., Christodoulakis, S.: Ontology-driven semantic ranking for natural language disambiguation in the ontoln framework. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519, pp. 443–457. Springer, Heidelberg (2007)
14. Kilgarriff, A., Grefenstette, G.: Introduction to the special issue on the web as corpus. *Computational Linguistics* 29(3), 333–348 (2003)
15. Miller, G.A., Charles, W.G.: Contextual Correlates of Semantic Similarity. In: *Language and Cognitive processes* (1991)
16. Motta, E., Sabou, M.: Next generation semantic web applications. In: 1st Asian Semantic Web Conference. LNCS. Springer, Heidelberg (2006)
17. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. In: 14th International Joint Conference on AI, Montreal (Canada) (1995)
18. Sabou, M., Gracia, J., Angeletou, S., d'Aquin, M., Motta, E.: Evaluating the semantic web: A task-based approach. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 423–437. Springer, Heidelberg (2007)
19. Sahami, M., Heilman, T.D.: A web-based kernel function for measuring the similarity of short text snippets. In: Proc. of the 15th Int. WWW Conference (2006)
20. Sheth, A., Arpinar, I., Kashyap, V.: Relationships at the Heart of Semantic Web: Modeling, Discovering and Exploiting Complex Semantic Relationships, vol. 139. Springer, Heidelberg (2003)
21. Strube, M., Ponzetto, S.P.: Wikirelate! computing semantic relatedness using Wikipedia. In: AAAI. AAAI Press, Menlo Park (2006)
22. Ted Pedersen, S.P., Banerjee, S.: Maximizing semantic relatedness to perform word sense disambiguation. Research Report UMSI 2005/25 (2005)
23. Trillo, R., Gracia, J., Espinoza, M., Mena, E.: Discovering the semantics of user keywords. *Journal on Universal Computer Science (JUICS)*. Special Issue: Ontologies and their Applications 13(12), 1908–1935 (2007)