

A Newspapers Expectations Indicator from Web Scraping

Authors : Pilar Rey del Castillo

Abstract : This document describes the building of an average indicator of the general sentiments about the future exposed in the newspapers in Spain. The raw data are collected through the scraping of the Digital Periodical and Newspaper Library website. Basic tools of natural language processing are later applied to the collected information to evaluate the sentiment strength of each word in the texts using a polarized dictionary. The last step consists of summarizing these sentiments to produce daily indices. The results are a first insight into the applicability of these techniques to produce periodic sentiment indicators.

Keywords : natural language processing, periodic indicator, sentiment analysis, web scraping

Conference Title : ICAMAS 2021 : International Conference on Applied Mathematics, Analysis and Statistics

Conference Location : Amsterdam, Netherlands

Conference Dates : December 02-03, 2021