# CHAPTER-3

# SYSTEM ANALYSIS

**EXISTING SYSTEM**

In Existing system is the manual web data extraction process has two major problems. Firstly, it can’t measure costs efficiently and can escalate it very quickly. The data collection increase as more data is collected from each website. Web Scraping can be done manually. All you need is the ability to copy/paste information and a spreadsheet to keep track of the extracted data.

**DISADVANTAGES**

* Arguably the slowest method of web scraping. Even at top speeds, a web scraping bot will be significantly faster than a human at scraping data.
* Manual web scraping can be quite expensive if only for the time investment. It can be even more expensive if you’re paying someone to do scrape the data.

**PROPOSED SYSTEM**

Web scraping is the process of extracting data from websites. Some data that is available on the web is presented in a format that makes it easier to collect and use it, for example in the form of downloadable comma-separated values (CSV) datasets that can then be imported in a spreadsheet or loaded into a data analysis script. Often however, even though it is publicly available, data is not readily available for reuse. For example it can be contained in a PDF, or a table on a website, or spread across multiple web pages. There are a variety of ways to scrape a website to extract information for reuse. In its simplest form, this can be achieved by copying and pasting snippets from a web page, but this can be unpractical if there is a large amount of data to be extracted, or if it is spread over a large number of pages. Instead, specialized tools and techniques can be used to automate this process, by defining what sites to visit, what information to look for, and whether data extraction should stop once the end of a page has been reached, or whether to follow hyperlinks and repeat the process recursively. Automating web scraping also allows to define whether the process should be run at regular intervals and capture changes in the data.

Automated web scraping tools have become increasingly popular due their ease of use and savings in time and costs. These tools also come in many different shapes and sizes, from simple browser extensions to more powerful software solutions.

**ADVANTAGES**

* AUTOMATION: The first and most important benefit of web scraping is developing tools that have simplified data retrieval from different websites to only a few clicks.
* COST EFFECTIVE: Data extraction by hand is an expensive task that necessitates a large workforce and large budgets.
* EASY IMPLEMENTATION: When a website scraping service begins gathering data, you should be confident that you are obtaining data from various websites, not just a single page. It is possible to have a large volume of data with a small investment to help you get the best out of that data.
* SPEED: Another feature worth mentioning is the speed with which web scraping services complete actions. Imagine that a scraping project that would typically take weeks is completed in a matter of hours. But of course, that depends on the complexity of the projects, resources, and tools used.

### **Feasibility Study:**

### Preliminary investigation examines project feasibility, the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All system is feasible if they are unlimited resources and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:

* Technical Feasibility
* Operational Feasibility
* Economic Feasibility

**Economic Feasibility:**

A system can be developed technically and that will be used if installed must still be a good investment for the organization. In the economic feasibility, the development cost in creating the system is evaluated against the ultimate benefit derived from the new systems. Financial benefits must equal or exceed the costs.

The system is economically feasible. It does not require any additional hardware or software. Since the interface for this system is developed using the existing resources and technologies available at NIC, there is nominal expenditure and economic feasibility for certain.

### **Operational Feasibility:**

Proposed projects are beneficial only if they can be turned into an information system. That will meet the organization’s operating requirements. Operational feasibility aspects of the project are to be taken as an important part of the project implementation. Some of the important issues raised are to test the operational feasibility of a project includes the following: -

* Is there sufficient support for the management from the users?
* Will the system be used and work properly if it is being developed and implemented?
* Will there be any resistance from the user that will undermine the possible applications?

This system is targeted to be in accordance with the above-mentioned issues. Beforehand, the management issues and user requirements have been taken into consideration. So, there is no question of resistance from the users that can undermine the possible application benefits. The well-planned design would ensure the optimal utilization of the computer resources and would help in the improvement of performance status.

**Technical Feasibility:**

The technical issue usually raised during the feasibility stage of the investigation includes the following:

* Does the necessary technology exist to do what is suggested?
* Do the proposed equipment have the technical capacity to hold the data required to use the new system?
* Will the proposed system provide adequate response to inquiries, regardless of the number or location of users?
* Can the system be upgraded if developed?
* Are there technical guarantees of accuracy, reliability, ease of access and data security? Earlier no system existed to cater to the needs of ‘Secure Infrastructure

Implementation System’. The current system developed is technically feasible. It is a web-based

User interface for audit workflow at NIC-CSD. Thus, it provides easy access to the users. The database’s purpose is to create, establish and maintain a workflow among various entities in order to facilitate all concerned users in their various capacities or roles. Permission to the users would be granted based on the roles specified. Therefore, it provides the technical guarantee of accuracy, reliability and security. The software and hard requirements for the development of this project are not many and are already available in-house at NIC or are available as free as open source.

**3.2 Functional Requirements:**

Functional requirements are the functions or features that must be included in any system to satisfy the business and be acceptable to users.

Based on this, the functional requirements that the system must require as follows:

Web scraping allows you to acquire non - tabular or poorly structured data from websites and convert it into a useable, structured format, such as csv file.

. Scraping is about more than just acquiring data.

. It can also help you archive data and track changes to data online.

The main four key parts of web scraping

. Data discovery

. Data extraction

. Extraction scale

. Data output

### **3.3 Non-Functional Requirement:**

The major non-functional Requirements of the system are as follows

* **Maintainability:** Components that are easily decoupled from each other can be easily replaced with better implementations. Basically, your engineering team will be more efficient in maintaining and updating the code within the iterative development workflow.
* **Performance:** Angular is, by default, a powerful and high performing front-end framework. Yet, unexpected challenges are bound to happen when you're building mission-critical web applications, apps that are content-heavy, and complex on the architectural side.
* **Readability:** Encapsulation also ensures that new developers – who’ve been recently onboarded to a project – can read code better and eventually reach their plateau of productivity faster.
* **Reusability:** Components of similar nature are well encapsulated, in other words, self- sufficient. Developers can reuse them across different parts of an application. This is particularly useful in enterprise-scope applications where different systems converge but may have many similar elements like search boxes, date pickers, sorting lists, etc.

**Performance:** This system is developing with the single page application in the high-level languages and using the advanced frontend and back-end technologies it will give response to the end user on client system with in very less time.

### **Minimum Hardware Requirements:**

* Processor - Intel Core i3
* Speed-2.2 Ghz
* RAM-8GB
* Hard Disk–500GB

### **3.5 Software Requirements:**

* Operating system **:** Windows 10
* Framework **:** Django
* Google-chrome : chrome browser
* IDE : jupyter Notebok