**DAILY ASSESSMENT FORMAT**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date:** | **30/05/2020** | **Name:** | **PADMINI M** |
| **Course:** | **LOGIC DESIGN** | **USN:** | **4AL17EC066** |
| **Topic:** | **Applications of Programmable logic controllers** | **Semester & Section:** | **6th Bsec** |
| **Github Repository:** | **Padmini** |  |  |

|  |
| --- |
| **FORENOON SESSION DETAILS** |
| **Image of session** |
| **Report – Report can be typed or hand written for up to two pages.** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date:** | **30/05/2020** | **Name:** | **PADMINI M** | |
| **Course:** | **PYTHON** | **USN:** | **4AL17EC066** | |
| **Topic:** | **Scrape Real Estate Property Data from the Web** | **Semester & Section:** | **6th Bsec** | |
| **AFTERNOON SESSION DETAILS** | | | |
| **Image of session** | | | |
| **Report – Report can be typed or hand written for up to two pages.**   * Web Scraping is a viable option to keep track of real estate listings available for sellers and agents. Being in possession of extracted information from real estate sites such as Zillow.com can help adjust prices of listings on your site or help you create a database for your business. * Scraping Logic-Construct the URL of the search results page from Zillow. We’ll have to create this URL manually to scrape results from that page. * Download HTML of the search result page using Python Requests – Quite easy, once you have the URL. We use python requests to download the entire HTML of this page. * Parse the page using LXML – LXML lets you navigate the HTML Tree Structure using Xpaths. We have predefined the XPaths for the details we need in the code. * Save the data to a CSV file.   PIP to install the following packages in Python   * Python Requests, to make requests and download the HTML content of the pages * Python LXML, for parsing the HTML Tree Structure using Xpaths * Information Extraction (IE) is a crucial cog in the field of Natural Language Processing (NLP) and linguistics. It’s widely used for tasks such as Question Answering Systems, Machine Translation, Entity Extraction, Event Extraction, Named Entity Linking, Coreference Resolution, Relation Extraction, etc. * Web scraping, often called web crawling or web spidering, or “programmatically going over a collection of web pages and extracting data,” is a powerful tool for working with data on the web. * With a web scraper, you can mine data about a set of products, get a large corpus of text or quantitative data to play around with, get data from a site without an official API, or just satisfy your own personal curiosity. * In this tutorial, you’ll learn about the fundamentals of the scraping and spidering process as you explore a playful data set. We’ll use BrickSet, a community-run site that contains information about LEGO sets. By the end of this tutorial, you’ll have a fully functional Python web scraper that walks through a series of pages on Brickset and extracts data about LEGO sets from each page, displaying the data to your screen. * The scraper will be easily expandable so you can tinker around with it and use it as a foundation for your own projects scraping data from the web. * Scrapy is one of the most popular and powerful Python scraping libraries; it takes a “batteries included” approach to scraping, meaning that it handles a lot of the common functionality that all scrapers need so developers don’t have to reinvent the wheel each time. It makes scraping a quick and fun process! * Scrapy, like most Python packages, is on PyPI (also known as pip). PyPI, the Python Package Index, is a community-owned repository of all published Python software. | | | |