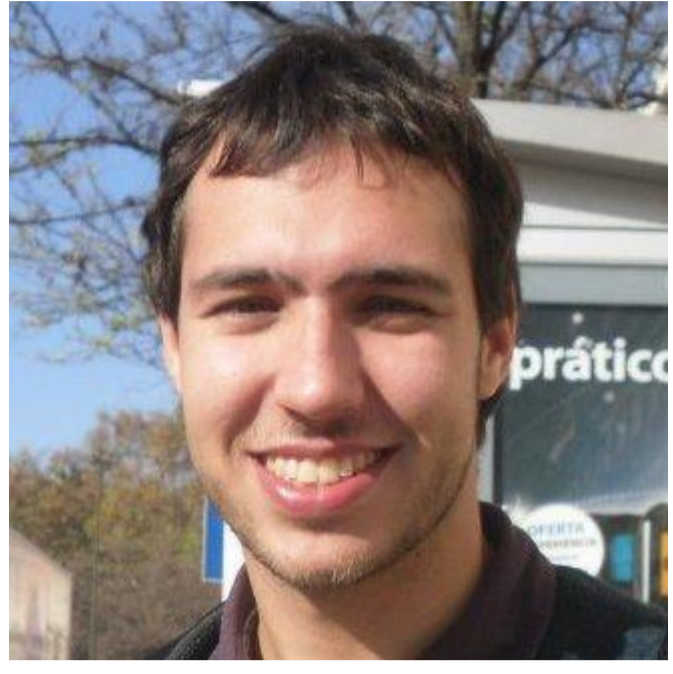




Aspect-based Sentiment Analysis for Portuguese: Exploring Syntactic and Semantic Features



balage@icmc.usp.br

Student: Pedro Paulo Balage Filho (PhD) - Scholarship: FAPESP

Supervisor: Prof. Dr. Thiago Alexandre Salgueiro Pardo

Interinstitutional Center for Computational Linguistics (NILC)

University of São Paulo - Brazil



Student Homepage

Research goals

Sentiment Analysis is a recent area of Natural Language Processing (NLP) which aims to process texts with opinion, sentiment and subjectivity. In this area, there are three levels of analysis: text level, sentence level and **aspect or entity level**. In the aspect-based sentiment analysis, feelings are extracted according to the aspects they modify. The focus of this work is on the **aspect-based sentiment extraction** for **product reviews** in **Portuguese**. The research goals are to explore **syntactic** and **semantic features** for such task.

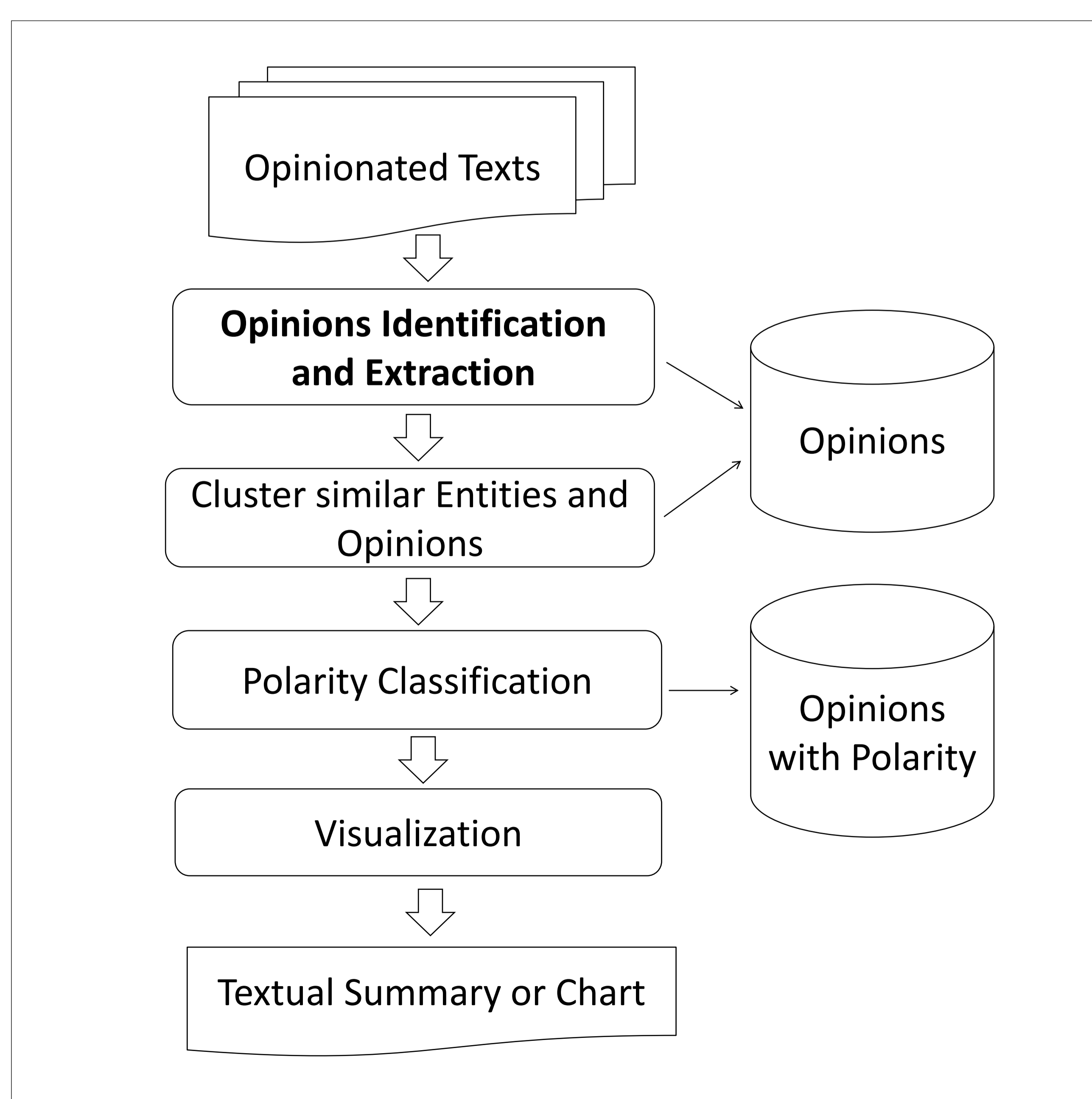
Task Example

O **iphone** é um **bom** aparelho. Sua **bateria** é **excelente**. A **qualidade** é **muito boa**, mas o **preço** é **exorbitante**.

(The iPhone is a good device. Its battery is excellent. The quality is very good, but the price is exorbitant.)

Aspect	Predicate	Sentiment
Aparelho (device)	Bom (good)	Positive
Bateria (battery)	Excelente (excellent)	Positive
qualidade (quality)	muito boa (very good)	Positive
preço (price)	Exorbitante (exorbitant)	Negative

Basic Architecture



Features to Investigate

- ✓ Lexical features
- ✓ Part-of-Speech constructions
- ✓ Dependency parser features
- ✓ Semantic roles
- ✓ Selection restriction semantics

Machine Learning Algorithm

- ✓ Conditional Random Fields

SemEval 2014 - Aspect Based Sentiment Analysis Task

WORD	POS	CHUNK	SRL	ASPECT?
Great	JJ	B-NP	B-A0	False
laptop	NN	E-NP	E-A0	False
that	WDT	S-NP	S-R-A0	False
offers	VBZ	S-VP	S-V	False
many	JJ	B-NP	B-A1	False
great	JJ	I-NP	I-A1	False
features	NNS	E-NP	E-A1	True

Features	Precision	Recall	F1-measure
Baseline	52.54	42.76	47.15
Word + POS	83.76	68.69	75.48
+ Chunk	83.38	68.16	75.01
+ NE	83.45	68.07	74.98
+ SRL	82.79	67.46	74.34
+ Frame	87.72	34.03	49.04

- Features provided by Senna SRL Tool and ARK Semafor Tool
- CRF provided by CRF++ Tool

Publications

Balage Filho, P. P., e Pardo, T. A. (2014). NILC_USP: Aspect Extraction using Semantic Labels. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 433–436. Dublin, Ireland, August 23-24.

Balage Filho, P. P., Avanço, L., Pardo, T. A. S., Nunes, M. G. V. (2014). NILC_USP: An Improved Hybrid System for Sentiment Analysis in Twitter Messages. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 428–432, Dublin, Ireland, August 23-24.

Hartmann, N. ; Avanco, L. ; Balage, P. P. ; Duran, M. ; Nunes, M. G. V.; Pardo, T. A. S. ; Aluisio, S. M. (2014). A Large Corpus of Product Reviews in Portuguese: Tackling Out-Of-Vocabulary Words. *Proceedings of the International Conference on Language Resources and Evaluation (LREC'14)*, 2014, pages 3865-3871, Reykjavik, Iceland.