Using K-Means Algorithm for Description Analysis of Text in RSS News Format

Paola Ariza-Colpas, Ana Isabel Oviedo-Carrascal, Emiro De-la-hoz-Franco

Abstract:

This article shows the use of different techniques for the extraction of information through text mining. Through this implementation, the performance of each of the techniques in the dataset analysis process can be identified, which allows the reader to recommend the most appropriate technique for the processing of this type of data. This article shows the implementation of the K-means algorithm to determine the location of the news described in RSS format and the results of this type of grouping through a descriptive analysis of the resulting clusters.

Keywords:

RSS news's format, Simple K-means, Bag of words, Stopwords, Text mining