

# A study on Linking Wikipedia categories to Wordnet synsets using text similarity\*

Antonio Toral\* Óscar Ferrández† Eneko Agirre<sup>◊</sup> Rafael Muñoz†

\*Istituto di Linguistica Computazionale, Consiglio Nazionale delle Ricerche  
Pisa, Italy

†Natural Language Processing and Information Systems Group  
University of Alicante, Spain

<sup>◊</sup>Informatika Fakultatea, University of the Basque Country  
Donostia, Basque Country

## Abstract

This paper studies the application of text similarity methods to disambiguate ambiguous links between WordNet nouns and Wikipedia categories. The methods range from word overlap between glosses, random projections, WordNet-based similarity, and a full-fledged textual entailment system. Both unsupervised and supervised combinations have been tried. The gold-standard with disambiguated links is publicly available. The results range from 64.7% for the first sense heuristic, 68% for an unsupervised combination, and up to 77.74% for a supervised combination.

## 1 Introduction

Human languages are extremely rich and ambiguous resulting in the fact that the same information can be expressed with different words and linguistic structures. Consequently, an ambiguous text might represent several distinct meanings and a concrete meaning might be expressed by different ways. Therefore, language ambiguity and variability are considered as essential blocks to solve in order to overcome the barrier that separates the human understanding from the computer understanding.

The task of detecting semantic similarity between texts addresses properly these language phenomena and also has a lot of potential applications for Natural Language Processing [15]; examples include word sense disambiguation [16], categorisation [11], summarisation [4], etc.

In the current research, semantic similarity is applied into a methodology devoted to the automatic construction of a Named Entity Repository [18]. This method exploits the knowledge available in already existing Language Resources (LR) to support procedures of lexico-semantic acquisition from Web 2.0 collaborative semistructured resources.

Our test case for English focuses on WordNet [5] as the LR and Wikipedia as the Web 2.0 resource. The first step consists in establishing links between

entries of both resources; the instantiable common nouns found in WordNet are mapped to Wikipedia categories. Obviously, these mappings are ambiguous for polysemous nouns. Another piece of research, YAGO [17], also addresses linking WordNet to Wikipedia. However, the authors do not deal with the ambiguity that arises when linking both resources (ambiguous mappings are simply manually disambiguated).

Our first attempt to resolve ambiguous mappings consisted in finding instances appearing both in a sense of the word in WordNet and in the mapped category in Wikipedia. This method offered perfect precision but suffered from low recall (39%) due to the small number of instances present in WordNet. The alternative solution we explore in this paper consists on applying semantic similarity between the LR definitions and the mapped Wikipedia abstracts.

Let us then consider our problem as a real-world testbed in which we will apply different methods for semantic similarity between contexts.

The rest of the paper is organised as follows. Next section summarises the different approaches to semantic similarity that are found in the literature. This is followed by the description of the different approaches that we have applied to perform semantic similarity. Afterwards, we present the evaluation and results. Finally, we close the paper by presenting conclusions.

## 2 Background

This section describes some of the most relevant works on obtaining similarities between short texts. Existing research on text similarity has focused mainly on whole documents or individual words, while short paragraphs - sentences - contexts have been mostly dismissed. Following paragraphs delve into techniques and/or approaches that were relevant for the development of this research work.

SimFinder [7] is a supervised system made up of 43 features extracted from text. It uses a log-linear regression model to determine the semantic similarity of two short text units from the evidence obtained from the different features.

[13] combines corpus-based (PMI-IR and Latent Semantic Analysis) and knowledge-based measures of

\*This work has been partially funded by the EU Commission (projects ICT-2007-211423 and FP6-IST-033860) and by the Spanish Government (project TIN2006-15265-C06-01).

word similarity. The results are then combined and used to derive a similarity metric between short texts.

Semantic Text Similarity [9] combines string and semantic similarity (a modified version of the longest common subsequence and Second order Co-occurrence PMI respectively) between words and common-word order similarity.

SenseClusters<sup>1</sup> is a language independent and unsupervised tool that clusters short contexts. It represents contexts using first or second order feature vectors. In order to reduce dimensionality it applies Singular Value Decomposition.

Apart from the aforementioned systems, it is worth mentioning two datasets that have been used to evaluate approaches to short text similarity. The first is the Microsoft paraphrase corpus [3], extracted from news sources. It is made up of 5,801 pairs of sentences, each together with a human judgement indicating whether the two sentences can be considered paraphrases or not. The second is the Pilot Short Text Semantic Similarity Benchmark Data Set [10], which contains 30 sentence pairs from the Collins Cobuild dictionary. In this case the judgements are not binary, but on a scale (from 0.0 for minimum similarity to 4.0 for maximum similarity).

### 3 Methods

The current section describes the different methods that we have applied. Approaches include Textual Entailment based on lexical and semantic inferences, a graph-based algorithm based on a LR and a Random projection algorithm to term-document matrices. We present also two baseline systems to which the aforementioned ones will be confronted. Finally, we introduce three combinations of the different approaches based on voting, unsupervised and supervised schemes.

#### 3.1 Textual Entailment

Textual Entailment has been defined as a generic framework for modeling semantic variability, which appears when a concrete meaning is described in different manners as proposed by [2]. Therefore, semantic similarity is addressed by defining the concept of Textual Entailment as a one-way meaning relation between two snippets. Moreover, a series of Workshops, called Recognising Textual Entailment (RTE<sup>2</sup>) challenges and the Answer Validation Exercise (AVE<sup>3</sup>) competitions, have been recently proposed with the objective of providing suitable frameworks to evaluate textual entailment systems.

To address the specific semantic similarity phenomenon we are dealing with in this research work (i.e. semantic similarity between WordNet glosses and Wikipedia categories), we used our in-house textual entailment system presented in [6]. This system has been previously used to support other NLP applications rather than puristic textual entailment tasks. For

instance, in Question Answering [14] and automatic text summarisation [12].

As a brief system overview, it is worth mentioning the most relevant inferences implemented aimed at solving entailment relations:

- Lexical inferences based on lexical distance measures. For instance, the Needleman-Wunsch algorithm, Smith-Waterman algorithm, a matching of consecutive subsequences, Jaro distance, Euclidean distance, IDF specificity based on word frequencies extracted from corpora, etc.
- Semantic inferences focused on semantic distances between concepts. These inferences implement several well-known WordNet-based similarity measures, verbs' similarities based on the relations encoded in VerbNet<sup>4</sup> and VerbOcean<sup>5</sup>, and reasoning about named entities correspondences between texts.

For the final application of the system to the target task of this work, we adapted it in order to manage bidirectional meaning relations. Linking WordNet glosses to Wikipedia categories is not a clear entailment phenomenon. It can occur that the gloss is implied by the category, the category is deduced by the gloss or the entailment appears in both directions. Therefore, to control these situations we opted for computing the average of the two system outputs regarding each unidirectional relation.

#### 3.2 Personalised PageRank over WordNet

WordNet is a lexical database of English, which groups nouns, verbs, adjectives and adverbs into sets of synonyms (synsets), each expressing a distinct concept. Synsets are interlinked with conceptual-semantic and lexical relations, including hypernymy, meronymy, causality, etc.

Given a pair of texts and a graph-based representation of WordNet, our method has basically two steps: We first compute the Personalised PageRank over WordNet separately for each of the texts, producing a probability distribution over WordNet synsets. We then compare how similar these two discrete probability distributions are by encoding them as vectors and computing the cosine between the vectors.

We represent WordNet as a graph  $G = (V, E)$  as follows:

- Graph nodes represent WordNet concepts (synsets) and dictionary words.
- Relations among synsets are represented by undirected edges.
- Dictionary words are linked to the synsets associated to them by directed edges.

For each text in the pair we first compute a personalised PageRank vector of graph  $G$  [8]. Basically,

<sup>1</sup> <http://senseclusters.sourceforge.net/>

<sup>2</sup> <http://pascallin.ecs.soton.ac.uk/Challenges/RTE/>

<sup>3</sup> <http://nlp.uned.es/clef-qa/ave/>

<sup>4</sup> <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

<sup>5</sup> <http://demo.patrickpantel.com/Content/verboclean/>

personalised PageRank is computed by modifying the random jump distribution vector in the traditional PageRank equation. In our case, we concentrate all probability mass in the words present in the target text.

Regarding PageRank implementation details, we chose a damping value of 0.85 and finish the calculation after 30 iterations. We have not optimised these values for this task. We used all the relations in WordNet 3.0<sup>6</sup>, including the disambiguated glosses<sup>7</sup>. This similarity method was used for word similarity [1] which report very good results on word similarity datasets.

### 3.3 Semantic Vectors

Semantic Vectors [19]<sup>8</sup> is an open source (BSD license) software package that creates WORDSPACE models from plain text. Its aim is to provide an easy-to-use and efficient tool which can fit both research and production users. It uses a random projection algorithm to perform dimension reduction as this is a simpler and more efficient technique than other alternatives such as Singular Value Decomposition.

It relies on Apache Lucene<sup>9</sup> for tokenisation and indexing in order to create a term document matrix. Once the reference corpus has been tokenised and indexed, Semantic Vectors creates a WORDSPACE model from the resulting matrix by applying random projection.

For the current task we have gathered a corpus made up of WordNet glosses and Wikipedia abstracts. On one hand, it contains the glosses of all the synsets present in WordNet 2.1., i.e. 117,598 glosses. On the other, it contains the abstracts of all the entries present in a Wikipedia dump obtained in January 2008, i.e. 2,179,275 abstracts. The final corpus has 1,292,447 terms.

Semantic Vectors provides a class (CompareTerms) that calculates the similarity between two terms (which can be words or texts). Thus we have directly used this in our experiments.

### 3.4 Baselines

We provide two baselines based on sense predominance and word overlap.

#### 3.4.1 First Sense

This baseline follows the assumption that senses in WordNet are ordered according to their usage predominance (i.e. the first sense is the most general). *First Sense* chooses always the first sense of WordNet as being the correspondent to the mapped Wikipedia category. Being Wikipedia a general resource, it is expected that the words that identify categories refer to their most common sense. E.g. it is really unexpected that the category called “Bishops” would refer to the sense “(chess) a piece that can be moved diagonally

over unoccupied squares of the same color” (third and last sense of the noun “bishop” in WordNet).

#### 3.4.2 Word overlap

This baseline calculates similarity between two texts by counting the number of overlapping words. In order to do this we have used the software package Text::Similarity<sup>10</sup>. This method has been applied both considering all the words that appear in the texts and discarding stop words. For the last, we have used the list of stop words of the English stemmer Snowball<sup>11</sup>.

### 3.5 Combinations

Due to the fact that the methods presented belong to different paradigms, we hypothesise that their results could be complementary and therefore we consider sensible to study possible combinations of them.

The first step in this direction has been the construction of an optimal combination, which we refer to as *oracle*. Given the outputs of the different systems and the gold standard, the oracle output sense/s for each instance is/are the sense/s present in the gold standard if any system return(s) it/them. The oracle represents then an optimal upper bound, the best result that could be obtained by combining the different systems.

Once we get an insight of the improvement that could be achieved by combining the diverse systems, we come up with three combination strategies:

- Voting. For each mapping it ranks senses according to the number of times they are returned by the different systems which are combined. Finally, it outputs the first ranked sense. Voting returns more than one sense if two or more senses are ranked first with the same score.
- Unsupervised combination. Within this combination, the methods taken into account have the same relevance computing a simple average function among the outputs of the considered methods (i.e. Textual entailment, WordNet-based method, Semantic Vectors and/or Word Overlap). As a result, the value returned by the average function is associated with its corresponding *Wikipedia category-WordNet sense* pair.
- Supervised combination. The whole set of inferences carried out by the Textual Entailment system together with the scores returned by the WordNet-based, Semantic Vectors and/or Word Overlap methods are computed as features for a machine learning algorithm. Specifically, we have used the BayesNet implementation provided by Weka<sup>12</sup>, and we obtained the 10-fold cross validation results over our gold standard corpus.

<sup>6</sup> Available from <http://wordnet.princeton.edu/>

<sup>7</sup> <http://wordnet.princeton.edu/glosses>

<sup>8</sup> <http://code.google.com/p/semanticvectors>

<sup>9</sup> <http://lucene.apache.org>

<sup>10</sup> <http://text-similarity.sourceforge.net>

<sup>11</sup> <http://snowball.tartarus.org/algorithms/english/stop.txt>

<sup>12</sup> <http://www.cs.waikato.ac.nz/ml/weka/>

## 4 Experiments and Discussion

### 4.1 Evaluation Framework

The evaluation data consists of a set of polysemous nouns from WordNet 2.1 which are mapped to Wikipedia categories. Additional information is provided both for nouns and categories; for the first their glosses while for the second their abstracts. The disambiguation task should then identify, for each noun, which of its senses, if any, corresponds to the mapped/s category/ies. The resulting gold standard files (corpus and key) are available for research purposes<sup>13</sup>. The corpus file follows the following format

```
<word id={id}>
  <sense number={num}>{sense gloss}</sense>
  [...]
  <sense number={num}>{sense gloss}</sense>
  <category id ={id}>{category abstract}</category>
  [...]
  <category id ={id}>{category abstract}</category>
</word>
```

while the key file is made up of lines with the format of Senseval-3 scorer<sup>14</sup>:

```
word category sense_number+
```

In order to build the corpus file we departed from a set of 254 nouns mapped to categories. 54 of them were discarded because the abstracts of the corresponding categories were empty. The final data-set contains 200 polysemous nouns mapped to 207 categories. Thus we have an evaluation set with 207 mappings.

Regarding Wikipedia abstracts, they are straightforwardly available for articles but not for categories. Therefore we developed a procedure to gather them:

```
if (category has referent_article)
  if(referent_article has abstract)
    return abstract of referent_article
if (category has article_with_same_lemma)
  if(article_with_same_lemma has abstract)
    return abstract of article_with_same_lemma
if (category_body longer than N characters)
  return category_body
return empty_string
```

Besides, we have manually created a key file. It contains the correct sense/s for each mapping. In most of the cases (154, 74,4%) there is a one to one correspondence. For 37 (17,9%) mappings, more than one sense corresponds to the mapped category, this usually occurs because the WordNet senses tend to be finer-grained than the Wikipedia categories. Concerning the remaining 16 (7,7%) mappings, no sense corresponds to the mapped category. Let's take a look at an example for each of these three cases:

- One sense corresponds to one category

```
<word id="admiral">
  <sense number="1">the supreme commander of a
  fleet; ranks above a vice admiral and below
  a fleet admiral</sense>
```

<sup>13</sup> <http://www.dlsi.ua.es/~atoral/#Resources>

<sup>14</sup> <http://www.senseval.org/senseval3/scoring>

```
<sense number="2">any of several brightly
colored butterflies</sense>
<category id="Admirals">Admiral is the rank,
or part of the name of the ranks, of the
highest naval officers. It is usually
considered a full admiral (equivalent to
full general) and four-star rank above Vice
Admiral and below Admiral of the Fleet/Fleet
Admiral. </category>
</word>
```

```
admiral Admirals 1
```

- More than one sense correspond to a category

```
<word id="communist">
  <sense number="1">a member of the communist
  party</sense>
  <sense number="2">a socialist who advocates
  communism</sense>
  <category id="Communists">This category lists
  people who have, at one time or another, been
  active in communist politics through either
  identifying themselves as communists or being
  members of parties identifying themselves as
  communist. It should not be taken for granted
  that inclusion in this category implies that
  figures remained their whole life or continue
  to be communists.
```

```
Note : communist activists should only be
featured in this category if no existing
subcategory (-ies) suits them better - the
comprehensive subcategory is :Category:Communists
by nationality . For more information on
categories, see: Wikipedia:Categorization .
</category>
</word>
```

```
communist Communists 1 2
```

- No sense corresponds to the category

```
<word id="chief_executive">
  <sense number="1">the person who holds the
  office of head of state of the United States
  government; "the President likes to jog every
  morning"</sense>
  <sense number="2">the office of the United
  States head of state; "a President is elected
  every four years"</sense>
  <category id="Chief_executives">Chief
  executives determine and formulate policies
  and provide the overall direction of companies
  or private and public sector organizations
  within the guidelines set up by a board of
  directors or similar governing body. They plan,
  direct, or coordinate operational activities at
  the highest level of management with the help
  of subordinate executives and staff managers.
  </category>
</word>
```

```
chief_executive Chief_executives 0
```

### 4.2 Result Analysis

Table 1 presents the scores obtained by the different systems and the baselines introduced in section 3. Regarding the application of the textual entailment sys-

tem, three different experiments were carried out, each one with a specific setting:

- *TE (trained AVE'07'08 + RTE-3)*: for this experiment the system was trained with the corpora provided in the AVE competitions (edition 2007 and 2008) and RTE-3 Challenge. This configuration uses a BayesNet algorithm, and it will show the capability of the system to solve the task when specific textual entailment corpora are used as training.
- *No training phase*: in order to assess whether the training corpora are appropriate to the final decision with regards to the task tackled in this work, we also decided to make an experiment without training phase. Therefore, the highest entailment coefficient returned by the system among all sense-category pairs for each word will be tagged as the correct link. These coefficients are obtained computing the set of lexical and semantic measures integrated into the system.
- *Supervised (10-fold cross-validation)*: a BayesNet algorithm was trained with the corpus described in section 4.1, which is intended to evaluate the task. We evaluated this experiment by 10-fold cross-validation using each textual entailment inference as a feature for the machine learning algorithm. This experiment shows the system behaviour when it is trained with a specific corpus for our task.

**Table 1: System Results**

Run	Accuracy
Baseline 1st sense	64.7%
Baseline Word overlap	56.3%
Baseline Word overlap (without stop words)	62.7%
Semantic Vectors	54.1%
Personalised PageRank	61.8%
Personalised PageRank (without stop words)	64.3%
TE (trained AVE 07-08 + RTE-3)	52.8%
TE (no training)	64.7%
TE (supervised)	77.74%

The first element that comes out is the high score obtained by the 1st sense baseline (64.7%). In fact, leaving aside supervision, only one system is able to reach its score, TE without training. It is also important the role of stop words. By filtering them, substantial better results can be obtained, as it can be seen both for the Word Overlap (62.7% vs. 56.3%) and the Personalised PageRank (64.3% vs. 61.8%) systems.

Results also point out that both the AVE and RTE corpora are not appropriate to this task (52.8%). This is due to the fact that the idiosyncrasies of each corpus are somewhat different resulting in a poor training stage. Nevertheless, computing the entailment coefficient returned by the system without training (64.7%),

a considerable improvement in accuracy is achieved. It proves the textual entailment inferences are suitable to support our research. Finally, as expected, the best TE result is when the dataset created for the evaluation is also processed as training and evaluated by 10-fold cross-validation (77.74%).

Table 2 presents the scores obtained by the different combinations introduced in section 3.5.

**Table 2: Combination Results**

Run	Accuracy
Oracle (PPR + SV + TE + WO)	84.5%
Voting (PPR + SV + TE + WO)	66.5%
Voting (PPR + SV + TE)	64.7%
Voting (PPR + SV + WO)	66.1%
Voting (PPR + TE + WO)	68%
Unsupervised (PPR + SV + TE + WO)	65.2%
Unsupervised (PPR + SV + TE)	64.7%
Unsupervised (PPR + SV + WO)	64.7%
Unsupervised (PPR + TE + WO)	65.7%
Supervised (PPR + SV + TE + WO)	77.24%
Supervised (PPR + SV + TE)	76.99%
Supervised (PPR + SV + WO)	75.12%
Supervised (PPR + TE + WO)	77.11%

The score achieved by the upper bound oracle (84.5%), nearly 7 points higher than the best supervised system (77.74%) seems to indicate that there is room for improving the performance by a supervised combination of the systems. However, none of the supervised combinations is able even to reach the score obtained by the supervised TE. The reason behind this is that the TE system implements some inferences which are somewhat similar to the knowledge reported by the other methods (e.g. the Smith-Waterman algorithm vs. word overlap, and WordNet Similarity measures vs. WordNet-based method). Therefore, the information gain supplied by the other methods is not enough in order to improve the TE performance.

Regarding unsupervised combinations, the best result (65.7%) is obtained by discarding Semantic Vectors, expected as out of the three systems this was the one that obtained the lowest score (54.1%). For some of these combinations the results improve the performance of the unsupervised TE configuration, but as we mentioned before, this is a slight improvement owing to the inner characteristics of the TE inferences.

Furthermore, it is worth noting though that the simple voting approach outperforms the unsupervised combination. The best result (68%) again is obtained when discarding Semantic Vectors.

## 5 Conclusions

This paper has presented an automatic approach to treat disambiguation when linking WordNet to Wikipedia. The proposal calculates semantic similarity between the definitions of WordNet senses and abstracts of Wikipedia categories in order to individuate

which of the senses corresponds to the category. We have applied different methods based on different approaches and explored also with their combination. In order to evaluate their performance we have manually annotated a gold standard. We have also considered two baselines. The results obtained are encouraging as compared to the accuracy of the best baseline (64.7%), an unsupervised combination obtains 68% while regarding supervised schemes, a Textual Entailment system applied bidirectionally achieves 77.74%.

Regarding future work, some research lines worth exploring emerge naturally from a first analysis. First, because of the fact that some inferences from the TE system overlap with the other methods (e.g. WordNet Similarity measures of the TE system vs. WordNet-based system), we plan to explore further combinations in which such inferences of the TE will be discarded. Second, we would like to study the cases in which none of the systems is able to obtain the correct disambiguation, i.e. the 15,5% of the dataset for which the oracle fails. This will give an insight of the kind of data that none of the methods is able to disambiguate. Finally, it would be interesting to measure the statistical significance between the scores obtained by the different systems.

## References

- [1] E. Agirre, A. Soroa, E. Alfonseca, K. Hall, J. Kravalova, and M. Pasca. A study on similarity and relatedness using distributional and wordnet-based approaches. In *Proceedings of NAACL*, 2009.
- [2] Ido Dagan and Oren Glickman. Probabilistic textual entailment: Generic applied modelling of language variability. In *Proceedings of the PASCAL Workshop on Learning Methods for Text Understanding and Mining*, Grenoble, France, 2004.
- [3] William B. Dolan and Chris Brockett. Automatically constructing a corpus of sentential paraphrases. In *IWP2005*, 2005.
- [4] G. Erkan and Dragomir R. Radev. Lexrank: Graph-based centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22:457–479, 2004.
- [5] Christian Fellbaum, editor. *WordNet: An Electronic Lexical Database (ISBN: 0-262-06197-X)*. MIT Press, first edition, 1998.
- [6] Óscar Ferrández, Daniel Micol, Rafael Muñoz, and Manuel Palomar. A perspective-based approach for solving textual entailment recognition. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 66–71, Prague, June 2007. ACL.
- [7] Vasileios Hatzivassiloglou, Judith L. Klavans, Melissa L. Holcombe, Regina Barzilay, Min yen Kan, and Kathleen R. Mckeown. Simfinder: A flexible clustering tool for summarization. In *NAACL Workshop on Automatic Summarization*, pages 41–49, 2001.
- [8] T. H. Haveliwala. Topic-sensitive pagerank. In *WWW '02: Proceedings of the 11th international conference on World Wide Web*, pages 517–526, New York, NY, USA, 2002. ACM.
- [9] Aminul Islam and Diana Inkpen. Semantic text similarity using corpus-based word similarity and string similarity. *ACM Trans. Knowl. Discov. Data*, 2(2):1–25, July 2008.
- [10] Yuhua Li, David McLean, Zuhair A. Bandar, James D. O’Shea, and Keeley Crockett. Sentence similarity based on semantic nets and corpus statistics. *IEEE Transactions on Knowledge and Data Engineering*, 18(8):1138–1150, 2006.
- [11] Tao Liu and Jun Guo. Text similarity computing based on standard deviation. In De-Shuang Huang, Xiao-Ping Zhang, and Guang-Bin Huang, editors, *ICIC (1)*, volume 3644 of *Lecture Notes in Computer Science*, pages 456–464. Springer, 2005.
- [12] Elena Lloret, Óscar Ferrández, Rafael Muñoz, and Manuel Palomar. A text summarization approach under the influence of textual entailment. In *Natural Language Processing and Cognitive Science, Proceedings of the 5th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2008)*, pages 22–31, Barcelona, Spain, 2008.
- [13] Rada Mihalcea, Courtney Corley, and Carlo Strapparava. Corpus-based and knowledge-based measures of text semantic similarity. In *In AAAI06*, pages 775–780, 2006.
- [14] Óscar Ferrández, Rubén Izquierdo, Sergio Ferrández, and José Luis Vicedo. Addressing ontology-based question answering with collections of user queries. *Information Processing and Management, In Press, Corrected Proof available online 31 October 2008*, 2008.
- [15] Ted Pedersen. Computational approaches to measuring the similarity of short contexts : A review of applications and methods. *CoRR*, abs/0806.3787, 2008.
- [16] Hinrich Schitze. Automatic word sense discrimination. *Journal of Computational Linguistics*, 24:97–123, 1998.
- [17] Fabian Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago - a large ontology from wikipedia and wordnet. *Elsevier Journal of Web Semantics*, 6(3):203–217, September 2008.
- [18] Antonio Toral, Rafael Muñoz, and Monica Monachini. Named entity wordnet. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC’08)*, Marrakech, Morocco, may 2008.
- [19] Dominic Widdows and Kathleen Ferraro. Semantic vectors: a scalable open source package and online technology management application. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC’08)*, Marrakech, Morocco, may 2008.