CTS API Introduction and Demo

Hubert Hickman

CTS API

The Cancer Clinical Trials Search API https://www.cancer.gov/syndication/api is a NCI supported API that provides a wide range of features including trial information and search capabilities. Much of the API content uses the NCI Thesaurus.

- CTS API Helpful links:
  - Swagger page - https://clinicaltrialsapi.cancer.gov/doc
  - You will need an API key - signup for a key here : https://clinicaltrialsapi.cancer.gov/signin
  - Rate limits for standard key 1,000 requests/day and 10 requests/sec ** (it is slower, IMHO)
  - Maximum number of trials returned per request is 50.

- NCI Thesaurus Helpful links:
  - EVS Explore (defaults to NCIt) https://evsexplore.semantics.cancer.gov/evsexplore/welcome
  - EVS API : https://evsexplore.semantics.cancer.gov/evsexplore/evsapi
    * No API key required
    * Will timeout if too much activity (need sleep/backoff timing code)

Simple Example

This example shows how to query the CTS API to get a count of active treatment trials. It retrieves one trial.

```python
import requests
import sys
import json
import os
```

```python
import pandas as pd
from dotenv import dotenv_values
import mercury as mr
config = dotenv_values('.env')
CTS_API_KEY=config['CTS_API_KEY']
cts_api_header = {"x-api-key": CTS_API_KEY,
                  "Content-Type": "application/json"}
includes = ['nct_id',
            'diseases',
            'biomarkers',
            'prior_therapy',
            'brief_title'
            ]

active_treatment_trials_that_are_recruiting = {'current_trial_status': 'Active',
        'sites.recruitment_status' : 'ACTIVE',
        'primary_purpose': 'TREATMENT',
        'size':1,
        'from':0,
        'include':includes
        }
r = requests.post('https://clinicaltrialsapi.cancer.gov/api/v2/trials',
                  json = active_treatment_trials_that_are_recruiting, headers=cts_api_header)

j = r.json()
mr.JSON(j)
```

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

Diseases

Expanding upon the above example, let us look at the diseases returned from the trial.

```python
import requests
import sys
import json
import os
import pandas as pd
```

```python
from dotenv import dotenv_values
import mercury as mr
import itables

config = dotenv_values('.env')
CTS_API_KEY=config['CTS_API_KEY']
cts_api_header = {"x-api-key": CTS_API_KEY,
                  "Content-Type": "application/json"}
includes = ['nct_id',
            'diseases',
            'biomarkers',
            'prior_therapy',
            'brief_title'
            ]

active_treatment_trials_that_are_recruiting = {'current_trial_status': 'Active',
        'sites.recruitment_status' : 'ACTIVE',
        'primary_purpose': 'TREATMENT',
        'size':1,
        'from':0,
        'include':includes
        }
r = requests.post('https://clinicaltrialsapi.cancer.gov/api/v2/trials',
                  json = active_treatment_trials_that_are_recruiting, headers=cts_api_header)

j = r.json()
diseases_df = pd.DataFrame(j['data'][0]['diseases'])
itables.show(diseases_df, column_filters="header")
```

TRIAL level diseases are those coded to the trial by clinical trial abstractors at NCI. TREE level diseases go 'up' the NCIt

Lead disease is/are the most focused trial level disease for the trial. Other trial level diseases are generally more broad or alternative matches.

Biomarkers

Biomarkers are abstracted as discrete data using NCIt codes. Biomarkers have been coded on new trials for a couple of years now – older trials may not have them even if the trial calls has biomarkers as inclucsion/exclusion criteria.

```python
import requests
import sys
import json
import os
import pandas as pd
from dotenv import dotenv_values
import mercury as mr
import itables


config = dotenv_values('.env')
CTS_API_KEY=config['CTS_API_KEY']
cts_api_header = {"x-api-key": CTS_API_KEY,
                  "Content-Type": "application/json"}
includes = ['nct_id',
            'diseases',
            'biomarkers',
            'prior_therapy',
            'brief_title'
            ]


active_treatment_trials_that_are_recruiting = {'current_trial_status': 'Active',
        'sites.recruitment_status' : 'ACTIVE',
        'primary_purpose': 'TREATMENT',
        'size':1,
        'from':0,
        'include':includes
        }
r = requests.post('https://clinicaltrialsapi.cancer.gov/api/v2/trials',
                  json = active_treatment_trials_that_are_recruiting, headers=cts_api_header)

j = r.json()
diseases_df = pd.DataFrame(j['data'][0]['biomarkers'])
itables.show(diseases_df, column_filters="header",
 buttons=["pageLength", "copyHtml5", "csvHtml5", "excelHtml5"])
```