Test assignment for Han

Do a test task to extract certain information from a document

1/29/2021

Your input: resumes with different formats (provided to you)

Your output: extract the names of the companies you have worked in the past, preferably, also extract the corresponding dates (time range) for each company.

Your tasks:

First, you may only write a brief summary for what you think will be your approaches to get the output from your own resume, and then we can discuss.

Secondly, if you feel comfortable to start with, you may write some python codes to just extract the output from resumes, and turn in the results with the codes.

Take resumes with different formats (provided to you), and see if you can write a code module, or write steps for a generally applicable method that can extract as many company names as possible (probably not all company names as it could be difficult for you at this point), not by stacking up code versions for each of the different formats, but by a more generalized rule-system that can catch as many company names as possible. You can have multiple conditional statements to handle certain typical formats, but your codes need to be generally applicable to different resumes. For example, try to write your method based on half of the sample resumes, and see if they can apply the codes to the other half of the resumes of different formats and see how many different resumes your code can cover.

Deadline: 1 week from today