Text analysis 1

The goal of this exercise is to create a running script, thus it is sufficient to upload a script file.

The provided data set consists of 14 annual reports of V4 presidencies. Texts are lemmatized, but not there is no additional pre-processing. The corpus consists of 171 940 individual words.

1. Prepare R for analysis (2pts)
   1. Use commands to set proper working directory
   2. Use commands to load proper packages
2. Use command to create corpus from texts (2pts)
3. Apply 3 chosen cleaning operations over the corpus of texts (2pts)
4. Create term-document matrix (2pts)
5. Obtain most frequent words and set reasonable cutoff (explore available functions and their outputs) (2pts)
6. Bonus quesiton (1pt) – convert term-document matrix to matrix object and save it as a CSV document

Suggestions

* Explore possibilities of packages using R help
* Google for ready-made solutions, if necessary
* If you want to make a note/comment in the script, use hashtag (#) and write text after that