

D-Search

Students: Kareem Shahhat Author, Omar Yousef, and Mostafa Taha

Supervisor: Assoc. Prof. Emad S. Othman

Senior Member IEEE - Region 8, High Institute for
Computers and Information Systems, AL-Shorouk Academy,
Cairo – Egypt, PH- 0020-010-25830256. E-mail: emad_othman@sha.edu.eg

Abstract:

The project, in short, is a development in the way of displaying the search results so that it does not display the sites or links that are related to the search word, but rather pulls the entire content of the sites and displays it to the user with all its information without opening it so that it makes you browse more than one site at the same time and access the largest amount of information about what you are looking for in Less time, but the big role of our site appears in searching for repetitive things, such as products or things that you need to compare, such as apartments, cars, phones, and so on. As it can download the extracted data in excel and jason file

Keywords: Data collection, Web Scraping, Python, Search, Data Analysis

1. Introduction:

The search engine in English is a program that is accessed through the Internet. In its work, it is based on a database that contains a lot of information, as well as on the principle of query. The search engine provides a list of results that match what the user is trying to search for on the Internet. The best Today there are billions of sites on the Internet, and there is a lot of information on it, so search engines make it easy to find these sites and information at present there are a lot of different search engines that are available on the Internet and each type of search engine has its own characteristics and the most well-known and used search engine is Google, and there are many other engines, such as: Yandex, Yahoo, Bing. Search engines make

finding what you are looking for easy with just the click of one button, but it is important to know how search engines work and

how to improve the appearance of your site in them. Learn the difference between organic search results and paid results and learn the importance and effectiveness of search engine advertising, when a user search there are thousands and sometimes millions of web pages with useful information, given that. Search discovers and indexes all content you Be available online from, web pages. PDFs, photos, videos, etc to this through a process known as “crawling” and “indexing They are arranged according to how closely they match the query in a process.

2. Statistical Analysis:

The reason we chose to build the project as a website because it can be operated on any electronic device that supports internet connection we are not limited to a specific operating system such as Android or Windows but it can be run on all devices and another reason is that it will not need to install and take additional space from the device or take from the resources of the device

3. What is web scraping?

Web scraping is the process of collecting structured web data in an automated fashion. It's also called web data extraction. Some of the main use cases of web scraping include price monitoring, price intelligence, news monitoring, lead generation, and market research among many others.

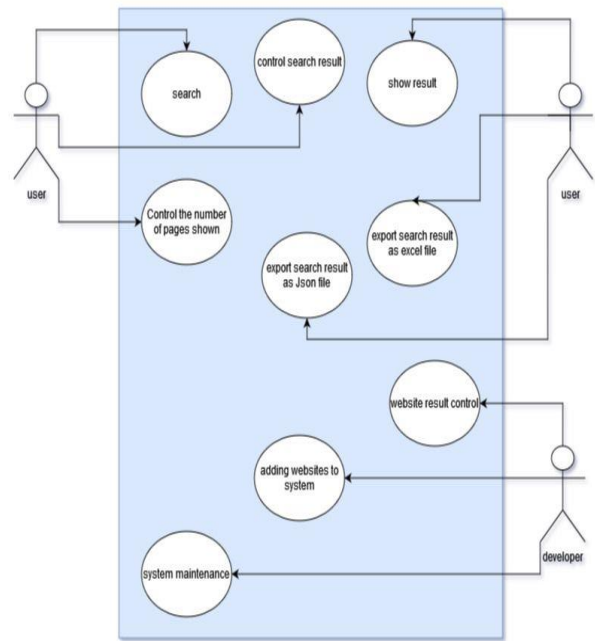
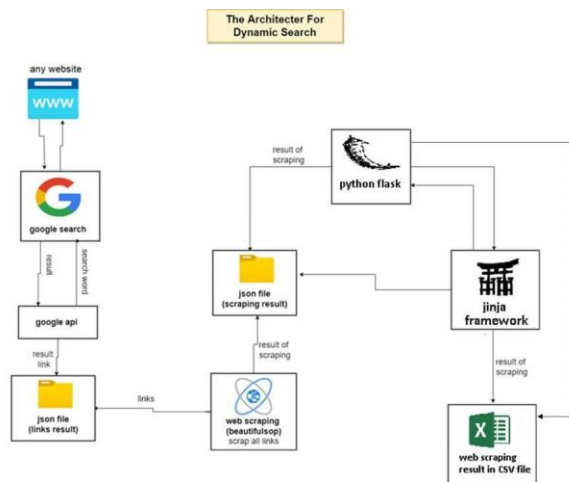
In general, web data extraction is used by people and businesses who want to make use of the vast amount of publicly available web data to make smarter decisions.

If you've ever copied and pasted information from a website, you've performed the same function as any web scraper, only on a microscopic, manual scale. Unlike the mundane, mind-numbing process of manually extracting data, web scraping uses intelligent automation to retrieve hundreds, millions, or even billions of data points from the internet's seemingly endless frontier.

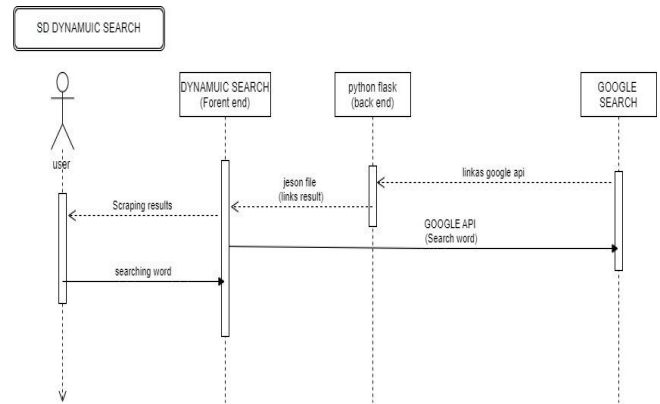
4. Problem Statement and Motivation:

- 1 - Regular search methods take more time and effort
- 2 - Difficulty accessing information
- 3 - Accessing more than one website at the same time
- 4 - Facilitating easier access to search results
- 5 - Save time and effort for the user during the search process
- 6 - Reaching the best search result with more clear details
- 7 - Access to the best way to display the largest number of results
- 8 - Display the results of more than one site at the same time
- 9 - think about the development of the form of search results and this will facilitate

5. System Architecture:



- Use case Diagram



- Sequence Diagram

6. Software Implementation:

Python Flask language definition:-

The latest stable version of Flask 0.12.2 released in May 2017 [4] Some applications that use Flask Pinterest,[5] Link Ann,[6] and the Flask community site.[7] Flask is called a mini framework because it does not require any tools or libraries in particular [8] It does not have a database abstraction layer, validator registration forms, or any other components but there are third party libraries that provide similar things anyway, Flask supports plugins that It can add

app features as if it were built in your flash. There are object-relationship matching plugins, validation of registration forms, upload handling, various open authentication techniques and several framework-related tools. These plugins are updated more regularly than the kernel frame flask.

Language features :-

- 1 - Contains a development server and debugger
- 2 - Full support for unit testing
- 3 - Restful year
- 4- Uses templates jinja2 4
- 5 - Support Tamil cookies
- 6 - Compatible with WSGI 1.0 .
- 7 - Uses Unicode .
- 8 - Validation included .
- 9 - Compatible with Google Play Engine .
- 10 - Injuries available to provide the required feat

i. High Performance

Many factors affect application performance, including CPU usage, frame number per second, average response time, request number per second, and more.

If we load the site on a powerful server, the issue will not be difficult, because most servers on the Internet have higher performance than a normal home computer, and this will affect the performance significantly.

ii. comprehensive search

The site helps users to access more than one information from more than one at the same time, but with a design that helps the user to see the products or search results better away from the crowds and advertisements that are found in regular sites

iii. Methodology

The private system contains a group of sites in most fields, which are the most famous in their field

Where for each field from one site to 3 sites, for example, a site or three for shopping, a site for books, a comprehensive shopping site, a research site, a real estate site, and so on, and each of these sites has its own code to withdraw information from within it, as our project cannot withdraw from any site unless it has a code to withdraw data from him modifications can be added based on the wishes of users to facilitate the process of using the program.

SYSTEM FUNCTIONS:

1. Display the results inside the sites:

After entering the search word and page number, the site displays to the user the results pulled from within the sites. This data is instant and updated. The system does not store any search results, but only displays them.

2. Load those results into an excel and json file:

By just a button click you can download all the data displayed in an excel or json file so that you can analyze that data more deeply

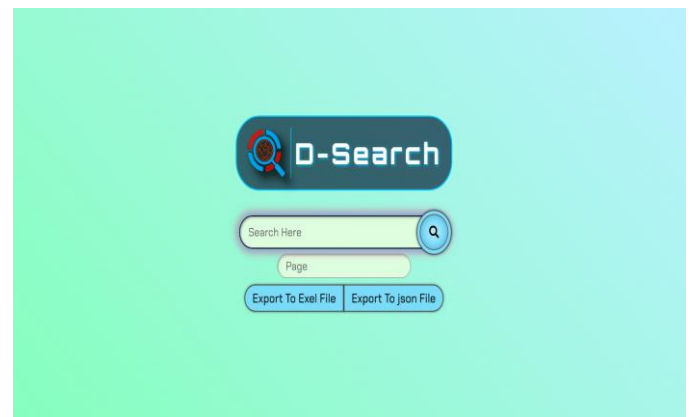


Fig. 1. D-search Screen

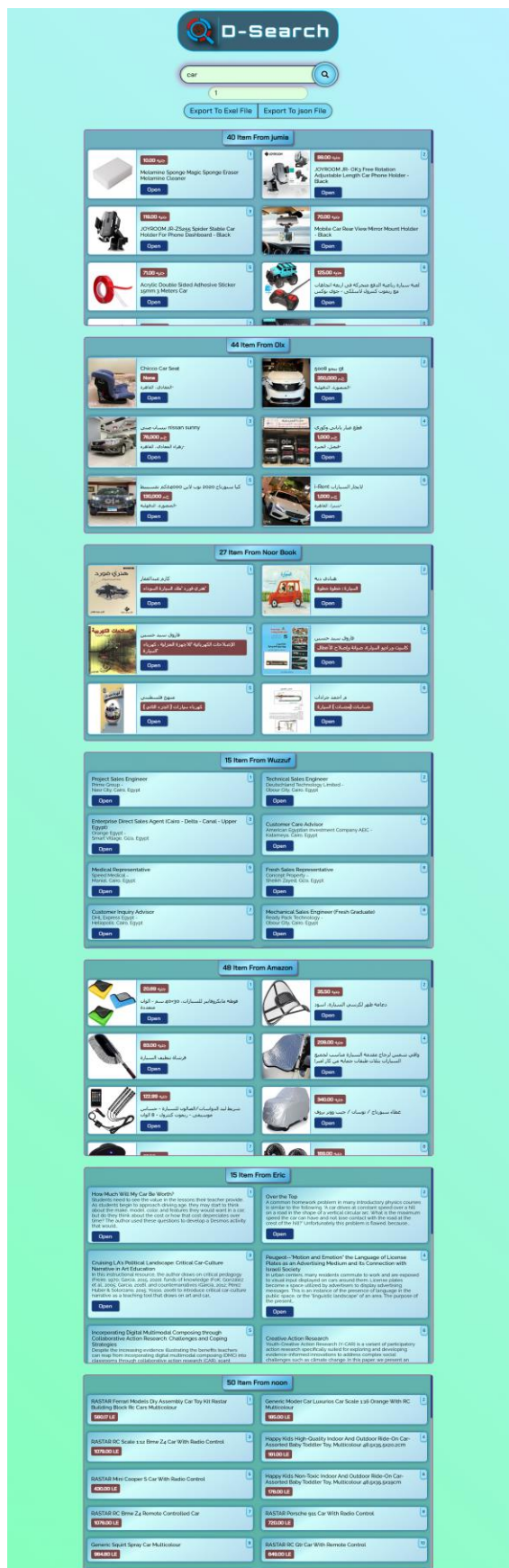


Figure2. Search result Screen

ACKNOWLEDGMENT

The authors are grateful to Prof. Emad Othman for his assistance and helpful discussions.

REFERENCES

1. Instant Flask Web Development ,Ron DuPlain , Esha Thakke ,2013
2. Flask Documentation (1.1.x)Release 1.1.4 ,Pallets ,Feb 21, 2022
3. Flask Web Development Developing Web Applications with Python, Miguel Grinberg ,2018
4. Discover your flask, Robert Pickard, 2017
5. Python Crash Course, Eric Matthes, 2016
6. Python Notes for Professionals @Notes for Professionals, official Python® group, 2020
7. Django Cookbook Web Development with Django, Beau Curtin, 2016