MINI - PROJECT REPORT

"Web scraping {flipkart web scraper}"

Class: **TY NS**
Group ID: **02**

BY
Anurag Chauhan
Ujjawal Agarwal
Abhinav Kadam
Palash Gajjar

Under the Guidance of

Prof Amol Dande

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MIT School of Engineering
Loni Kalbhor, Pune

CERTIFICATE

This is to certify that the Mini- Project report entitled

"***Web scraping {flipkart web scraper}***"

submitted by

GROUP MEMBERS	ENROLLMENT NO.	ROLL NO.	CONTACT NO
Abhinav kadam	MITU20BTCS0005	2203121	7756917309
Anurag Chauhan	MITU20BTCS0042	2203333	9313950989
Ujjawal Agarwal	MITU20BTCS0323	2203288	8239121758
Palash Gajjar	MITU20BTCS0187	2203145	8320937114

is a record of Bonafide work carried out by them, under my guidance, in partial fulfilment of the requirement for the Third Year of Engineering (Computer) at M.I.T. School of Engineering, Pune under MIT Art, Design & Technology University.

Date: Place:

prof. Amol Dande
Guide,
Department of CSE,
M.I.T. School of Engineering
Loni Kalbhor, Pune

Dr. Shraddha Phansalkar
Head,
Department of CSE
MIT School of Engineering
Loni-Kalbhor, Pune

ACKNOWLEDGEMENT

We would like to express our deepest appreciation to all those who provided us the possibility to complete this project. A special gratitude we give to our Guide *Prof. Amol Dande* whose contribution in stimulating suggestions and encouragement, helped us to coordinate our project especially in writing this report.

Furthermore, we would also like to acknowledge with much appreciation the crucial role of the staff of Mini Project, who gave the permission to use all required equipment and the necessary materials to complete this Project.

Group Members:
Abhinav Kadam
Anurag Chauhan
Ujjwal Agarwal
Palash Gajjar

TABLE OF CONTENTS

PROJECT CATEGORY

Web Scraping comes the category of: -

- *Machine learning*
- *Data Science*

INTRODUCTION

Web scraping is an automatic method to obtain large amounts of data from websites. Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications. There are many different ways to perform web scraping to obtain data from websites. These include using online services, particular API's or even creating your code for web scraping from scratch. Many large websites, like Google, Twitter, Facebook, StackOverflow, etc. have API's that allow you to access their data in a structured format.

PROBLEM STATEMENT

Imagine you have to pull a large amount of data from websites, and you want to do it as quickly as possible. How would you do it without manually going to each website and getting the data?

ABSTRACT

Web scraping is an automated method used to extract large amounts of data from websites. The data on the websites are unstructured. Web scraping helps collect these unstructured data and store it in a structured form. This project is specifically for scraping product information like product name, prices, product ratings about desired product, product range from flipkart. The scraped data it then automatically stored into an excel file (.csv file).

TECHNOLOGY AND RESOURCES

Language: - *Python*

Python Libraries

- *Tkinter*
- *requests*
- *bs4*
- *pandas*
- *PIL*
- *os*

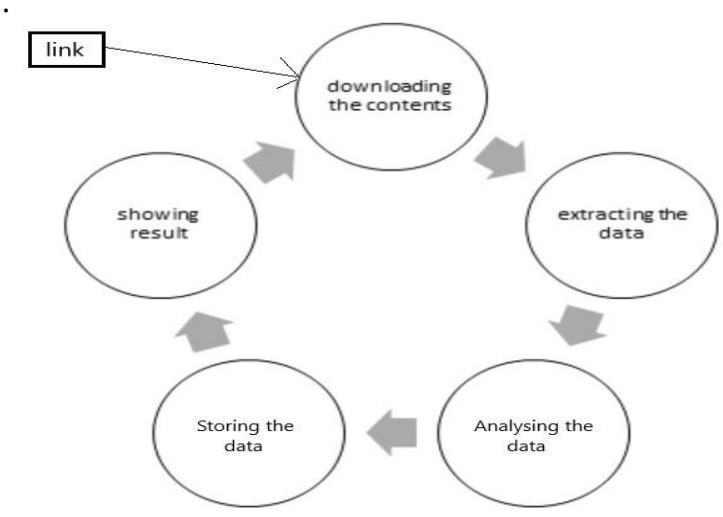
Software Used: *PyCharm*

PROCESS DESCRIPTION

The working of flipkart web-scraper in simple steps

- Find the URL that you want to scrape {flipkart URL}
- Run Code
- Scraper will ask user to Enter URL
- Scraper will ask user to Enter the file name
- It will extract the data from given URL
- It then converts the extracted unstructured data into a structured format.
- It will create an excel(.csv) file with provided filename Store the data in it.

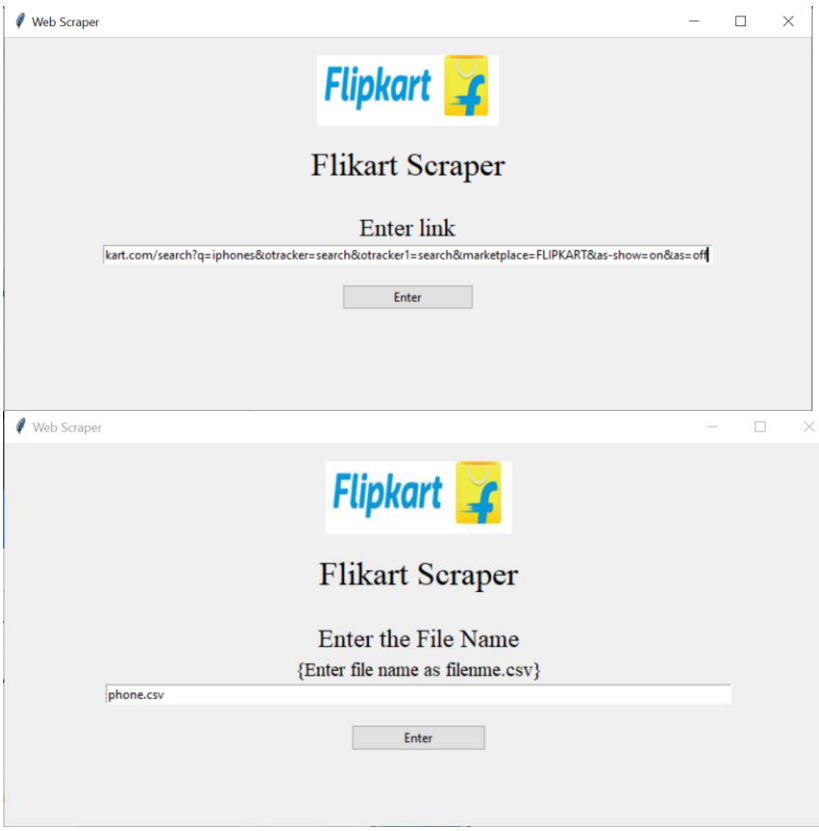
PROPOSED SYSTEM [BLOCK DIAGRAM]



LITERATURE SURVEY

S no.	Source	Title	Date
1.	ICECA	Data Analysis by Web Scraping using Python	02 September 2019
2.	IEEE	Web Scraping: From Tools to Related Legislation and Implementation Using Python	03 February 2021
3.	Getting Structured Data from the Internet {book}	Web Scraping in Python Using Beautiful Soup Library	13 November 2020
4.	Research paper	pandas: a Foundational Python Library for Data Analysis and Statistics	08 November 2012
5.	**Python** GUI Programming with **Tkinter** **{book}**	Develop responsive and powerful GUI applications with **Tkinter**	May 2018

OUTPUT SCREENSHOTS



	A	B	C	D
1	Product Name	Prices	Ratings	
2	APPLE iPhone 13 (Starlight, 128 GB)	Rs. 65,999	4.7	
3	APPLE iPhone 13 (Midnight, 128 GB)	Rs. 65,999	4.7	
4	APPLE iPhone 13 (Pink, 128 GB)	Rs. 65,999	4.7	
5	APPLE iPhone 13 (Green, 128 GB)	Rs. 65,999	4.7	
6	APPLE iPhone 13 (Blue, 128 GB)	Rs. 65,999	4.7	
7	APPLE iPhone 11 (White, 64 GB)	Rs. 39,999	4.6	
8	APPLE iPhone 11 (Black, 64 GB)	Rs. 39,999	4.6	
9	APPLE iPhone 11 (Black, 128 GB)	Rs. 48,900	4.6	
10	APPLE iPhone 11 (White, 128 GB)	Rs. 48,900	4.6	
11	APPLE iPhone 11 (Purple, 64 GB)	Rs. 39,999	4.6	
12	APPLE iPhone 12 (Black, 64 GB)	Rs. 48,999	4.6	
13	APPLE iPhone 12 (Blue, 64 GB)	Rs. 48,999	4.6	
14	APPLE iPhone 11 (Green, 64 GB)	Rs. 39,999	4.6	
15	APPLE iPhone 12 Mini (Blue, 64 GB)	Rs. 43,999	4.5	
16	APPLE iPhone 11 (Purple, 128 GB)	Rs. 44,999	4.6	
17	APPLE iPhone 12 Mini (Black, 64 GB)	Rs. 43,999	4.5	
18	APPLE iPhone 13 ((PRODUCT)RED, 128 GB)	Rs. 65,999	4.7	

Conclusion

This project will help users to easily scrape data about products from any flipkart webpage and obtain the output in excel file(.csv).

FUTURE SCOPE AND FURTHER ENHANCEMENTS

Features like file encryption can be added to the existing project. This project can also be made as an extension for a browser for easier use. Web scraping is used in a variety of digital businesses that rely on data harvesting. So, this project can be used by individuals, businesses for data harvesting.

REFERENCES

1. Broucke, Seppe Vanden, and Bart Baesens. *Practical Web Scraping for Data Science: best practices and examples with Python*. CreateSpace, 2017.
2. *https://www.scrapingbee.com/blog/web-scraping-101-with-python/*
3. Lawson, Richard. *Web scraping with Python*. Packt Publishing Ltd, 2015.
4. O'Reilly, S. (2006). *Nominative fair use and Internet aggregators: Copyright and trademark challenges posed by bots, web crawlers and screen-scraping technologies. Loyola Consumer Law Review, 19, 273.*
5. Mitchell, Ryan. *Web scraping with Python: Collecting more data from the modern web*. " O'Reilly Media, Inc.", 2018.

6.4 Form A

MIT School of Engineering
Department of Computer Science and Engineering
Mini Project III Topic Approval

Class: TYNS
Group ID: 02
Group Members:

Date: 22/09/2022

Enrollment No	Name of Student	E-mail Address	Contact
MITU20BTCS0042	Anurag Chauhan	anurag1000chauhan@gmail.com	9313950989
MITU20BTCS0323	Ujjawal Agrawal	Ujjawalagrawal05@gmail.com	8239121758
MITU20BTCS0005	Abhinav Kadam	iamonly2001@gmail.com	7756917309
MITU20BTCS0187	Palash Gajjar	palashgajjarsh@gmail.com	8320937114

Project Title Evaluation Parameters:

Sr. No.	Parameters	Topic 1	Topic 2	Topic 3
1.	Title	Web scraping {flipkart scraper}	Attendance management system with facial recognition	Parking management system
2.	Domain Expertise	Data science and machine learning	Artificial intelligence	Artificial intelligence
3.	Technical Feasibility	Any pc which can run python(3.6)	Camera and any pc which can run python(3.6)	Camera and any pc which can run python(3.6)
4.	Future Scope	Can be used by individuals, businesses for data harvesting.	Can be used in ATMs, Offices, schools, etc.	Can be used in big parking lots.
5.	Applicability	Applicable	Applicable	Applicable
	Approved (✓) (Filled by the Examiner)	Approved	Not approved	Not approved
Remark:				

Sr. No. Name of Subject Expert Signature

1. Prof. Amol Dande

2.

Form B

Date: 22/09/2022

Class: TY CSE NS
Project Group ID: 02
Project Title: Web Scraping {flipkart webscraper}

Project Title Evaluation Parameters:

Sr. No.	Parameters	Description About Project	Marks(5)
1.	Business Ideas and Implementation from project	This GUI based web scrapping tool scrapes data like product name, prices, ratings from flipkart page and store output in a csv file.	
2.	Market Survey (competitors, substitute products, potential market, etc.)	Although the market for this is not huge, a very small number of digital data harvesting businesses along with some individuals may have use for it. There are no simple scraping tools that just scrape data for flipkart but there are some advanced scrapers that can do the job.	
3.	Market Acceptability of Product	Although there are sophisticated web scrapping tools available for flipkart webscrapping but there is a need for simple, cheap, and easy to use web scrapping tools in the market.	
4.	Emerging Trends about Project and Product	One of the most exciting trends in web scraping at the moment is the growing capabilities of AI data parsers and crawlers.	

5.	Income Generation ideas through Project	We can license the tools for an annual fee, or we sell the tool where the customer pays for the tool while buying and can use it for lifetime.	
6.	Project Profitability	Since there is no cost of development for this tool all the money made is profit.	
7.	Cost Benefit Analysis	There's no initial cost incurred for the tool. All the resources to create the scrapping tool will be free.	
8.	Any Other Point	————————	
Remark:			

Commercial Feasibility of the project is evaluated based on the above parameters.

Project Approval Status: Approved / Not Approved

(Name & Designation of Market Expert)
Signature with Date.

GROUP NO: - 02

Topic: - Web Scraping {flipkart Web Scraper}

DATE: - 22/09/2022

Group Members:

Enrolment Number	ROLL NO.	NAME OF STUDENT	EMAIL ID	CONTACT NUMBER
MITUBTCS0005	2203121	ANHINAV KADAM	iamonly2001@gmail.com	77569617309
MITU20BTCS0042	2203333	Anurag Chauhan	anurag1000chauhan@gmail.com	9313950989
MITU20BTCS0187	2203145	Palash Gajjar	palashgajjarsh@gmail.com	8320937114
MITU20BTCS0323	2203288	Ujjawal agrawal	UJJAWALAGRAWAL05@gmail.com	8239121758

Guided by: Prof. Amol Dande

Problem Statement: - Imagine you have to pull a large amount of data from websites, and you want to do it as quickly as possible. How would you do it without manually going to each website and getting the data?

Abstract: -

Web scraping is an automated method used to extract large amounts of data from websites. The data on the websites are unstructured. Web scraping helps collect these unstructured data and store it in a structured form. This project is specifically for scraping product information like product name, prices, product ratings about desired product, product range from flipkart. The scraped data it then automatically stored into an excel file (.csv file).

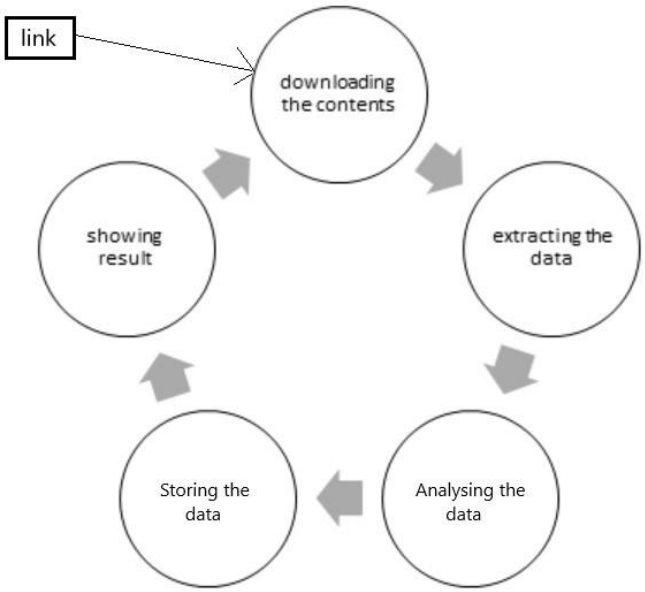
The working of flipkart web-scraper in simple steps

- Find the URL that you want to scrape {flipkart URL}
- Run Code
- Scraper will ask user to Enter URL
- Scraper will ask user to Enter the file name
- It will extract the data from given URL
- It then converts the extracted unstructured data into a structured format.
- It will create an excel(.csv) file with provided filename Store the data in it.

Literature Survey: -

S no.	Source	Title	Date
1.	ICECA	Data Analysis by Web Scraping using Python	02 September 2019
2.	IEEE	Web Scraping: From Tools to Related Legislation and Implementation Using Python	03 February 2021
3.	Getting Structured Data from the Internet {book}	Web Scraping in Python Using Beautiful Soup Library	13 November 2020
4.	Research paper	pandas: a Foundational Python Library for Data Analysis and Statistics	08 November 2012
5.	**Python** GUI Programming with **Tkinter {book}**	Develop responsive and powerful GUI applications with **Tkinter**	May 2018

Proposed System (Block Diagram): -



Conclusion: -

This project will help users to easily scrape data about products from any flipkart webpage and obtain the output in excel file(.csv).

References: -

1. Broucke, Seppe Vanden, and Bart Baesens. *Practical Web Scraping for Data Science: best practices and examples with Python*. CreateSpace, 2017.
2. *https://www.scrapingbee.com/blog/web-scraping-101-with-python/*
3. Lawson, Richard. *Web scraping with Python*. Packt Publishing Ltd, 2015.
4. O'Reilly, S. (2006). *Nominative fair use and Internet aggregators: Copyright and trademark challenges posed by bots, web crawlers and screen-scraping technologies. Loyola Consumer Law Review, 19, 273.*
5. Mitchell, Ryan. *Web scraping with Python: Collecting more data from the modern web*. " O'Reilly Media, Inc.", 2018.