

# **Software Requirements Specification for The Hindu News Scraper**

**Version 1.0 approved**

**Prepared by Azeemushan**

**SIGNZY**

**2020-6-28**

## Table of Contents

1. Introduction	1
1.1 Purpose	1
1.2 Document Conventions	1
1.3 Intended Audience and Reading Suggestions	1
1.4 Product Scope	1
1.5 References	1
2. Overall Description	2
2.1 Product Perspective	2
2.2 Product Functions	2
2.3 User Classes and Characteristics	2
2.4 Operating Environment	2
2.5 Design and Implementation Constraints	2
2.6 User Documentation	2
2.7 Assumptions and Dependencies	2
3. External Interface Requirements	3
3.1 User Interfaces	3
3.2 Hardware Interfaces	3
3.3 Software Interfaces	3
3.4 Communications Interfaces	3
4. System Features	4
5. Non-functional Requirements	5
5.2 Safety Requirements	5
5.3 Security Requirements	5
5.4 Software Quality Attributes	5
5.5 Business Rules	5
6. Other Requirements	6

## Table of Tables

Table 1.1 shows document conventions	0
Table 1.2 shows intended readers of the document	0
Table 1.3 shows the operating environment	1
Table 1.4 shows the communication interface	2
Table 1.5 shows Operating System feature	3
Table 1.6 shows performance requirements	4
Table 1.7 shows safety requirements	4
Table 1.8 shows the security requirements	4
Table 1.9 shows the quality attributes	4
Table 2.0 shows business rules	4
Table 2.1 shows the other requirements	5

## 1. Introduction

### 1.1 Purpose

The SRS describes the full The Hindu News Scraper This SRS describes, the constraints, document conventions, product scope, perspective, operating environment, system features, non-functional requirements and all other dependencies needed by the application. I am hereby attaching the required setup files for further smoothing of the project..

### 1.2 Document Conventions

The following conventions have been considered in this SRS;

Table 1.1 shows document conventions

Convention	Reason for using the convention
------------	---------------------------------

### 1.3 Intended Audience and Reading Suggestions

These are the intended readers of this document;

Table 1.2 shows intended readers of the document

Reader	Use of the document to the reader
News Reader	It will give all news at one place.
Historian	The person who has keen intrest in studying all the news of past.
Author	Author can check all his contributions till date by one click.

### 1.4 Product Scope

#### 1.4.1 Goal

The overall goal of News Scraper is to study all the past News Trends across the globe at one single click.

#### 1.4.2 Objectives

The Objective is to get the pre processed news at one step to be analysed for the getting articles related to one person. It will give all the articles related to one single person which could further be analysed and studied upon.

#### 1.4.2 Benefits

The benefit of using this tool would be that one person could get to know all about the news happened in a year.This could further be used by Researchers,Machine Learner,Data Analyst or by Historian.

### 1.5 References

1

## 2. Overall Description

### 2.1 Product Perspective

This system is a desktop system . It is a new self contained product

### 2.2 Product Functions

#### makeURL

This function will provide all the accessible urls of The Hindu News website.it will generate all the links to be accessed by the program to scrape.It uses calender module in order to make list of URLs.

#### DayNewsLink

This function will provide us the option to scrape all the news articles of the all days of the 2010 year. It will take one argument and that is the list of URLS of all the days of the year which are to be scraped.

#### searchNews

It will take name of the author and news content dictionary as argument.and returns us the final articles related to that author.

### 2.3 User Classes and Characteristics

The following are the different users and their Characteristics for the product;

#### Scraper

The user view will use this product To scrape news article's link,title and news using the Selenium

The user will be required to Should be knowledgeable in using Python,Selenium

### 2.4 Operating Environment

Table 1.3 shows the operating environment

Operating Environment	Reason
Internet	For accessibility on the web.
OS	For accessing Python over the OS.
Python	For running this tool
Selenium	For Scraping all the news article's urls, news headings and news content.

### 2.5 Design and Implementation Constraints

#### Design Constraint

The system is to be developed using Python language enriched with the Selenium,Calendar and Pandas packages as the basic package for the system.These packages were used for further analysis in this tool.

### 2.6 User Documentation

The help files, tutorials, will provide users with the help they need to effectively and efficiently use the product.

## **2.7 Assumptions and Dependencies**

### **Assumptions**

Scraping will usually take 1 hours. Time may vary depending on your Internet Speed.

### **Dependencies**

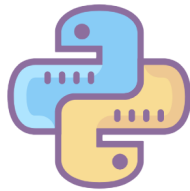
2

### 3. External Interface Requirements

#### 3.1 User Interfaces

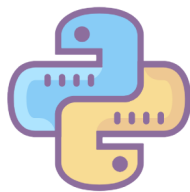
##### Chrome Webdriver

For accessibility on the web to further scrape the data.



##### Selenium

For Scraping news articles.



#### 3.2 Hardware Interfaces

##### Google Colab

This device will provide NA and take in Some Commands to install selenium and chrome driver from the system.It will communicate usingIt will be used to increase the computation rate as we can get TPU runtime environment here.

This device will provide

#### 3.3 Software Interfaces

##### Python 3.X

This device will provide NA and take in Install Selenium through "pip install selenium" . from the system.It will communicate using

##### Chrome + Chrome Webdriver

This device will provide NA. and take in NA. from the system.It will communicate using

This device will provide

**3.4 Communications Interfaces**

Table 1.4 shows the communication interface

Channel	Reason
---------	--------

3

**4. System Features**  
**4.1 Operating System**

This is a medium priority feature.

Table 1.5 shows Operating System feature

RQ code	Requirement
---------	-------------



5. Non-functional Requirements

5.1 Performance Requirements

Table 1.6 shows performance requirements

RQ code	Requirement
---------	-------------

5.2 Safety Requirements

5.3 Security Requirements

Table 1.7 shows safety requirements

RQ code	Requirement
---------	-------------

Table 1.8 shows the security requirements

RQ code	Requirement
---------	-------------

5.4 Software Quality Attributes

Table 1.9 shows the quality attributes

RQ code	Requirement
---------	-------------

5.5 Business Rules

Table 2.0 shows business rules

RQ code	Requirement
	It is to be noted that the computation time is very high hence to try this tool run test.py Else wait for an hour to scrape.

6. Other Requirements

Table 2.1 shows the other requirements

RQ code	Requirement
---------	-------------