

ScrapGen Requirements

Target Release	1.0
Expected Release Date	October 24, 2018
Development Method	Scrum Methodology
Project Owner	Akhand Mishra

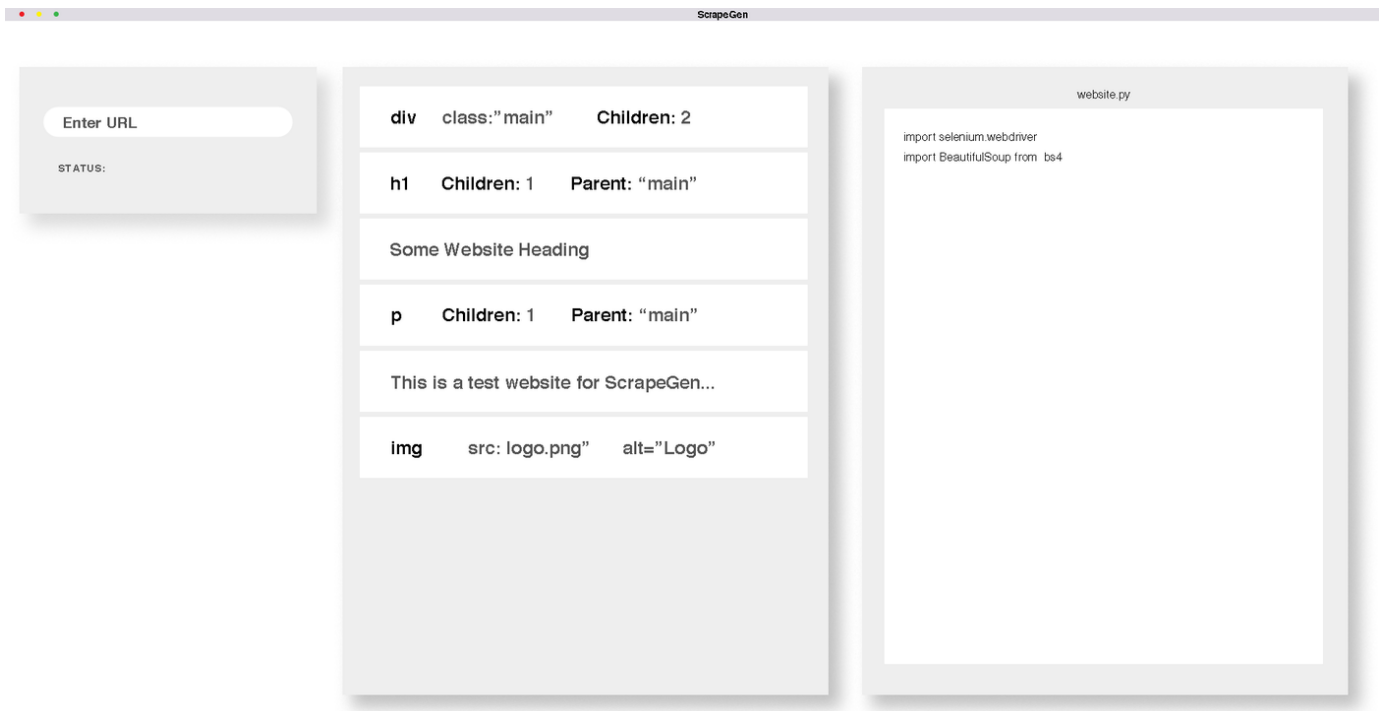
Background and Strategy fit

Enter the world of Artificial Intelligence, data has become valuable to organisations and developers. It is apparent that the biggest source of data is the web and hence, the need for scraping different websites has arisen. Developers need an automated way out of the problem of building separate scripts for different website. Thus, this software is a must.

Requirements

#	User Story Title	User Story Description	Priority	Notes
1	Visualized Scraping	A user wants a visualized and less complex way to recognize the HTML element. Formatted Nodes are easy to figure out.	Must Have	<ul style="list-style-type: none">The DOM Tree builder serves this purpose.ScrapGen will build a GUI based on the parent-child relationship of elements in the retrieved HTML document.The user can see the type, name, attributes and children of the current node.
2	Automated Scripts generation	A user wants to avoid the redundant work of writing same code over and over just to scrape one element from the website.	Must Have	<ul style="list-style-type: none">The Script Generator module shall fulfill this requirement.The user only needs to click the element s/he needs to scrape.
3	Real-time Editing	A user wants to edit the scripts in real-time.	Should Have	
4	JavaScript Execution	A user suggests that not only static but even dynamic websites could be scraped using ScrapGen.	Must Have	<ul style="list-style-type: none">Usage of Selenium module of Python can fetch the HTML after the scripts have been executed.Headless Chrome driver shall be required.This may slow down the software.
5	Cross Language	A user suggests that ScrapGen should generate scripts in both: Python 2.x and Python 3.x	Should Have	
6	Dashboard & Management	A developer suggests to let users create account which let them save and manage different scraping scripts.	Should Have	

User Interface Design



Questions

Below is a list of questions to be addressed as a result of this document

Questions	Outcome
Is the application cross platform?	Yes, the UI shall be built on Electron.js and back-end is on Python, thus it'll run similar on any OS.
Is ScrapGen offline?	Yes and No, you can open and edit the saved DOM Trees in offline mode but to fetch HTML for a new website, you must be connected to the internet.
Do we need to install selenium or any other dependencies?	As far as fetching, generating or editing script goes, you do not need any dependencies, however to run the generated script in a Python IDLE, you may need those.

Not Doing

- Connecting and letting users share script among different users and platforms.
- Optimization of scripts generated.
- Letting user select the elements through the browser itself.