

A Novel of Product Price Analysis Using Web Scraping

SOUMYA B PEDDI¹, SANGEETA², VAISHALI GADED³, VEDHIKA⁴

^{1, 2, 3, 4} Dept of Computer Science and Engineering, Poojya Doddappa Appa College of Engineering, Kalaburagi

Abstract— Web scraping is an application used to extract information from various web services. E-commerce websites in today's world have become one of the most important sources for buying all kinds of products. As there are many e-commerce websites available it has become difficult for users to choose best website for desired product amongst these websites. Comparison of E-commerce products using web scraping enables users to compare prices and get desired product at reasonable price. Users can also select multiple products that belong to same category for comparing its features.

Indexed Terms— Web Scraping, Data Analysis, Visualization

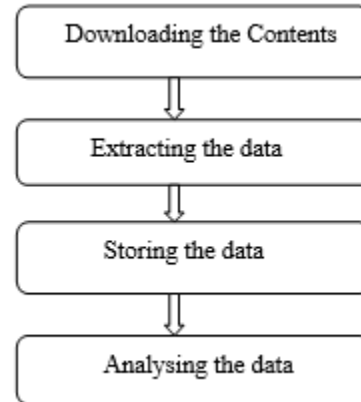
I. INTRODUCTION

Product analysis: It involves examining product features, cost, availability, quality, appearance and other aspects. This system uses the following technologies:

Web Crawler: This is the best way to create a web crawler that will redirect to e-commerce websites. The fetched URL's are send to scraper for scraping process.

Web Scraper: It is used to extract HTML data from URL's and use it for personal purpose. In this system, Scraping is done using python libraries like Requests and BeautifulSoup4.

BLOCK DIAGRAM



In today's world the need for online shopping over the Traditional Shopping is being increased day by day. To perform this people are utilizing electronic gadgets such as tablets, mobiles and PC. Web scraping is generally a new strategy for gathering the web information. The proposed solution helps online users to grab best deal for their product from multiple e-commerce websites on single web interface. This will in turn save users time, money and efforts to find the same product prices on different e-commerce websites.

Existing system is the manual web data extraction process. It has two major problems. Firstly, it can't measure costs efficiently and can escalate it very quickly. The data collection costs increase as more data is collected from each website. In order to conduct a manual extraction, businesses need to hire large number of staffs, this increases the cost of labour significantly. Secondly, each manual extraction is known to be error prone. Further, if any business process is very complex then cleaning up the data can get expensive and time consuming.

Web Scraping is a software technique to extract information from websites. Usually, such

programming programs recreate human investigation of the World Wide Web by either executing low-level Hyper content Transfer Protocol (HTTP), or installing a completely fledged internet browser, like Internet Explorer or Mozilla Firefox. Web Scraping is the Internet contains the most helpful arrangement of information at any point collected, generally openly open free of charge. Web scratching is getting progressively valuable as a way to effortlessly assemble and sort out the plenty of data accessible on the web. Utilizing a straightforward language like Python, you can creep the data out of complex sites utilizing basic programming.

How Price Comparison Websites Work

Price comparison websites extract useful details such as product prices, reviews, features, and descriptions from multiple websites. These details are then used on the price comparison website and tailored accordingly for easy access. So, when a buyer searches for a product on the website, the site quickly compares and lists similar products from a number of retailers. This process simplifies the buying decision of the buyer since they can compare factors such as price deals, shipping costs, and other features. However, the algorithms involved depend on massive data. As expected, data extraction in real-time is time-consuming. As if that wasn't enough, the dynamic pricing system employed by e-commerce websites makes it difficult to keep track of price changes. Amazon, for instance, is approximately 417 hours faster than its competitors in adapting price changes.

Importance Of Price Comparison

Price scraping is used by Price Comparison Sites and E-Commerce sites of all sizes and types. It could be a simple Price Scraper using a chrome extension, a python script to scrape data from competing e-commerce websites like Walmart or Best Buy, or a full-fledged web scraping service. Retailers need to perform continuous price monitoring for several reasons. First, your competitors can sell the same product at a lower price or offer a more ludicrous discount. But Amazon managed to surpass this barrier by using web scraping and continued price monitoring. So let us understand Amazon's secret formula. Amazon leverages excellent price scraping to monitor competitors and offers products at competitive prices.

II. METHODOLOGY

This work comprises of three phases. The first phase is to scrap the product details like product name, product price and their ratings from e-commerce websites – Flipkart, Amazon and store it in a csv format. The second phase of the work is data analysis. The third phase is visualization.

ALGORITHM

1. WEB SCRAPING
2. DATA ANALYSIS
3. VISUALIZATION



Figure.2: Modules

WEB SCRAPING

The usage of Web Scrapping is to extract HTML data from URL,,s and use it for personal purpose. In this work python library requests and beautifulsoup4 are used for performing web scraping. To parse html pages a python library BeautifulSoup4 is used. The product details from various website are scrapped and stored in csv file with the help of the library. There are various approaches to scrap sites, such as online services, APIs or composing own code. A few sites permit web scratching and some don't. To know whether a website permits web scraping or not, one can come across at the site's –robots.txt file.

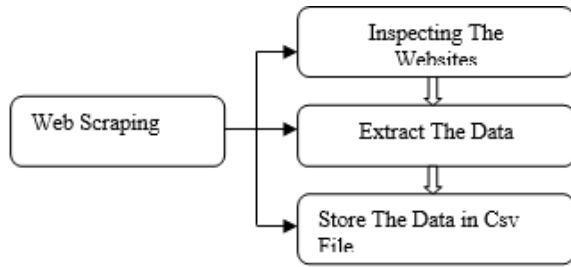


Figure.3: webscraping

WORKING

```

{% extends 'price/base.html' %} {% load static %} {%
block content %}
<section class="container-section">
<nav class="navbar-section">
<a href="#">
<h1><imgsrc="{% static 'images/cart2.png' %}"
/>Ecompare</h1></a>
<button>
  class="navbar-toggler"
  type="button"
  data-bs-toggle="collapse"
  data-bs-target="#navbarSupportedContent"
  aria-controls="navbarSupportedContent"
  aria-expanded="false"
  aria-label="Toggle navigation"
</button>
<ul class="nav-menu">
<li class="nav-item" aria-current="page">
<a class="nav-link menu-item .current-menu-item"
href="#home">
<i class="bi bi-house-fill"></i>HOME</a>
</li>
<div class="wee"></div>
<li class="nav-item">
<a class="nav-link menu-item" href="#contact">
<i class="bi bi-person-fill">
</i>ABOUT</a>
</li>
</ul>
<div class="hamburger">
<span class="bar"></span>
<span class="bar"></span>
<span class="bar"></span>
</div>
</nav>
<div class="Home-section" id="home">
<div class="btn-form">
<h2>

```

```

  You want lower prices, <br />
  we hear you
</h2>
<form action="/next" method="GET" style="margin-
top: 50px">
<input
  class="btn-input px-3"
  type="text"
  name="text"
  placeholder="Enter Product Name"/>
<button class="btn-search" href="">Search</button>
</form>
</div>
<div class="back-img">
<imgsrc="{% static '/images/home.svg' %}" />
</div>
</div>
</section>
<footer id="contact" class="container-fluid
size_footer about-section">
<div class="row py-5">
<div class="col-12 align-items-center justify-content-
center mb-5">
<h2>About us</h2>
<h4><a href="#">Sangeeta</a></h4>
<br>
<h4><a href="#">Vaishali</a></h4>
<br>
<h4><a href="#">Vedhika</a></h4>
</div>
<div class="col-12 align-items-center justify-content-
center">
<div class="d-flex justify-content-end">
<a href="#"><i class="bi bi-arrow-up-circle-
fill"></i></a>
</div>
</div>
<hr />
<p class="lead">
<strong>
  Copyright &copy; PDA College | Made with
  ♥ for the people</strong>
</p>
</div>
</div>
</footer>
{% endblock content %}

```

III. RESULTS

We have created web interface as shown in fig 7.1. In which we can search the products.

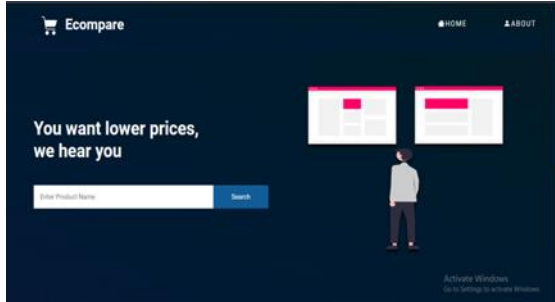


Fig 4: Searching for a Product

For example: Here iphone x product is searched after searching we got all the websites prices of iphone x as shown in

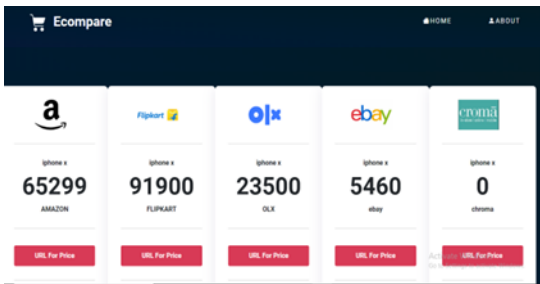


Fig 5: Display of Websites

In this figure we are switched to e commerce websites we want.

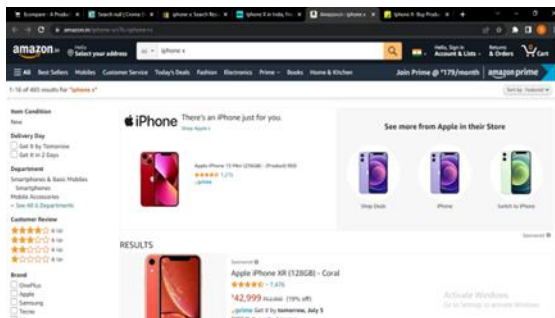


Fig 6: Visit to a particular website

IV. CONCLUSION

Web scraping and techniques confronting numerous difficulties as the extraction of the information are not excessively simple. In this work the e-commerce

websites are scraped to extract the products price. The extracted data has been visualized. Then created a single web interface where in we can searched for single product and compare their prices with respect to different e-commerce websites.

REFERENCES

- [1] Renita Crystal Pereira – “Web Scrapping Summary, techniques and tools” (2020). Supporting the company and profiting from it. Getting fast shopping and also save time.
- [2] Kaushal Parikh – “Web Scrapping detection with the help of Machine Learning” (2020). Every time a company places its data on internet. Web solutions have significant impact on the result of the cause.
- [3] Sameer Padghan – “Web pages in assistance with web scraping easily and framework details obtained” (2020). The scraping used would increase significantly. This method should be treated as blessing.
- [4] E. Ferrara, P. De Meo, G. Fiumara, and R. Baumgartner, “Web data extraction, applications and techniques: A survey,” *Knowl.-Based Syst.*, vol. 70, pp. 301–323, Nov. 2014.
- [5] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, “Deep learning applications and challenges in big data analytics,” *J. BigData*, vol. 2, no. 1, p. 1, Dec. 2015.
- [6] E. Uzun, H. V. Agun, and T. Yerlikaya, “A hybrid approach for extracting informative content from Web pages,” *Inf. Process. Manage.*, vol. 49, no. 4, pp. 928–944, Jul. 2013.
- [7] E. Uzun, E. S. Güner, Y. Kılıçaslan, T. Yerlikaya, and H. V. Agun, “An effective and efficient Web content extractor for optimizing the crawling process,” *Softw., Pract. Exper.*, vol. 44, no. 10, pp. 1181–1199, Oct. 2014.
- [8] E. Uçar, E. Uzun, and P. Tüfekci, “A novel algorithm for extracting the user reviews from Web pages,” *J. Inf. Sci.*, vol. 43, no. 5, pp. 696–712,
- [9] E. Uzun, H. N. Bulus, A. Doruk, and E. Ozhan, “Evaluation of Hap, AngelSharp and HTMLDocument in Web content extraction,” in

Web

[10] content extraction,” in Proc.Int. Sci. Proc.Int. Sci. Conf. (UNITECH), vol.2,2017, pp. 275-278.