jupyter notebook to load data from onet

author: Jan Jörg date: 13.03.2024

In [ ]:
```python
# imports
import requests
from requests.auth import HTTPBasicAuth
import json
from db import get_database
```

In [ ]:
```python
# import authentication data
with open('./infos.json') as f:
    infos = json.load(f)
    onet = infos['onet']
    onetUsername = onet['username']
    onetPassword = onet['password']
```

In [ ]:
```python
# get all available jobs from onet
occupationsUrl = "https://services.onetcenter.org/ws/online/occupations/"
headers = {'Accept': 'application/json'}
initialRequest = requests.get(occupationsUrl, auth=HTTPBasicAuth(onetUser

if initialRequest.status_code == 200:
    data = initialRequest.json()
    # the initial request only returns 20 jobs, so we need to make anothe
    # the initial request includes the total of available jobs
    total = data['total']
else:
    print(f"Request failed with status code {initialRequest.status_code}"

getAllOccupations = requests.get(occupationsUrl, auth=HTTPBasicAuth(onetU

if getAllOccupations.status_code == 200:
    data = getAllOccupations.json()
    occupations = data['occupation']
else:
    print(f"Request failed with status code {getAllOccupations.status_cod
```

In [ ]:
```python
# Fetch details for each occupation
summaries=[]

for occupation in occupations:
    summaryUrl = f"https://services.onetcenter.org/ws/online/occupations/
    summaryRequest = requests.get(summaryUrl, auth=HTTPBasicAuth(onetUser

    if summaryRequest.status_code == 200:
        occupation_data = summaryRequest.json()

        summaries.append(occupation_data)
    else:
```

```
        print(f"Request for occupation {occupation['code']} failed with s
```

In [ ]:
```python
dbname = get_database()

collection_name = dbname["onet"]

# Convert summaries to a list of dictionaries
summaries_dict = [summary for summary in summaries]

# Insert the dictionaries into the MongoDB collection
collection_name.insert_many(summaries_dict)

print("total jobs: ", len(summaries_dict))
```

total jobs: 1016