Topic modelling for value change

This notebook can be used to analyse values based on a set of documents. As values in documents are often latent (i.e. they are not explicitly mentioned), we use probabilistic topic models to capture values by means of distributions of words instead of keywords. More literature about the difficulty to cope with latent values can be found in the following paper: <https://doi.org/10.1016/j.apenergy.2018.06.082>

## Requirements

To use this notebook, you need:

* A google account
* An input file containing the documents you want to analyse in CSV, xlsx or txt format. Each document much contain or be associated with a date. The quality of the analysis increases when the number of documents in the file is large. Although there are no strict rules, the minimum number of documents is probably around 500.

## Preparation of the input

The input file must contain two types of information for each document: a text

Before using the notebook, the input file you aim to use might need to be adjusted so it can be read in the notebook. Here are some instructions to do so.

### CSV and xlsx formats

CSV and xlsx are preferred formats as the file typically requires less data preparation. Make sure that the file contains one document per row. Name all the column(s) containing the texts you want to use for the analysis '[Text\_for\_analysis]'. In the notebook, texts of multiple columns will be combined into one single text. Name the column containing the dates '[Date]'.

### Txt format

Using docx format is possible too but this requires more preparation work as the notebook may not know what refers to the text and what refers to the dates. To do so, open the docx file and paste ‘[New\_document]’ before every new document. Paste '[Date]' before the date of each document. Try to automatise this process as much as possible by using the Find and Replace function provided by Microsoft Word.

## Launching the notebook

* Open google drive using a web browser (<https://drive.google.com/drive/u/0/my-drive>). Unpack 'Topic\_modelling\_analysis.zip' and paste it in your google drive. It is important that you do not embed the ‘Topic\_modelling\_analysis’ folder into another folder. It should be at the root of your google drive.
* Open the ‘Topic\_modelling\_analysis’ folder and place your input file containing the texts you want to analyse in the 'data' folder. This folder already contains some datasets in case you do not have one yet.
* In the ‘Topic\_modelling\_analysis’ folder you will find a file called 'Topic\_modelling\_for\_value\_change\_workshop.ipynb': double-click on it. You have now opened the notebook.

Further instructions to run the notebook and analyse value change can be found in the notebook.

Good luck!