

ANÁLISE DE SENTIMENTO EM NÍVEL DE ASPECTO COM BASE EM ONTOLOGIAS DE DOMÍNIO E ANÁLISE DE DEPENDÊNCIA SINTÁTICA

FRANCISCO DIAS FRANCO¹; LARISSA ASTROGILDO DE FREITAS²;
ULISSES BRISOLARA CORRÊA³

¹ Universidade Federal de Pelotas – fdfranco@inf.ufpel.edu.br

² Universidade Federal de Pelotas – larissa.freitas@ufpel.edu.br

³ Universidade Federal de Pelotas – ulisses@inf.ufpel.edu.br

1. DATASET

O *dataset*^{1 2} que será utilizado durante presente trabalho foi composto por *reviews* de usuários do TripAdvisor³, sendo que cada *review* pode, ou não, conter uma ou mais opiniões.

O *dataset* é composto por 3.111 amostras, provenientes de 847 avaliações sobre o setor hoteleiro. Sendo 2.112 amostras positivas, 527 amostras negativas e 472 amostras neutras. Das 3.111 amostras, 255 contém aspectos compostos por mais de uma palavra e 2856 contém aspectos simples.

As informações são distribuídas em 5 colunas: *review*, *polarity*, *aspect*, *start_position* e *end_position*. A coluna *review* contém a *review* avaliada, a coluna *polarity* contém a polaridade do aspecto, a coluna *aspect* contém o aspecto encontrado na *review*, a coluna *start_position* contém a posição inicial do aspe to na *review* e a coluna *end_position* contém a posição final do aspe to na *review*.

¹O *dataset* foi construído juntando os *dataset* desenvolvidos anteriormente por Freitas (2015) e Corrêa (2021). O *dataset* de Freitas (2015). O Dataset usado durante o estudo de Freitas (2015) está disponível publicamente, através do link <<https://www.inf.pucrs.br/linatural/wordpress/recursos-e-ferramentas/tripadvisor/>>, enquanto que o Dataset de Corrêa (2021) é privado. Ambos os recursos foram anotados seguindo as diretrizes de anotação descritas por Freitas (2015). Para fazer as anotações foi utilizando uma ontologia de domínio, chamada de HOntology, ontologia multi-língua para o setor de acomodações (CHAVES; FREITAS; VIEIRA, 2012).

²O *dataset* usado durante a competição ABSAPT-2022 (SILVA et al., 2022) pode ser baixado entrando em contato com <absapt2022@inf.ufpel.edu.br>, após o contato você receberá por e-mail com o corpus compactado e a senha. Mais informações podem ser obtidas pelo link <<https://sites.google.com/inf.ufpel.edu.br/absapt2022/>>.

³O TripAdvisor é o maior site de viagens do mundo, possuindo mais de 859 milhões de avaliações e opiniões sobre 8,6 milhões de acomodações, restaurantes, experiências, companhias aéreas e cruzeiros. Dessa forma, ele se torna uma importante fonte de informações para treinamento de modelos de SA para serviços de hospedagens.

5. REFERÊNCIAS

CHAVES, M. S.; FREITAS, L.; VIEIRA, R. Hontology: A multilingual ontology for the accommodation sector in the tourism industry. In: INSTICC. **Proceedings of the 4th International Conference on Knowledge Engineering and Ontology Development (KEOD 2012)**. Barcelona, Espanha: SciTePress, 2012. p. 149–154. ISBN 978-989-8565-30-3. ISSN 2184-3228.

CORRÊA, U. B. **Análise de sentimento baseada em aspectos usando aprendizado profundo: uma proposta aplicada à língua portuguesa**. 123 p. Tese (Doutorado em Computação) — Universidade Federal de Pelotas, Pelotas, 2021.

FREITAS, L. A. d. **Feature-level sentiment analysis applied to brazilian portuguese reviews**. 94 p. Tese (Doutorado em Ciência da Computação) — Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2015.

SILVA, F. L. V. da; XAVIER, G. da S.; MENSENBURG, H. M.; RODRIGUES, R. F.; ARAÚJO, L. P. dos Santos y R. M.; CORRÊA, U. B.; FREITAS, L. A. de. Absapt 2022 at iberlef: Overview of the task on aspect-based sentiment analysis in portuguese. **Procesamiento del Lenguaje Natural**, v. 69, n. 0, p. 199–205, 2022. ISSN 1989-7553. Disponível em: <<http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6440>>.