

### Test Question CIZADO131DD

Write a python code to extract emails found within the website content with their relevance. See the input and output format.

#### A. Input

Get 50% off on every purchase. contact marketing team at market@qq.com. Find all your linkedin contacts for free, jeff.peterson@b2bsearch.com. qq.com partnership program apply at market@qq.com

#### B. Expected Output

```
{ "market@qq.com" : {"Occurance":2, "EmailType": "Non-Human"},  
  "jeff.peterson@b2bsearch.com" : {"Occurance":1, "EmailType": "Human"}  
}
```

#### C. Explanation:

The output must be in a nested json format.

"Occurance" : No of times the email is repeated in the text.

"EmailType" : Type of the email. You can have more complex logic to identify human and non-human emails but in this exercise, Just try the logic given below:

Finding human emails: If the email is of format firstname.lastname@email.com then you can assume that the email is human.

Finding non-human emails: If the email format is text@email.com where text is less than 8 characters, then you can assume that the email is likely to be non-human.

**D. Testing your code:** Test your code against the dataset websiteData.txt and find the results. Please upload your code and results in a github repository and make it public. Share your github repository link in the form below. Please structure your project folder and file naming as shown below.

```
--Project root folder  
----emailScrap.py  
----websiteData.txt  
----result.json
```