NILC at ABSAPT 2022: Aspect Extraction for Portuguese

Mateus Tarcinalli Machado[1], Thiago Alexandre Salgueiro Pardo[1]

[1]*University of São Paulo, Interinstitutional Center for Computational Linguistics (NILC), Institute of Mathematical and Computer Sciences, Av Trabalhador São-carlense, 400, São Carlos, São Paulo, Brazil*

Abstract

In this paper, we describe the adopted method to extract aspects of the Aspect-Based Sentiment Analysis in Portuguese (ABSAPT) task proposed in the context of the evaluation campaign of IberLEF 2022. We used the Conditional Random Fields (CRF) machine learning algorithm combined with a post-processing step achieving an accuracy of 0.97. We performed a detailed analysis of the erroneously detected and non-detected aspects, with both lists formed only by implicit aspects. For the analysis, we used a refined typology of the kind of knowledge necessary to identify implicit aspects. Finally, we conclude that with a simpler machine learning algorithm like CRF it is possible to achieve good results with a lower processing cost than deep learning methods. The main aspect identification errors are related to implicit aspects, evidencing the greater difficulty in finding this type of aspect.

Keywords

Aspect-based sentiment analysis, Aspect Extraction, Implicit aspects

1. Introduction

We can define sentiment analysis as an area of study at the intersection between Computer Science and Linguistics that aims to automatically analyze text [1] to understand people's opinions, feelings, assessments, attitudes, and emotions concerning products, services, organizations, individuals, issues and topics [2].

One of the first areas of application of sentiment analysis was the evaluation of products, conducted through the analysis of opinion texts obtained from social networks or specialized websites. This analysis can be done at a more general (about the product) or more specific level (about the characteristics of the product). From this need to qualify specific characteristics of each item or object analyzed, the area known as aspect-based sentiment analysis (ABSA) was born. ABSA performs deeper analysis on opinion texts, identifying and relating sentiments and aspects [3] of a given entity.

Most systems that perform aspect-based sentiment analysis divide the problem into three steps [4]:

- Extraction of aspect terms: extracts the characteristics (aspects) about the analyzed entity (product). For example, about a hotel, we may have aspects related to its facilities, location,

IberLEF 2022, September 2022, A Coruña, Spain.
mateusmachado@usp.br (M. T. Machado); taspardo@icmc.usp.br (T. A. S. Pardo)
0000-0003-1848-3082 (M. T. Machado); 0000-0003-2111-1319 (T. A. S. Pardo)
© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3202/absapt-paper2.pdf

and employees, among others [2, 5, 6, 7].

- Aspect term sentiment estimation: this step estimates the polarity (positive, negative, or neutral) and possibly the intensity (e.g., strongly negative, moderately positive) of opinions for each aspect term of the target entity [2, 8, 9].
- Aspect aggregation: some systems group aspect terms that are synonymous ("price", "cost") or, more generally, group aspect terms to obtain aspects of a more general granularity ("concierge", "manager" and "receptionist" can be replaced by "employees") [2, 5, 10, 11].

Something that directly influences aspect-based sentiment analysis results is how aspects appear in texts. They can be explicitly mentioned, as in the example "The location of the hotel is excellent". In other situations, the aspect is implicitly mentioned, as in "It's close to the city center", where the same location aspect is indirectly mentioned, making the aspect detection task more challenging.

This paper concerns our participation at IberLEF 2022 [12], a shared evaluation campaign for Natural Language Processing (NLP) systems In Spanish and other Iberian languages. In particular, we have participated in the Aspect-Based Sentiment Analysis in Portuguese (ABSAPT) task in its Aspect Term Extraction sub-task. We describe our aspect extraction method using a classifier based on the Conditional Random Fields algorithm and a post-processing step. We also present here a detailed analysis of the errors found by the proposed method, especially regarding the implicit aspects and a refined typology of necessary knowledge to identify them.

The rest of the paper is organized as follows. Section 2 presents the data and methods used for the task. Section 3 shows the results and analysis of errors. Section 4 presents some final remarks.

2. Data and methods

In our most recent work [13], we analyzed fourteen methods for aspect extraction, with a focus on implicit aspects. We processed opinion texts from camera, book, smartphone, and hotel domains. In the hotel domain (which is the one related to the ABSAPT evaluation), we analyzed the dataset of the work of [14] in conjunction with the annotation of implicit aspects that we conducted [15].

For the task of extraction of aspects, we tested frequency-based (FreqBaseline [16]) and frequency- and rule-based hybrid methods (Hu & Liu [16]), both extended with the use of a pruning step to remove unrelated aspects using the Word2Vec [17] distributional model [4, 18]. We also applied the Multinomial Naive Bayes [19], Passive Aggressive [20], Perceptron [21], and Stochastic Gradient Descent [22] machine learning algorithms using the bag-of-words model as features. For the Conditional Random Fields [23] algorithm, we evaluated the tokens and their neighbors with their respective POS tags as features. In the last experiment, we tested pre-trained BERT models, where we performed in-domain and cross-domain experiments with multilanguage[1] [24] and Portuguese[2] [25] models.

[1] https://huggingface.co/bert-base-multilingual-uncased
[2] https://huggingface.co/neuralmind/bert-base-portuguese-cased

Table 1
Results for the hotel domain

Method	**Precision**	**Recall**	**F1**	**% implicit aspects**
FreqBaseline	0.74	0.51	0.61	0.12
FreqBaseline + Word2Vec	0.72	0.52	0.60	0.15
Hu & Liu	0.78	0.46	0.58	0.08
Hu & Liu + Word2Vec	0.79	0.47	0.59	0.09
Hu & Liu + Infrequent Word2Vec	0.82	0.45	0.58	0.08
Multinomial Naive Bayes	0.49	0.56	0.52	0.14
Passive Aggressive	**0.98**	0.64	0.77	0.18
Perceptron	0.94	0.64	0.76	0.18
SGD	0.97	0.58	0.73	0.13
Conditional Random Fields	0.94	0.71	**0.81**	0.23
BERT	0.89	**0.72**	0.79	**0.26**
BERT-cross	0.88	0.68	0.77	0.24
BERT-pt	0.90	0.70	0.78	0.25
BERT-pt-cross	0.90	**0.72**	0.80	0.25

Table 1 presents the results obtained in these experiments for the hotel domain, and we can see that the Passive Aggressive machine learning method had the best precision, the pre-trained BERT language model had the best recall, and the Conditional Random Fields (CRF) had the best F-measure. The last column shows the percentage of implicit aspects that each method identified. As the F-measure is a harmonic average between the two previous measures, we chose the CRF to carry out the experiments of ABSAPT task, in addition to having the possibility of analyzing other combinations of features in the training of models with this algorithm. CRF was also the best method for extracting the implicit aspects.

2.1. Corpus

For the ABSAPT task, the organizers made available a corpus developed by Freitas [26] and Correa [27], containing reviews about accommodation service companies, written in Portuguese. The authors annotated the aspects found and their respective polarities in each review, following the same guidelines in both works. The final dataset contains 847 reviews, comprising 3,090 aspects (with 77 unique aspects).

2.2. Pre-processing

In the pre-processing step, we prepare the texts for applying the machine learning method. In this experiment, we divided the reviews into sentences and then into tokens. We also did the morphosyntactic analysis using the part of speech tagger of the spaCy [28] module with the model pt_core_news_lg[3]. After this step, we converted this data to the IOB format (short for

[3]https://spacy.io/models/pt

Inside, Outside, and Beginning annotation format). To identify the aspects, we used the tags "B-ASP" for Begin of Aspect, "I-ASP" for Inside of Aspect, and "O" for Outside of Aspect.

2.3. Method for aspect extraction

As we explained earlier, we chose the Conditional Random Fields algorithm for this experiment. It uses probabilistic graph models that consider characteristics of words and their surroundings, suitable for segmenting and labeling data sequences [23], which are desired characteristics for the aspect identification task.

In the experiment, we used the module sklearn-crfsuite [29]. As features for training the model, we choose the tokens, their neighbors and respective morphosyntactic labels. We divided the dataset into a training set with 70% of the sentences and a test set with the remaining 30%. For optimization, we used the random search method, given the continuous nature of the parameters c_1 and c_2 of the algorithm implementation. We performed the training and validation steps 1,000 times with random values for these variables and selected the best result by the accuracy measure, which was 0.93 for c1 of 0.3914 and c2 of 0.0593, all approximated values.

The next step was to divide the dataset into five parts to perform and validate the tests in 5-folds. Using the parameters found with the random search, we found an average accuracy of 0.96 as shown in Table 2. We then analyzed the sets of aspects found, not found, and those erroneously found, as follows.

- Correctly identified aspects : '*aeroporto*', '*almoço*', '*apartamento*', '*apartamentos*', '*atendimento*', '***banheira***', '*cafeteira*', '*café da manhã*', '*cama*', '*camas*', '*carpete*', '*carpetes*', '*casal*', '*cassino*', '*cassinos*', '*centro da cidade*', '***chuveiro***', '***cidade***', '*colchão*', '*conforto*', '*corredor*', '*corredores*', '*cozinha*', '*custo-benefício*', '*ducha*', '*duchas*', '*elevador*', '*elevadores*', '*escadas*', '***estacionamento***', '*família*', '*famílias*', '*farmácia*', '*farmácias*', '*frigobar*', '*funcionários*', '*garagem*', '***gerente***', '*horário*', '*horários*', '*hotel*', '*instalações*', '*internet*', '***jantar***', '*lavanderia*', '*limpeza*', '*localização*', '*lojas*', '*luxo*', '***museu***', '*móveis*', '*padrão*', '*pia*', '*piscina*', '*piscinas*', '*praia*', '***praça***', '*preço*', '*preços*', '*quarto*', '*quartos*', '*recepção*', '*rodoviária*', '*roupa de cama*', '*rua*', '*ruas*', '***secador***', '*serviço*', '*serviço de quarto*', '*serviços*', '***shopping***', '*suíte*', '*suítes*', '*teatro*', '*teatros*', '*telefone*', '*televisão*', '*toalha*', '*toalhas*', '*tv*'
- Incorrectly identified aspects: '*centro de cidade*', '*colchão da cama*', '*serviço de café da manhã*'
- Aspects not found: '*aquecimento*', '***banheiras***', '*calefação*', '***chuveiros***', '***cidades***', '*cortinas*', '*escada*', '*estabelecimento*', '***estacionamentos***', '***gerentes***', '*gerência*', '*iluminação*', '***jantares***', '*jardim*', '*motel*', '***museus***', '*parque*', '*porteiro*', '***praças***', '***secadores***', '***shoppings***', '*tapete*', '*tomada*', '*tomadas*', '*torneira*', '*torneiras*', '*travesseiros*', '*turistas*'

In the list of aspects that were not found, we could find some terms in plural and, looking at the list of correctly identified aspects, we realized that ten of them were there in their singular form (aspects in bold in the lists). We then carried out some tests that did not prove to be effective. In a first attempt to identify these aspects, we added the lemma of each token as a feature to the classifier, so that plural and singular words would be converted to the same lemma. In a second attempt, we replaced the token feature with its respective lemma. Finally, we added

Table 2
Results with average and standard deviation in the experiments with 5 folds.

Method	Precision		Recall		F Measure		Accuracy	
	Avg.	Dev.	Avg.	Dev.	Avg.	Dev.	Avg.	Dev.
CRF	**0.9933**	0.0027	0.9579	0.0058	0.9753	0.0021	0.9579	0.0058
CRF + Pos-Processing	0.9914	0.0025	**0.9663**	0.0068	**0.9787**	0.0029	**0.9663**	0.0068

a post-processing step comparing the lemmas of the words of the texts with the lemmas of the aspects found by the classifier. Although not effective, this last approach had the best result, which led us to implement a post-processing step analyzing the plural form of the aspects found. To do this, we converted all aspects that were in singular form to their respective plural form and included them as possible aspects to be checked to compose the list of aspects found. This way, we could reduce the list of aspects not found (indicated below in strikethrough style), but three new aspects were added to the incorrect list (aspects in bold).

- Incorrectly identified aspects: '*centro de cidade*', '*colchão da cama*', '***lavanderias***', '***luxos***', '*serviço de café da manhã*', '***tvs***'
- Aspects not found: '*aquecimento*', '~~***banheiras***~~', '*calefação*', '~~***chuveiros***~~', '~~***cidades***~~', '*cortinas*', '*escada*', '*estabelecimento*', '~~***estacionamentos***~~', '~~***gerentes***~~', '*gerência*', '*iluminação*', '~~***jantares***~~', '*jardim*', '*motel*', '***museus***', '*parque*', '*porteiro*', '~~***praças***~~', '~~***secadores***~~', '~~***shoppings***~~', '*tapete*', '*tomada*', '*tomadas*', '*torneira*', '*torneiras*', '*travesseiros*', '*turistas*'

3. Results and error analysis

We calculated the average and standard deviation of the metrics found in the 5 folds based on the obtained results. We computed precision, recall, f-measure, and accuracy. We can see that, despite our post-processing method having located nine new aspects, the difference between the results was minimal, as we can see in Table 2. Anyway, when we deal with large volumes of data, all performance gain makes a difference.

After we analyzed the results of the CRF, we conduct an error analysis of the aspects not found and of those erroneously marked at the end of the post-processing step. A first important observation is that all aspects of the two lists fit into the classification of implicit aspects that we presented, a fact that highlights the difficulty of finding such aspects. This fact made it possible to use our typology of implicit aspects [15] to carry out this analysis. In this typology, we aim to characterize the implicit aspects in categories and subcategories according to the knowledge necessary for their identification. Table 3 presents a summary of the categories and subcategories in the typology.

In Tables 4 and 5, we present the aspect terms with their respective aspects, the categories and subcategories found in the typology and the possible types of errors that led to the incorrect classification. The aspects followed the annotation performed in Freitas and Vieira[14].

The first case of incorrectly detected aspect was the expression '*centro de cidade*'. This classification was certainly produced because we have the expression '*centro da cidade*' often

Table 3
Typology of knowledge necessary to identify implicit aspects: categories, subcategories, aspects and examples

Category	Subcategory	Aspect	Example
Event	Non-verbal form	Price	*pagamento* (payment)
	Verb	Meal	*comer* (to eat)
Feature	Attribute	Design	*material* (material)
	Equivalence	Employee	*equipe* (team)
	Is-a	Employee	*recepcionista* (receptionist)
	Part-of	Facilities	*banheiro* (bathroom)
Qualification	Adjective	Installation	*mobiliado* (furnished)
	Equivalence	Installation	*hotel simples* (simple hotel)
	Nominal form	Location	*proximidade* (proximity)
Contextual	Location	Location	*fácil acesso ao centro* (easy access to downtown)
	Related	Cleanliness	*cheira a mofo* (musty smell)

Table 4
Incorrect aspects.

Aspect term	Aspect	Category	Subcategory	Error
centro de cidade (center of a city)	Location	Contextual	Local	Classifier
colchão da cama (bed mattress)	Sleep quality	Contextual	Related	Classifier
lavanderias (laundries)	Location	Contextual	Local	Annotation
luxos (luxuries)	Facilities/Room	Qualification	Equivalence	Annotation
serviço de café da manhã (breakfast service)	Meal	Feature	is-a	Classifier

appearing related to the hotel's aspect location. Interestingly, both forms appear in the same sentence: "*Fica bem localizado, no* ***centro da cidade****, o que é bom para quem vai resolver coisas por perto, mas ruim por ser uma área bagunçada como todo bom* ***centro de cidade*** *brasileira.*" (It is well located, in the **center of the city**, which is good for those who are going to solve things nearby, but bad for being a messy area like all good **city center** in Brazil.). The first occurrence refers to the location aspect, and the second to the appearance of the center of Brazilian cities in general. The error is related to the classifier, and it would be difficult for the classifier to be able to differentiate the two expressions without understanding the context of each one.

The next case was the expression *"colchão da cama"*. Here the expression is formed by two distinct aspects, *"colchão"* and *"cama"*, which led the classifier to make the mistake of uniting both in a single aspect. The same happened with the expression *"serviço de café da manhã"*, where the term *"café da manhã"* was connected to the word *"serviço"* that is found in other aspects.

As possible dataset annotation errors, we have the aspects *"lavanderias"*, *"luxos"*, and *"tvs"*.

All of them were annotated only in their singular forms. The word *"lavanderia"* is an interesting case because the term appears related to two distinct aspects: location (hotel close to a laundry) and the hotel's laundry service. Given that the word in the singular form appears as an aspect, we believe that its respective plural form should also have been annotated, as it was the case with other aspects. It is worth mentioning that these three errors were made in the post-processing step. The classifier possibly did not make such errors due to the lower frequency of plural forms.

Table 5
Aspects not found.

Term aspect	QTY	Aspect	Category	Subcategory	Error
aquecimento (heating)	8	Facilities	Feature	Part-of	Classifier
calefação (calefaction)	2	Facilities	Feature	Part-of	Classifier
cortinas (curtains)	4	Facilities	Feature	Part-of	Classifier
escada (stairs)	3	Facilities	Feature	Part-of	Classifier
estabelecimento (establishment)	2	Facilities	Feature	Equivalence	Classifier
gerência (management)	1	Employees	Feature	Is-a	Classifier
iluminação (lighting)	4	Facilities	Feature	Part-of	Classifier
jardim (garden)	4	Facilities	Feature	Part-of	Classifier
motel (motel)	1	Facilities	Contextual	Related	Annotation
museus (museums)	3	Location	Contextual	Local	Classifier
parque (park)	4	Location	Contextual	Local	Classifier
porteiro (concierge)	1	Employees	Feature	Is-a	Classifier
tapete (carpet)	2	Facilities	Feature	Part-of	Classifier
tomada (socket)	2	Facilities	Feature	Part-of	Classifier
tomadas (sockets)	1	Facilities	Feature	Part-of	Classifier
torneira (faucet)	3	Facilities	Feature	Part-of	Classifier
torneiras (faucets)	1	Facilities	Feature	Part-of	Classifier
travesseiros (pillows)	5	Facilities	Feature	Part-of	Classifier
turistas (tourists)	8	Facilities	Contextual	Related	Classifier

Analyzing the table of aspects not found (Table 5), we can see that most cases were related to classifier errors. This is probably related to a smaller number of these aspects. A caveat for the aspect "*museus*": this aspect appears in its singular form, and probably should not have been detected due to validation in the 5 folds, appearing in a different set in its singular form, thus making its detection impossible.

As exceptions to the issue of the number of occurrences, we have the aspects "*aquecimento*" and "*turistas*". Regarding the "*aquecimento*" aspect, we have no hypothesis about the fact that it was not located, as it appears eight times in a similar context to other aspects related to the hotel's facilities, which in theory would not be an obstacle to its detection. Regarding the aspect '*turistas*', it is a more difficult aspect to detect since it is necessary to understand the context in which it appears. It is necessary to understand that, for some people, a hotel full of tourists is something negative, as in the example "*Muito cheio de turistas*" (Too full of tourists), as well as its positive absence "*já que não tem muitos turistas por lá*" (since there are not many tourists there). Regarding the facilities, it is necessary to know that a hotel that attracts tourists must have certain appealing characteristics, as in the sentence "*Local ideal para turistas*" (Ideal place for tourists).

One last aspect not found that deserves to be commented on is the aspect “motel”. In our view, it seems more like a case of an error in the annotation. The word appears ironic in the sentence “*Economize o seu dinheiro e vá para um motel*” (Save your money, and go to a motel), not dealing with a motel itself. Therefore, we believe that it would be necessary to mark the entire sentence to be able to understand the negative criticism presented to the hotel’s facilities.

4. Final remarks

Our experiments show that a good performance in the aspect extraction task may be achieved with a relatively simple strategy. Our error analysis reveals the difficulty in detecting implicit aspects, especially those that depend on additional information about the context in which they are inserted. In general, classifier errors were related to terms not annotated as aspects but very similar to other aspects found in the dataset. The aspects that were not found were mostly due to their low frequency in the dataset, which harms most of the machine learning-based strategies.

Acknowledgments

The authors of this work would like to thank the Center for Artificial Intelligence (C4AI-USP) and the support from the São Paulo Research Foundation (FAPESP grant #2019/07665-4) and from the IBM Corporation.

References

[1] M. Taboada, Sentiment analysis: An overview from linguistics, Annual Review of Linguistics 2 (2016) 325–347.

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

[3] T. T. Thet, J.-C. Na, C. S. Khoo, Aspect-based sentiment analysis of movie reviews on discussion boards, Journal of information science 36 (2010) 823–848.

[4] J. Pavlopoulos, I. Androutsopoulos, Aspect Term Extraction for Sentiment Analysis: New Datasets, New Evaluation Measures and an Improved Unsupervised Method, in: Proceedings of the 5th Workshop on Language Analysis for Social Media, Association for Computational Linguistics, Gothenburg, Sweden, 2014, pp. 44–52.

[5] C. Long, J. Zhang, X. Zhu, A review selection approach for accurate feature rating estimation, in: Coling 2010: Posters, 2010, pp. 766–774.

[6] B. Snyder, R. Barzilay, Multiple aspect ranking using the good grief algorithm, in: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, 2007, pp. 300–307.

[7] J. Yu, Z.-J. Zha, M. Wang, T.-S. Chua, Aspect ranking: identifying important product aspects from online consumer reviews, in: Proceedings of the 49th annual meeting of

the association for computational linguistics: human language technologies, 2011, pp. 1496–1505.
[8] B. Pang, L. Lee, Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales, in: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005, pp. 115–124.
[9] M. Tsytsarau, T. Palpanas, Survey on mining subjective data on the web, Data Mining and Knowledge Discovery 24 (2012) 478–514.
[10] Z. Zhai, B. Liu, H. Xu, P. Jia, Grouping product features using semi-supervised learning with soft-constraints, in: Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), 2010, pp. 1272–1280.
[11] Z. Zhai, B. Liu, H. Xu, P. Jia, Clustering product features for opinion mining, in: Proceedings of the fourth ACM international conference on Web search and data mining, 2011, pp. 347–354.
[12] M. M. y Gómez, J. Gonzalo, F. Rangel, M. Casavantes, M. Ángel Álvarez Carmona, G. Bel-Enguix, H. J. Escalante, L. Freitas, A. Miranda-Escalada, F. Rodríguez-Sánchez, A. Rosá, M. A. Sobrevilla-Cabezudo, M. Taulé, R. Valencia-García (Eds.), Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022), CEUR Workshop Proceedings, CEUR-WS.org, 2022.
[13] M. T. Machado, T. A. S. Pardo, Evaluating methods for extraction of aspect terms in opinion texts in portuguese - the challenges of implicit aspects, in: Language Resources and Evaluation Conference, ELRA, 2022, pp. 3819–3828.
[14] L. A. de Freitas A, R. Vieira, Exploring resources for sentiment analysis in portuguese language, in: Proceedings of the 4th Brazilian Conference on Intelligent Systems, IEEE Computer Society, Natal, Brazil, 2015, pp. 152–156.
[15] M. T. Machado, T. A. S. Pardo, E. E. S. Ruiz, A. D. Felippo, F. Vargas, Implicit opinion aspect clues in portuguese texts: Analysis and categorization, in: Proceedings of the 15th International Conference on Computational Processing of the Portuguese, Springer, Fortaleza, Brazil, 2022, pp. 68–78.
[16] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, United States of America, 2004, pp. 168–177.
[17] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems, volume 2, Curran Associates Inc., Lake Tahoe, United States of America, 2013, p. 3111–3119.
[18] M. T. Machado, T. A. S. Pardo, E. E. S. Ruiz, Analysis of unsupervised aspect term identification methods for portuguese reviews, in: Proceedings of the 14th National Meeting on Artificial and Computational Intelligence, Brazilian Computing Society, Uberlândia, Brazil, 2017, pp. 239–249.
[19] C. D. Manning, P. Raghavan, H. Schütze, Introduction to information retrieval, Cambridge University Press, 2008.
[20] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, Y. Singer, Online passive-aggressive algorithms, Journal of Machine Learning Research 7 (2006) 551–585.
[21] Y. Freund, R. E. Schapire, Large margin classification using the perceptron algorithm,

Machine learning 37 (1999) 277–296.
[22] L. Bottou, Stochastic gradient descent tricks, in: Neural networks: Tricks of the trade, Springer Berlin Heidelberg, 2012, pp. 421–436.
[23] H. Gandhi, V. Attar, Extracting aspect terms using crf and bi-lstm models, Procedia Computer Science 167 (2020) 2486–2495.
[24] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186.
[25] F. Souza, R. Nogueira, R. Lotufo, BERTimbau: pretrained BERT models for Brazilian Portuguese, in: Proceedings of the 9th Brazilian Conference on Intelligent Systems, Springer, 2020, pp. 403–417.
[26] L. A. de Freitas, Feature-level sentiment analysis applied to brazilian portuguese reviews, Ph.D. thesis, Pontifícia Universidade Católica do Rio Grande do Sul, 2015.
[27] U. B. Corrêa, Análise de sentimento baseada em aspectos usando aprendizado profundo: uma proposta aplicada à língua portuguesa, Ph.D. thesis, Universidade Federal de Pelotas, 2021.
[28] M. Honnibal, I. Montani, S. Van Landeghem, A. Boyd, spaCy: Industrial-strength Natural Language Processing in Python, 2020.
[29] M. Korobov, sklearn-crfsuite, 2015. URL: https://github.com/TeamHG-Memex/sklearn-crfsuite.

A. Online Resources

More information and the source codes for the NILC participation are available via

- GitHub,
- POeTiSA, POrtuguese processing - Towards Syntactic Analysis and parsing.