PROJECT README

May 15, 2020

Virtualenv is included, however if something goes wrong try installing requirements from *requirements.txt*

1. Activate virtualenv with *source/venv/bin/activate*

2. Run *main.py* and when prompted insert as an integer value to specify the number of topics to generate for every year.

Please beware that every time the program launches the content of the subfolders is deleted.

NOTE: Sometimes the program crashes almost instantly, because of the keyword *43.80.+p*. Pyplot thinks it is a command. Anyways, as it works at second/third try, I decided not to correct this issue to not alter the graph view.

Directories: **results** directory contains all the tasks required for the completion of the project.

results/TASK1 and **results/TASK2** are relative to *T1* and *T2* in the project requests.

1 Task 1

Folder **TASK1** lists all years in [2000, 2018]. Each year contains:

- a picture representing the execution of the Spreading of Influence algorithm for the node that generated the biggest topic (basically the node that influenced more other nodes that year).

- a **spreading of influence** folder containing folders in the format "*score_generating keyword*" relative to the starting node fo the spreading of influence algorithm. This folder contains a text file "*influenced_nodes.txt*" that is the list of nodes influenced by the starting keyword. The list of influenced nodes is a topic.

- a **merges** folder containing subfolders in the form of "rank.generating keywords" that did not merge to other topics, but had other topics merged to them instead. *rank* indicates the size of the topic, therefore if rank=1 the number of topics contained is the biggest.

  - Each subfolder contains "merge_tracking.txt" and "merged_topics.txt" files. The former is the list of topics merged to that topic written as generating keywords, the latter are the actual merged topics.

2   Task 2

Folder **TASK2** contains:

- **dots** folder: to ignore, used only to save graph images

- **images** folder, listing years in $2000 - 2018$, each containing a tracing graph starting from a keyword in that year. Please note that some years may be missing because all of the nodes in that year have already been merged to nodes in the previous year.

- a folder **remarkable_merges**, containing the top-10 biggest tracing graphs, along with the nodes merged:

  - Each subfolder is relative to the year and the name of the generating keyword from which the tracing graph started. Subfolders are ordered from bigger to smaller topic graph.