Exercise 1: Text Tokenization

Text Tokenization

- URL: *https://news.ifeng.com/c/89TNORdIths*
- Crawler: www.topcoder.com/thrive/articles/web-crawler-in-python
- Chinese word tokenization: https://github.com/fxsjy/jieba

Text Tokenization

- Submission
  - Code file
  - Excel file with tokenized Chinese words:

URL	Title	Content
https://news.ifeng.com/c/89TNOI	我/爱/中山大学	中山大学/是/一所/985/高校/...

  - Zip the file with filename “2021NLP-exercise 1-学号-姓名.zip/rar”
  - Send the result to
    - Email: sysucsers@163.com
    - Subject: 2021NLP-exercise 1-学号-姓名，例如“2021NLP-exercise 1-xx-xxx”；
  - Deadline: 2021-9-26, 24:00