BoW\_v2.ipynb - READMe

-----------------------

Description: This Google Colab notebook "BoW\_v2.ipynb" implements the Bag-of-Words model

-----------------------

Prerequisites:

- Logged in with a Google account, with access to Google Colab, Google Colab's CPUs and GPUs (subscription needed to have prioritised access GPU's for a significant longer time) (for this notebook, you only need CPUs)

- Granting permission to run each of our Google Colab notebooks and python files.

- Mount drive at /content/drive/ [This is done when running the notebook, see below]

- The folders/files regarding Preprocessed training and validation datasets are at the described location.

-----------------------

How to run:

Before running, the user can specify

- which preprocessed datasets they want to work with

In the first section, there is a variable

PREPROCESSING\_CHOICE defined by PREPROCESSING\_OPTIONS[INDEX], where INDEX is an integer [0;8]. The user can choose INDEX from that range, which corresponds to the preprocessed dataset.

After selecting INDEX:

Run all sections, top to bottom.

-----------------------

Output:

There are print statements for stats. Other than that no file to save stats.

-----------------------

-----------------------

Source: This notebook was adapted for our purposes from the Bag-of-Words implemenation given by CIL. (see link below)

<https://colab.research.google.com/github/dalab/lecture_cil_public/blob/master/exercises/2021/Project_2.ipynb>