
# Google Scholar's-Research Paper

Programming Languages Lab

Naima Farooqi 16803005

Aditya Gupta 16803004

Submitted to: Mr. Raghu Vamsi Potukuchi
Date: 09/05/2019

## Abstract

This project is about accessing the information of all the research papers written by teachers/Professors of JIIT, Noida. All the necessary details such as the number of citations in a particular year, the name of the research papers, etc are displayed in a very user-friendly way on the GUI designed by us.

## Scope of the project

- **Functional Requirements:**
  **a)** ParseHub: A software called ParseHub is used to scrape the data from the website. Necessary information is marked on the website and then it is extracted.
  **b)** Python: The API-key from parseHub is written in our python code, so whenever we run the python code, automatically ParseHub would extract information.
  **c)** Tkinter: Tkinter is used to design the GUI.
  **d)** Exception Handling
  **e)** Regular Expressions
  **f)** JSON Parsing

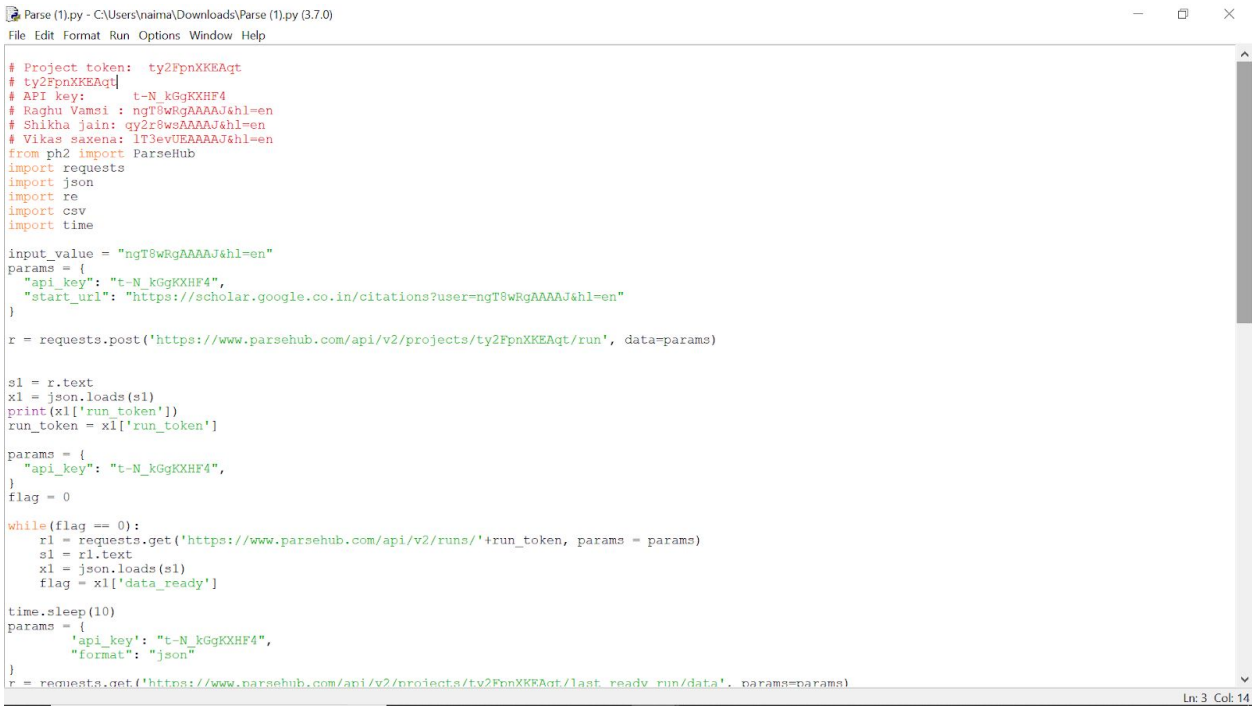
- **Non-Functional Requirements:**
  **a)** The input provided should always be a URL.
  **b)** The ParseHub API-key is always fixed and depends on the the project saved on ParseHub account you are currently logged in.
  **c)** Accuracy is 100% but Time complexity is not very impressive

- **Description of the Module:**

## Design of the Project

In this project, we have extracted the details of all the research papers written by teachers of JIIT, Noida from Google scholars website using web data scraping. All the essential details being scrapped using a proper python program. We have programmed a Graphic User Interface which would give us the specific details needed by the user. The user only needs to input the URI of a particular teacher and then select what and how he wants the output to be displayed (research paper wise, group by year or citations, etc). The needful would be displayed on our GUI's screen.

# Implementation



Parse (2).py - C:\Users\naima\Downloads\Parse (2).py (3.7.0)

File Edit Format Run Options Window Help

```
}
r = requests.get('https://www.parsehub.com/api/v2/projects/ty2FpnXKEAqt/last_ready_run/data', params=params)
s = r.text
with open('scholars_'+input_value+'.csv','w') as writeFile:
        writer = csv.writer(writeFile)
        writer.writerow(["Paper_Name","Total_cite_count","Publish_year","Authors","Publication_date","Conference","Pages","Publisher","cite_in_2000","cite_in
        data = json.loads(s)
        count = 0
        flag = ""
        for i in range(len(data["Paper_name"])):

                paper_name = ""
                total_citation = ""
                publish_year = ""
                authors = ""
                publication_date = ""
                conference = ""
                pages = ""
                publisher = ""
                cite_value = ["0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0","0"]
                k = data["Paper_name"]
                # s = k[i]
                paper = k[i]
                paper_name = paper['name']
                if flag == paper_name:
                        break
                if count == 0:
                        flag = paper_name
                        count += 1
                try:
                        total_citation = paper['Total_citation']
                except KeyError:
                        total_citation = ""
                try:
                        publish_year = paper['Published_year']
                except KeyError:
                        publish_year = ""
                authors = paper['Authors']

                try:
                        publication_date = paper['publicationDate']
                except KeyError:
                        publication_date = ""

                try:
```

Ln: 2 Col: 30

Parse (2).py - C:\Users\naima\Downloads\Parse (2).py (3.7.0)

File Edit Format Run Options Window Help

```
                        flag = paper_name
                        count += 1
                try:
                        total_citation = paper['Total_citation']
                except KeyError:
                        total_citation = ""
                try:
                        publish_year = paper['Published_year']
                except KeyError:
                        publish_year = ""
                authors = paper['Authors']

                try:
                        publication_date = paper['publicationDate']
                except KeyError:
                        publication_date = ""

                try:
                        conference = paper['Conference']
                except KeyError:
                        conference = ""

                try:
                        pages = paper['pages']
                except KeyError:
                        pages = ""

                try:
                        publisher = paper['publisher']
                except KeyError:
                        publisher = ""

                try:
                        citation = paper['Citation_count']
                except KeyError:
                        citation = []
                for j in range(len(citation)):
                        cite = citation[j]
                        y = cite["url"]
                        sp = (re.split(r'=',y))
                        year = sp[-1]
                        index = int(year[2:])
                        cite_value[index] = cite['Cite']
                writer.writerow([paper_name,total_citation,publish_year,authors,publication_date,conference,pages,publisher,cite_value[0],cite_value[1],cite_
writeFile.close()
```

Ln: 2 Col: 30