Analysis of unsupervised aspect term identification methods for Portuguese reviews

Mateus Tarcinalli Machado[1,2,4], **Thiago Alexandre Salgueiro Pardo**[2,4], **Evandro Eduardo Seron Ruiz**[3,4]

[1]Programa de Pós-Graduação em Computação Aplicada

[2]Departamento de Ciências da Computação, ICMC

[3]Departamento de Computação e Matemática, FFCLRP

[4]Universidade de São Paulo, Brazil.

[mateusmachado, taspardo, evandro]@usp.br

Abstract. *Automatic term identification is a fundamental task for text mining and many other natural language processing applications. Based on the tradition of statistical and symbolic methods, whether adopting or not machine learning techniques, new hybrid proposals for term extraction have arisen. In this paper we investigate the above approaches in an unsupervised way to identify relevant terms in Portuguese, focusing on book reviews. Terms, in this context, usually represent opinion aspects, whose identification is a challenging and an equally important task. We show that current methods have varied performances, but that we may significantly improve their results by considering the distributional properties of words.*

Resumo. *A identificação automática de termos é fundamental na mineração de textos e em aplicações de processamento de linguagem natural. Com base em métodos estatísticos e simbólicos, adotando ou não técnicas de aprendizado de máquina, surgiram novas propostas híbridas para extração de termos. Neste trabalho, investigamos essas abordagens de forma não supervisionada para identificar termos relevantes em Português, com foco em revisões de livros. Os termos, neste contexto, geralmente representam aspectos de opinião, cuja identificação é uma tarefa desafiadora e igualmente importante. Mostramos que métodos atuais têm desempenhos variados, mas que seus resultados podem ser melhorados considerando as propriedades distribucionais das palavras.*

1. Introduction

Besides programming machines with basic linguistic knowledge and communication skills, one of the main obstacles for a comprehensive human-computer interaction is the lack of ability to fully understand human language in a particular context. This is true even for human communication, as many different daily situations (from simpler ones, as going to a social meeting or sharing a comment about a product on the web; to more complex ones, as negotiating the details of a business) demand specific knowledge about the circumstances and the usual communication protocol to follow. For computers to perform well in such conditions, they need to acquire the necessary domain knowledge, as humans do.

A small but significant step towards this aspiration may be given by the field of Automatic Term Extraction, also known by the ATE acronym. Term extraction is the task of identifying a vocabulary, or relevant words in a domain [Frakes et al. 2014]. Recently, ATE has gained special attention in several tasks in the Natural Language Processing (NLP), Machine Learning and related areas, including, e.g., applications in biomedical informatics, where ATE is used for translation of technical terms [Lossio-Ventura et al. 2016], and applications that use ATE on electronic textbooks for creating educational ontologies for learning systems [Conde et al. 2016]. In particular, ATE is very relevant in the more recent area of opinion mining and sentiment analysis, more specifically, in the area of aspect-based sentiment analysis, which aims at analyzing a large amount of data and synthesizing relevant information about aspects evaluated by users (e.g., the technical aspects of smartphones, as screen quality and battery life). In this context, terms usually represent aspects, and the precise identification of aspects of the evaluated entity is crucial to accurately assign them their corresponding sentiments/polarities [Schouten and Frasincar 2016, Pavlopoulos and Androutsopoulos 2014].

The importance of opinion mining and sentiment analysis has increased with the Web 2.0 and the upload of user generated text in various on-line services, such as chat/messengers, blogs and micro-blogs, social media communication, e-commerce and product review sites. All these services contributed to such an amount of unstructured data that demanded new ways to extract, load, transform and analyze the information content [Ravi and Ravi 2015]. Nowadays, understanding the opinions of the public on distinctive matters, such as political engagement, product preferences and social events, to name only a few, is an industry concern and also an academic research opportunity. It is also a relevant task to the final customer in situations where s/he needs supporting information for taking a decision about buying a product, reading a book, or subscribing to a service, among other situations. Correctly identifying the *aspect terms* is very important, therefore.

This paper investigates three different methods for ATE, as well as some variations and combinations. We investigate: 1) a frequency-based baseline method (which we refer by `FREQ`); 2) the Hu & Liu hybrid method [Hu and Liu 2004b]; and 3) the enhancement of the two previous methods with a word distributional approach, the word2vec method [Mikolov et al. 2013] (referred by W2V from here on). We evaluate how good the methods are in finding the aspects in book reviews written in Portuguese. This paper presents the following novelties related to the state-of-the-art in ATE: a) it evaluates the three above ATE approaches in the first and only known Portuguese annotated corpus for aspects; and b) the Hu & Liu method was enhanced by processing infrequent terms with the word2vec approach. We show that the three researched methods for ATE present varied performances, but that the distributional properties of words may improve the results.

Section 2 reports some related research efforts. In Section 3, the data and the ATE methods that we apply are explained. Experimental results are reported and discussed in Section 4. Some conclusions and final remarks are presented in Section 5.

2. Related work

One of the first comprehensive researches about ATE dates back to middle 1990, when Kageura and colleagues [Kageura and Umino 1996] report about a number of approaches to what they called automatic term recognition (ATR). At the time, the techniques adopted word statistics to generate a list of index terms for a given document. Statistical approaches explore the fact that the words composing a term tend to be found closer to each other. Later, Castellvi and collaborators [Castellví et al. 2001] analyzed twelve terminology extraction tools, including symbolic and hybrid ATE approaches.

Frantzi and Ananiadou [Frantzi and Ananiadou 1997] suggest a C-value method, using statistical weights, for the extraction of multi-word terms. We know the majority of domain specific terms consist of multiple words, also known as uninterrupted collocations. The C-value method incorporates context information for the evaluation of candidate terms. It was reported later [Frantzi et al. 2000] that the C-value method increased the precision on selected terms from figures of 30% to figures in a range of 69–74%. About 85% of domain specific terms are said to be compound nouns [Nakagawa and Mori 2002].

Besides the use of statistical approaches for terminology identification, there are also the symbolic approaches based on syntactic description of terms. Hu & Liu [Hu and Liu 2004b] adopted the symbolic approach for ATE with a subtle use of statistics. They specifically look for noun phrases, which usually produce a high recall, limiting their number by their frequency of occurrence, which is the basis for pruning the term possibilities. De Clercq and colleagues [De Clercq et al. 2015] have also applied a hybrid ATE strategy. They built a system called TExSIS. The system applies a POS tagger, a lemmatizer and a chunker. All the nouns and noun phrases are considered candidate terms. Then, they use statistical filters to determine the specificity of and cohesion between these candidate terms.

A more closely related work to the one reported here has been the one of Pavlopoulos & Androutsopoulos [Pavlopoulos and Androutsopoulos 2014]. Their work focus is ATE, but it comes as one of the steps towards their research for aspect-based sentiment analysis (ABSA). To perform ATE, the authors applied both, a statistical (`FREQ` Baseline) and a symbolic method (Hu & Liu), and they have also used the distributional approach for word representation W2V [Mikolov et al. 2013] to prune these prior two techniques. This paper differs form the previous one in two points: (i) we applied the W2V method not only as a pruning strategy after the Hu & Liu method, but also for the resulting infrequent terms; and (ii) we have tested this sequence of ATE methods for book reviews in Portuguese, which shows to be a very challenging domain.

A survey on ATE for the Portuguese language [da Silva Conrado et al. 2014] has shown few NLP research groups are investigating the task of term extraction for Portuguese [Lopes et al. 2009, Lopes et al. 2010, da Silva Conrado et al. 2013, Conrado et al. 2013] and none of them has used the approach that is presented here, which is detailed in what follows.

3. Material and methods

3.1. The dataset

All the methods cited before will be detailed below and have been applied to the ReLi corpus [Freitas et al. 2014], a Portuguese book review corpus consisting of 1,600 reviews for 14 distinctive books, containing 12,470 Portuguese sentences and 259,978 words. ReLi had its sentences manually tagged by linguistic experts for their aspects and opinions/polarities. Among these sentences, there are those which express opinions about either a book or a particular aspect of the book. There are also neutral sentences, without polarity. Overall, there are 2,883 positive opinion sentences, 596 negative opinion sentences and 212 neutral sentences.

The evaluation of the methods was performed by comparing the returned targets (the identified aspects) to the corresponding reference/gold targets annotated in the corpus. The evaluation methodology presented in this article is similar to the one used for the SemEval workshops (see, e.g., [Pontiki et al. 2015]). The targets were extracted with their respective starting and ending offsets locations in the sentence. Only distinct targets were considered, and NULL targets were discarded. Precision and recall were computed, using a micro-averaging method [Sebastiani 2002]. F-measure, a harmonic mean of precision and recall, were also computed.

Considering a contingency table and the classic information retrieval notions of precision (P) and recall (R), an evaluation measure is based on the number of true positives (tp), true negatives (tn), false positives (fp) and false negatives (fn). In this way, micro-averaging P^{μ} and micro-recall R^{μ} may be achieved by summing all individual decisions, such as:

$$P^{\mu} = \frac{\sum_{c=1}^{C} \text{tp}_i}{\sum_{c=1}^{C} (\text{tp}_i + \text{fp}_i)} \qquad R^{\mu} = \frac{\sum_{c=1}^{C} \text{tp}_i}{\sum_{c=1}^{C} (\text{tp}_i + \text{fn}_i)}$$

where C is the number of c classes under evaluation and i represents individual instances in these classes.

3.2. The methods

We implemented and evaluated three approaches for ATE, which are summarized below.

`FREQ` Baseline This ATE method (referred simply by `FREQ` here) is perhaps the simplest form of term identification. Is is entirely based on the frequency of occurrence of nouns and noun phrases. It returns the most frequent nouns and noun phrases in a text after a pruning mechanism, keeping only the words with a certain frequency to limit the eventual error of labeling every noun or noun phrase as a relevant term. This method has been reported as very effective by various researchers [Hu and Liu 2004a, Wei et al. 2010, Liu 2012].

Hu & Liu This method may be categorized as a hybrid term detection method because it is based on the frequency strategy over tokens previously tagged. This method has the following advantages over the `FREQ` method: it detects longer and compound terms; it may detect infrequent terms; and it tries to correct itself by pruning these incorrect terms.

The Hu & Liu algorithm may be summarized in eight steps, synthesized below:

1. Frequent term detection: nouns and noun phrases should be extracted from sentences using a POS tagger and a parser. This method ignores terms below a frequency less than 1% in the corpus. In our implementation, we vary this frequency between 1 and 10% to choose a more suitable value.
2. Identification of compound noun terms: noun bigrams and triples of nouns in a sentence shall be extracted in the same order they appear. The resulting list of nouns and noun phrases includes the *candidate terms*. For example: the word 'campo' may appear many times in a text, but this word comes always inside the expression "O apanhador no campo de centeio". This way, the triple 'apanhador campo centeio' will be extracted as a compound candidate term.
3. p-support (pure support) calculation: the p-support of a candidate term t is the number of sentences containing t, excluding sentences where another candidate t' appears and the t candidate is a subset of t'.
4. Compactness score calculation: in a sentence that contains a compound candidate term, the distance in words between each pair of terms of the candidate is verified. If this distance is greater than 3 words, the score is increased.
5. Compactness pruning: term candidates that appear more than once in a non-compact form are removed.
6. p-support pruning: candidates with a p-support value less than three are also removed.
7. Extraction of Opinion words: all the adjectives that appear close to a term candidate shall be extracted.
8. Infrequent term detection: all sentences without any term are reanalyzed and, if an opinion word is found, its closest noun is added to the list of candidate terms.

We have also applied the Word2Vec model. It was used as a subsidiary pruning method to the previous `FREQ` and Hu & Liu methods. The W2V model was designed by Mikolov and his colleagues at Google [Mikolov et al. 2013] as an efficient method for learning word vector representations from large corpora. The W2V method is based on an earlier distributed representation of words, the vector space model (VSM), in which words are represented as dense vectors in a continuous vector space. In the W2V model, the vectors are semantic representations of words and are treated as having latent variables, each one representing a "dimension" of the word meaning. In this representation, semantically similar words have vector representations close to one another. One may also say that semantically similar words are *embedded nearby each other*. Before W2V, the adopted Distributional Hypothesis was based on the statistics of how often some word co-occurred with its neighbor words in a large text corpus. In a different way, W2V uses a predictive model, a computationally efficient predictive neural probabilistic language model. Predictive models directly attempt to predict a word from its neighbors in terms of learned embedded vectors.

Formally, neural probabilistic language models are trained using the maximum likelihood principle to maximize the probability of the next word, w_{t+j}, in a neighborhood $|c|$, given the previous words w_j. Given a sequence of training words $(w_1, w_2, w_3, \ldots, w_T)$ in a corpus with T words, the objective of the W2V model is to maximize the following average $\log$ probability:

$$J_{ML} = \frac{1}{T}\sum_{t=1}^{T}\sum_{-c<j<c, j\neq 0}\log \mathrm{p}(w_{t+j} \mid w_j)$$

In the above formulation, c refers to the size of the training context for the center word w_t. Larger values of c result in more training samples, which, as a consequence, may lead to a higher accuracy model. In other words, the aim of this method is to try to correlate words in a vocabulary by measuring their compatibility via probability values.

Each term vector may be constructed by training a probabilistic language model from a corpus. For Portuguese texts, we used the model trained by Rodrigues and his colleagues [Rodrigues et al. 2003], using data from Wikipedia, scientific and journalistic texts, lectures, movie subtitles, technical texts, and other documents.

As the W2V approach is used for pruning unrelated term lists, it was necessary to calculate a couple of specific term vectors, known as the "context" vector and the "common" vector. The former is a term vector extracted from the top 10 most frequent terms of the corpus under analysis. It is meant to be a representative vector of the context information in the ReLi corpus. The latter is a term vector created from the top 20 most frequent terms in a corpora representing the Portuguese language as a whole. Its purpose is to serve as a representative vector of the common terms of a language. Both, the context and the common vectors, are centroid vectors. The distinction in the number of most frequent terms to create both term vectors is based on the intuition that the candidate terms are surrounded by more general concept terms than the domain-specific terms. In order to know if one term belongs to the book context, its representative vector must be compared to both vectors using the traditional cosine similarity. If the result is closer to the context vector the term is maintained, if it is closer to the common vector the term is discarded.

`FREQ` Baseline + W2V Pruning In the `FREQ` model for single terms, words are treated as discrete atomic symbols. In a different direction, the W2V model steers words towards their meanings. We believe that, using the W2V model, the `FREQ` method may have its threshold pruning step extended or even enhanced.

Besides the frequency pruning of the `FREQ` method, an extra pruning may be used to further reduce the number of non-related detected terms. This extra pruning step is sustained by the concept of *term vector*, which is a vector representative of a term on its context. The terms found by the `FREQ` Baseline method were pruned based on their distance to each context vector, as explained before.

Hu & Liu + W2V Pruning We have also modified the Hu & Liu method to add W2V pruning. This method has already a standard pruning step in its first stage, called *Frequent term detection*. After this first pruning, the W2V method was also applied as suggested by [Pavlopoulos and Androutsopoulos 2014]. With this second pruning, the intention was to remove the non-context terms.

Hu & Liu + W2V Pruning + Infrequent term processing with W2V For each term located near an opinion word, a W2V analysis is performed. If the term belongs to the context under study (books), it is added to the infrequent term list; otherwise, the algorithm moves on to the analysis of the next closest term. This process is repeated until a contextualized term is found and, if it happens that no term is found, the algorithm considers as infrequent term the one closest to the opinion word.

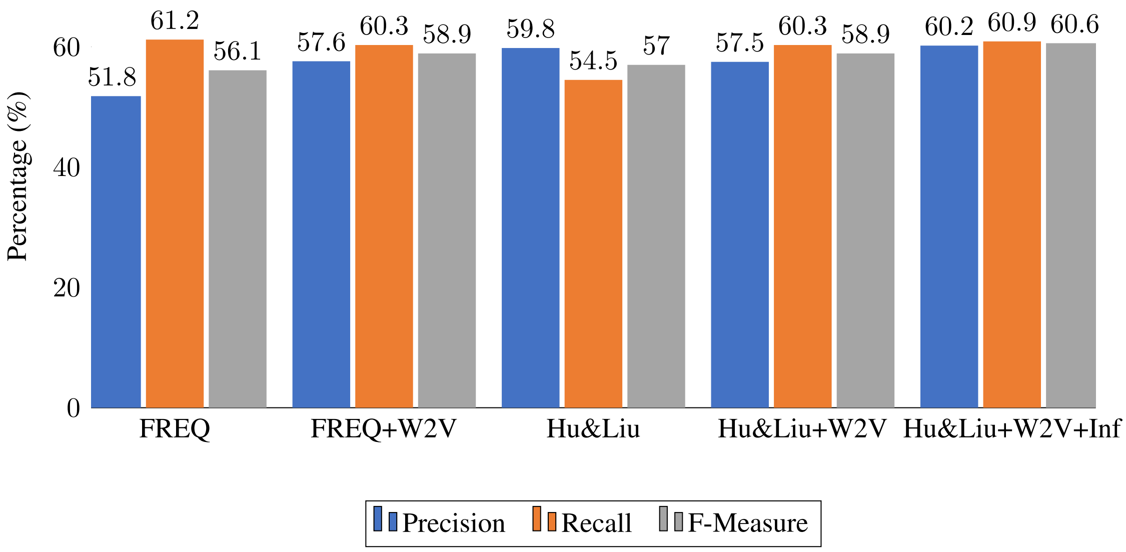
4. Results

We analyzed the behavior of the algorithms in two situations: a) with a relatively small dataset containing 2,233 opinion phrases from ReLi corpus, and b) with an extended dataset with additional 51,149 sentences obtained from book reviews collected from the skoob[1] site, the same site that originated the ReLi corpus.

It is worth recalling that, as a preliminary step, both corpora have been submitted to a pre-processing phase, when a tokenizer [Bird et al. 2009] was applied at the sentence level. For part-of-speech tagging, the `nlpnet` POS tagger [Fonseca and Rosa 2013] was applied.

The results obtained for the statistical, symbolic and distributional methods are shown in Figure 1 for the ReLi Corpus and in Figure 2 for the Extended ReLi Copus. A detailed description of how the methods were applied and their results are presented below.

Figure 1. ReLi Corpus results.

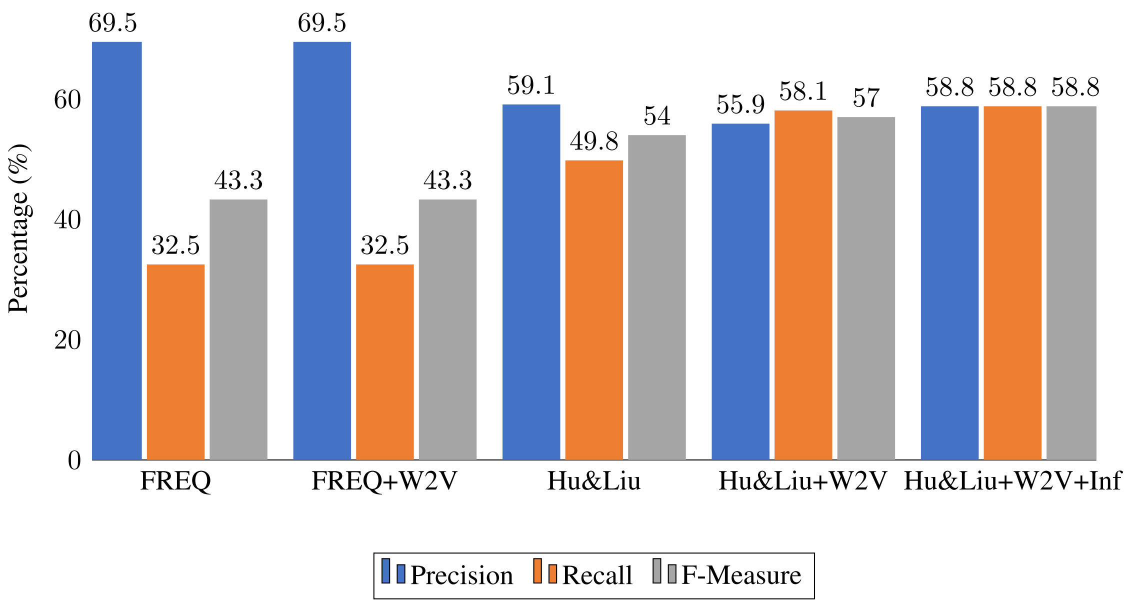


4.1. `FREQ` Baseline

All nouns found on the pre-processing phase were organized in descending order by their frequencies. All terms below a certain frequency were excluded. To identify this cutoff

[1]https://www.skoob.com.br/

Figure 2. Extended ReLi Corpus results.



frequency, the datasets were processed with different values, in a range of 1% to 10% to find the best fit value for the F-measure. Figures 1 and 2 show very contrasting values for both corpora, the standard ReLi and its extended version. While recall showed better results for the ReLi corpus, the same did not happen for the extended ReLi, where we obtained the highest precision value and also the lowest values for recall and f-measure of all the tests. This result may be explained by the fact that the `FREQ` method is purely a frequency based method. When the `FREQ` is executed over a large dataset, the amount of resulting terms tends to be smaller than if processed for a smaller corpus, but the results usually comes with a better quality, thus implying higher values for precision and lower values for recall.

4.2. `FREQ` Baseline + W2V Pruning

The W2V pruning of non-context terms enhanced the precision (from $\approx$51.8 to $\approx$57.6) and f-measure (from $\approx$56.1 to $\approx$58.9) when applied only to the standard ReLi corpus. The words ‘mundo’, ‘tempo’ and ‘forma’ are examples of non-context terms that have been correctly excluded from the final set of detected aspects. With the extended corpus, the pruning step had no effect. See Figures 1 and 2. In a larger corpus, non-context terms were naturally eliminated due to their relative lower frequency, thus nulling the effect of the W2V pruning stage.

4.3. Hu & Liu

The Hu & Liu method has the advantage of detecting compound terms, having three pruning stages (by frequency, compactness, and p-support) and detecting infrequent terms. Words like ‘titulo’ and ‘texto’, that are not usually so frequent, have been now correctly identified. For the ReLi corpus, precision values increased. Although this method is a more sophisticated ATE method compared to the `FREQ` Baseline, some non-context terms still appeared, which led to a decrease in recall. For the extended ReLi corpus, recall increased but precision had a considerable decrease.

4.4. Hu & Liu + W2V Pruning

The effect of this pruning step is clear in finding terms such as 'personagem', 'Holden' and 'obra'. As also seen in Figures 1 and 2, the W2V pruning step enhanced the prior recall results of the stand-alone Hu&Liu and decreased precision values for both corpora. However, the value of F-measure improved, indicating that, in general, the use of this new pruning approach was positive.

4.5. Hu & Liu + W2V Pruning + Infrequent term processing with W2V

Differently from [Pavlopoulos and Androutsopoulos 2014], we tested the W2V pruning in the last step of the Hu & Liu method, on the infrequent term list, as explained before. The improvement with this is step over the infrequent terms may be exemplified for discarding the term 'justificativa' which is not a related infrequent aspect. It also correctly chooses 'texto' instead of 'sacadas' and 'enredo' instead of 'adolescente'. Figures 1 and 2 shows the obtained improvements. The best values were obtained for all the measurements, disregarding the one obtained in the `FREQ` with the extended ReLi, result explained on Section 4.1.

5. Conclusions and final remarks

Our results show that the tested methods obtain varied performances for the two corpora, the standard ReLi and its extended version. However, one may notice that incorporating the distributional properties of words (the W2V strategy) generally improves the results. Overall, we may see that the distributional approach of W2V has caused an impact in both the statistical `FREQ` Baseline method and the hybrid Hu & Liu method. Applying the word2vec has increased the precision of these methods up to 10 percent points (pp), showing the advantages of considering distributional semantics in the process.

A great advantage of pruning with W2V is the removal of context unrelated terms. When this removal was applied only to the infrequent term candidates, one may see a general decrease in the measures, showing that most of these infrequent terms are also part of the context under study. The removal of infrequent term candidates after W2V increased the results by a very small margin for the ReLi corpus.

To the interested reader, examples of correctly detected aspects on ReLi corpus are: 'livro', 'história', 'narrativa', 'personagem', and 'escrita'. Some examples of non-detected aspects are 'o apanhador no campo de centeio' and 'adolescente do livro', which were expressions that appeared very few times.

Acknowledgements

This research was partially supported by FAPESP.

References

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc.

Castellví, M. T. C., Bagot, R. E., and Palatresi, J. V. (2001). Automatic term detection: A review of current systems. *Recent advances in computational terminology*, 2:53–88.

Conde, A., Larrañaga, M., Arruarte, A., Elorriaga, J. A., and Roth, D. (2016). Litewi: A combined term extraction and entity linking method for eliciting educational ontologies from textbooks. *Journal of the Association for Information Science and Technology*, 67(2):380–399.

Conrado, M. S., Rossi, R. G., Pardo, T. A. S. P., and Rezende, S. O. (2013). Applying transductive learning for automatic term extraction: The case of the ecology domain. In *2013 Second International Conference on Informatics Applications (ICIA)*, pages 264–269.

da Silva Conrado, M., Felippo, A. D., Salgueiro Pardo, T. A., and Rezende, S. O. (2014). A survey of automatic term extraction for Brazilian Portuguese. *Journal of the Brazilian Computer Society*, 20(1):12.

da Silva Conrado, M., Pardo, T. A. S., and Rezende, S. O. (2013). A machine learning approach to automatic term extraction using a rich feature set. In *HLT-NAACL*, pages 16–23.

De Clercq, O., Van de Kauter, M., Lefever, E., and Hoste, V. (2015). Applying hybrid terminology extraction to aspect-based sentiment analysis. In *International Workshop on Semantic Evaluation (SemEval 2015)*, pages 719–724. Association for Computational Linguistics.

Fonseca, E. R. and Rosa, J. L. G. (2013). Mac-morpho revisited: Towards robust part-of-speech tagging. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, pages 98–107.

Frakes, W. B., Kulczycki, G., and Tilley, J. (2014). *A Comparison of Methods for Automatic Term Extraction for Domain Analysis*, pages 269–281. Springer International Publishing, Cham.

Frantzi, K., Ananiadou, S., and Mima, H. (2000). Automatic recognition of multi-word terms:. the C-value/NC-value method. *International Journal on Digital Libraries*, 3(2):115–130.

Frantzi, K. T. and Ananiadou, S. (1997). Automatic Term Recognition using Contextual Cues. In *In Proceedings of 3rd DELOS Workshop*.

Freitas, C., Motta, E., Lilidiú, R. L., and César, J. (2014). Sparkling Vampire...lol! Annotating Opinions in a Book Review Corpus. In Aluísio, S. and Tagnin, S. E. O., editors, *New Language Technologies and Linguistic Research: A Two-Way Road*, chapter 8, pages 128–146. Cambridge Scholars Publishing, Cambridge.

Hu, M. and Liu, B. (2004a). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168–177. ACM.

Hu, M. and Liu, B. (2004b). Mining opinion features in customer reviews. In *AAAI*, volume 4, pages 755–760.

Kageura, K. and Umino, B. (1996). Methods of automatic term recognition: A review. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 3(2):259–289.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.

Lopes, L., Fernandes, P., Vieira, R., and Fedrizzi, G. (2009). EχATOLP–An Automatic Tool for Term Extraction from Portuguese Language Corpora. In *Proceedings of the 4th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 427–431.

Lopes, L., Oliveira, L. H., and Vieira, R. (2010). Portuguese term extraction methods: Comparing linguistic and statistical approaches. In *International Conference on Computational Processing of Portuguese Language*, volume 6.

Lossio-Ventura, J. A., Jonquet, C., Roche, M., and Teisseire, M. (2016). Biomedical term extraction: overview and a new methodology. *Information Retrieval Journal*, 19(1-2):59–99.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Nakagawa, H. and Mori, T. (2002). A simple but powerful automatic term extraction method. In *COLING-02 on COMPUTERM 2002: Second International Workshop on Computational Terminology*, volume 14, pages 1–7. Association for Computational Linguistics.

Pavlopoulos, J. and Androutsopoulos, I. (2014). Aspect term extraction for sentiment analysis: New datasets, new evaluation measures and an improved unsupervised method. *Proceedings of LASMEACL*, pages 44–52.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado. Association for Computational Linguistics.

Ravi, K. and Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89:14–46.

Rodrigues, J., Branco, A., Neale, S., and Silva, J. (2003). LX-DSemVectors: Distributional Semantics Models for Portuguese. *6th International Workshop PROPOR'2003, Faro, Portugal, June 2003*, 8775(2721):214–219.

Schouten, K. and Frasincar, F. (2016). Survey on Aspect-Level Sentiment Analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):813–830.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1–47.

Wei, C.-P., Chen, Y.-M., Yang, C.-S., and Yang, C. C. (2010). Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews. *Information Systems and E-Business Management*, 8(2):149–167.