**Project Overview:**

The aim of the project is to verify on an empirical basis and to analyse the features of the “revival” of Confucianism ideals in CPC political rhetoric.

1. Creating a dataset: I was able to find digital copies of some of the key-writings for my project, I will upload one in word format as an example of length and style. I will continue with the gathering of data in the upcoming days.

Issue: which format better suits python?

1. Text analysis:

a) defining a “Confucian dictionary” of words related to Confucian concepts. I have an appointment via Skype on Thursday with a professor of Ca’ Foscari whose field of research is Chinese political language and who will discuss with me the part related to the Confucian basic terms I am searching for in these writings;

b) start with the coding part of the project: figure out which library better suits word segmentation process in Chinese (either jieba or Standford Word Segmenter) and start writing the code for the word frequency count.

c) using the database I created for counting occurrences of words and identify trends

1. Sentiment Analysis: compatibly with my still poor coding skills, I would like to conduct a sentiment analysis on some words, for examples “Confucius”, in order to see if the context in which it appears is positive or negative.

**Project Deliverable:**

* Final product and data visualisation: the final output will be a report including word-count tables and graphs and the meaning of the data in light of CPC political context. Moreover, the graphs and other data visualisation techniques can be displayed and be available on specialized websites. Lastly, the database I create for the project can be available and public for further research, compatibly with copyright issues.

**Project schedule:**

**Milestone 1:** (approximately half of November)gatherall data; define the dictionary; narrow down the rest of text analysis work;

**Milestone 2:** test python libraries and start writing the code for the text analysis part (word frequency);

**Milestone 3:** Sentiment analysis (second semester)

**Milestone 4:** Data visualisation (second semester)