Random-Walk Models of Term Semantics: An Application to Opinion-Related Properties

Andrea Esuli and Fabrizio Sebastiani

Istituto di Scienza e Tecnologie dell'Informazione
Consiglio Nazionale delle Ricerche
Via Giuseppe Moruzzi, 1 – 56124 Pisa, Italy
{andrea.esuli,fabrizio.sebastiani}@isti.cnr.it

Abstract

It has recently been proposed that term senses can be automatically ranked by how strongly they possess a given opinion-related property, by applying PageRank, the well known random-walk algorithm lying at the basis of the Google search engine, to a graph in which nodes are represented by eXtended WordNet synsets and links are represented by the binary relation $s_i \blacktriangleright s_k$ ("the gloss of synset s_i contains a term belonging to synset s_k"). In other words, these properties are seen as "flowing" through this graph, from the *definiendum* (i.e., the synset being defined) to the *definiens* (i.e., a synset which occurs in the gloss of the *definiendum*), with PageRank controlling the "hydraulics" of this flow. In this paper we contend that two other random-walk algorithms may be equally adequate to this task, and provide an intuitive justification to support this claim. The first is a random-walk algorithm different from PageRank which we apply to the "inverse" graph, i.e., with properties flowing from the *definiens* to the *definiendum*. The second algorithm is a bidirectional random-walk algorithm, which assumes that properties may flow from the *definiens* to the *definiendum* and viceversa. We report results which significantly improve on the ones obtained by simple PageRank.

1. Introduction

The automatic annotation of lexical items by means of opinion-related properties (ORPs), such as positivity and negativity, has recently attracted a lot of interest, due to increased applicative interest in sentiment analysis and opinion mining. While early works in this field (see e.g. (Hatzivassiloglou and McKeown, 1997)) operated at the term level (i.e., by assuming that ORPs qualify terms) and assumed a binary model of annotation (i.e., lexical items either possess or do not posses the property), more recent works draw finer-grained distinctions, e.g., by working at the term sense level (i.e., by assuming that ORPs qualify term senses, or synsets) (Andreevskaia and Bergler, 2006a; Esuli and Sebastiani, 2006; Ide, 2006; Wiebe and Mihalcea, 2006), and by assuming a graded model of annotation (i.e., lexical items may possess the property only to a certain degree) (Andreevskaia and Bergler, 2006b; Grefenstette et al., 2006; Kim and Hovy, 2004; Subasic and Huettner, 2001).

In a recent paper (Esuli and Sebastiani, 2007) we have contributed to this literature by proposing a method for automatically ranking term senses by how strongly they possess a given ORP. The method consists in applying PageRank (Brin and Page, 1998), the well known random-walk algorithm that lies at the basis of the Google search engine, to a graph in which nodes are represented by WordNet synsets and links are represented by the binary relation $s_i \blacktriangleright s_k$ ("the gloss of synset s_i contains a term belonging to synset s_k"). The fact that the $\blacktriangleright$ relation is not explicit in WordNet is circumvented by actually using eXtended WordNet (Harabagiu et al., 1999), a publicly available, automatically sense-disambiguated version of WordNet in which every term occurring in a gloss is replaced by the synset it is deemed to belong to. The central hypothesis on which (Esuli and Sebastiani, 2007) relies is that the glosses of positive (resp. negative) synsets will mostly contain terms belonging to positive (negative) synsets, and that these properties may thus be seen as "flowing" through the graph induced by the $\blacktriangleright$ relation, from the *definiendum* (i.e., the synset s_i being defined) to the *definiens* (i.e., a synset s_k which occurs in the gloss of the *definiendum*), with PageRank controlling the logic of this flow.

In this paper we contend that two other random-walk algorithms, that we illustrate in the paper, may also be plausible choices for controlling the logic of ORP flow. The first is a random-walk algorithm different from PageRank which is applied to the "inverse" graph, i.e., the graph defined by the binary relation $\blacktriangleleft$ with properties flowing from the *definiens* to the *definiendum*. The second algorithm is a bidirectional random-walk algorithm, which assumes that properties may flow from the *definiens* to the *definiendum* and viceversa. We report results which significantly improve on the ones of (Esuli and Sebastiani, 2007).

The rest of the paper is organized as follows. In Section 2. we briefly summarize the PageRank-based model of ORP flow proposed in (Esuli and Sebastiani, 2007). In Section 3. we present our modifications of this model, resulting in two random-walk models each departing from the purely PageRank-based in a different direction. Section 4. describes the structure of our experiments and discusses the results we have obtained, comparing them to the results obtained in (Esuli and Sebastiani, 2007). Section 5. concludes, pointing at avenues of future research.

2. The PageRank model of ORP flow

Let $G = \langle N, L\rangle$ be a directed graph, with N its set of nodes and L its set of directed links; let $\mathbf{W}_0^{\blacktriangleright}$ be the $|N| \times |N|$ *adjacency matrix* of G, i.e., the matrix such that $\mathbf{W}_0^{\blacktriangleright}[i,j] = 1$ iff there is a link from node n_i to node n_j. We will denote by $B(i) = \{n_j \mid \mathbf{W}_0^{\blacktriangleright}[j,i] = 1\}$ the set of the *backward neighbours* of n_i, and by $F(i) = \{n_j \mid \mathbf{W}_0^{\blacktriangleright}[i,j] = 1\}$ the set of the *forward neighbours* of n_i. Let $\mathbf{W}^{\blacktriangleright}$ be the *row-normalized adjacency matrix* of G, i.e., the matrix such that $\mathbf{W}^{\blacktriangleright}[i,j] = \frac{1}{|F(i)|}$ iff $\mathbf{W}_0^{\blacktriangleright}[i,j] = 1$ and $\mathbf{W}^{\blacktriangleright}[i,j] = 0$ otherwise.

The input to PageRank is $\mathbf{W}^{\blacktriangleright}$ matrix (plus a vector $\mathbf{e}$ to be discussed later), and its output is a vector $\mathbf{a} = \langle a_1, \ldots, a_{|N|} \rangle$, where a_i represents the "score" of node n_i, which in our application measures the degree to which n_i has the ORP of interest. PageRank iteratively computes $\mathbf{a}$ based on the formula

$$a_i^{(k)} \leftarrow \alpha \sum_{j \in B(i)} \frac{a_j^{(k-1)}}{|F(j)|} + (1-\alpha)e_i \quad (1)$$

where $a_i^{(k)}$ denotes the value of a_i at the k-th iteration, e_i is a constant such that $\sum_i e_i = 1$, and $0 \leq \alpha \leq 1$ is a control parameter. In vectorial form, Equation 1 can be written as

$$\mathbf{a}^{(k)} = \alpha \mathbf{a}^{(k-1)} \mathbf{W}^{\blacktriangleright} + (1-\alpha)\mathbf{e} \quad (2)$$

The value of e_i amounts to an *internal source of score* for n_i that is constant across the iterations and independent from its backward neighbours. For instance, attributing a null e_i value to all but a few Web pages that are about a given topic can be used in order to bias the ranking of Web pages in favour of this topic (Haveliwala, 2003).

In (Esuli and Sebastiani, 2007) two different and independent rankings are produced, one for positivity and one for negativity. The e_i values are used as internal sources of positivity (resp. negativity) by attributing a null e_i value to all but a few "seed" synsets of renowned positivity (negativity). Through its iterations PageRank will thus make positivity (negativity) flow from the seed synsets, from which positivity flows out at a rate constant throughout the iterations, into other synsets along the ▶ relation (by using the $\mathbf{W}^{\blacktriangleright}$ matrix defined on this relation), until a stable state is reached; at this point the a_i values can be used to rank the synsets in terms of positivity (negativity).

(Esuli and Sebastiani, 2007) lists two main reasons why PageRank seems to be a good model of ORP flow:

1. If terms contained in synset s_k occur in the glosses of many positive synsets, and if the positivity scores of these synsets are high, then it seems likely that s_k is itself positive (the same happening for negativity), which justifies the summation of Equation 1.

2. If the gloss of a positive synset that contains a term in synset s_k also contains many other terms, this seems a weaker indication that s_k is itself positive (which justifies dividing by $|F(j)|$ in Equation 1).

3. Revising the PageRank model

In the following we refer to the model proposed in (Esuli and Sebastiani, 2007) as the *direct flow* model, and we present two other alternative models.

3.1. The inverse flow model

The basic intuition in (Esuli and Sebastiani, 2007) is that if a synset possesses a given ORP, then also the synsets occurring in its gloss are likely to possess the same ORP; this is modelled as a network flow in which the ORP flows from the *definiendum* to the *definiens* along the ▶ relation. However, starting from the same intuition it is equally plausible to hypothesize an *inverse flow* model, in which it is the synsets that occur in the gloss of the *definiendum* that influence the *definiendum* itself, and not viceversa. In this model the ORP thus flows from the *definiens* to the *definiendum*, along the ◀ relation (defined as the symmetric relation of ▶).

We formalize the inverse flow model by the equation

$$a_i^{(k)} \leftarrow \frac{\alpha}{|B(i)|} \sum_{j \in B(i)} a_j^{(k-1)} + (1-\alpha)e_i \quad (3)$$

where $B(i)$ is now derived from the adjacency matrix $\mathbf{W}_\mathbf{0}^{\blacktriangleleft}$ defined by the ◀ relation.

We stress that the inverse flow model is characterized not only by a different incidence matrix wrt the direct flow model, but by a very different equation of the "hydraulics" of ORP flow. In fact, Equation 3 states that node a_i receives the *average*, and not the sum, of the scores of the nodes that point to a_i, modulo α and e_i. In the case of inverse flow we consider this a reasonable assumption since:

1. If the gloss of a synset s_k contains many terms that belong to positive synsets, and if the positivity scores of these synsets are high, then it seems likely that s_k is itself positive (the same happening for negativity), which justifies the summation of Equation 3.

2. If the gloss of a synset s_i that contains a term belonging to a positive synset s_k also contains many other terms, this seems a weaker indication that s_i is itself positive (which justifies dividing by $|B(i)|$ in Equation 3).

In order to write Equation 3 in matrix form we may exploit the fact that $\mathbf{W}_\mathbf{0}^{\blacktriangleleft}$ happens to be equal to $(\mathbf{W}_\mathbf{0}^{\blacktriangleright})^T$, the transpose of $\mathbf{W}_\mathbf{0}^{\blacktriangleright}$, and that applying the normalization factor $|B(i)|$ in Equation 3 is equivalent to performing column normalization on $\mathbf{W}_\mathbf{0}^{\blacktriangleleft}$. Thus $\mathbf{W}^{\blacktriangleleft} = (\mathbf{W}^{\blacktriangleright})^T$, and Equation 3 can be written in matrix form as

$$\mathbf{a}^{(k)} = \alpha \mathbf{a}^{(k-1)} (\mathbf{W}^{\blacktriangleright})^T + (1-\alpha)\mathbf{e} \quad (4)$$

where $\mathbf{W}^{\blacktriangleright}$ is the row-normalized adjacency matrix used for the direct flow model in Equation 2. This indicates that, even if Equation 3 is very different from the equation that originates PageRank (Equation 1), the inverse flow model can anyway be computed by using PageRank, with the only difference that the $\mathbf{W}^{\blacktriangleright}$ matrix of the direct model needs to be replaced by its transpose $(\mathbf{W}^{\blacktriangleright})^T$.

3.2. The bidirectional model

We have argued that both the direct flow and the inverse flow models are reasonable models of how ORPs flow between synsets. Actually, in this analysis no argument has been put forward that either model is better than the other, or that the two models are mutually incompatible. It seems thus plausible that a third, *bidirectional flow* model could be hypothesized, in which ORPs flow from the *definiendum* to the *definiens* and vice versa, pretty much as in an electrical network. A synset s_k is thus seen to distribute its positivity score both to the synsets which occur in its gloss (the ► relation) and to the synsets whose glosses contain it (the ◄ relation). The binary relation ◄► according to which ORPs flow in the bidirectional model is thus defined as ◄►≡◄ ∪ ►. we formalize the bidirectional flow model by the following equation:

$$a_i^{(k)} \leftarrow \alpha \sum_{j \in B^{\blacktriangleright}(i)} \frac{a_j^{(k-1)}}{|F^{\blacktriangleright}(j)|} + \frac{\alpha}{|B^{\blacktriangleleft}(i)|} \sum_{j \in B^{\blacktriangleleft}(i)} a_j^{(k-1)} + (1-\alpha)e_i \quad (5)$$

where the $B^{\blacktriangleright}$, $F^{\blacktriangleright}$, $B^{\blacktriangleleft}$ and $F^{\blacktriangleleft}$ are the neighbourhood functions of the direct and inverse flow models.

The vectorial form of Equation 5 can be easily derived by observing that the normalized matrix $\mathbf{W}^{\blacktriangleleft\blacktriangleright}$ for the bidirectional flow model induced by Equation 5 is equal to $\mathbf{W}^{\blacktriangleright} + (\mathbf{W}^{\blacktriangleright})^T$; we thus obtain

$$\mathbf{a}^{(k)} = \alpha \mathbf{a}^{(k-1)}(\mathbf{W}^{\blacktriangleright} + (\mathbf{W}^{\blacktriangleright})^T) + (1-\alpha)\mathbf{e} \quad (6)$$

Again, this formula shows that also the bidirectional flow model can be computed using PageRank, with the only difference that the $\mathbf{W}^{\blacktriangleright} + (\mathbf{W}^{\blacktriangleright})^T$ matrix needs to be used in place of the $\mathbf{W}^{\blacktriangleright}$ matrix of the direct model and of the $(\mathbf{W}^{\blacktriangleright})^T$ of the inverse model.

4. Experiments

In order to provide a fair comparison with the results of the direct model as presented in (Esuli and Sebastiani, 2007), the gold standard, evaluation function, choice of $\mathbf{e}$ vectors, and experimental methodology discussed below are exactly the same as in that paper, which can thus be consulted for more detail on the experimental setting.

4.1. The $\mathbf{e}$ vector

As in (Esuli and Sebastiani, 2007), several alternative choices of the $\mathbf{e}$ vector have been tested.

The first vector (hereafter dubbed $\mathbf{e}1$) consists of all values uniformly set to $\frac{1}{|N|}$. This is the $\mathbf{e}$ vector that was originally used in (Brin and Page, 1998), and brings about an unbiased (that is, wrt particular properties) ranking of WordNet. Of course, it is not meant to be used for ranking by positivity or negativity; we have used it simply in order to evaluate the impact of ORP-biased vectors for positivity (negativity) ranking.

The first sensible, minimalistic definition of $\mathbf{e}$ (dubbed $\mathbf{e}2$) is that of a vector with uniform non-null e_i scores for the synsets that contain the adjective "good" ("bad"), and null scores for all other synsets. A further, still fairly minimalistic definition (dubbed $\mathbf{e}3$) is that of a vector with uniform non-null e_i scores for the synsets that contain at least one of the seven "paradigmatic" positive (negative) adjectives used as seeds in (Turney and Littman, 2003), and null scores for all other synsets.

We have also tested a more complex version of $\mathbf{e}$, with e_i scores equal to the positivity (negativity) scores assigned to synsets s_i in release 1.0 of SentiWordNet, normalized so that $\sum_{i=1}^{|N|} e_i = 1$. SentiWordNet (Esuli and Sebastiani, 2006)[1] is an automatically constructed lexical resource which assigns to each WordNet synset a positivity score, a negativity score, and a neutrality score. In a similar way we also produced a further $\mathbf{e}$ vector (dubbed $\mathbf{e}5$) by normalizing the scores, which we obtained from the authors, of release 1.1 of SentiWordNet, which was generated through a slight modification of the approach that had brought about release 1.0.

Finally, we have tested a way of combining the direct and the inverse flow models which is inherently different from the combination implemented in the bidirectional model: we feed to the direct (resp., inverse) flow model a vector $\mathbf{e}$ (dubbed $\mathbf{e}6$) resulting from the best run (i.e., with the optimal value of α – see Section 4.3.) of the inverse (resp., direct) flow model. This method of combination amounts to "concatenating" the flows of the direct and inverse flow model, rather than letting the two flows occur at the same time.

4.2. The gold standard and the effectiveness measure

In order to evaluate the quality of the rankings produced by our three alternative random-walk models we have used the Micro-WNOp corpus as a gold standard[2], according to the same experimental protocol as used in (Esuli and Sebastiani, 2007). Micro-WNOp consists in a set of approximately 1,000 WordNet synsets, each of which was manually assigned (by a research group different from our own) a triplet of scores, one of positivity, one of negativity, one of neutrality. We use a first group of 496 synsets (Group1) as a validation set, i.e., in order to perform parameter optimization, and a second group of 499 synsets (Group2) as a test set, i.e., in order to evaluate our method once all the parameters have been optimized.

The result of applying our three models to the various graphs induced by the ►, ◄ and ◄► relations, given a vector $\mathbf{e}$ of internal sources of positivity (negativity) score and a value for the α parameter, is a ranking of all WordNet synsets in terms of positivity (negativity). By using different $\mathbf{e}$ vectors and different values of α we obtain different rankings, whose quality we evaluate by comparing them against the ranking obtained from Micro-WNOp.

To evaluate the rankings produced by our models we have used the *p-normalized Kendall* τ *distance* (noted τ_p – see e.g., (Fagin et al., 2004)) between the gold standard rankings and the predicted rankings. For a prediction which perfectly coincides with the gold standard, τ_p equals 0; for a prediction which is exactly the inverse of the gold standard, τ_p is equal to 1.

See the full paper for more details on both Micro-WNOp and τ_p.

4.3. The results

Table 1 shows the results obtained by our three models with the different choices for the $\mathbf{e}$ vector as detailed in Section 4.1.. PageRank is iterated until the cosine of the

[1] http://swn.isti.cnr.it/

[2] http://www.unipv.it/wnop/

Table 1: τ_p values obtained by the three proposed models; ▶, ◀ and ◀▶ indicate the direct, inverse, and bidirectional models, respectively; Δ indicates the improvement of each model wrt the baseline ("B"), consisting of the ranking obtained by the corresponding $\mathbf{e}$ vector before the application of any ORP flow algorithm. Boldface indicates the best result obtained.

Ranking by positivity							
$\mathbf{e}$	B	▶	Δ	◀	Δ	◀▶	Δ
$\mathbf{e}1$	0.500	0.496	-0.8%	0.479	-4.2%	0.489	-2.1%
$\mathbf{e}2$	0.500	0.467	-6.7%	0.435	-13.0%	0.457	-8.7%
$\mathbf{e}3$	0.500	0.471	-5.8%	0.424	-15.1%	0.477	-4.7%
$\mathbf{e}4$	0.349	0.325	-6.8%	**0.292**	-16.4%	0.312	-10.7%
$\mathbf{e}5$	0.400	0.380	-4.9%	0.345	-13.6%	0.374	-6.4%
$\mathbf{e}6$	–	0.292	0%	0.318	-2.1%	–	–
Ranking by negativity							
$\mathbf{e}$	B	▶	Δ	◀	Δ	◀▶	Δ
$\mathbf{e}1$	0.500	0.549	9.8%	0.461	-7.7%	0.506	1.2%
$\mathbf{e}2$	0.500	0.502	0.3%	0.416	-16.8%	0.475	-5.1%
$\mathbf{e}3$	0.500	0.495	-0.9%	0.387	-22.7%	0.452	-9.5%
$\mathbf{e}4$	0.296	0.284	-4.3%	**0.222**	-25.0%	0.248	-16.4%
$\mathbf{e}5$	0.407	0.393	-3.5%	0.270	-33.6%	0.319	-21.7%
$\mathbf{e}6$	–	0.222	0%	0.241	-15.1%	–	–

angle between the vectors $\mathbf{a}^{(k)}$ and $\mathbf{a}^{(k+1)}$ generated by two subsequent iterations is above a predefined threshold χ (we use $\chi = 1 - 10^{-9}$). However, in order to limit the amount of processing, we stop PageRank whenever this condition has not been reached in 1000 iterations.

The results indicate the performance obtained on the test set with the value of α that was determined optimal by experimentation on the validation set; different values of α may thus be used for different choices of $\mathbf{e}$. The "B" (baseline) column contains the values of τ_p as computed directly on the $\mathbf{e}$ vector, i.e., before the application of PageRank. The Δ values shown to the right of each column denote the relative improvement obtained by the method indicated against the baseline (since low values of τ_p are better, an improvement is indicated by a negative value).

Table 1 clearly indicates that the inverted flow model always produces the best results, irrespectively of the choice of the $\mathbf{e}$ vector. Moreover, the best absolute values for positivity (0.292) and negativity (0.222) show a large improvement wrt their original $\mathbf{e}$ vectors (-16.4% for positivity and -25.0% for negativity). This is relevant, since they were obtained with vectors $\mathbf{e}4$ (the ones derived from SentiWordNet 1.0); in other words, the improvement is obtained with respect to an already high-quality lexical resource for ORPs, obtained by the same techniques that, at the term level, are still the best-known performers for polarity detection on the widely used General Inquirer benchmark (Esuli and Sebastiani, 2005).

Although the direct flow model of (Esuli and Sebastiani, 2007) also improves with respect to the baseline, the inverted flow model is largely superior to it (the latter improving on the former by 10.1% on positivity and by 21.8% on negativity). Concerning the bidirectional flow model, while it also outperforms the direct flow model, it does so less markedly than the inverse model does; in the light of the previously discussed results this is unsurprising, given that it is a combination of the other two models.

The superiority of the inverse flow model is also apparent from the results of the $\mathbf{e}6$ experiments. Here, the inverse flow model as applied to the best vector resulting from the direct flow model manages to improve the quality of this vector (by 2.1% on positivity and by 15.1% on negativity), but still underperforms wrt the best result it has obtained (on $\mathbf{e}4$). On the contrary, the direct flow model as applied to the best vector resulting from the inverse flow model leaves the vectors unchanged. A closer inspection of this latter result shows that the value of α that performed optimally in this case was $\alpha = 0$, which corresponds to …leaving the $\mathbf{e}$ vector unchanged, i.e., renouncing to let ORPs flow through the network. All values of $\alpha > 0$ managed instead to obtain an *inferior* performance wrt the best performance obtained by the inverse model.

4.4. Anecdotal evaluation

An analysis of the top-ranked synsets returned by each model according to positivity and negativity (not reported for reasons of space – see the full paper for details) shows that some of the top-ranked synsets for the direct flow model, especially for the ranking by positivity, contain function words, such as the verbs "to be" and "to have", or words that simply occur frequently within glosses, such as "quality" or "capable". These synsets receive many incoming links in the direct flow model, and this pushes them up in the ranking[3].

This phenomenon does not appear in the inverse flow model. For example, the synsets that appear in the glosses of verbs such as "to be" are unlikely to be ORP-loaded; such verbs thus obtain a low score. In the full paper we show that the inverse flow model top-ranks those glosses which are almost exclusively composed of semantically oriented terms. Again, the bidirectional flow model trades off between the other two models, producing a ranking which appears to mix the characteristics of the other two.

5. Conclusion

We have presented two novel random-walk models for ranking WordNet synsets according to how strongly they possess a given ORP; the difference between the two models and the direct flow model proposed in (Esuli and Sebastiani, 2007) lies not only in the (obviously different) incidence matrix, but also in the different equations that determine the "hydraulics" of ORP flow. However, by exploiting properties of the row-normalized incidence matrix of the inverse flow model, all the three models can be recast in terms of the application of PageRank to different matrixes.

We have shown that the inverse flow model here proposed is significantly superior to the direct flow model

[3] In order to solve this problem we have also tested a version of the direct flow model in which synset s_k receives the *average*, and not the sum, of the contributions of the synsets s_i such that s_i ▶ s_k; however, this has produced inferior results wrt the standard direct flow model.

proposed in (Esuli and Sebastiani, 2007). We have presented comparative results that show, both in a quantitative and qualitative way, the superiority of the inverse flow model. We can thus confidently assert that ORPs may best be seen as flowing from *definiens* to *definiendum*, and not vice versa, as instead hypothesized in (Esuli and Sebastiani, 2007).

We have applied and discussed our models in the context of *opinion-related* properties of synsets. However, we conjecture that these models can be of more general use, i.e., for the determination of other semantic properties of term senses, such as membership in a domain (Magnini and Cavaglià, 2000).

In the future we plan to re-apply the same algorithms to the forthcoming manually sense-disambiguated version of WordNet. This will allow to eliminate the effect of the noise introduced in eXtended WordNet by the automatic sense disambiguation phase, and test whether the results of this paper are valid also when "correct", manually disambiguated glosses are used.

Acknowledgments

This work was partially funded by the 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.

6. References

Andreevskaia, Alina and Sabine Bergler, 2006a. Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'06)*. Trento, IT.

Andreevskaia, Alina and Sabine Bergler, 2006b. Sentiment tag extraction from WordNet glosses. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*. Genova, IT.

Brin, Sergey and Lawrence Page, 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117.

Esuli, Andrea and Fabrizio Sebastiani, 2005. Determining the semantic orientation of terms through gloss analysis. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM'05)*. Bremen, DE.

Esuli, Andrea and Fabrizio Sebastiani, 2006. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*. Genova, IT.

Esuli, Andrea and Fabrizio Sebastiani, 2007. PageRanking WordNet synsets: An application to opinion mining. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL'07)*. Prague, CZ.

Fagin, Ronald, Ravi Kumar, Mohammad Mahdiany, D. Sivakumar, and Erik Veez, 2004. Comparing and aggregating rankings with ties. In *Proceedings of ACM International Conference on Principles of Database Systems (PODS'04)*. Paris, FR.

Grefenstette, Gregory, Yan Qu, David A. Evans, and James G. Shanahan, 2006. Validating the coverage of lexical resources for affect analysis and automatically classifying new words along semantic axes. In James G. Shanahan, Yan Qu, and Janyce Wiebe (eds.), *Computing Attitude and Affect in Text: Theories and Applications*. Heidelberg, DE: Springer, pages 93–107.

Harabagiu, Sanda H., George A. Miller, and Dan I. Moldovan, 1999. WordNet 2: A morphologically and semantically enhanced resource. In *Proceedings of the ACL SIGLEX Workshop on Standardizing Lexical Resources*. College Park, US.

Hatzivassiloglou, Vasileios and Kathleen R. McKeown, 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL'97)*. Madrid, ES.

Haveliwala, Taher H., 2003. Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):784–796.

Ide, Nancy, 2006. Making senses: Bootstrapping sense-tagged lists of semantically-related words. In *Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Text Processing (CICLING'06)*. Mexico City, MX.

Kim, Soo-Min and Eduard Hovy, 2004. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING'04)*. Geneva, CH.

Magnini, Bernardo and Gabriela Cavaglià, 2000. Integrating subject field codes into WordNet. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC'2000)*. Athens, GR.

Subasic, Pero and Alison Huettner, 2001. Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems*, 9(4):483–496.

Turney, Peter D. and Michael L. Littman, 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4):315–346.

Wiebe, Janyce and Rada Mihalcea, 2006. Word sense and subjectivity. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL'06)*. Sydney, AU.