

CS410 Final Project

This project aims to crawl data from webpage by chrome extensions and then use text ranking method to analyze them.

How to use

1. chrome extension:
just load the file folder into chrome://extensions/ in developer mode.
Find a webpage you interest and click the extension icon
A file named crawling.txt will be downloaded, including all text data in that webpage.
2. Text ranking
Just run "python main.py -data <filename>"

How to implement

Chrome extension just grab text data from webpage.

The implementation of TextRank uses cosine similarity of word vectors from pre-trained Word2Vec embeddings as the similarity metric.