

Research Paper Search Automation using a Google API

CS 6463 - Parallel and Distributed Software Systems

Project Documentation

Dependencies:

Google Account

Python Package Manager

Google API package for Python

Steps for Setting up the project:

1- Create a Google account if needed

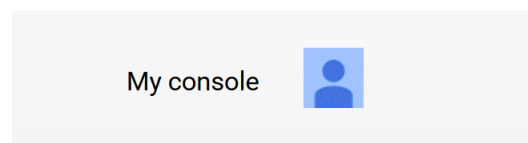
A new Google Account can be created here:

<https://accounts.google.com/SignUp?hl=en>

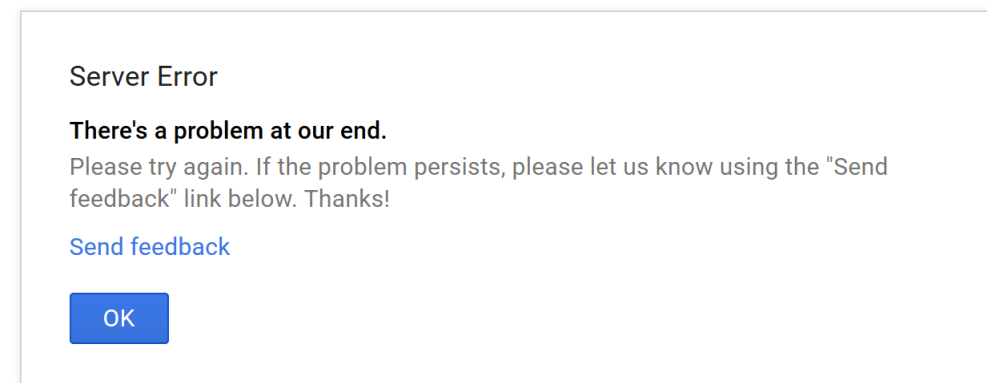
2- Create a custom search project in the Google cloud platform:

Go to <https://cloud.google.com/>

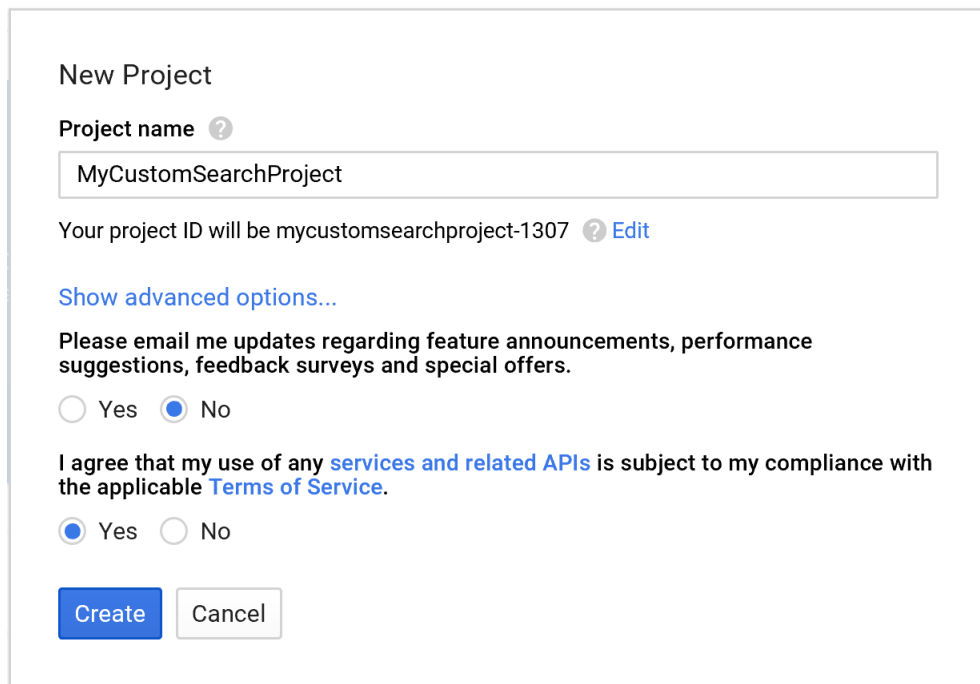
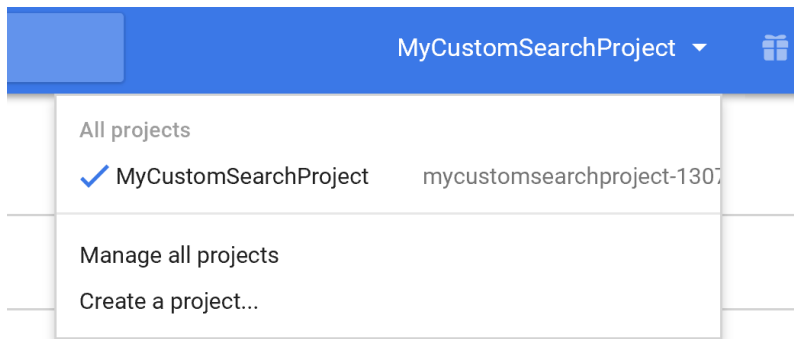
Click on My Console on the right hand side



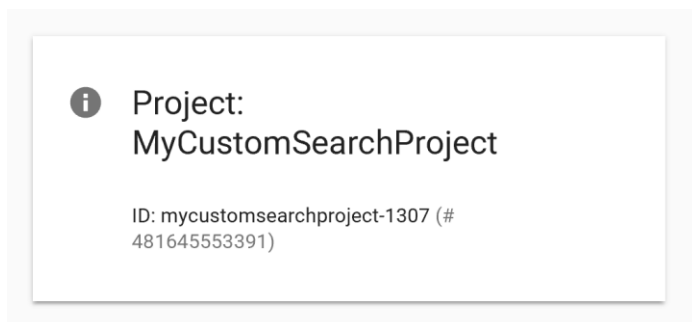
If no project has been created previously it might throw an error when using FireFox so preferably use Google Chrome for this step.



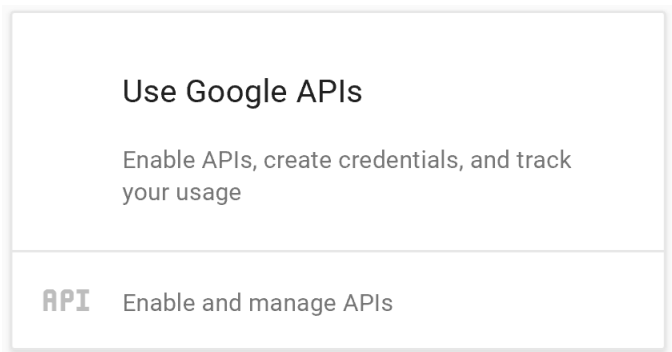
The option for creating a project will be available on the top menu bar, select “Create a project”

A screenshot of a 'New Project' form. The form has a title 'New Project'. Below it is a 'Project name' field with a question mark icon, containing the text 'MyCustomSearchProject'. Below the name field is a message: 'Your project ID will be mycustomsearchproject-1307' followed by a question mark icon and a blue 'Edit' link. Below this is a link 'Show advanced options...'. Then there is a section for email updates: 'Please email me updates regarding feature announcements, performance suggestions, feedback surveys and special offers.' with radio buttons for 'Yes' and 'No' (where 'No' is selected). Below that is a section for terms of service: 'I agree that my use of any services and related APIs is subject to my compliance with the applicable Terms of Service.' with radio buttons for 'Yes' and 'No' (where 'Yes' is selected). At the bottom are two buttons: 'Create' (blue) and 'Cancel' (white with a grey border).

It will take a few seconds and then the project will appear in the dashboard



Enable the custom search API by clicking on the dashboard that has the custom search Widget



Then click on custom search under other API's



Other popular APIs

[Analytics API](#)

[Translate API](#)

[Custom Search API](#)

[URL Shortener API](#)

[PageSpeed Insights API](#)

[Fusion Tables API](#)

[Web Fonts Developer API](#)

Click the enable button, once enable it will ask you to set up credentials, so click on go to credentials



Custom Search API

The link will try to walk you through the best type of key for your needs but for this project I have used an API key so you can just click on API key and select Server Key and click Create

Create server API key

This key should be kept secret on your server

Every API request is generated by software running on a machine that you control. Per-user limits will be enforced using the address found in each request's `userIp` parameter, if specified. If the `userIp` parameter is missing, your machine's IP address will be used instead. [Learn more](#)

Name

Accept requests from these server IP addresses (Optional)

Examples: 192.168.0.1, 172.16.0.0/12, 2001:db8::1 or 2001:db8::/64

Note: It may take up to 5 minutes for settings to take effect

You will get an API key; save it this will be the first key you will need for the project.

3 - Create a Custom Search

Go to this site

<https://cse.google.com>

On the right hand side click on the button create a custom search

Add the dblp site as the site for the custom search - <http://dblp.uni-trier.de/>

Enter the site name and click "Create" to create a search engine for your site. [Learn more](#)

Sites to search

Click Create.

Congratulations!

You've successfully created your Custom search engine.

Add it to your site

Get code

Click on get the code and look for the line of code that has the following:

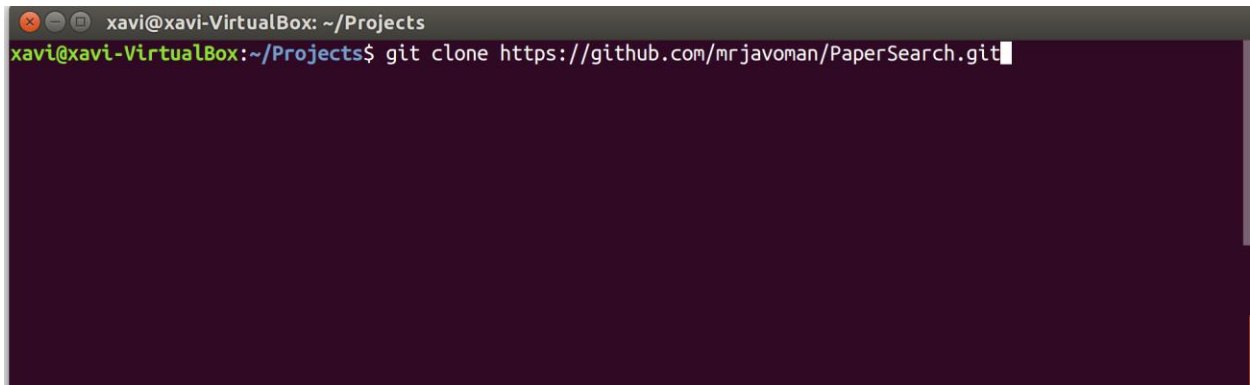
```
var cx = "xxxxxxxxx ..."
```

Save the number since this will be used in the configuration file of the script.

Additionally, this code can also be obtained when clicking on the main menu and then clicking on the custom search list item.

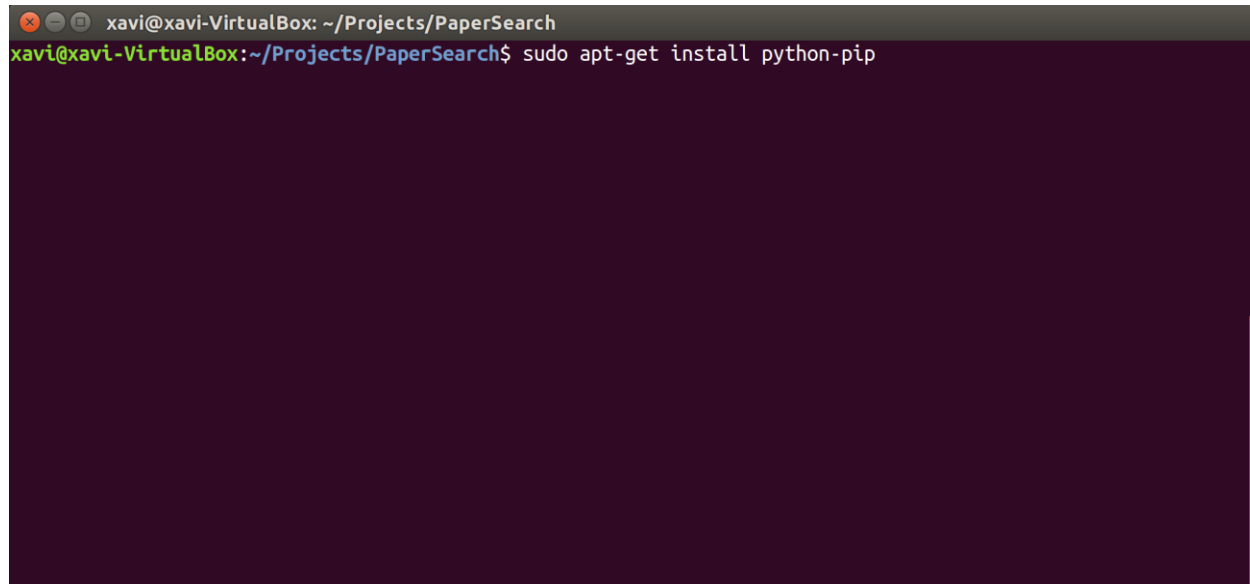
4 - Install the script

Download the project from GitHub

A screenshot of a terminal window with a dark background. The title bar at the top reads 'xavi@xavi-VirtualBox: ~/Projects'. The terminal shows the command 'xavi@xavi-VirtualBox:~/Projects\$ git clone https://github.com/mrjavoman/PaperSearch.git' with a cursor at the end of the line.

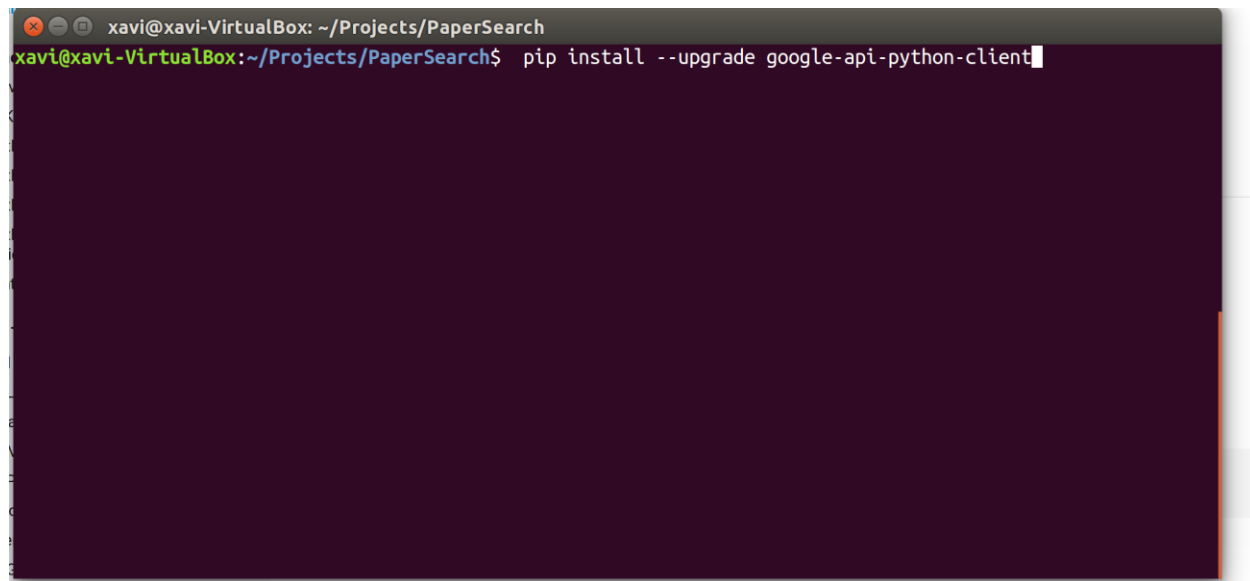
```
xavi@xavi-VirtualBox: ~/Projects
xavi@xavi-VirtualBox:~/Projects$ git clone https://github.com/mrjavoman/PaperSearch.git
```

Install the python package manager if not already installed

A terminal window with a dark purple background. The title bar shows 'xavi@xavi-VirtualBox: ~/Projects/PaperSearch'. The command prompt is 'xavi@xavi-VirtualBox:~/Projects/PaperSearch\$' and the command entered is 'sudo apt-get install python-pip'.

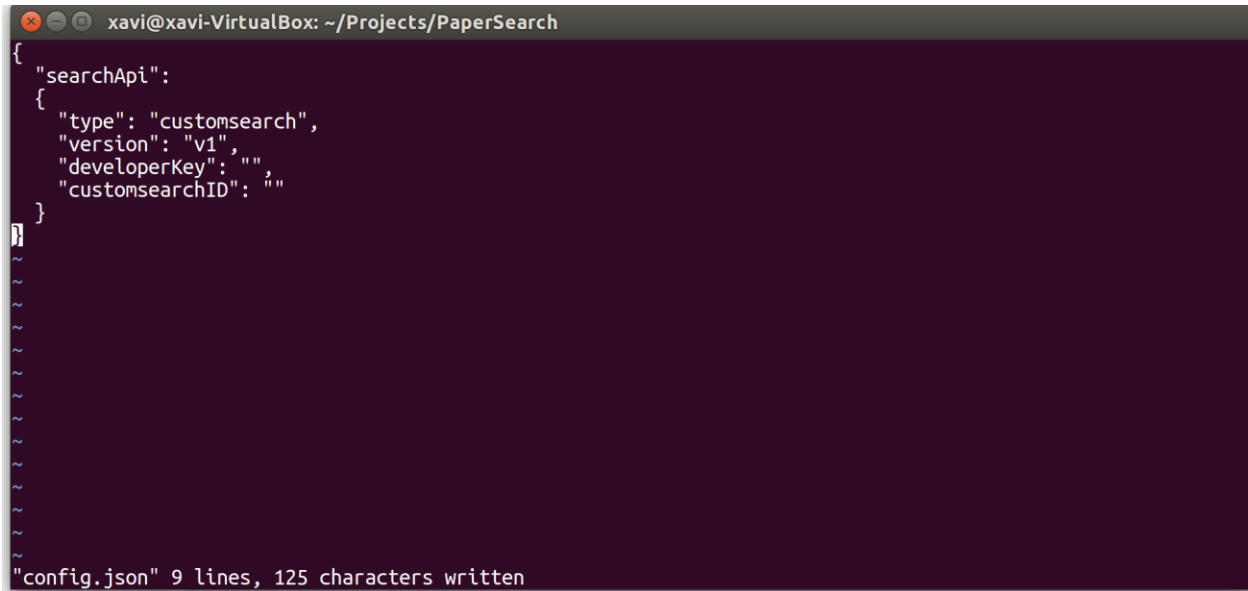
```
xavi@xavi-VirtualBox: ~/Projects/PaperSearch
xavi@xavi-VirtualBox:~/Projects/PaperSearch$ sudo apt-get install python-pip
```

Install the Google API package

A terminal window with a dark purple background. The title bar shows 'xavi@xavi-VirtualBox: ~/Projects/PaperSearch'. The command prompt is 'xavi@xavi-VirtualBox:~/Projects/PaperSearch\$' and the command entered is 'pip install --upgrade google-api-python-client'.

```
xavi@xavi-VirtualBox: ~/Projects/PaperSearch
xavi@xavi-VirtualBox:~/Projects/PaperSearch$ pip install --upgrade google-api-python-client
```

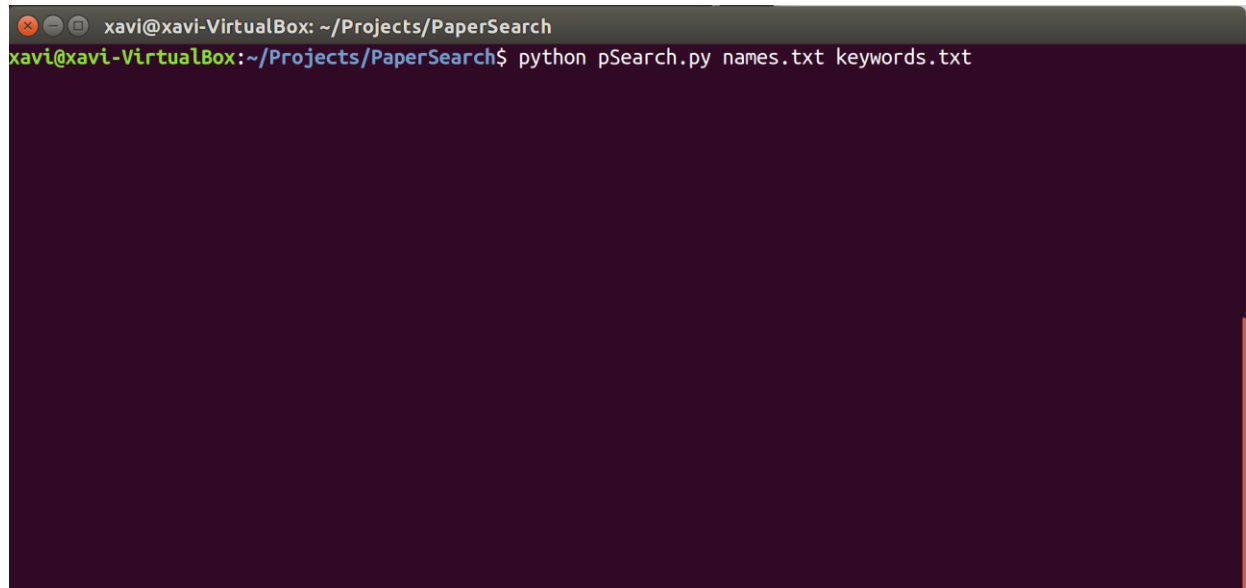
Setup the developer key and the custom search engine id in the config.json file, the developer key will be the Server Key that was generated in step 2 and the Search Engine ID will be the one that was generated in step 3

A terminal window titled 'xavi@xavi-VirtualBox: ~/Projects/PaperSearch' shows the creation of a file named 'config.json'. The file content is a JSON object with a 'searchApi' key containing an object with 'type', 'version', 'developerKey', and 'customsearchID' fields. The terminal shows the file being created and its size (9 lines, 125 characters).

```
xavi@xavi-VirtualBox: ~/Projects/PaperSearch
{
  "searchApi":
  {
    "type": "customsearch",
    "version": "v1",
    "developerKey": "",
    "customsearchID": ""
  }
}

"config.json" 9 lines, 125 characters written
```

Run the script with the following command:

A terminal window titled 'xavi@xavi-VirtualBox: ~/Projects/PaperSearch' shows the execution of a Python script. The command 'python pSearch.py names.txt keywords.txt' has been entered and executed, resulting in a blank terminal screen.

```
xavi@xavi-VirtualBox: ~/Projects/PaperSearch
xavi@xavi-VirtualBox:~/Projects/PaperSearch$ python pSearch.py names.txt keywords.txt
```