

CFC Insight Technical Challenge

This technical challenge is designed to take 2 to 4 hours.

Your submission should be uploaded to a github repository and include your python code, a requirements.txt file and any setup or running instructions.

The submission should be portable – another person should be able to follow the setup and running instructions, and successfully run your code on their platform.

All code should be written for python version ≥ 3.6 . You are allowed to use external libraries and binaries if you choose.

The Challenge

Produce a program that:

1. Scrape the index webpage hosted at ``cfcunderwriting.com``
2. Writes a list of **all externally loaded resources** (e.g. images/scripts/fonts not hosted on cfcunderwriting.com) to a JSON output file.
3. Enumerates the page's hyperlinks and identifies the location of the "Privacy Policy" page
4. Use the privacy policy URL identified in step 3 and scrape the pages content. Produce a case-insensitive word frequency count for all of the visible text on the page. Your frequency count should also be written to a JSON output file..

Assessment Criteria

Treat the challenge as you would a real project so keep in mind readability, performance, code quality and comments.

Submission

Please send back, by email, the URL to the github repository that you have checked your work into.

Follow Up

Successful entries will be discussed at interview. We would love to understand what you would do if you had more time, and how you would expand the solution e.g. Adding support for international languages or making the solution more robust.