



*16ª Conferência Ibérica de
Sistemas e Tecnologias
de Informação*

23 a 26
JUNHO
2021
Chaves/PT

A Multilingual Lexicon-based Approach for Sentiment Analysis in Social and Cultural Information System Data

Sandra Jardim and Carlos Mora
Smart Cities Research Center
Polytechnic Institute of Tomar
Tomar, Portugal

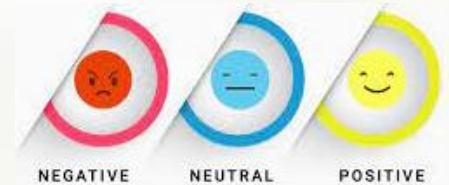
Tiago Santana
Technology Superior School of Abrantes
Polytechnic Institute of Tomar
Tomar, Portugal

Decision Making Process



Sentiment Analysis

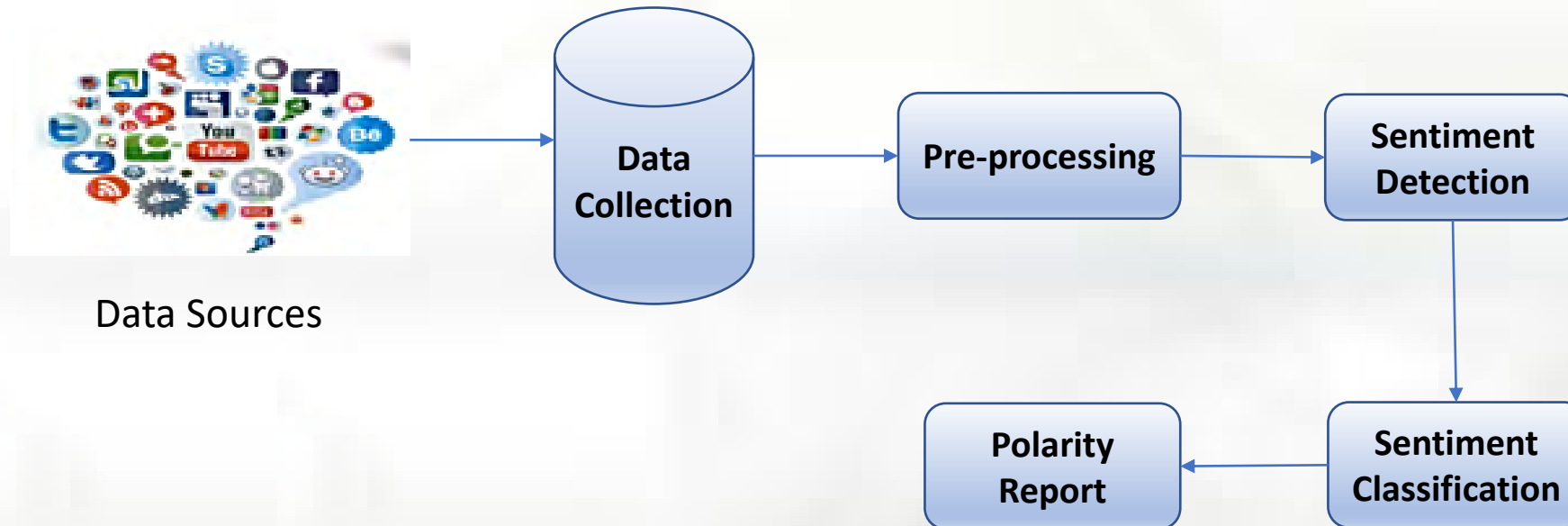
A textual information classification automated process of detecting, extracting and classifying opinions, according to the data polarity



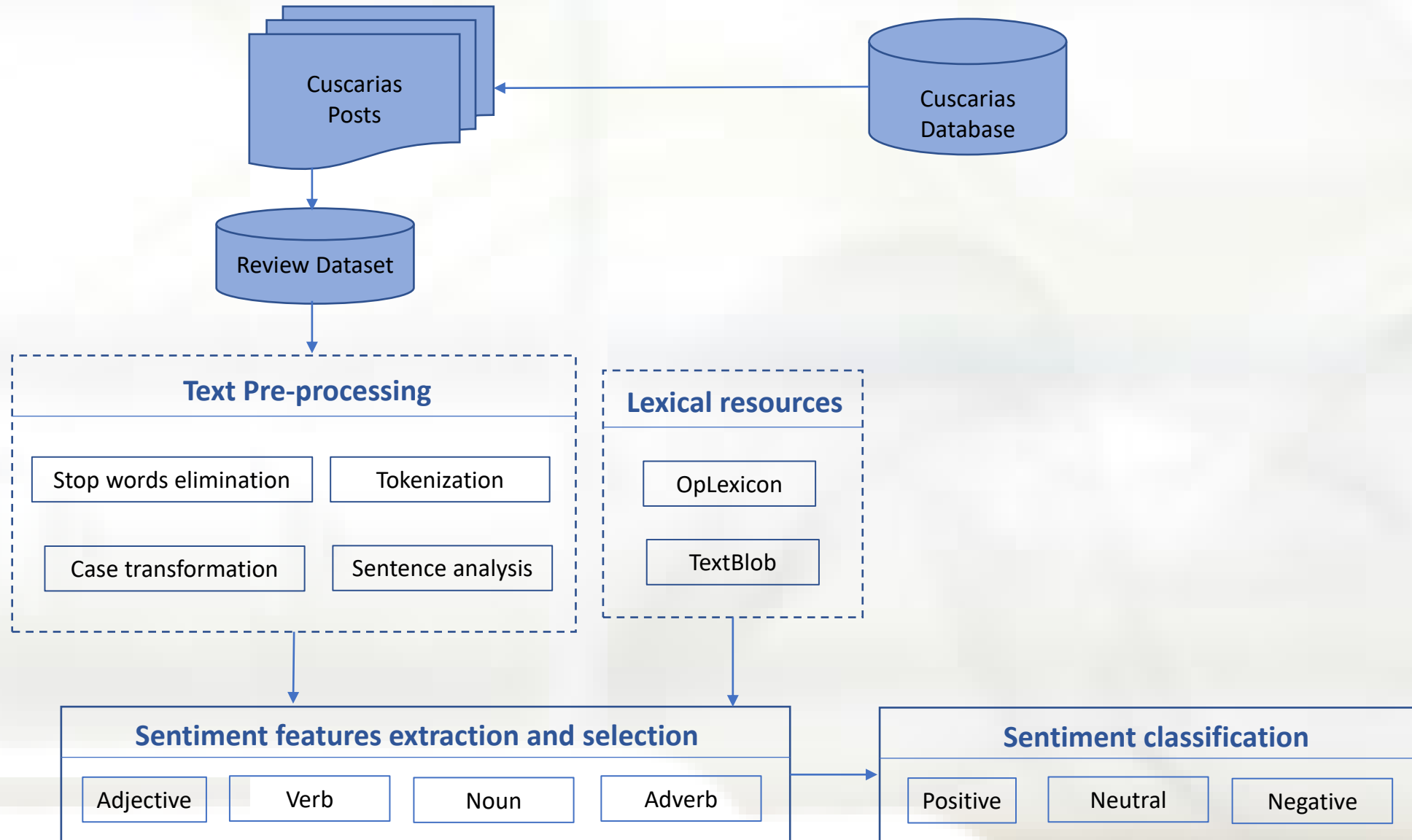
In business



Sentiment Analysis Process



System core architecture



Developed System

- Unsupervised automatic method of multilingual lexicon-based sentiment analysis algorithm
 - Using the Python programming language
 - The solution architecture follows a file organization made up of 5 Python files
 - API_sentiment, Cron_sentiment, EN_sentiment, PT_sentiment, and Translation
- Dataset
 - The dataset is built from the posts recorded on the Cuscarias cultural and social information system
- Multilingual Lexicon-Based Sentiment Analysis
 - Detection of the language of the text to be processed
 - Sentiment classification method applied according text language

Developed System - Portuguese Texts

- Dataset OpLexicon
 - Data structure with the polarities of adjectives and verbs used in the Portuguese language, with more than 32000 entries
- Creation of a dictionary with the keywords and respective polarities
- Text pre-processing
 - Pre-processing techniques were applied to prepare de texts for the sentiment features extraction and selection
 - Elimination of non-relevant and non-text information and the case transformation
- Sentiment features extraction and selection
 - Identification of nouns, adjectives, verbs and adverbs, for later classification.
- Sentiment classification
 - Performed by adding the polarity values of each word in the text

Developed System – English & Other Languages

- English Texts
 - TextBlob text processing library
 - The sentence construction in English is not as complex nor enjoys as much variability as in Portuguese, which means that the algorithm has a lower number of exceptions
- Texts in other languages
 - Translation of the original text into English, using the Python library *googletrans*

Experiments

Original Text	Pre-processed Text
This is a great place, with a wonderful view	This is a great place, with a wonderful view
Not a very nice hotel. The bedroom was very tiny but the breakfast was very good	Not a very nice hotel. The bedroom was very tiny but the breakfast was very good
No, I don't agree. Paris is a beautiful city	No, I don't agree. Paris is a beautiful city

Pre-processed Text	Sentiment Polarity Value
great (1) place wonderful (1) view	2
Not nice (-1) hotel bedroom tiny (-1) breakfast good (1)	-1
don't agree (-1) Paris beautiful (1) city	0

Algorithm performance evaluation

Dataset

2.100 Portuguese posts
1.450 English posts
820 Spanish and French posts

Sentiment classification accuracy

87% for Portuguese posts
92 % for English posts
72% for Spanish posts
79% for French posts

Conclusions and Future Work

- An unsupervised automatic method of multilingual lexicon-based sentiment analysis algorithm
- The architecture of the proposed solution consists of three main levels (pre-processing of the texts to be analyzed and classified; sentiment identification and extraction, using the dictionaries corresponding to the language in which the texts are written; classification of the identified sentiments and the calculation of the global polarity of the expressed sentiment)
- The implemented solution was tested in the analysis of user's publications in a digital tourism platform, written in four different languages, showing a high accuracy for sentiment classification.
- Since the submission of this paper, some improvements have already been made to the system, namely, the clustering of users according to the polarity of their opinions regarding each Cuscaria
- As future work, we intend to enhance the system by extracting and identifying the evaluated features of each Cuscaria along with the information of the polarity results



*16ª Conferência Ibérica de
Sistemas e Tecnologias
de Informação*

23 a 26
JUNHO
2021
Chaves/PT

Thank You!

sandra.jardim@ipt.pt

This work has been funded by national funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., under the Project UIDB/05567/2020.