Document-search project to check if particular documents are accessible on Google using Rake implementation and Google Search API.

Prerequisites

These instructions will get you a copy of the project up and running on your local machine for development and testing purposes.

1) Get Google API key -> <https://developers.google.com/api-client-library/python/guide/aaa_apikeys>

2) Setup Google Custom Search Engine -> <https://support.google.com/customsearch/answer/2630963?hl=en> In order to search our documents on the entire page, follow these instructions-<https://support.google.com/customsearch/answer/2631040>

3) Install Google API client for Python pip install google-api-python-client

4) Install PDFMiner to convert pdf files to text <http://www.unixuser.org/~euske/python/pdfminer/>

Deployment

1) Save rake.py, GoogleSearch.py, SmartStoplist.txt in the same folder.

2) Set values for my\_api\_key and my\_cse\_id in GoogleSearch.py as per the values generated from step 1 and 2 of prerequistes.

3) Set my\_dir in GoogleSearch.py to the directory path of your documents.

4) Run GoogleSearch.py

Acknowledgments

Python implementation of RAKE algorithm: <https://github.com/zelandiya/RAKE-tutorial>

This project has been supervised by Prof. Craig Knoblock. Thanks go to him for giving such an interesting topic to work on and for all the helpful advice during project implementation. Contact GitHub API Training Shop Blog About © 2016 GitHub, Inc. Terms Privacy Security Status Help