

Text Mining : Home Work 3

Due on June 18, 2023

This exercise asks you to perform text mining for Chinese. You shall find [https :
//alvinntnu.github.io/NTNU_ENC2036_LECTURES/chinese-text-processing.
html](https://alvinntnu.github.io/NTNU_ENC2036_LECTURES/chinese-text-processing.html) very useful.

1. Scrap 10 most recent Chinese news articles and form a corpus.
2. Perform word segmentation, check the results and edit a use-defined dictionary accordingly.
3. Perform word segmentation again with use-defined dictionary.
4. Remove stop words and then perform word segmentation.
5. Perform POS Tagging and then word segmentation.
6. Plot a quick overview of the corpus.
7. Perform the concordance analysis with `kwic()`
8. Create a word frequency list.
9. Create a word cloud.