

Alex Marin Felices	219305	u162444
Mireia Palou Tort	219331	u162439
Blanca Ruano Miranda	205722	u148716

Lab 1: Text Processing

Github link:

<https://github.com/alex11marin11/IRWA-2021-final-project-u162439-u148716-u162444/releases/tag/IRWA-2021-final-project-part-1>

In this lab our data consisted of a set of tweets from the World Health Organization (@WHO) account.

Our task consisted of reading the file that contained these tweets and processing them.

As in class we had done something very similar we decided to adapt our code to this situation.

We saved the tweets in a pandas dataset. Then we created a function to process correctly the text that we passed through all the columns that had the full text of the tweet.

This function has 5 steps: first transform the text into lowercase, then remove the punctuation, afterwards tokenize it, then remove the stopwords and finally stemming.

Now we have the data ready for the next part.