Kathmandu University

Department of Computer Science and Engineering

Dhulikhel, Kavre

A Project Proposal

On

Scrappy

[Code No: COMP 206]

(For the partial fulfillment of 2nd Year/1st Semester in Computer Science)

Submitted by:

Usta Adhikari (68)

Rajeshwor Niroula (71)

Milan Dhamala (73)

Bikramaditya Prasad Subedi (75)

Submitted to:

Mr. Nabin Ghimire

Department of Computer Science and Engineering

Submission Date: 12th November, 2019

Cover Letter

November 12, 2019

Respected Sir,

This proposal has been submitted for the partial fulfillment of the COMP 206 course, which consists of a group project to be submitted under the directions and monitoring of a supervisor. Following are the topics covered under this proposal.

- Introduction to our project……..
- Requirements……..
- System Description and Work Breakdown……

Thank you for giving us this opportunity.

Sincerely,
Usta Adhikari (68)
Rajeshwor Niroula (71)
Milan Dhamala (73)
Bikramaditya Prasad Subedi (7)

Abstract

Scrappy focuses on creating subtle product comparing tool which allows for easier product purchases saving enormous comparison and purchasing times. Our project is a web application based tool with an objective of providing simpler way of browsing products from several online marketplace. In scrappy, we scrape information about products such as images, prices, etc. These information will be organized in pages making product easier to compare.

Keywords: Web Application, Scrape

Table of Contents:

Title **Page No.**

List of Figure

Figure **Page No.**

List of Table

Table **Page No.**

Acronyms

HTML: Hypertext Markup Language

CSS: Cascading Style Sheets

MySQL: Database using Structured Query Language

Chapter 1: Introduction

Project Scrappy is a web tool used for extracting data from websites. Basically, web scrapping typically refers to automated processes implemented using a bot or web crawler. This form of extracted datasets can be used for later retrieval or analysis.

1.1: Background

If we observe todays online market system, there are too many online shopping platforms and from site to site prices for even the same product varies making it difficult to find the best price so an easy product comparing website feels very much needed. Even though we can find plenty of web scrappers for sites like amazon and eBay there aren't any user friendly product comparing websites in the local market. Therefore, people tend to seek multiple sources and opinions before they jump into buying any products. As a result, people have to scour through multiple online marketplaces just to get hold onto a single product. With the objective to simplify this problem, our project Scrappy will strive to accumulate multiple products from multiple sources on a single screen, served in a single platter to devour.

Scrappy is a web-based product scraping tool, where each user can visit the site and type the desired keyword and get the keyword related products listings from multiple sources on a single screen. It not only, will provide product option but also decrease unnecessary time we have to spend on internet to get the right product pricing. With the global reach in mind, our project will scrape national and international products that will increase the inclusivity of information in the site which will not only be limited to local products but also cover international outreaches.

1.2. Objectives

- An all in one platform for price analysis and product purchase
- Market analysis at scale
- Increase inclusivity of information
- User intuitive web design

1.3 Motivation and Significance

Websites like Hagglezon, Trivago, Flightscanner, Expedia etc. was our motivation for our project. Hagglezon compares amazon products, prices throughout most of the European countries that helps customers to compare prices of the same amazon service throughout several European countries. Similarly, Trivago displays prices of hotel rooms in certain destination that allows you to pick hotel on your budget. Similarly, Flightscanner compares flight tickets across multiple services. Thus, we decided to design a web-application that allows customers to compare prices from multiple sites.

Chapter 2: Related Works/Existing Works

Web scraping is used in a variety of digital businesses that rely on data harvesting. Legitimate use cases include: Search engine bots crawling a site, analyzing its content and then ranking it. We researched and collected some data on other application that work on similar principle. Some existing programs are listed below:

Web Scraping 2.0

Web Scraping 2.0 is a web application that turns any website to an API. This scraper works even when the website updates.

ParseHub

Another relatively unknown website called "ParseHub" also works on a similar goal. It is also a free scrapper that turns websites to API or spreadsheet.

Trivago

Trivago also compares 200+ booking website and gives best deal to customers.

Chapter 3: Procedure and Methods

For completion of our website we have created certain procedure which we will be following strictly. They are:

1) Study and research- Firstly, we have planned to do some research related to front-end and backend development needed for our project by visiting similar kind of websites, by consulting with seniors and teachers.

2) Front-end Development: Front End part i.e.; Visual parts of website is the first thing that will be noticed by user and is responsible for having first impression on a user. So, we will be making our website attractive and user-friendly using HTML, CSS, Bootstrap.

3) Back-end Development: Back End part is the backbone of our website. Without it our website will not function as needed. We will be doing backend programming with the help of Java and MySQL, where MySQL is used for designing database and Java will perform web-scraping tasks and store the collected data on database, generate web-page content, or send and receive request.

4) Testing and Debugging: There can be some syntactical and logical errors. Our website might not be as user friendly as we expected. So, we will be performing tests and debugging to improve it.

5) Documentation: After completion of our above listed task, we will prepare work report and present our project.

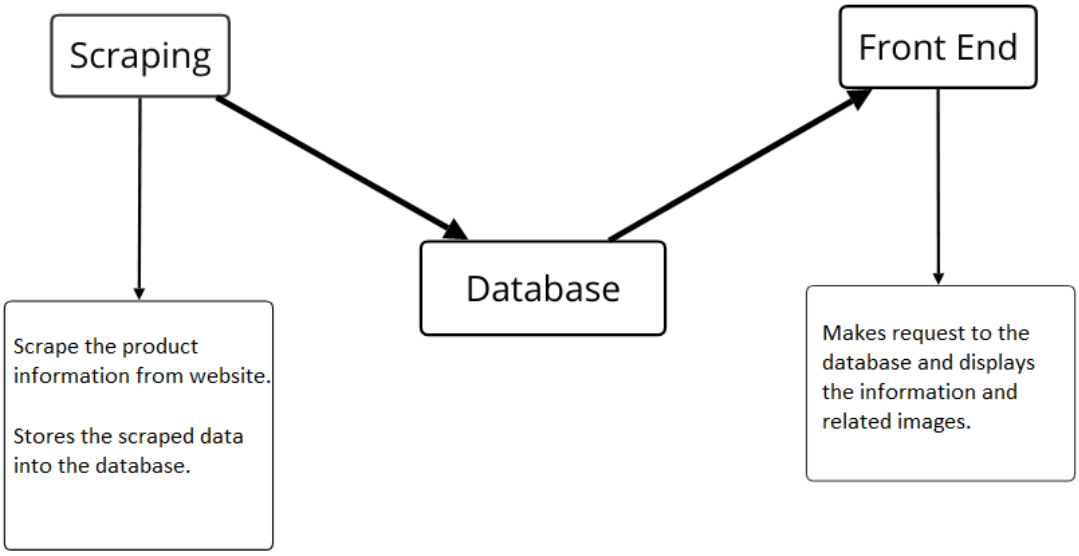

Fig 3.1: Workflow Diagram

Chapter 4: System Requirement Specification

4.1 Software Specification

4.1.1 Front End Tools: HTML, CSS, Bootstrap

4.1.2 Back End Tools: Java web scraping and web development libraries

4.1.3 Database: MySQL

4.1.4 IDE: IntelliJ IDEA

4.1.5 Browsers: Chrome, Microsoft Edge

4.2 Hardware Specification

Since the website will be simplistic, sophisticated hardware is not required. Any modern PC capable of running a modern OS should suffice.

Chapter 5: Project Planning and Scheduling

The work breakdown and time in weeks required to complete the specific task are shown as in the Gantt chart below: -

Task	1	2	3	4	5	6	7	8	9	10	11	12
Research and study	■	■	■									
Graphic Designing				■	■	■						
Core Programming						■	■	■	■	■		
Program testing										■	■	
Documentation											■	■

Table 5.1: Gantt Chart

Tasks:

1. Research and Study
2. Graphic Designing
3. Core Programming
4. Program Testing
5. Documentation

References

https://www.hagglezon.com/

Hagglezon
As an Amazon Associate Hagglezon earns from qualifying purchases.