DECLARATION

We declare that this written submission represents us ideas is our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources.

We also declare that we have adhered to all principles of academic honesty

and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when

needed.

Table of Contents

1. Introduction ………………………………………………………………………………

1.1 Scope of the document…………………………………………………………………

1.2 Intended Audience………………………………………………………………………

1.3 System Overview………………………………………………………………………..

2. Project Briefing……………………………………………………………………………….

3. Problem Statement…………………………………………………………………………..

4. Problem Solution……………………………………………………………………………..

5. Objective of Project…………………………………………………………………………..

6. Scope of Project………………………………………………………………………………

7. Requirements Gathering…………………………………………………………………….

8. Analysis……………………………………………………………………………………….

9. Final Screenshot of Project Output…………………………………………………………

1. Introduction
   1. Scope of the document

* This section will cover the details regarding scope of the document.
* Web scraping is a crawling method used to collect information.
  1. Intended Audience
* This section will cover the intended audience.
  1. System overview

>This section will capture overview of system application i.e for what system is being developed

> Who are the stake holders of system?

> What are other external Systems through which this will be interacting.

1. Project Briefing

In this project we were free to choose any website. The websites chosen by the rest of the cohort ranged from e-commerce to news websites showing the different applications of web scraping.

We chose a website for scientific articles because we thought it would be interesting to see what kind of data we could obtain and furthermore what insights we could gather as a result of this data. We have chosen to keep the website anonymous. In any case the goal of this post is to outline how to build a pipeline for any website of interest.

In this post, we show our main steps and challenges along the way. We have included code snippets and recommendations on how to create an end to end pipeline for web scraping. The code snippets we show here are not OOP (Object Oriented Programming) for the sake of simplicity, but we highly recommend to write OOP code in your web scraper implementation.

1. Problem Statement

We can’t get information easily throughout the world news. We cant get the news exceptionally.

1. Problem Solution

Here we came with the solution. We have created a universal web scraper using python and its beautiful dictionary. This will help the people to easily gather the news for business man and normal people.

1. Objective of the project

Objective of this project is gathering news through the help of scraping the sites.

1. Scope of the project

In this document we have explained the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. This project will run with help of python language.

1. Requirements gathering

<) Linux or windows operating system.

<) Team members

<) Python IDE and library.

<) Github account.

1. Analysis

In this project we have done the things with python language and the beautiful soup dictionary. The data we have gathered with the help of google and the github. The information is always a wealth. The project took some long days to complete. The team cooperation and the effort made this work so easy.so here we are, we have successfully created and done the project with the help of all the sources.

9.Final screenshot of output

