Now, with the processed data is time to create the index so we can have faster searching, to create the index we will start by obtaining the terms using the build\_terms function from the first part of the project. Then, for each term we will store the document and the positions where the term can be found in that document, we will also store a data structure with the most important information from each tweet (tweet, username, date, hashtags, likes, retweets, URL), this way, we can access faster to the most relevant information about a tweet.

After having the index, we have implemented a search method which given a query and an index it pre-processes the query and looks in the index for the tweets where the term appears and returns a list with these documents.

Now, we have to evaluate this search engine, we will do it by testing 5 queries, the queries chosen were:

* Covid pandemic
* International disaster
* Ritmo de vacunacion
* Percentage de hospitalizados
* Mental health

We chose these queries because …

To apply a TF-IDF ranking to the results we created the method create\_index\_tfidf which given the collection of tweets and the total number of tweets returns the index, the term frequency (tf), document frequency (df) and the inverse document frequency (idf) for each term and also tweet\_index containing the most relevant info for each tweet. Then, the ranking will be performed by the function rank\_documents which will …