

TERM PROJECT ABSTRACT

Year: II B. Tech – CSM

Date: 16/09/2024

Section:

Project Category No

Name of the student: KASA HARENDRA

Regd. No: 23BQ1A4274

Project Category :WEB SCRAPPING

Title of the Project :FLIPKART_SCRAPPER

Abstract: (not more than 300 words):

The project is about scraping the data and description about the products available in Flipkart using the SCRAPY FRAMEWORK. The Scrapy framework is an open source tool that helps to scrape data from any website through the usage of HTML tags.

The scraping is done using spiders that crawl across webpages which are available in allowed domain that is "flipkart.com". It navigates to the part of HTML document of the webpage and acquires the requires text and data and yields it. The yielded data is stored in the CSV files which can be used for analysis of data.

First all the requirements like Scrapy and requests are installed.

1. The scrapy project is created using a command "scrapy startproject <project_name>" that is run in the command line prompt or terminal.
2. A spider that done all these crawling and scraping work is generated.
3. The spider is instructed to crawl in a website to specific locations and obtain data to yield it in the form of a python dictionary.
4. The "search term" is obtained from the user and then it is used to generate URL which locates to the webpage including the products related to the search term.
5. The spider crawls across all the products in the URL and sometimes crawl to subsequent webpages.

The data yielded by the spider is stored in a CSV file which in stored on user's local machine on user's prompt.



Signature of the student



Signature of the Guide