

Information Retrieval Homework, Report I

R08725002 資管碩二 黃文鴻

1. Environment

VSCode

2. Programming Language

Python 3.8

3. Usage

```
pip3 install nltk  
python3 ./pal.py
```

4. Introduction

In line 14, tokenize document by remove "\n", ",", ".", "'", and split into tokens using blanks

In line 16, lower case all the tokens

In line 18-19, stemming tokens using Porter's algorithm with PorterStemmer in nltk package

In line 21-22, download stopwords and filter words in stopwords list

In line 24-26, save the filtered tokens into result.txt