# Building a Sentiment Lexicon for Social Judgement Mining


Mário J. Silva[1], Paula Carvalho[2], and Luís Sarmento[3]

[1] IST/INESC-ID
mjs@inesc-id.pt
[2] INESC-ID
pcc@inesc-id.pt
[3] SAPO/Portugal Telecom
las@co.sapo.pt



**Abstract.** We present a methodology for automatically enlarging a Portuguese sentiment lexicon for mining social judgments from text, i.e., detecting opinions on human entities. Starting from publicly-availabe language resources, the identification of human adjectives is performed through the combination of a linguistic-based strategy, for extracting human adjective candidates from corpora, and machine learning for filtering the human adjectives from the candidate list. We then create a graph of the synonymic relations among the human adjectives, which is built from multiple open thesauri. The graph provides distance features for training a model for polarity assignment. Our initial evaluation shows that this method produces results at least as good as the best that have been reported for this task.


## 1 Introduction

Synonyms have identical semantic orientation when used in the same context, which has motivated different authors to explore synonymy in language resources for automatically enlarging their sentiment lexicons. In general, lexicon-based approaches start from a confined list of manually annotated polar words, which are then used as seeds for finding new polar items. Classical methods typically explore WordNet [7] and other lexical resources publicly available to determine the polarity of the words with which known polar words are semantically related.

Regarding synonymy, the basic solution consists in propagating the polarity information of known polar words to the elements belonging to the same synset. However, such procedure is fallible because identical lexical forms (or homographs) can make part of different synsets presenting a diversity of meanings and polarities. Additionally, the majority of publicly available lexical resources describing synonymy only provide an inventory of potential synonyms for a given word, without defining the context where they have an identical interpretation. For example, the adjective *fresh* can be used as modifier of a human noun, and it can be replaced by adjectives such as *impertinent* or *impudent*, which have a negative semantic orientation. Also, it can modify non-human nouns, exhibiting


H. Caseli et al. (Eds.): PROPOR 2012, LNAI 7243, pp. 218–228, 2012.

in this case an opposite polarity. For example, when combined with an abstract noun such as *portrayal*, *fresh* is interpretable as *new* or *novel*, conveying a positive semantic orientation.

The polarity of predicates can change according to the syntactic-semantic nature of the nouns they relates to. This means that the polarity assignment will only be successful if such combination constraints will be taken into account.

We present a methodology for automatically enlarging a sentiment lexicon for mining social judgments from text, which takes into account the syntactic-semantic nature of predicate's arguments. We use a two-step approach to accomplish such goal: first, we find which adjectives can be used as human modifiers, and, next, assign them a polarity attribute. The identification of human adjectives is performed through the combination of a linguistic-based strategy, for extracting human adjective candidates from corpora, and a machine learning strategy, for automatically selecting the best human candidates. Like Rao and Ravichandran [15], we treat polarity labeling as a propagation problem in a graph, built from several open thesauri, but we restrict the polarity propagation to synonyms sharing similar syntactic-semantic properties.

We address human adjective predicates (i.e. adjectives modifying human nouns), because we are interested in extracting opinions targeting human entities, but the proposed methodology is language independent and could be applied to a wider range of syntactic-semantic predicates. In fact, we have used the methodology to create a sentiment lexicon that has been applied in opinion mining tasks involving social judgements of human targets, based on public linguistic resources available for Portuguese.

The remainder of the paper is organized as follows: Section 2 presents some related work; Section 3 describes the linguistic resources used in our experiments and for generating the sentiment lexicon; Sections 4 and 5 detail the methodology we used for identifying human adjectives and classifying their polarity, respectively; evaluation results are discussed in Section 6; some concluding remarks are, finally, presented in Section 7.

## 2 Related Work

First approaches on sentiment lexicon construction aim at identifying the subjective lexical units in general language and determining their semantic orientation (or polarity). The pioneer work of Hatzivassiloglou and McKeown [9] tackles the problem of determining the semantic orientation of adjectives by exploring the co-occurrence of positive and negative adjectives with other expressions in corpora, namely in the scope of copulative constructions (whose constituents tend to be coherent in terms of polarity) and adversative constructions (whose constituents tend to exhibit different polarity). After combining these constraints across many adjectives, the authors use a clustering algorithm that separates the adjectives into different groups, which are then labeled as positive or negative.

Turney and Littman [18] propose a bootstrapping method for inferring the semantic orientation of new polar words by computing the pointwise mutual information (PMI) between each target word and a few set of positive and negative paradigm words (or seeds) previously classified. On the other hand, [16] use two bootstrapping algorithms that exploit extraction patterns to learn sets of subjective nouns. The authors show that *Meta-Bootstrapping* and *Basilisk* algorithms, typically used for automatically generating extraction patterns to identify words belonging to a semantic class, can also be effective in identifying subjective words. This approach is based on the principle that words of the same semantic class or category tend to occur in similar contexts.

Significant research on sentiment lexicon construction has been also exploring WordNet (e.g. [10,6,8]) and other lexical resources for languages where WordNet-like resources are not available [15], to acquire new polar words. For example, Kamps et al. [10] try to determine sentiments of adjectives in WordNet by measuring the relative distance of each adjective to the reference positive and negative words "good" and "bad", respectively. Kim and Hovy [12] built a sentiment classifier that also uses WordNet, but performing a tri-polarity classification (positive, negative and neutral), based on a manually annotated dataset composed of verbs and adjectives.

Rao and Ravichandran [15] treat polarity detection as a semi-supervised label propagation problem in a graph. Takamura et al. [17] exploit the gloss information associated to words in dictionaries, for determining their semantic orientation. They construct a lexical network by linking two words whenever one of them appears in the gloss of the other word. Semantic orientations are regarded as spins of electrons, and the mean field approximation is used to compute the approximate probability function of the system.

Despite the availability of a number of sentiment lexicons, especially for English, it has being argued that their use is frequently unsatisfactory, because they do not reflect domain-specific lexical usage [5]. Hence, different approaches have been proposed to create domain-dependent polarity lexicons, instead of general-purpose lexicons. To face this problem, Fahrni and Klenner propose a two-stage method for determining the sentiment of adjectives in a given domain [1]. First, they use the Wikipedia for automatically detecting candidate targets associated to adjectives in a given domain, followed by a bootstrapping approach to determine the target-specific adjective polarity. Choi and Cardie propose integer linear programming to adapt an existing lexicon into a new specific one [5]. They consider the relations among words to derive the most likely polarity of each lexical item for a given domain. Kanayama and Nasukawa use manually-crafted syntactic patterns to identify polar atoms in a given domain [11].

In this paper, we propose a methodology for developing a fine-grained sentiment lexicon, which can be seen as an alternative to general-purpose and domain-specific lexicons. Polarities in this lexicon are assigned based on the syntactic-semantic category of targets of sentiment, applying to texts from any domain. It only handles human adjectives presently, but it can be adapted to other syntactic and semantic categories. As Riloff et al. [16] and Kanayama and Nasukawa [11], we make use of manually-crafted lexico-syntactic patterns, but they do not require any POS tagging or parsing. Similar to Kim and Hovy [12], we use a manually annotated lexicon for assigning positive, negative and neutral polarity to adjectives linked to these ones by a synonymy relationship. As suggested by Rao and Ravichandran [15], we used information aggregated from several publicly-available thesauri, since Portuguese, like many other languages, does not have comprehensive and publicly available WordNets.

## 3 Input Linguistic Resources

The identification of human adjective candidates relies on a small set of lexical resources, namely a lexicon of adjectives (Section 3.1), a list of names (Section 3.2), and a dictionary of profession and position names (Section 3.3). These are applied to a large n-gram collection (Section 3.5). The lexicon of adjectives is also used, together with a dictionary of synonyms (Section 3.4), in the polarity assignment stage.

### 3.1 Lexicon of Adjectives

We started with a lexicon of 24,792 adjectival lemmas, which are partially annotated with their semantic category and polarity. In detail, 4,034 adjectives were manually assigned to the human attribute and the remaining 511 lemmas to the non-human attribute. Human adjectives are characterized as co-occurring with a human subject (e.g. *the prime-minister is popular*). On the contrary, this type of subject is interdicted with non-human adjectives (e.g. *the prime-minister is sporadic*).

Human adjectives were also manually labeled with their prior polarities, which may be positive (1), negative (-1) or neutral(0). In terms of polarity distribution, 56% of the entries were labeled as negative (2,242 lemmas), 19% as positive (785 lemmas) and the remaining 25% as neutral (1,009 lemmas).

Polar adjectives were mostly collected from a lexico-syntactic database of human intransitive adjectives available in contemporary European Portuguese [3].

### 3.2 Dictionary of Names

In addition to this lexicon, the patterns described below make also use of a dictionary of names and surnames. These were collected from the public lists of placed secondary teacher names in the 2009 recruitment, available from the Portuguese Ministry of Education website. We obtained a list of 562 proper names and 1,388 surnames. First names correspond to the first element in name combinations. After removing all possible prepositions and conjunctions, we extracted all the tokens from name combinations after the first two (in Portuguese countries, many people have two given names), and classified them as surnames.

### 3.3 Dictionary of Profession and Position Names

Finally, we use a dictionary composed by 383 lemmas ( $\sim$ 1200 inflected forms) denoting a professional or official position. Dictionary entries were semi-automatically compiled from news corpora, by exploring syntactic structures where such type of nouns typically occurs (e.g. in apposition to a human named entity).

### 3.4 Synonym Dictionaries

To expand the original polarity lexicon of human adjectives, we explored synonymy among adjectives in different publicly-available thesauri for Portuguese. We specifically used PAPEL 2.0 [14], TeP [13] and DicSin[1]. The previously mentioned resources comprise 87,327 different lemmas; distributed in 136,913 pairs of synonyms, 36,326 involving adjectives.

### 3.5 N-gram Corpus

In our experiments, we explore WPT05, a collection of over 10 million documents from the Portuguese web [2]. We used the n-grams (and their frequencies) generated from the documents in the collection with language automatically identified as Portuguese ($\sim$ 7 million documents, 26 Gb of text). We filtered the tokens with $length > 32$, but did not exclude n-grams from the set with low frequency. Given that we will be looking for the occurrence of multiple patterns with a given lemma in our classification process, low frequency n-grams can combine to produce high frequency patterns. In this research, we explored 8 million unigrams, 501 million trigrams, 984 million tetragrams and 1,321 million pentagrams. This corpus contains a large and representative sample of the Portuguese documents available on the Web, including a comprehensive range of types and genres of texts.

## 4 Identification of Human Adjectives

To identify human adjective candidates, we create a library of hand-crafted lexico-syntactic patterns representing elementary copular and adnominal constructions where such predicates can be found. These are then applied to WPT05, to gather evidence about adjective behavior. The adjective and pattern frequencies in the n-gram corpus are then used as input features to a binary classifier that we have trained and tested using the manually labeled adjectives.

### 4.1 Pattern Recognition

A total of 29 distinct lexico-syntactic patterns were formalized and applied to WPT05 trigrams, quadrigrams and pentagrams (Table 1). These patterns apply only to masculine and feminine singular forms, and some of them differ very slightly from each other, depending, for example, on the nature of the subject involved.

[1] http://www.dicsin.com.br/

**Table 1.**

| ID | Pattern | Matches |
|---|---|---|
| 301 | N COP ADJ | 14,474 |
| 302 | IHREF COP ADJ | 27,655 |
| 303 | HREF COP ADJ | 9,157 |
| 304 | ERGO COP ADJ | 23,337 |
| 305 | tu COP ADJ | 1,412 |
| 306 | IND HREF ADJ | 11,489 |
| 307 | IND ERGO ADJ | 31,933 |
| 308 | IND ADJ HREF | 1,206 |
| 309 | IND ADJ ERGO | 13,497 |
| 401 | N S COP ADJ | 7,240 |
| 402 | S S COP ADJ | 4,466 |
| 403 | N COP MODIF ADJ | 2,755 |
| 404 | N COP IND ADJ | 2,218 |
| 405 | ART ADJ DO N | 2,659 |
| 406 | ART ADJ DO ERGO | 5,846 |
| 407 | tu COP MODIF ADJ | 616 |
| 408 | HREF COP MODIF ADJ | 2,238 |
| 409 | ERGO COP MODIF ADJ | 4,170 |
| 410 | IREF COP MODIF ADJ | 11,994 |
| 411 | IREF COP IND ADJ | 3,957 |
| 412 | HREF COP IND ADJ | 1,102 |
| 413 | ERGO COP IND ADJ | 2,146 |
| 414 | IND HREF MODIF ADJ | 4,822 |
| 415 | IND ERGO MODIF ADJ | 4,030 |
| 501 | N S COP MODIF ADJ | 805 |
| 502 | S S COP MODIF ADJ | 465 |
| 503 | N S COP IND ADJ | 1,088 |
| 504 | S S COP IND ADJ | 667 |
| 505 | ART ADJ DO S S | 698 |
| | Total | 191,782 |

In these patterns, we match the subject position against dictionaries of Portuguese person names, and profession and position names (ERGO), such as *primeiro-ministro* (prime-minister) and *professor* (professor). Regarding names, we explored first names (N) and combinations of *first-name surname* (N S) and *surname surname* (S S). The subject position may also be filled by a human generic noun (HREF), such as *pessoa* (person) or *indivíduo* (individual) and by the person singular pronouns *ele* (he), *ela* (she), *você* (you), and *tu* (you), coded as IREF.

Within these constructions, adjectives can relate to their subject through the elementary copulative verbs (COP) *ser* and/or *estar* (both translatable by to be) and other aspectual variants (namely, *andar, continuar, ficar, permancer, encontrar-se, mostrar-se, reveler-se, tornar-se* and *viver*). We confine the tense in predicative constructions to simple present, past and future (third-person singular). The constructions invoking the second-person singular (tu, you) only consider simple present verb forms, the most representative when using direct address pronouns.

Furthermore, adjectives can be found in adnominal position, post-modifying or pre-modifying the noun they are correlated, which may be preceded by an indefinite article (IND). We also account for the possibility of adjectives filling the head of a cross-construction, linking to the noun they modify by the preposition de (of). Pre-modification of adjectives is also considered. We use of a small list of quantifier and intensifier adverbs (MODIF) that usually co-occur with human adjectives (e.g. *very, particularly, truly*).

Together, the patterns used in our experiments matched 191,782 different sequences, containing 8,579 different adjectival lemmas.

### 4.2 Classification

To refine the results provided by the lexico-patterns and filter out potential erroneous cases, we first explored, for each candidate human adjective, (i) the number of matches in the corpus, and (ii) the number and type of instantiated patterns.

For example, the adjective *ventilado* (*ventilated*) only matches once, while *impotente* (*impotent*) has a total of 97 matches, instantiating 9 different patterns. It is reasonable to infer that the adjective impotent is more prone to be considered as a valid human adjective than *ventilado*. It can also be assumed that the adjectives not matching any pattern in the entire collection are either rare in language or they have not a human behavior.

We train a statistical classifier to automatically distinguish high-human evidence (HHE) adjectives from low-human evidence (LHE) adjectives. The automatic classification is performed based on the following identified features: (i) frequency of each pattern, (ii) total number of instantiated patterns, (iii) frequency of matches, and (iv) lemma frequency in the n-gram collection. These attributes are associated to each recognized adjective from our original lexicon, including those whose semantic category is already known.

In our experiments, the training set was composed by 4,042 entries, distributed in the following proportion: 2,580 HHE adjectives, and 1,462 LHE adjectives. The HHE adjectives correspond to the lemmas in the corpus labeled as human in the original lexicon. LHE adjectives include both the adjectives recognized in the corpus that are assigned to the not-human attribute in the original lexicon, and the adjectives that do not occur in the n-gram collection, regardless of their prior semantic classification in the original lexicon.

## 5 Polarity Assignment

Once the human adjectives are identified in the lexicon, we focus on assigning polarities to the adjectives with high human evidence (HHE). The procedure starts by deriving a synonym graph, called a *syngraph*, where the nodes are the previously identified human lemmas and the edges represent synonymy relationships between lemmas. Each node in the syngraph is named as the concatenation of a lemma, its grammatical category and semantic class. This combination, which we designate henceforth as a qualified lemma, *qualiflemma*, was previously applied to the normalization of entries in dictionaries. By having a network of qualiflemmas, instead of just lemmas, we can prevent the propagation of synonymy relations between lemmas of distinct syntactic-semantic categories (homographs), and enables the assignment of different polarities to such lemmas.

To automatically assign polarities to the unlabelled qualiflemmas, we train a statistical classifier, which explores a feature vector extracted from the syngraph. We used 80% of the polar qualiflemmas for generating the syngraph, and saved the remaining 20% for learning a model to assign polarities to lemmas.

In the syngraph we have nodes with polarities,$-1, 0, 1, null$, where null designates unassigned polarity. The goal is to learn a model that predicts the polarity of a node with null polarity given the polarity information of its neighborhood. A qualiflemma in the syngraph with unassigned polarity may possibly have adjacent nodes exhibiting the four distinct polarities, making the decision complex. A situation where all the adjacent nodes have null polarity is quite common. However, we can attempt to observe across the adjacent nodes and assign polarities based on the polarities of the cloud of the connected qualiflemmas in the synonym graph. We capture that information by computing the shortest-paths and distances to the nearest nodes with assigned polarity.

The distances are computed using Dijkstra's shortest-path algorithm on a modified syngraph, to which we added three start nodes, labeled "1", "-1" and "0", each representing a polarity value. These are directly connected to all the nodes representing qualiflemmas with the same assigned polarity. The distances from each qualiflemma $q$, to each of these three start nodes correspond to the $dpos_q$, $dzer_q$ and $dneg_q$ features used in the subsequent statistical classification.

Besides these features, we also calculate three polarity weights ($wpos_q$, $wzer_q$, $wneg_q$) as the sum of the inverses of the distances of each node to the corresponding start node. For $wpos_q$, we have:

$$wpos_q = \sum_i \frac{1}{1 + dpos_i}$$

where the $i$ represent the nodes adjacent to $q$.

## 6 Evaluation

We use in all experiments the C4.5 (J48) classifier implementation of the Weka toolkit with default parameters and 5-fold cross validation [19]. There is no particular reason for picking this algorithm other than it enables the identification of the features that contribute to polarity assignment in a learning algorithm. To reduce the impact of unbalanced data in automatic classification, we use the SMOTE filter implemented in Weka, which creates new minority class examples by interpolating between existing minority instances. In the polarity assignment experiments, we used 36,326 pairs of synonyms, which were obtained from the publicly-available resources previously described.

The derived syngraph contains 5,063 nodes, of which 1,989 have a prior polarity (500 positive, 380 neutral, 1,109 negative). An inspection of this graph of human adjectives shows that deciding on the polarities based on the synonyms is not trivial. The graph is highly connected: its order is 7.15 and the counts of nodes directly connected to a node of positive/neutral/negative polarity are 4,340, 4,460 and 3,731, respectively.

### 6.1 Human Classifier

The generated models are able to correctly predict the semantic category of adjectives in 94% of the cases, with a recall also of 94%. These evaluation results reinforce both the adequacy of methodology proposed, and the pertinence of the previously identified features for classifying adjectives as having LHE or HHE.

### 6.2 Lexical Expansion

We started with a large set of human adjective lemmas with initial polarities and restricted the expansion to those lemmas with high evidence of usage in social judgments in our corpus. However, we were still able to uniformly expand the lexicon by 70% for all polarity classes with good accuracy.

### 6.3 Polarity Classifier

The learned model has an accuracy of 87%. The Chi-square test of independence for this contingency table indicates a significance level of 0.1% ($p - value < 0,001$). The effect of the SMOTE filter in the training of the model is a higher precision for the classification of positive and negative adjectives (88 and 89%, respectively). The most problematic cases involve, as expected, the neutral polarity, which is, in average, correctly assigned only in 82% of the cases. The highest recall is obtained for positive polarities (93%), and the lowest for neutral polarities (70%). For negatives, recall is 92%. We intentionally trained the model to improve prediction performance with the positive lemmas because we have found that positive opinions represent the best predictor of politicians' popularity [4].

The quality of our classification appears to be at least as good as the best recently reported results. Rao and Ravichandran [15] report F-measure values of 93.00% with their label-propagation method on Hindi (relations among lemmas extracted from WordNet) and 82.46% on French (relations extracted from the OpenOffice French Thesaurus). In their evaluation, however, they classified all the adjective lemmas instead of just the human lemmas. In addition, their evaluation only assigned two polarity values (positive and negative). To compare the performance of our method we reassigned the results of our classifier into two classes, by dividing the neutral lemmas between the two other classes proportionally to their frequency. The resulting average F-measure of our method is higher: 97.03%.

Our method appears to perform better, despite the differences between the two evaluation settings, namely different languages and different resources for synonymy extraction. Given that performance is lower when the synonymic relationships are not as well accurate (French OpenOffice thesaurus), we conjecture that the performance of our methods may too be highly sensitive to the accuracy of the relationships present in the *syngraph* upon which the polarity classification algorithm is based. The strategy that we adopted of restricting the classification to human lemmas and discarding the relationships between non-human lemmas in the graph may be the reason for the observed improved F-measure.

## 7 Conclusions and Future Work

The experiments show that the automatic identification and classification of human adjectives in a lexicon can be carried out successfully by obtaining evidence about their use in large corpora. This can be performed using specific sets of handcrafted lexico-syntactic patterns describing the contexts where such categories are expected to occur. The frequency of matches recognized by those patterns, together with the adjective frequency in corpora, proved to be important and a distinctive feature for automatically distinguishing high-human evidence adjectives from low-human evidence adjectives.

Other analyses need to be conducted with the proposed method in order to assess its robustness, such as measuring how accuracy is affected by the size of the initial lexicon or by the percentage of lemmas in that lexicon reserved for the synonyms graph construction. An updated version of the lexicon described in this paper is available from our website (`http://dmir.inesc-id.pt/reaction/SentiLex-PT_02_in_English`).

**Acknowledgements.** We are grateful to Carlos Costa for programming and validating part of the software. This work was partially supported by FCT (Portuguese re- search funding agency) under grant UTA-Est/MAI/0006/2009 (REACTION project), and scholarship SFRH/BPD/45416/2008. We also thank FCT for its INESC-ID multi-annual support.

## References

1. Fahrni, A., Klenner, M.: Old wine or warm beer: Target-specific sentiment analysis of adjectives. In: Symposion on Affective Language in Human and Machine, AISB Convention (2008)
2. Batista, D., Silva, M.J.: A statistical study of the wpt05 crawl of the portuguese web. In: FALA 2010: VI Jornadas en Tecnología del Habla and II Iberian SLTech Workshop, Vigo, Spain (2010)
3. Carvalho, P.: Análise e Representaço de Construçes Adjectivais para Processamento Automático de Texto: Adjectivos Intransitivos Humanos. PhD thesis, University of Lisbon (2007)

4. Carvalho, P., Sarmento, L., Teixeira, J., Silva, M.J.: Liars and saviors in a sentiment annotated corpus of comments to political debates. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 564–568. Association for Computational Linguistics, Portland (2011)
5. Choi, Y., Cardie, C.: Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In: EMNLP (2009)
6. Esuli, A., Sebastiani, F.: Determining term subjectivity and term orientation for opinion mining. In: E-ACL, Trento, Italy (2006)
7. Fellbaum, C.: WordNet: An Electronic Lexical Database (Language, Speech, and Communication). The MIT Press (1998)
8. Godbole, N., Srinivasaiah, M., Skiena, S.: Large-scale sentiment analysis for news and blogs. In: International Conference on WebLogs and Social Media (ICWSM 2007), Colorado, USA (2007)
9. Hatzivassiloglou, V., McKeown, K.: Predicting the semantic orientation of adjectives. In: ACL (1997)
10. Kamps, J., Marx, M., Mokken, R.J., de Rijke, M.: Using wordnet to measure semantic orienta-tion of adjectives. In: LREC (2004)
11. Kanayama, H., Nasukawa, T.: Fully auto-matic lexicon expansion for domain-oriented sen-timent analysis. In: ACL (2006)
12. Kim, S.M., Hovy, E.: Extracting opinions, opinion holders, and topics expressed in online news media text. In: ACL/COLING Workshop on Sentiment and Subjectivity in Text (2006)
13. Maziero, E., Pardo, T., Felippo, A., da Silva, B.D.: A base de dados lexical e a interface web do tep 2.0 - thesaurus eletrônico para o português do brasil. In: TIL (2008)
14. Oliveira, H.G., Santos, D., Gomes, P.: Extracção de relações semânticas entre palavras a partir de um dicionário: o papel e sua avaliação. Linguamática 2(1) (2010)
15. Rao, D., Ravichandran, D.: Semi-supervised polarity lexicon induction. In: EACL (2009)
16. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: EMNLP (2003)
17. Takamura, H., Inui, T., Okumura, M.: Extracting semantic orientation of words using spin model. In: ACL (2005)
18. Turney, P., Littman, M.: Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems 21(4) (2003)
19. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd edn. Morgan Kaufman, San Francisco (2005)