Filtering words

If you still do not feel confident with python, you can work on a very nice tutorial of Neal Caren on how to process text with python.

http://nealcaren.web.unc.edu/an-introduction-to-text-analysis-with-python-part-1/
http://nealcaren.web.unc.edu/an-introduction-to-text-analysis-with-python-part-2/

1. Installation of NLTK

We are going to process natural languages. We need to install several softwares to filter our texts:

```
sudo pip install -U nltk
sudo pip install stop-words
```

The main software, we need is called gensim: https://radimrehurek.com/gensim/:

```
sudo apt-get install libblas-dev liblapack-dev
sudo easy_install pattern
sudo easy_install gensim
```

A full tutorial on topics modeling for human using Latent Dirichlet Allocation (LDA) is available here:

http://bit.ly/2ffakzH

Question 1: To work with LDA, you need to detect the language of a text. How are you going to solve this issue ?

Question 2: LDA modeling requires to finely tune several parameters. Explore the parameters to find the ones that are the most significant.

Question 3: What is the best option to recover the topics of interest of a target:

- work independently on each text or
- concatenate all the texts to work on a single text ?