CPS3320W01

Project 2\_1\_FlashText

Yue He

1063785

Write-up

1. Explain its central purpose: FlashText is a Python library created specifically for the purpose of searching and replacing words in a document. Now, the way FlashText works is that it requires a word or a list of words and a string. The words which FlashText calls keywords are then searched or replaced in the string. When the number of keywords to be searched is below 500 then, regular expressions provide a little edge over FlashText. But as soon as number of keywords cross 500, FlashText surpasses regular expression performance by a wide margin. FlashText is any day better than regular expressions for replacing keywords in a document. FlashText can improve the speed of running.
2. describes useful and/or interesting ways in which it might be used (with at least two practical examples): I think that the functions of search and replace in this library are very important in our work and study or in our daily life. In the first example, I searched keyword “New York” in the introduction of New York, and determined the position of keyword “New York”. The output shown positions of every keywords automatically and I did not need to read the whole paragraph. In the example two, I used function to replace New York with NY and New York City with NYC in these sentences, and I did not need to modify it one by one in the sentences. The output shows sentences which had been changed. In a word, using this library can greatly improve the efficiency and will not make mistakes.
3. provides a high-level overview of the different functions included in the library as found in its official documentation (and provide a link to the documentation): add\_keyword (<unclean name>, <standardized name>) function is used to classify the unclean name into the standardized name. Once it matches unclean name, it displays standardized name; extract\_keywords funtion returns the matched keywords; replace\_keywords function directly returns an entire sentence; remove\_keyword function is used to remove all keywords in a whole sentence; KeywordProcessor is a trie tree.