

SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining

Andrea Esuli* and Fabrizio Sebastiani†

*Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche
Via Giuseppe Moruzzi 1, 56124 Pisa, Italy
E-mail: andrea.esuli@isti.cnr.it

†Dipartimento di Matematica Pura e Applicata, Università di Padova
Via Giovan Battista Belzoni 7, 35131 Padova, Italy
E-mail: fabrizio.sebastiani@unipd.it

Abstract

Opinion mining (OM) is a recent subdiscipline at the crossroads of information retrieval and computational linguistics which is concerned not with the topic a document is about, but with the opinion it expresses. OM has a rich set of applications, ranging from tracking users' opinions about products or about political candidates as expressed in online forums, to customer relationship management. In order to aid the extraction of opinions from text, recent research has tried to automatically determine the "PN-polarity" of *subjective* terms, i.e. identify whether a term that is a marker of opinionated content has a *positive* or a *negative* connotation. Research on determining whether a term is indeed a marker of opinionated content (a *subjective* term) or not (an *objective* term) has been, instead, much more scarce. In this work we describe SENTIWORDNET, a lexical resource in which each WORDNET synset s is associated to three numerical scores $Obj(s)$, $Pos(s)$ and $Neg(s)$, describing how objective, positive, and negative the terms contained in the synset are. The method used to develop SENTIWORDNET is based on the quantitative analysis of the *glosses* associated to synsets, and on the use of the resulting vectorial term representations for semi-supervised synset classification. The three scores are derived by combining the results produced by a committee of eight ternary classifiers, all characterized by similar accuracy levels but different classification behaviour. SENTIWORDNET is freely available for research purposes, and is endowed with a Web-based graphical user interface.

1. Introduction

Opinion mining (OM – also known as “sentiment classification”) is a recent subdiscipline at the crossroads of information retrieval and computational linguistics which is concerned not with the topic a text is about, but with the opinion it expresses. Opinion-driven content management has several important applications, such as determining critics' opinions about a given product by classifying online product reviews, or tracking the shifting attitudes of the general public towards a political candidate by mining online forums or blogs. Within OM, several subtasks can be identified, all of them having to do with tagging a given text according to expressed opinion:

1. *determining text SO-polarity*, as in deciding whether a given text has a factual nature (i.e. describes a given situation or event, without expressing a positive or a negative opinion on it) or expresses an opinion on its subject matter. This amounts to performing binary text categorization under categories **Subjective** and **Objective** (Pang and Lee, 2004; Yu and Hatzivassiloglou, 2003);
2. *determining text PN-polarity*, as in deciding if a given **Subjective** text expresses a **Positive** or a **Negative** opinion on its subject matter (Pang and Lee, 2004; Turney, 2002);
3. *determining the strength of text PN-polarity*, as in deciding e.g. whether the **Positive** opinion expressed by a text on its subject matter is **Weakly Positive**, **Mildly Positive**, or **Strongly Positive** (Pang and Lee, 2005; Wilson et al., 2004).

To aid these tasks, several researchers have attempted to automatically determine whether a *term* that is a marker of opinionated content has a **Positive** or a **Negative** connotation (Esuli and Sebastiani, 2005; Hatzivassiloglou and McKeown, 1997; Kamps et al., 2004; Kim and Hovy, 2004; Takamura et al., 2005; Turney and Littman, 2003), since it is by considering the combined contribution of these terms that one may hope to solve Tasks 1, 2 and 3. The conceptually simplest approach to this latter problem is probably Turney's (Turney, 2002), who has obtained interesting results on Task 2 by considering the algebraic sum of the orientations of terms as representative of the orientation of the document they belong to; but more sophisticated approaches are also possible (Hatzivassiloglou and Wiebe, 2000; Riloff et al., 2003; Whitelaw et al., 2005; Wilson et al., 2004).

The task of determining whether a term is indeed a marker of opinionated content (i.e. is **Subjective** or **Objective**) has instead received much less attention (Esuli and Sebastiani, 2006; Riloff et al., 2003; Vegnaduzzo, 2004). Note that in these works no distinction between different senses of a word is attempted, so that the term, and not its senses, are classified (although some such works (Hatzivassiloglou and McKeown, 1997; Kamps et al., 2004) distinguish between different POSs of a word).

In this paper we describe SENTIWORDNET (version 1.0), a lexical resource in which each synset of WORDNET (version 2.0) is associated to three numerical scores $Obj(s)$, $Pos(s)$ and $Neg(s)$, describing how **Objective**, **Positive**, and **Negative** the terms contained in the synset are. The assumption that underlies our switch from terms to synsets is that different senses of the same term may

have different opinion-related properties. Each of the three scores ranges from 0.0 to 1.0, and their sum is 1.0 for each synset. This means that a synset may have nonzero scores for all the three categories, which would indicate that the corresponding terms have, in the sense indicated by the synset, each of the three opinion-related properties only to a certain degree¹. For example, the synset `[estimable(3)]`², corresponding to the sense “*may be computed or estimated*” of the adjective `estimable`, has an *Obj* score of 1.0 (and *Pos* and *Neg* scores of 0.0), while the synset `[estimable(1)]` corresponding to the sense “*deserving of respect or high regard*” has a *Pos* score of 0.75, a *Neg* score of 0.0, and an *Obj* score of 0.25.

A similar intuition had previously been presented in (Kim and Hovy, 2004), whereby a term could have both a **Positive** and a **Negative** PN-polarity, each to a certain degree. A similar point has also recently been made in (Andreevskaia and Bergler, 2006), in which terms that possess a given opinion-related property to a higher degree are claimed to be also the ones on which human annotators asked to assign this property agree more. Non-binary scores are attached to opinion-related properties also in (Turney and Littman, 2003), but the interpretation here is related to the confidence in the correctness of the labelling, rather than in how strong the term is deemed to possess the property.

We believe that a graded (as opposed to “hard”) evaluation of opinion-related properties of terms can be helpful in the development of opinion mining applications. A hard classification method will probably label as **Objective** any term that has no strong SO-polarity, e.g. terms such as `short` or `alone`. If a sentence contains many such terms, a resource based on a hard classification will probably miss its subtly subjective character, while a graded lexical resource like SENTIWORDNET may provide enough information to capture such nuances.

The method we have used to develop SENTIWORDNET is based on our previous work on determining the opinion-related properties of *terms* (Esuli and Sebastiani, 2005; Esuli and Sebastiani, 2006). The method relies on the quantitative analysis of the *glosses* associated to synsets, and on the use of the resulting vectorial term representations for

semi-supervised synset classification. The three scores are derived by combining the results produced by a committee of eight ternary classifiers, each of which has demonstrated, in our previous tests, similar accuracy but different characteristics in terms of classification behaviour.

SENTIWORDNET is freely available for research purposes, and is endowed with a Web-based graphical user interface.

2. Building SENTIWORDNET

The method we have used to develop SENTIWORDNET is an adaptation to synset classification of our method for deciding the PN-polarity (Esuli and Sebastiani, 2005) and SO-polarity (Esuli and Sebastiani, 2006) of *terms*. The method relies on training a set of ternary classifiers³, each of them capable of deciding whether a synset is **Positive**, or **Negative**, or **Objective**. Each ternary classifier differs from the other in the training set used to train it and in the learning device used to train it, thus producing different classification results of the WORDNET synsets. Opinion-related scores for a synset are determined by the (normalized) proportion of ternary classifiers that have assigned the corresponding label to it. If all the ternary classifiers agree in assigning the same label to a synset, that label will have the maximum score for that synset, otherwise each label will have a score proportional to the number of classifiers that have assigned it.

2.1. Training a classifier

Each ternary classifier is generated using the *semi-supervised* method described in (Esuli and Sebastiani, 2006). A semi-supervised method is a learning process whereby only a small subset $L \subset Tr$ of the training data Tr have been *manually* labelled. In origin the training data in $U = Tr - L$ were instead unlabelled; it is the process itself that has labelled them, automatically, by using L (with the possible addition of other publicly available resources) as input. Our method defines L as the union of three *seed* (i.e. training) sets L_p , L_n and L_o of known **Positive**, **Negative** and **Objective** synsets, respectively⁴.

L_p and L_n are two small sets, which we have defined by manually selecting the intended synsets⁵ for 14 “paradigmatic” **Positive** and **Negative** terms (e.g. the **Positive** term `nice`, the **Negative** term `nasty`) which were used as seed terms in (Turney and Littman, 2003).

L_p and L_n are then iteratively expanded, in K iterations, into the final training sets Tr_p^K and Tr_n^K . At each iteration step k two sets Tr_p^k and Tr_n^k are generated, where $Tr_p^k \supset Tr_p^{k-1} \supset \dots \supset Tr_p^1 = L_p$ and $Tr_n^k \supset Tr_n^{k-1} \supset \dots \supset Tr_n^1 = L_n$. The expansion at iteration step k consists

¹Note that associating a graded score to a synset for a certain property (e.g. **Positive**) may have (at least) three different interpretations: (i) the terms in the synset are **Positive** only to a certain degree; (ii) the terms in the synset are sometimes used in a **Positive** sense and sometimes not, e.g. depending on the context of use; (iii) a combination of (i) and (ii) is the case. Interpretation (i) has a *fuzzy* character, implying that each instance of these terms, in each context of use, have the property to a certain degree, while interpretation (ii) has a *probabilistic* interpretation (of a frequentistic type), implying that membership of a synset in the set denoted by the property must be computed by counting the number of contexts of use in which the terms have the property. We do not attempt to take a stand on this distinction, which (to our knowledge) had never been raised in sentiment analysis and that requires an in-depth linguistic study, although we tend to believe that (iii) is the case.

²We here adopt the standard convention according to which a term enclosed in square bracket denotes a synset; thus `[poor(7)]` refers not just to the term `poor` but to the synset consisting of `{inadequate(2), poor(7), short(4)}`.

³An *n*-ary classifier is a device that attaches to each object exactly one from a predefined set of n labels.

⁴A WORDNET synset represent a unique sense, which is defined by a unique gloss and is associated to a set of terms all with the same POS, each one associated to a sense number, (e.g. the adjectives `blasphemous(2)`, `blue(4)`, `profane(1)` are all contained in the same synset, whose sense is defined by the gloss “*characterized by profanity or cursing*”).

⁵For example, for the term `nice` we have removed the synset relative to the French city of Nice. The process has resulted in 47 **Positive** and 58 **Negative** synsets.

- in adding to Tr_p^k (resp. Tr_n^k) all the synsets that are connected to synsets in Tr_p^{k-1} (resp. Tr_n^{k-1}) by WORDNET lexical relations (e.g. *also-see*) such that the two related synsets can be taken to have *the same* PN-polarity;
- in adding to Tr_p^k (resp. Tr_n^k) all the synsets that are connected to synsets in Tr_n^{k-1} (resp. Tr_p^{k-1}) by WORDNET lexical relations (e.g. *direct antonymy*) such that the two related synsets can be taken to have *opposite* PN-polarity.

The relations we have used in (Esuli and Sebastiani, 2005; Esuli and Sebastiani, 2006) are synonymy and direct antonymy between terms, as is common in related literature (Kamps et al., 2004; Kim and Hovy, 2004; Takamura et al., 2005). In the case of synsets, synonymy cannot be used because it is the relation that defines synsets, thus it does connect different synsets. We have then followed the method used in (Valitutti et al., 2004) for the development of WORDNET-AFFECT, a lexical resource that tags WORDNET synsets by means of a taxonomy of affective categories (e.g. *Behaviour, Personality, Cognitive state*): after hand-collecting a number of labelled terms from other resources, Valitutti and colleagues generate WORDNET-AFFECT by adding to them the synsets reachable by navigating the relations of *direct antonymy, similarity, derived-from, pertains-to, attribute, and also-see*, which they consider to reliably preserve/invert the involved labels. Given the similarity with our task, we have used exactly these relations in our expansion. The final sets Tr_p^K and Tr_n^K , along with the set Tr_o^K described below, are used to train the ternary classifiers.

The L_o set is treated differently from L_p and L_n , because of the inherently “complementary” nature of the **Objective** category (an **Objective** term can be defined as a term that does *not* have either **Positive** or **Negative** characteristics). We have heuristically defined L_o as the set of synsets that (a) do not belong to either Tr_p^K or Tr_n^K , and (b) contain terms not marked as either **Positive** or **Negative** in the General Inquirer lexicon (Stone et al., 1966). The resulting L_o set consists of 17,530 synsets; for any K , we define Tr_o^K to coincide with L_o .

We give each synset a vectorial representation, obtained by applying a standard text indexing technique (cosine-normalized $tf * idf$ preceded by stop word removal) to its gloss, which we thus take to be a textual representation of its semantics. Our basic assumption is that terms with similar polarity tend to have “similar” glosses: for instance, that the glosses of *honest* and *intrepid* will both contain appreciative expressions, while the glosses of *disturbing* and *superfluous* will both contain derogative expressions.

The vectorial representations of the training synsets for a given label c_i are then input to a standard supervised learner, which generates two binary classifiers. One of them must discriminate between terms that belong to the **Positive** category and ones that belong to its complement (**not Positive**), while the other must discriminate between terms that belong to the **Negative** category and ones that belong to its complement (**not Negative**). Terms that have

been classified *both* into **Positive** by the former classifier and into (**not Negative**) by the latter are deemed to be positive, and terms that have been classified *both* into (**not Positive**) by the former classifier and into **Negative** by the latter are deemed to be negative. The terms that have been classified (i) into both (**not Positive**) and (**not Negative**), or (ii) into both **Positive** and **Negative**, are taken to be **Objective**. In the training phase, the terms in $Tr_n^K \cup Tr_o^K$ are used as training examples of category (**not Positive**), and the terms in $Tr_p^K \cup Tr_o^K$ are used as training examples of category (**not Negative**). The resulting ternary classifier $\hat{\Phi}$ is then applied to the vectorial representations of all WORDNET synsets (including those in $Tr^K - L$), to produce the sentiment classification of the entire WORDNET.

2.2. Defining the committee of classifiers

In (Esuli and Sebastiani, 2006) we point out how different combinations of training set and learner perform differently, even though with similar accuracy. The main three observations we recall here are the following:

- Low values of K produce small training sets for **Positive** and **Negative**, which produces binary classifiers with low recall and high precision for these categories. By increasing K these sets get larger, and the effect of these larger numbers is to increase recall and but to also add “noise” to the training set, which decreases precision.
- Learners that use information about the prior probabilities of categories, e.g. naive Bayesian learners and SVMs, which estimate these probabilities from the training sets, are sensitive to the relative cardinalities of the training sets, and tend to classify more items into the categories that have more positive training items. Learners that do not use this kind of information, like Rocchio, do not exhibit this kind of behaviour.
- The variability described in the previous points does not affect the overall accuracy of the method, but only the *balance* in classification between **Subjective** and **Objective** items, while the accuracy in discriminating between **Positive** and **Negative** items tends to be constant.

Following these considerations, we have decided to combine different configurations of training set and learner into a committee, to produce the final SENTIWORDNET scores. Specifically, we have defined four different training sets, by choosing four different values of K (0, 2, 4, 6), and we have alternatively used two learners (Rocchio and SVMs)⁶; this yields a total of eight ternary classifiers. With $K = 0$ and the SVM learner we have obtained very “conservative” binary classifier for **Positive** and **Negative**, with very low recall and high precision. For $K = 6$ SVMs produced instead

⁶The Rocchio learner we have used is from Andrew McCallum’s *Bow* package (<http://www-2.cs.cmu.edu/~mccallum/bow/>), while the SVMs learner we have used is version 6.01 of Thorsten Joachims’ *SVMlight* (<http://svmlight.joachims.org/>).

“liberal” binary classifiers for these two labels, that classify many synsets as **Positive** or **Negative** even in the presence of very little evidence of subjectivity. The Rocchio learner has a similar behaviour, although not dependent on the prior probabilities of categories. SENTIWORDNET is thus obtained by combining, for each synset, the scores produced by the eight ternary classifiers and normalizing them to 1.0.

2.3. Some statistics

Table 1 shows some statistics about the distribution of scores in SENTIWORDNET. The first remarkable fact is that the synsets judged to have some degree of opinion-related properties (i.e. not fully **Objective**) are a considerable part of the whole WORDNET, i.e. 24.63% of it. However, as the objectivity score decreases, indicating a stronger subjectivity score (either as **Positive**, or as **Negative**, or as a combination of them), the number of the synsets involved decreases rapidly, from 10.45% for **Objective** ≤ 0.5 , to 0.56% for **Objective** ≤ 0.125 . This seems to indicate that there are only few terms that are unquestionably **Positive** (or **Negative**), where “unquestionably” here indicates widespread agreement among different automated classifiers; in essence, this is the same observation which has independently been made in (Andreevskaia and Bergler, 2006), where agreement among human classifiers is shown to correlate strongly with agreement among automated classifiers, and where such agreement is strong only for a small subset of “core”, strongly-marked terms.

Table 1 reports a breakdown by POS of the scores obtained by synsets. It is quite evident that “adverb” and “adjective” synsets are evaluated as (at least partially) **Subjective** (i.e. $Obj(s) < 1$) much more frequently (39.66% and 35.7% of the cases, respectively) than “verb” (11.04%) or “name” synsets (9.98%). This fact seems to indicate that, in natural language, opinionated content is most often carried by parts of speech used as modifiers (i.e. adverbs, adjectives) rather than parts of speech used as heads (i.e. verbs, nouns), as exemplified by expressions such as a disastrous appearance or a fabulous game. This intuition might be rephrased by saying that the most frequent role of heads is to denote entities or events, while that of modifiers is (among other things) to express a judgment of merit on them.

3. Visualizing SENTIWORDNET

Given that the sum of the opinion-related scores assigned to a synset is always 1.0, it is possible to display these values in a triangle whose vertices are the maximum possible values for the three dimensions observed. Figure 1 shows the graphical model we have designed to display the scores of a synset. This model is used in the Web-based graphical user interface through which SENTIWORDNET can be freely accessed at <http://patty.isti.cnr.it/~esuli/software/SentiWordNet>. Figures 2 and 3 show two screenshots of the output for the terms *estimable* and *short*.

4. Evaluating SENTIWORDNET

How reliable are the opinion-related scores attached to synsets in SentiWordNet? Testing the accuracy of our

Score	Positive	Negative	Objective
Adjectives			
0.0	65.77%	62.81%	0.08%
0.125	12.12%	7.32%	2.14%
0.25	8.81%	8.68%	7.42%
0.375	4.85%	5.19%	11.73%
0.5	3.74%	5.63%	9.50%
0.625	2.94%	5.53%	7.65%
0.75	1.28%	3.72%	9.21%
0.875	0.47%	1.07%	7.57%
1.0	0.03%	0.04%	44.71%
Avg	0.106	0.151	0.743
Names			
0.0	90.80%	89.25%	0.00%
0.125	4.53%	3.93%	0.23%
0.25	2.37%	2.42%	0.87%
0.375	1.25%	1.54%	1.84%
0.5	0.62%	1.35%	2.32%
0.625	0.24%	0.91%	2.57%
0.75	0.14%	0.48%	3.27%
0.875	0.05%	0.12%	5.40%
1.0	0.00%	0.00%	83.50%
Avg	0.022	0.034	0.944
Verbs			
0.0	89.98%	87.93%	0.00%
0.125	4.43%	4.94%	0.21%
0.25	2.66%	2.95%	0.64%
0.375	1.55%	1.81%	1.35%
0.5	0.84%	1.24%	2.67%
0.625	0.36%	0.63%	3.40%
0.75	0.10%	0.42%	4.57%
0.875	0.07%	0.08%	6.11%
1.0	0.00%	0.00%	81.05%
Avg	0.026	0.034	0.940
Adverbs			
0.0	43.70%	76.99%	0.00%
0.125	6.25%	9.66%	0.57%
0.25	6.17%	5.32%	3.00%
0.375	14.44%	2.51%	12.83%
0.5	22.63%	2.70%	23.91%
0.625	5.70%	1.72%	13.56%
0.75	1.06%	0.82%	6.11%
0.875	0.05%	0.27%	7.04%
1.0	0.00%	0.00%	32.97%
Avg	0.235	0.067	0.698
All parts of speech			
0.0	85.18%	84.45%	0.02%
0.125	5.79%	4.77%	0.54%
0.25	3.56%	3.58%	1.97%
0.375	2.28%	2.19%	3.72%
0.5	1.85%	2.07%	4.20%
0.625	0.87%	1.64%	3.83%
0.75	0.35%	1.00%	4.47%
0.875	0.12%	0.27%	5.88%
1.0	0.01%	0.01%	75.37%
Avg	0.043	0.054	0.903

Table 1: Breakdown by part of speech of the scores obtained by WORDNET synsets, and average scores obtained for each part of speech.

tagging method experimentally is impossible, since for this we would need a full manual tagging of WORDNET according to our three labels of interest, and the lack of such a manually tagged resource is exactly the reason why we are interested in generating it automatically.

A first, approximate indication of the quality of SENTIWORDNET can be gleaned by looking at the accuracy obtained by our method in classifying the General Inquirer (Stone et al., 1966), a lexicon which is instead fully tagged according to three opinion-related labels we have been discussing; the results of this classification exercise are reported in (Esuli and Sebastiani, 2006). The reader should however bear in mind a few differences between the method used in (Esuli and Sebastiani, 2006) and the one used here: (i) we here classify entire synsets, while in (Esuli and Sebastiani, 2006) we classified terms, which can sometimes be ambiguous and thus more difficult to classify correctly; (ii) as discussed in Section 2.1., the WORDNET lexical relations used for the expansion of the training set are different. The effectiveness results reported in (Esuli and Sebastiani, 2006) may thus be considered only approximately indicative of the accuracy of the SENTIWORDNET labelling.

A second, more direct route to evaluating SENTIWORDNET is to produce a human labelling of a subset of WORDNET, and to use this subset as a “gold standard” against which to evaluate the scores attached to the same synsets in SENTIWORDNET. We are currently producing this labelled corpus⁷, which will consist of 1000 WORDNET synsets tagged by five different evaluators; for each synset each evaluator will attribute, through a graphical interface we have designed, a score for each of the three labels of interest such that the three scores sum up to 1.0. Comparisons among the scores assigned by different evaluators to the same synsets will also allow us to obtain inter-indexer inconsistency results for this task; the five evaluators have initially gone through a training session in which the meaning of the labels has been clarified, which should keep inter-indexer inconsistency within reasonable bounds.

Note that 1000 synsets correspond to less than 1% of the total 115,000 WORDNET synsets; this points at the fact that, again, the accuracy obtained on this benchmark may be considered only as indicative of the (unknown) level of accuracy with which SentiWordNet has been produced. Notwithstanding this fact this benchmark will prove a useful tool in the comparative evaluation of future systems that, like ours, tag WordNet synsets by opinion, including possible future releases of SentiWordNet.

5. Conclusion and future research

We believe that SentiWordNet can prove a useful tool for opinion mining applications, because of its wide coverage (*all* WordNet synsets are tagged according to *each* of the three labels Objective, Positive, Negative) and because of its fine grain, obtained by qualifying the labels by means of numerical scores.

⁷This work is being carried out in collaboration with Andrea Sansò from the University of Pavia, whose help we gratefully acknowledge.

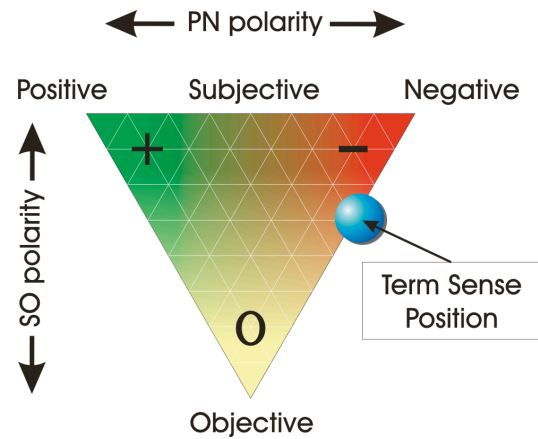


Figure 1: The graphical representation adopted by SENTIWORDNET for representing the opinion-related properties of a term sense.

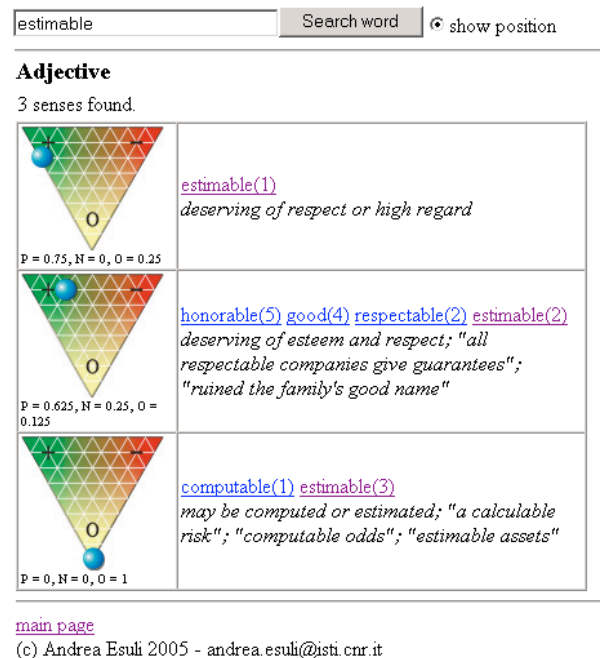


Figure 2: SENTIWORDNET visualization of the opinion-related properties of the term *estimable*.

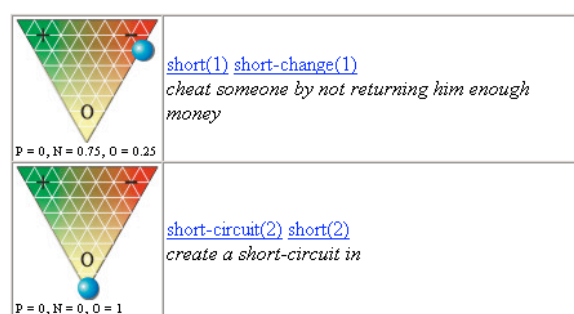
We are currently testing new algorithms for tagging WordNet synsets by sentiment, and thus plan to continue the development of SentiWordNet beyond the currently released “Version 1.0”; once developed, the gold standard discussed in Section 4. will contribute to guiding this development, hopefully allowing us to make available to the scientific community more and more refined releases of SentiWordNet.

6. Acknowledgments

This work was partially supported by Project ONTOTEXT “From Text to Knowledge for the Semantic Web”, funded by the Provincia Autonoma di Trento under the 2004–2006 “Fondo Unico per la Ricerca” funding scheme.

Verb

2 senses found.



Adjective

15 senses found.

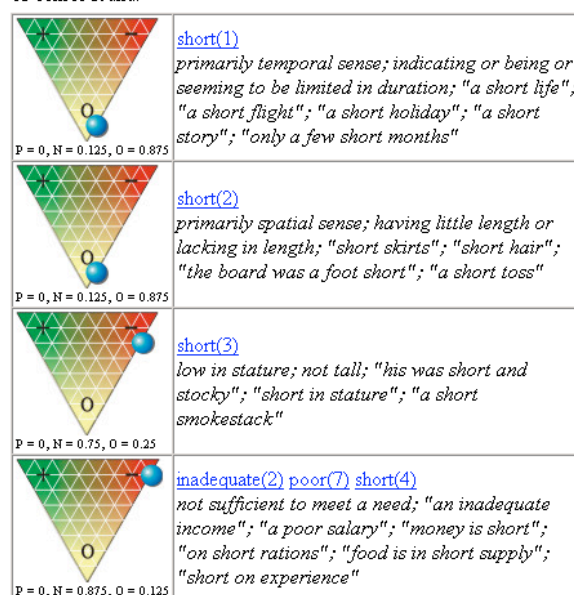


Figure 3: SENTIWORDNET visualization of the opinion-related properties of the term short.

7. References

- Alina Andreevskaia and Sabine Bergler. 2006. Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In *Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, IT. Forthcoming.
- Andrea Esuli and Fabrizio Sebastiani. 2005. Determining the semantic orientation of terms through gloss analysis. In *Proceedings of CIKM-05, 14th ACM International Conference on Information and Knowledge Management*, pages 617–624, Bremen, DE.
- Andrea Esuli and Fabrizio Sebastiani. 2006. Determining term subjectivity and term orientation for opinion mining. In *Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, IT. Forthcoming.
- Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics*, pages 174–181, Madrid, ES.
- Vasileios Hatzivassiloglou and Janyce M. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of COLING-00, 18th International Conference on Computational Linguistics*, pages 174–181.
- Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten De Rijke. 2004. Using WordNet to measure semantic orientation of adjectives. In *Proceedings of LREC-04, 4th International Conference on Language Resources and Evaluation*, volume IV, pages 1115–1118, Lisbon, PT.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of COLING-04, 20th International Conference on Computational Linguistics*, pages 1367–1373, Geneva, CH.
- Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of ACL-04, 42nd Meeting of the Association for Computational Linguistics*, pages 271–278, Barcelona, ES.
- Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of ACL-05, 43rd Meeting of the Association for Computational Linguistics*, pages 115–124, Ann Arbor, US.
- Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of CONLL-03, 7th Conference on Natural Language Learning*, pages 25–32, Edmonton, CA.
- P. J. Stone, D. C. Dunphy, M. S. Smith, and D. M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press, Cambridge, US.
- Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting emotional polarity of words using spin model. In *Proceedings of ACL-05, 43rd Annual Meeting of the Association for Computational Linguistics*, pages 133–140, Ann Arbor, US.
- Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4):315–346.
- Peter Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of ACL-02, 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, Philadelphia, US.
- Alessandro Valitutti, Carlo Strapparava, and Oliviero Stock. 2004. Developing affective lexical resources. *PsychNology Journal*, 2(1):61–83.
- Stefano Vegnaduzzo. 2004. Acquisition of subjective adjectives with limited resources. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, Stanford, US.
- Casey Whitelaw, Navendu Garg, and Shlomo Argamon. 2005. Using appraisal groups for sentiment analysis. In *Proceedings of CIKM-05, 14th ACM International Conference on Information and Knowledge Management*, pages 625–631, Bremen, DE.
- Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. 2004. Just how mad are you? Finding strong and weak opinion clauses. In *Proceedings of AAAI-04, 21st Conference of the American Association for Artificial Intelligence*, pages 761–769, San Jose, US.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of EMNLP-03, 8th Conference on Empirical Methods in Natural Language Processing*, pages 129–136, Sapporo, JP.