**Chrome Extension to generate Content Relevant Data through Intelligent Scanning**

**Akash Sareen**

**Department of computer science,**

**Sathyabama institute of Science and technology,**

**Chennai, India**

[**Akash.sareen7@gmail.com**](mailto:Akash.sareen7@gmail.com)

**Tushar Choudhary**

**Department of computer science,**

**Sathyabama institute of Science and technology,**

**Chennai, India**

[**ash.tushar28@gmail.com**](mailto:ash.tushar28@gmail.com)

**Dr. T Sasikala,**

**Dean of School of computing,**

**Sathyabama institute of Science and technology,**

**Chennai, India**

[**dean.computing@gmail.com**](mailto:dean.computing@gmail.com)

**Abstract-**

**Google Chrome and Google Search are one of the most popular tools available for literature study and surfing around the world used by millions of researchers and academicians. With the rise of Search Engines, Extensions have also been developed to aid people in various activities. Some popular extensions can spell check whatever we type on Internet while others put special emphasis on productivity by cutting down futile time being spent on less productive website. In our project, we have designed a State-of-the-Art Chrome Extension which can intelligently scan a website and give back user relevant data. The Chrome Extension has been developed using Web Development Tools which would allow the user to get content relevant data without performing multiple Search Queries, in a single click.**

**Keywords: Chrome Extension, Scanning, Search, Results**

1. INTRODUCTION

With the introduction of Google Search in 1997 developed by Google, its popularity and usage has seen leaps and bounds and it has been widely used around the world, handling more than 5 Billion Search Queries every day. On 2008, Google also launched its first cross-platform Browser, supported by its open source Chromium Project which soon became the most popular browser in the market. The advent of Google Search and Google Chrome were phenomenal in the view of the ever rising usage of Internet worldwide and contributed heavily to the world of development and research.

With the rise of Google Chrome, various extensions were also developed which are small software modules that can be used to customize a Web-Page and included in the browser. While plugins have been used by browsers for a long time, extensions are majorly source code compared to the plugins which are object code. An extension is developed to further enhance the feature and functionality of the browser. It utilizes JavaScript Application Programming Interfaces (API) to complement the website a further add tools and features to the website which the user uses more often.

Currently extensions are widely used across multiple Browsers like Google Chrome, Opera and Mozilla Firefox [6] where they utilize an add-on to deliver additional in-browser features or information from any website. These Extensions are also designed to manipulate the functionality of the website and help the user utilize many built-in features. For example, many of the extensions today allow the user to block unwanted intrusive advertisements and also reformat our page to a new User-Interface (UI) as the user wishes for.

Many of these extensions have seen popular use among the users for the versatility and the robustness they provide. These extensions are helpful for multiple academicians and researchers where these extensions can offer them spell and grammatical check in real time, provide new words to improve the vocabulary and much more. These features have made these Extensions a popular way for Developers to quickly reach to users without prompting the user to reach to their respective Websites for anything they need.

In our paper, we propose a Chrome Extension that has been developed and deployed using Web Development Tools on Google Chrome Web Browser, named as “Deep Crawler” which has been developed to scan the website completely and return websites and links that are relevant to the particular website where the user has invoked the browser. This helps increase the efficiency of the Google Search Operation and hence reduce the human effort by searching through your webpage to give you the right results relevant to the context that the client is searching.

The current methodology functions by accepting some inputs from the user and then displaying result in a format which is not very intuitive. On Google Search, after searching the result displayed is in a unformatted way where the user feel very inconvenient to navigate and proceed. This has led many websites to utilize tools like Search Engine Optimization (SEO) which are displayed first when searched over the web To avoid this, our application “Deep Crawler” comes into the picture which can intelligently scan a website and return content relevant links. Now, if someone is reading about an article or blog and wishes to read more about the same article our deep crawler helps them to find such pages. With just one click on deep crawler it will be displaying all related sites in a popup.

1. LITERATURE SURVEY

Yuila Volvoka et all proposed a Self-Control Chrome Extension [3] which allows people to control their addiction towards addictive websites and help them manage their time more effectively. In their approach, they developed a Chrome Extension which allows them to manage the time effectively which is being spent on watching Online Television Series. Rather than blocking the feature, the extension works towards cooperating with the user to effectively reduce the time spent on wasteful activities.

Lei Liu et all have proposed yet another Chrome Extension [4] which makes it possible to stimulate Bot activities and implement Spamming, Password Sniffing and DDoS Attacks which is made surprisingly feasible aided by rich Application Programming Interfaces (API). Through their approach, they have demonstrated how Desktop Browsers can be utilized for such activities and how features that are being utilized for providing users browser-rich features can also be used for performing malicious activities.

In the research paper [1] by Lujo Bauer et all, the dangers posed by Chrome Extensions have been explored on how these extensions can prove fatal for the user and how they can steal personal data from the users, given the privileges provided to them by the user and the browser. The paper further explores on how websites are vulnerable to such attacks and how extensions can carry out such attacks.

In the approach [2] stated by Arjun Guha et all, it has been explored how a verified security layer can be added which can stand guard against Chrome Extensions and how Data Logs can be used to fine-grain the access control and the extension controls can be limited to help protect the user from malicious activities and threats posed by modern Extensions.

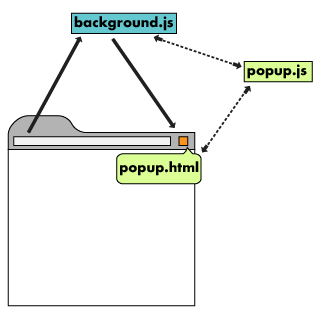
Justin Joseph et all took up a different approach [5] by including a Cookie Based Protocol which can defend against malicious Browser-extensions. In the browser, cookies are continuously used for authentication which secures the data stored in the browser like saved password and form details. This helps protect from session hijacking and helps secure the data further protecting the data from replay attacks.

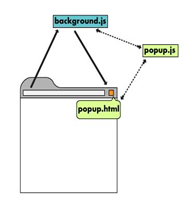
Through these Literature Surveys, it was made possible to understand how Extensions work and how they might be prone to attacks by attackers and how they can be made secure using various methods. It also helps us understand the rise of malicious Browser extensions and how through third-party, they can mask their activities of providing additional functionalities by performing attacks, especially clickjacking attacks, where they can steal personal data and how safe and secure measures can be implemented in our own approach to allow the user hassle-proof and fully secure usage in a convenient manner.

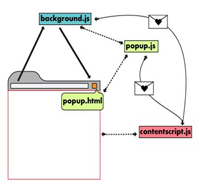
1. PROPOSED SYSTEM

In our project, we have developed a Chrome Extension using HTML, CSS, JavaScript and JSON. Hyper-Text Markup Language (HTML) is the standard language of the Internet and is used for Front-End Web Development while Cascading Style Sheet (CSS) is used for styling the Web Pages and to give it a proper User-Interface. Javascript is an Open-Source Language that is utilized for creating Web Applications on both Front-End and Server-Side purposes. In addition we are using here a manifest file named manifest.json. This is a metadata file in JSON format that contains properties like your extension’s name, description, version number and so on. In this file we tell Chrome what the extension is going to do, and what permissions it requires.

The project architecture has been articulated below:







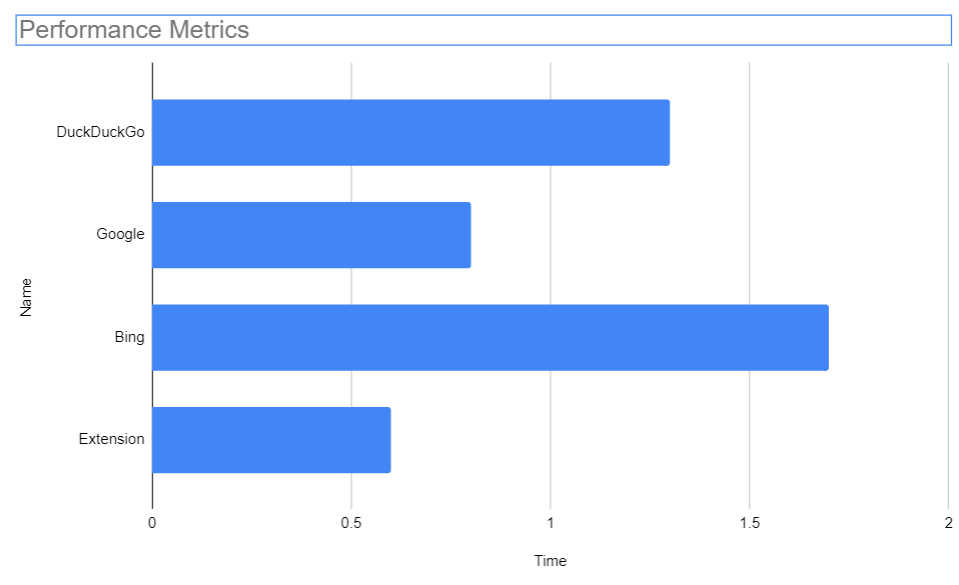
In our project we have the files: background\_main.js which includes Event Listeners that utilizes GET requests to search for queries by scanning the web page using the keywords that have been used in the website. After getting these domains, they are set and displayed on the screen.

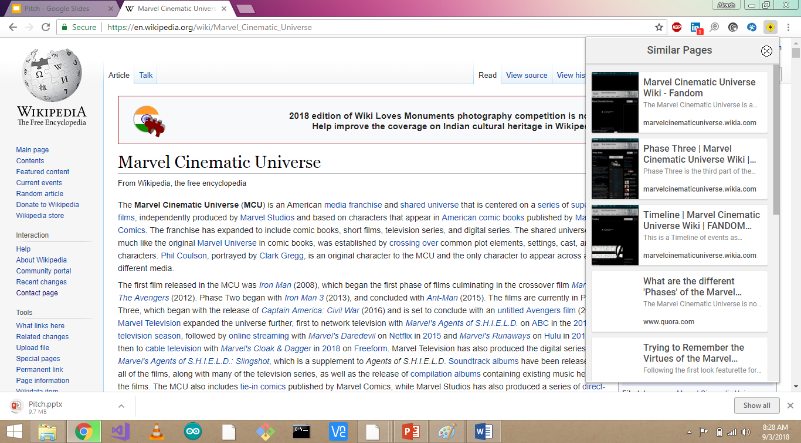
The Chrome Extension will automatically generate all links relevant to the website that we are currently on after being invoked with a click. The user will not need to search for any relevant information and can go through the links generated by the Extension.

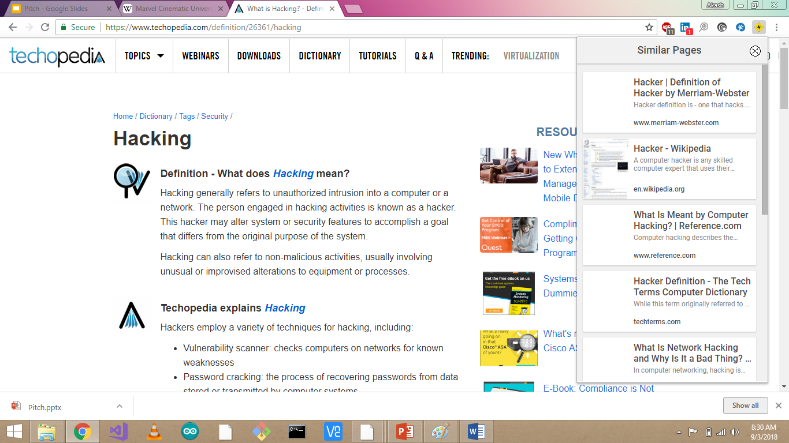
1. PERFORMANCE ANALYSIS

The Chrome Extension was tested on Google Chrome Cross-Platform Browser and positive results were identified with the same. The extension once invoked was able to garner the links that have been intelligently searched and scrapped from the website. The Extension was tested on more than 80 Websites and the results gathered from them pointed out to similar content that can be found on other websites.

Some of the Test Images while invoking the Extension on various websites has been displayed below:







1. Conclusion

In our approach, we developed a Chrome Extension that can intelligently scrap and scan a website to generate relevant results without the user needing to perform repeated Search Queries hence making the whole process automated and much more convenient for the User. Recent smarted Meta search includes the data clustering and analysis of the Meta data and then producing to the browser. With the Emergence of the Web 2.0 services like Social Networking and Web-Logs and feeds increases the non-textual information like podcasting, and online-video and convergence of the non-conventional forms of communication such as mobile phones with Internet will require easy information retrieval in these areas.

VI. Acknowledgements

We are very thankful to the guide Dr. T Sasikala M.E.,

Ph.D., Dean, School of Computing, for her valuable guidance, suggestions and encouragement throughout this work. We convey our thanks to Dr. S. Vigneshwari M.E., Ph.D and Dr. L Lakshmanan M.E., Ph.D., Heads of the Department, Computer Science and Engineering, for full support during the reviews. I also thank all the teaching staff of the dept. for their support.

1. References

[1] Analyzing the dangers posed by Chrome extensions - IEEE Conference Publication. [Online]. Available: <https://ieeexplore.ieee.org/document/6997485>.

[2] Verified Security for Browser Extensions - IEEE Conference Publication. [Online]. Available: <https://ieeexplore.ieee.org/document/5958025/>.

[3] Striving with online addiction with a self-control chrome extension - IEEE Conference Publication. [Online]. Available: <https://ieeexplore.ieee.org/document/7850190/>.

[4] Botnet with Browser Extensions - IEEE Conference Publication. [Online]. Available: <https://ieeexplore.ieee.org/document/6113263/>.

[5] Cookie Based Protocol to Defend Malicious Browser Extensions - IEEE Conference Publication. [Online]. Available: <https://ieeexplore.ieee.org/document/8888425/>.

[6] “Browser Extensions,” MDN Web Docs. [Online]. Available: [https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtension](https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions)

[7] “B.Bharti, D.Hemalatha, “Analytical Schemes to Optimize The Mining Result using Incremental map reduce “

Global Journal of Pure and Applied Mathematics(GJPAM) Paper Code : 53572, vol. 13 (8). To be indexed SCOPUS.

[8] M.D.Kamlesh and B.Bharti (2014), “Slicing and Efficient transaction data Publication and for data publishing” -Indian Journal of Science and Technology, Vol.10, ISS.8, pp 306-309- SCOPUS Indexed