SCTR's Pune Institute of Computer Technology Dhankawadi, Pune

A.Y. 2023-24

WADL MINI PROJECT REPORT ON

"WebEase : An Online Shopping Guide"

Submitted By

33383-Dhiraj Sontakke

33384-Swaranjali Lanjulkar

33390-Poonam Vetal

33391-Dhiraj Wadile

Under the guidance of

Mrs.Deepali Salapurkar

DEPARTMENT OF INFORMATION TECHNOLOGY
ACADEMIC YEAR 2023-24

ABSTRACT

Web scraping, the process of extracting data from websites, presents businesses with various challenges. Deciphering complex website structures and ensuring compliance with legal and ethical standards are primary concerns. Additionally, effective troubleshooting and error handling are necessary to maintain scraping tools' reliability.

To address these challenges, developers can employ tools like Puppeteer along with Bright Data's web unlocker feature. These technologies enable seamless extraction from complex websites while adhering to legal and ethical standards. Moreover, meticulous attention to detail and proactive problem-solving skills are essential for effective troubleshooting and error handling.

Despite the challenges, web scraping remains indispensable for businesses. It revolutionizes data acquisition and analysis, providing insights into market trends, consumer behavior, and competitor strategies. By integrating web scraping with technologies like Next.js for dynamic web applications and Cron jobs for automated tasks, businesses can gain a competitive edge by transforming raw data into strategic intelligence. Through strategic integration and adherence to best practices, organizations can harness the power of web scraping to navigate the complexities of the modern marketplace and emerge as leaders in their industries.

INTRODUCTION

In the digital era, the role of web development in shaping businesses competitive edge cannot be overstated. By strategically harnessing web technologies, companies can build dynamic and responsive online platforms that not only captivate customers but also streamline operations and foster growth. However, amidst this digital landscape, web scraping emerges as a powerful ally to web development, offering a means to extract invaluable insights from the vast expanse of online data. This report aims to delve into the symbiotic relationship between web development and web scraping, emphasizing how their integration can drive informed decision-making and provide a crucial competitive advantage.

This report sets out several clear objectives to illuminate the fusion of web development and web scraping. Firstly, it seeks to underscore the pivotal role of web development in today's fiercely competitive environment, stressing the necessity of crafting engaging online experiences for sustained success. Secondly, it aims to explore the multifaceted applications of web development in facilitating web scraping endeavors, from tracking product prices to monitoring competitor strategies. Lastly, the report endeavors to furnish practical insights on seamlessly integrating web scraping into existing web development workflows, empowering businesses to leverage these technologies synergistically as part of their overarching digital strategies.

Subsequent sections of the report will venture deeper into the core tenets of web development, elucidating advanced techniques and tools for harmonizing web scraping seamlessly. Through tangible examples and hands-on guidance, businesses will gain actionable insights into leveraging web development principles and methodologies to amplify their web scraping capabilities. By bridging theoretical concepts with practical applications, the report aims to equip businesses with the acumen and proficiency needed to innovate, refine strategies, and seize a competitive advantage in today's dynamic digital milieu through the strategic fusion of web development and web scraping.

LITERATURE SURVEY

Domain	Key Technologies/Methodologies
Web Scraping	- Puppeteer for headless browser automation
	- Bright Data for overcoming website blocks
Database Management	- MongoDB for effective storage and organization
Full-Stack Development	- Next.js for building dynamic web applications
	- Responsive layout design
	- Navigation bars
	- Modals
	- Integration with React components and CSS frameworks
	- Server-side rendering
	- API routes
	- Automatic code splitting

Domain	Key Technologies/Methodologies
Web Scraping	- Python as the primary language for web scraping
	- Inbuilt libraries and extensive third-party support
	- High speed and accuracy compared to other languages
	- Various Python libraries available for different applications
Analysis	- Statistical analysis of web scraping tools and techniques
Conference Details	- Published in 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT)
	- Conference date: 11-12 February 2023
	- Date added to IEEE Xplore: 20 March 2023
	- DOI: 10.1109/ICITIIT57246.2023.10068666
	- Publisher: IEEE
	- Conference Location: Kottayam, India

IMPLEMENTATION DETAILS

Web Technologies Used:

- Puppeteer for headless browser automation.
- Bright Data for unblockable scraping.
- Mongoose for MongoDB interaction.
- Next.js for full-stack development.
- Axios and Cheerio for web scraping.
- Tailwind CSS for frontend styling.

Frontend Development:

- Responsive layouts using flexbox.
- Navigation bars with logos and icons.
- Integration of modals for interactive user experiences.

Backend Development:

- Utilization of Mongoose with MongoDB for database management.
- Setting up Cron jobs for automated periodic price checks and updates.
- Implementation of email notifications for price alerts.

Integration:

- Creation of API routes for interacting with the database
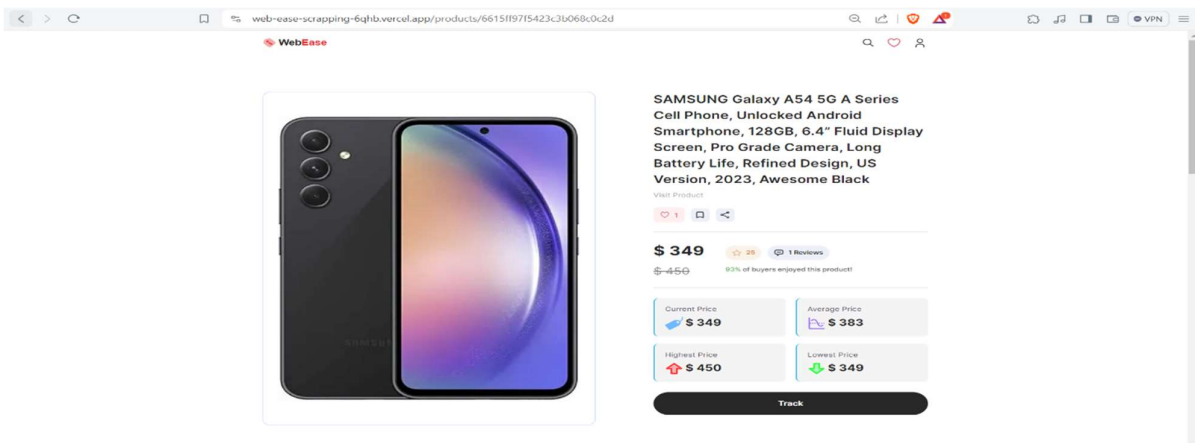- Automation of tasks such as data scraping and email notifications
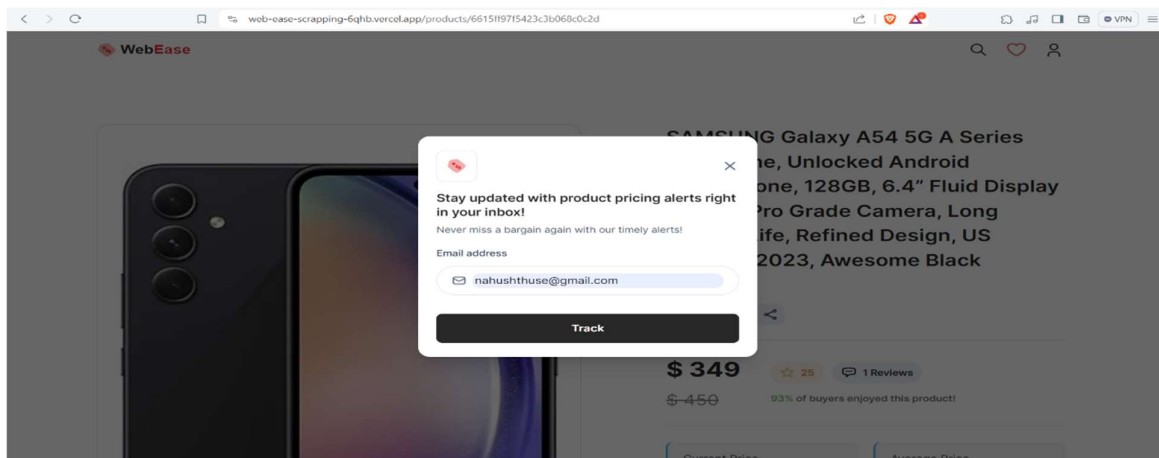
OUTPUT

- **Landing Page**
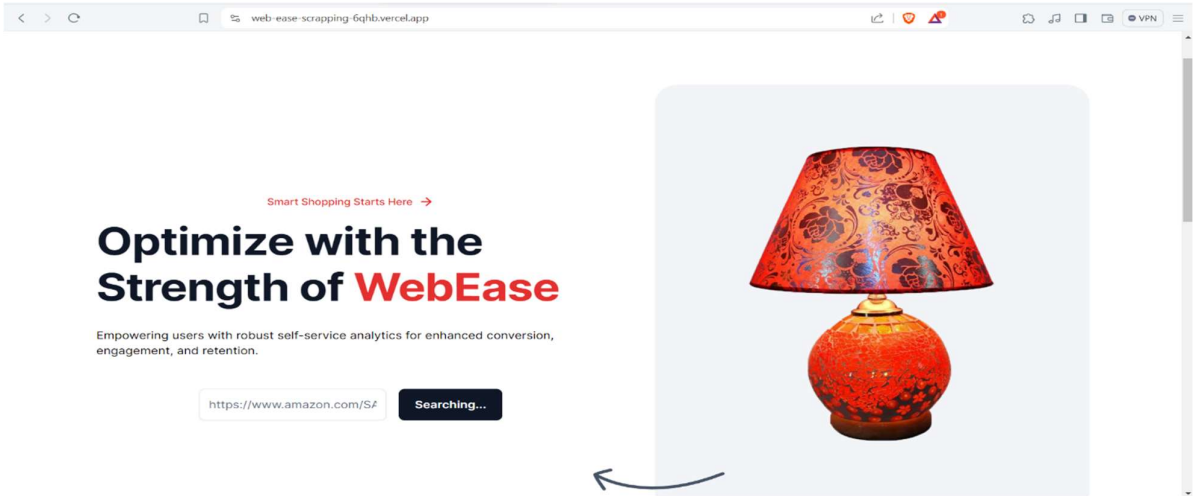
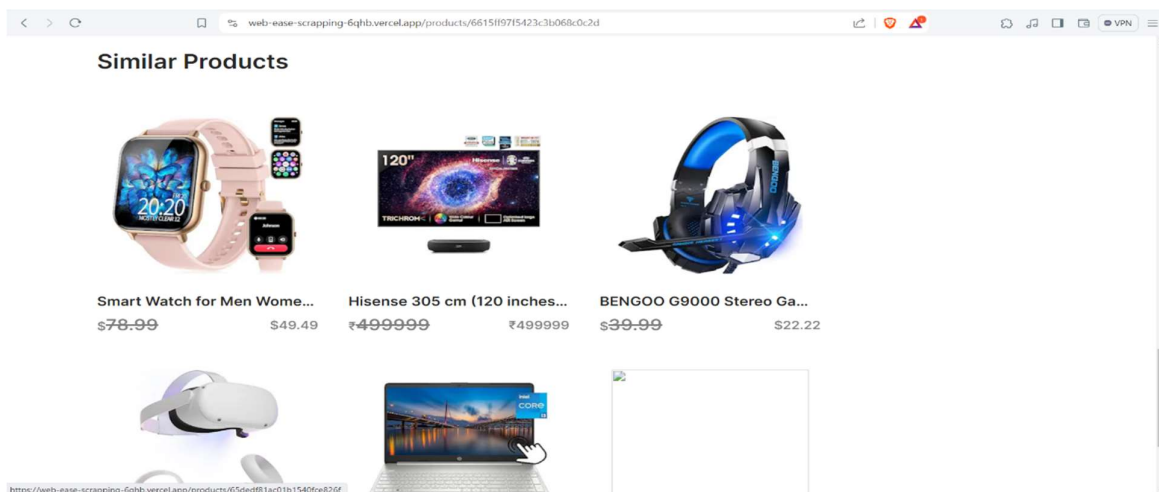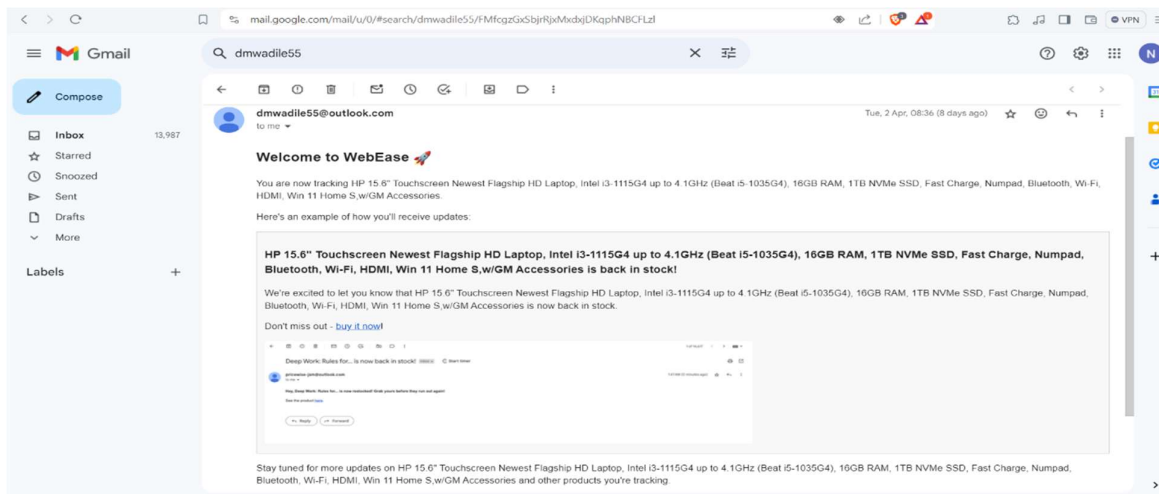- **Trending Items Page**
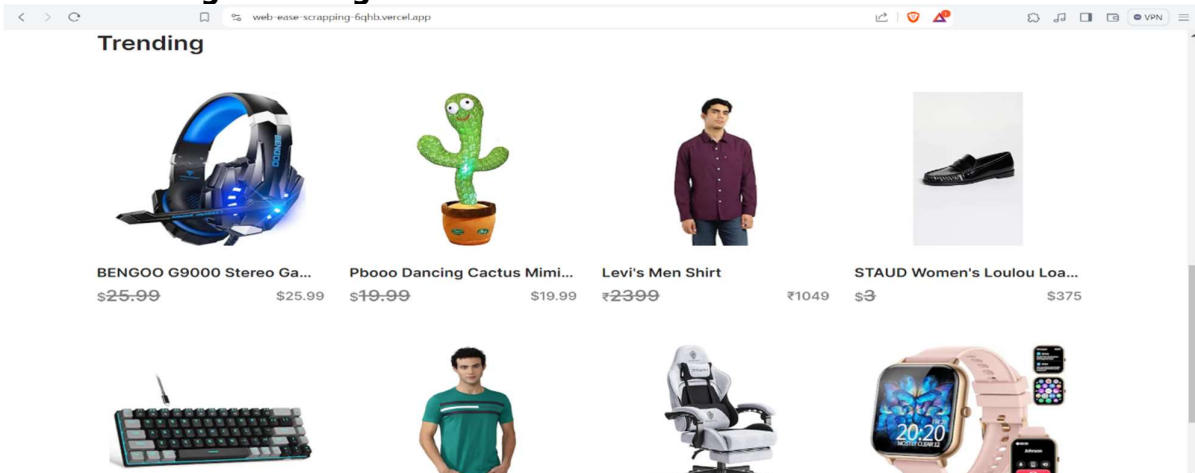
- **Item Scrapping Page**

- **Email Tracking Page**

- **Similar Product Page**

- **Email Notification with all Scrapped Data**

CONCLUSION

The report accentuates the indispensable role of web scraping in diverse domains and offers actionable insights for its effective deployment. It advocates for leveraging cutting-edge technologies such as Puppeteer and Bright Data to streamline data extraction processes, while also stressing the importance of adhering to best practices in database management, particularly with MongoDB. Furthermore, it delves into the realm of full-stack development using Next.js, underscoring the significance of creating responsive and user-friendly interfaces that enhance the overall web scraping experience.

Ethical considerations remain a focal point throughout the report, emphasizing the necessity of responsible data collection and usage. By prioritizing ethical practices, businesses can maintain trust and credibility while harnessing the power of web scraping to drive informed decision-making and gain a competitive edge in today's digital landscape. Looking ahead, the report suggests continual monitoring of advancements in web scraping technologies and ethical frameworks to ensure sustained effectiveness and compliance with evolving standards and regulations.

REFERENCES

[1] Bright Data. "Web Scraping with Bright Data." Accessed at: Web Scraper IDE - Free Trial (brightdata.com)

[2] MongoDB. "Mongoose - MongoDB Object Modeling for Node.js." Accessed at: Mongoose ODM v8.2.1 (mongoosejs.com)

[3] Puppeteer. "Headless Chrome Node.js API." Accessed at: Puppeteer | Puppeteer (pptr.dev)

[4] Next.js. "The React Framework for Production." Accessed at: Next.js by Vercel - The React Framework (nextjs.org)

[5] Tailwind CSS. "A Utility-First CSS Framework for Rapid UI Development." Accessed at: Tailwind CSS - A Utility-First CSS Framework for Rapidly Building Custom Designs

[6] Axios. "Promise Based HTTP Client for the Browser and Node.js." Accessed at: Getting Started | Axios Docs (axios-http.com)

[7] Cheerio. "Fast, Flexible, and Lean Implementation of Core jQuery Designed Specifically for the Server." Accessed at: cheerio - npm (npmjs.com)