Extracting Information and Analysing Sentiment in Twitter

Lab. Assignment III

Sergi Linares Ortigas NIA: 183669
Víctor Jordán García NIA: 183747
Kevin Romero Salgado NIA: 183712

Index

Introduction

In this assignment we are going to create a program with which we are going to be able to analyse the sentiment of a huge number of tweets about a specific topic, with that kind of code we are able to process that kind of big data and extract valuable information for companies.

Creating the new Gapp

First of all, for the creation of the new gap we needed a dictionary of positive, negative and neutral words from different languages (in our case Catalan, Spanish and English). This dictionary was created using a code made by us (it will be added in the delivery) in which we read a xml file with a lot of words and depending on the value of the pol attribute we classify the word in one of the three categories mentioned before using the following criteria:

Pol >= 0.5 → Positive word

Pol <= -0.5 → Negative word

-0.5 < Pol < 0.5 → Neutral word

Then, with the words classified in a correct way, we saved it in a different list files (.lst) and we modified the file main.def for adding that list to the file.

Finally, we opened the GATE Devoloper 8.0 and we charged in the TwitIE Application in which they give us an specific gazetteer, we must change it by one in which the .def associated is the one which we have created before.

Now, we have our Application Gapp ready to be implemented with the Java program.

Explanation of Files

AIW2019_P3_G12

It is our main class where we have an intuitive menu in which we call each function

CallGATEConditionalCorpusPipeline

In this class we process all JSON. We also start everything that has to do with GATE and our GAPP

Sentiment

In Sentiment we extract the kind of sentiment that tweet has, to put it into the XML and be able to work it later on the web

HashtagObject

To create objects of type Hastag

TweetObject

To create objects of type Tweet

TweetTreatment

With this class we treat the tweet xml to turn it into something that we can use for the web

ExtractSentiment

Secondary program that we use to extract all XML and build them using JSONs

KeyConstants

We will use this file to declare the constants we have (google API, gate cloud, ...) and be able to use them in the future when generating the web scripts

MapsUtils

Class provided by Horacio to be able to create the maps.

ReadFile

Separate all tweets in different files.

CreateMapToWorkOut

Class provided by Horacio to be able to create the maps.

SimpleHTMLConstructor

Functions to create our website. It is part of practice 2 so it is not developed for this lab

SimpleHTMLExtractor

Functions to create our website. It is part of practice 2 so it is not developed for this lab

Problems and Conclusions

We believe it is a lab worth trying. It is complicated to do if you don't have good planning and clear concepts. Our main problem has been that we have not been able to generate the web in a dynamic way. We have had many implementation problems and errors with our Gapp. That has made us waste a lot of time, and deliver with days of delay and without the web module made.

Our projection and organization of the components of the group and of the project itself has been good, but our execution has not.