

RA Test Instructions

In the CSV file provided ("RA_test.csv") you will find a table containing the following information for the prospectuses of 6 mutual funds:

- Name of the fund family (i.e. the company): "groupname"
- Type of filing: "submissiontype"
- Date of the submission as of month end: "monthend"
- A link to a text file that contains the full submission on the EDGAR website: "link"

Based on this information, please write a Python program to carry out the following steps:

1. Access the CSV file
2. Retrieve the 6 prospectuses of interest directly from the EDGAR website, using the link provided in the CSV file
3. Write a simple text parser that performs the following actions for each file:
 - a. Extract the HTML header
 - b. Extract the following information from the header: the company name, the submission type and the date of filing
 - c. Compare the information extracted from the header to the information provided in the CSV file (i.e. compare the company name to "groupname", the submission type to "submissiontype" and the date of filing to "monthend")
 - d. Generate output stating which file is currently being processed, and whether each piece of information from the header matches the information provided in the CSV
4. Create an SQL table to store all the information for each fund (groupname, submissiontype, monthend, and full text of the prospectus) when all three variables in part 3c match.

Please note that:

- In order to extract the relevant information, you will need to examine the 6 text files yourself to identify the sections of the HTML header that describe the company name, the submission type and the date of filing
- The groupname provided in the CSV file is always lowercase and has been stripped of all special characters and extra whitespaces, while the name of the company in the prospectus may include uppercase letters, symbols or extra whitespaces.
- The date provided in the CSV file has been changed to the the last day of the month, while the date in the prospectus might fall on any day of the month. You will need to take this into account when comparing the dates (i.e., if the prospectus was filed in the same month and year as the date provided in the CSV file, you should consider that a match).
- For parsing in Python, I generally use Beautiful Soup (bs4) but you may use whatever you find most useful.

Submission:

If you wish to proceed with your application, please send a copy of your CV and a file containing a functioning version of your python code to columbiafinancera@gmail.com.