**Project overview**

Confucianism as a system of thought has deeply influenced Chinese culture and society. After having been the official ideology of the majority of imperial dynasties, since 1911 Chinese people and Chinese government’s relation with Confucianism has become increasingly troublesome. Due to its conservative nature, Confucianism was seen as the main cause of China cultural backwardness and was thus condemned. This tendency reached its peak during the Maoist era, when Confucius was the target of a campaign of harsh criticism, sponsored by Mao Zedong: the core values of Confucian philosophy were incompatible with Communist party ideals. Only recently has a Confucian “revival” been identified in Chinese leaders’ official speeches and statements, particularly during Hu Jintao and Xi Jinping presidencies.

The aim of the project is to verify on an empirical basis and to analyse the features of this “revival” of Confucianism ideals. In order to do so, I would like to conduct a text analysis on selected speeches and articles by the five key figures and presidents of the People’s Republic of China, namely Mao Zedong, Deng Xiaoping, Jiang Zemin, Hu Jintao and Xi Jinping. The first part of the project will focus on the creation of a database of speeches and articles in Chinese. Among the ones I’m interested in, are the reports of the nineteen CPC National Congresses held up to now and other collections of articles published by official government press.

For what concerns the coding part of the project, I would like to conduct a text analysis using Python. One of the biggest issue of the project is the fact that of working with a non-alphabetical language. Up to now, I was able to find and install a Python library called “jieba”, a word segmentation tool which enables Python to read and understand words (as conceived in alphabetical languages) in Chinese.

Once the language issue is solved, I will create a dictionary of Confucian words to check their occurrences in the speeches of the different presidents. Hopefully, the results will show the evolution of the controversial relation between communist ideology and Confucian philosophy.

Finally, I will try to grasp a more accurate insight by means of a sentiment analysis on some keywords like “Confucius”. This will allow me to know not only the occurrences of such words, but also whether the context in which they appear is positive or negative.

**Data requirements**

I haven’t found a dataset of speeches or articles available on the Internet. The dataset that I will create for the project will include collections of speeches and articles of the presidents. Among the titles I have identified, I would like to include in the project: *Selections from Deng Xiaoping*, *Selections from Hu Jintao*, *Selections from Jiang Zemin* and *The Governance of China*.

One major issue is to figure out which format best suits the code for the text analysis project, be it txt or something else.

**Final output**

The aim of the project is to show a shift in the rhetoric of China Communist Party propaganda with particular regards to its relation to Confucian ideals. The results of the text analysis and the sentiment related to certain words could be shown through some graphs, which are useful for visualising occurrences of Confucian terms under each leadership, while the sentiment of such terms could be represented through different colours. These graphs can be part of a broader and wider report on the trend the project was able to analyse and its implications.

User interaction or/and audience intended for the project are still a work in progress. It could be useful to publish the result on a website along with possibly enhanced and richer graphs displaying the knowledge gained through the data analysis.

**Next steps**

1. Getting in touch with experts of the field and figure out a) which libraries are needed to complete the task b) which data format is more suitable for the text analysis I intend to carry out;
2. Creating the database, at least starting from one of the abovementioned texts, and figure out how to access these collections, which have been officially published, in a digital form;