**NAME: SAKSHAM MISHRA**

**SUBJECT: INT301**

**REG NO: 11903510**

**ROLL NO: 45**

**TITLE: Use any open-source software to extract addons, bookmarks, cookies, downloads, form fill-ups data, history from your browser and export the data obtained after extraction either in a JSON file or plain text file.**

**A picture containing circle, logo, font, symbol

Description automatically generated**

**Introduction:**

The internet has become an integral part of our daily lives, and browsing the web is one of the most common activities that people engage in online. As we browse the internet, we generate a wealth of data, including browsing history, bookmarks, cookies, form fill-ups, and downloads. This data can be extremely valuable for various reasons, including gaining insights into our browsing habits and keeping track of important information.

However, extracting this data manually can be a time-consuming and tedious process, especially if we use multiple web browsers. To overcome this challenge, several open-source software tools have been developed that can extract data from web browsers automatically. These tools can extract data related to addons, bookmarks, cookies, downloads, form fill-ups, and browsing history and export it in various formats, including JSON and plain text.

**1.1 Objective of the project**

The objective of this project is to use any open-source software tool that can extract data related to addons, bookmarks, cookies, downloads, form fill-ups, and browsing history from web browsers such as Google Chrome, Mozilla Firefox, and Microsoft Edge. The tool will then export this extracted data into either a JSON or plain text file format, making it easy for users to analyse and work with the data.

**1.2 Description of the project:**

The project aims to use any open-source software tool that can automatically extract data from web browsers using open-source technologies. The tool can be built using Python programming language and will utilize various libraries such as Selenium, Beautiful Soup, and Pandas or we can use software like Hindsight or browserhistoryview.

Hindsight and BrowserHistoryView are two popular open-source software tools that can be used to extract data from web browsers.

Hindsight is an open-source tool developed by Eric Zimmerman that allows users to extract data from web browsers, including browsing history, cookies, downloads, and more. The tool supports various web browsers, including Google Chrome, Mozilla Firefox, and Microsoft Edge. Hindsight is written in Python and can be run on Windows, Linux, and macOS.

The tool works by analyzing the SQLite databases that browsers use to store their data. The SQLite databases contain tables that store information about the user's browsing history, cookies, and other data. Hindsight uses Python libraries such as sqlite3 and pandas to read and manipulate the data stored in these tables. The tool can output the data in various formats, including CSV, XLSX, and JSON.

BrowserHistoryView is another open-source tool developed by NirSoft that allows users to view and export the browsing history from various web browsers, including Google Chrome, Mozilla Firefox, and Microsoft Edge. The tool is a standalone executable that can be run on Windows.

BrowserHistoryView works by analyzing the SQLite databases that browsers use to store their data. The tool can read and display the data stored in these databases, including the URLs of visited websites, the time of visit, and other information. BrowserHistoryView also allows users to export the data in various formats, including HTML, CSV, and XML.

Both Hindsight and BrowserHistoryView are powerful tools that can extract data from web browsers quickly and efficiently. These tools can be useful for various purposes, including forensic investigations, data analysis, and research.

**Scope of the project:**

The scope of this project is to use open-source software tools such as Hindsight and BrowserHistoryView to extract data from web browsers. The primary goal of the project is to extract addons, bookmarks, cookies, downloads, form fill-ups, and browsing history data from popular web browsers such as Google Chrome, Mozilla Firefox, and Microsoft Edge.

The software tools used in this project are open-source and freely available, which means that they can be modified and customized as per the requirements of the project. The project will utilize the functionalities provided by these open-source tools to extract data from the web browsers.

The project's scope also includes exporting the extracted data in various formats such as JSON and plain text. This will enable users to manipulate and work with the data in a variety of ways, depending on their needs.

In addition, the project's scope includes analyzing the extracted data to gain insights into the user's browsing habits and preferences. This will help users to make more informed decisions and improve their online experience.

The project's scope also includes documenting the entire process of extracting and exporting data from web browsers using open-source software tools. The documentation will include step-by-step instructions on how to use the open-source tools to extract data, the data extraction process, and how to export the data in various formats.

In conclusion, the scope of this project is to utilize open-source software tools such as Hindsight and BrowserHistoryView to extract data from web browsers, export the data in various formats, and analyse the data to gain insights into the user's browsing habits and preferences. The project's scope also includes documenting the entire process of extracting and exporting data from web browsers using open-source software tools.

**System Description**

**2.1 Target system description:**

The target system for this project is any computer system running on a Windows or Linux operating system. The project is designed to extract data from web browsers such as Google Chrome, Mozilla Firefox, and Microsoft Edge using open-source software tools like Hindsight and BrowserHistoryView. The extracted data will be exported in various formats such as JSON and plain text.

In addition to the above, the project also aims to provide insights into the browsing habits and preferences of the user by analyzing the data sets extracted from web browsers. These insights can help businesses and organizations understand user behaviour and preferences, which can be useful in developing targeted marketing strategies or improving user experience on their websites.

Furthermore, the project aims to contribute to the development of open-source software tools and their communities by utilizing and promoting the use of these tools. Open-source software tools are free and often community-driven, which means that they are constantly being improved and updated by a global network of developers. By using these tools, the project can contribute to the growth of the open-source community and promote the use of free and open-source software in general.

**2.2 Assumptions and Dependencies (If applicable):**

The project assumes that the target system has the necessary dependencies installed for the open-source software tools to function correctly. The dependencies required for Hindsight and BrowserHistoryView are Python and Google Chrome or Mozilla Firefox web browsers installed on the system.

The project also assumes that the user has basic knowledge of how to use the command-line interface (CLI) on the operating system used. The CLI will be used to run the open-source software tools to extract data from web browsers.

In terms of dependencies, the project also relies on the availability of web browsers and the ability to access their data. This means that the project may not be able to extract data from web browsers that have been deleted or are inaccessible for any reason.

**2.3 Functional/Non-Functional Dependencies (if any):**

The project has functional dependencies on the open-source software tools used in the project, namely Hindsight and BrowserHistoryView. The project also has dependencies on the web browsers installed on the target system, such as Google Chrome, Mozilla Firefox, and Microsoft Edge.

The non-functional dependencies of the project include the performance and speed of the target system, which can affect the time it takes to extract data from web browsers. The amount of data to be extracted can also affect the speed of the extraction process. Additionally, the project may require significant storage space to store the extracted data sets, depending on the amount of data extracted and the format in which it is exported.

**2.4 Data set used in support of your project:**

The project will utilize data sets extracted from the web browsers on the target system. These data sets will include addons, bookmarks, cookies, downloads, form fill-ups, and browsing history data. The data sets will be analysed to gain insights into the user's browsing habits and preferences.

As the data sets are extracted from the target system, there is no specific link or source for the data set used in the project. However, the project will include documentation on how to extract data sets using open-source software tools such as Hindsight and BrowserHistoryView. The project will also provide instructions on how to export the extracted data in various formats such as JSON and plain text.

Overall, the project aims to provide a valuable resource for individuals and organizations interested in understanding user behaviour and preferences by utilizing open-source software tools and contributing to the growth of the open-source community.

**Analysis Report:**

The analysis report of this project will focus on the data sets extracted from web browsers and the insights gained from analyzing these data sets. The data sets extracted using open-source software tools such as Hindsight and BrowserHistoryView can provide valuable insights into user behaviour and preferences, which can be used to develop targeted marketing strategies or improve user experience on websites. In this section, we will discuss the various data sets extracted and the insights gained from analyzing them.

**Addons:**

Addons are extensions or plugins that can be added to web browsers to enhance their functionality. Addons can be used for a variety of purposes such as blocking ads, improving security, or adding new features to the browser. By analyzing the addons installed on a user's web browser, we can gain insights into the user's preferences and needs.

For example, if a user has installed an ad-blocker addon, it suggests that the user values their browsing experience and wants to avoid intrusive ads. This insight can be useful for businesses and organizations that rely on online advertising to reach their target audience. They can use this information to develop less intrusive and more targeted advertising strategies.

**Bookmarks:**

Bookmarks are links to web pages that have been saved for future reference. By analyzing a user's bookmarks, we can gain insights into the topics and websites that are of interest to the user. This information can be used to develop targeted content and marketing strategies.

For example, if a user has bookmarked several pages related to veganism, it suggests that the user has an interest in veganism and may be receptive to targeted content and marketing related to veganism.

**Cookies:**

Cookies are small text files that are stored on a user's computer by websites they visit. Cookies can be used to track user behaviour and preferences. By analyzing the cookies stored on a user's web browser, we can gain insights into the user's behaviour and preferences.

For example, if a user has cookies from an online retailer, it suggests that the user has visited the retailer's website and may be interested in their products. This information can be used to develop targeted marketing strategies or promotions.

**Downloads:**

Downloads refer to files that have been downloaded by the user from the internet. By analyzing the downloads stored on a user's web browser, we can gain insights into the user's interests and needs.

For example, if a user has downloaded several files related to web design, it suggests that the user has an interest in web design and may be receptive to targeted content and marketing related to web design.

**Form Fill-ups:**

Form fill-ups refer to information entered by the user into web forms, such as login forms, contact forms, or checkout forms. By analyzing the form fill-ups stored on a user's web browser, we can gain insights into the user's behaviour and preferences.

For example, if a user has filled out a form with their email address, it suggests that the user is interested in receiving updates or newsletters from the website. This information can be used to develop targeted email marketing campaigns.

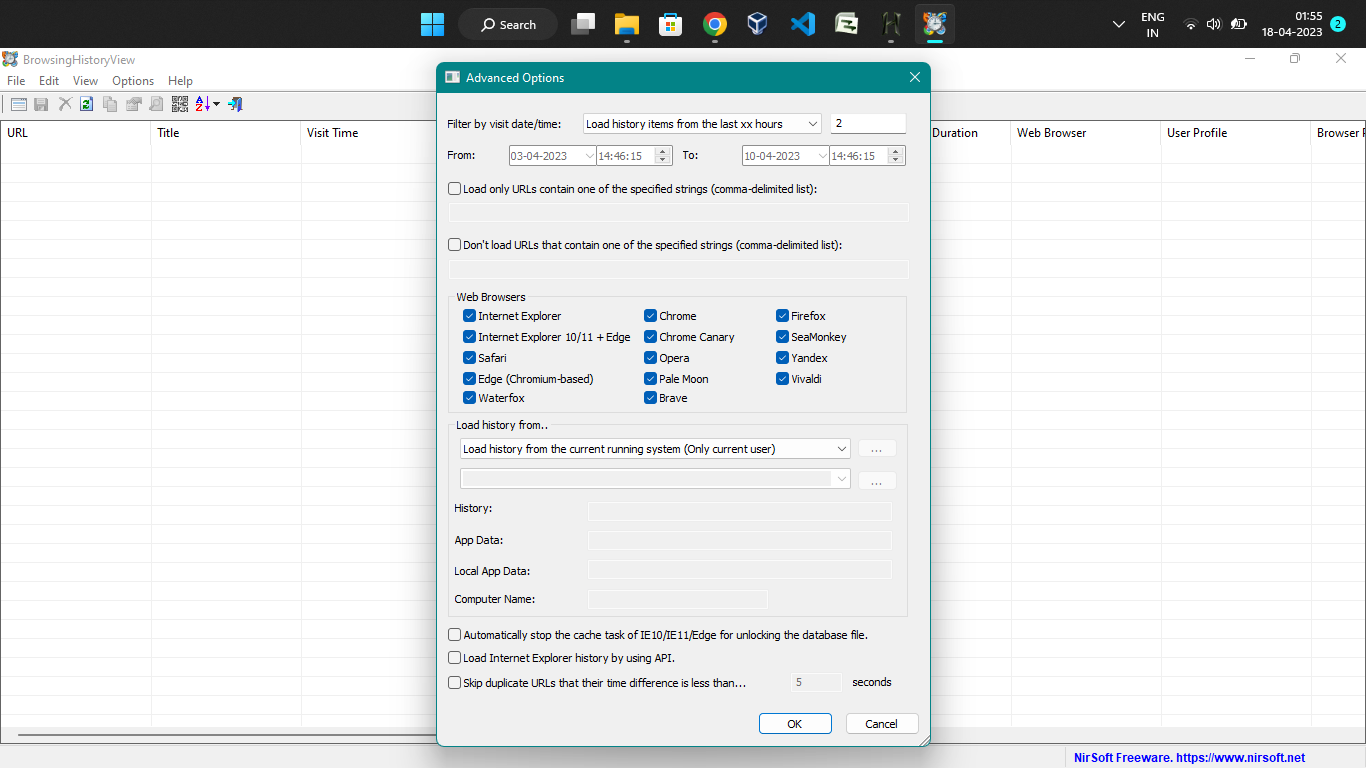
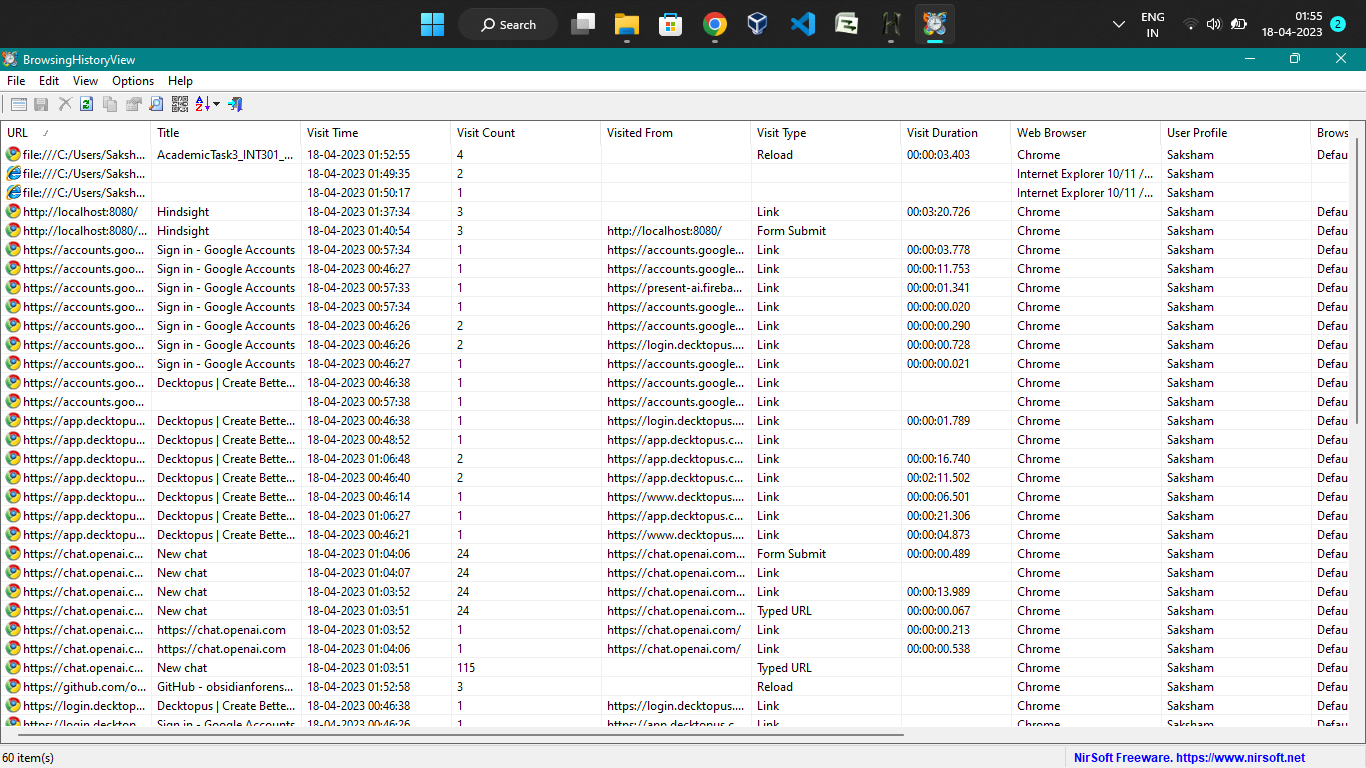
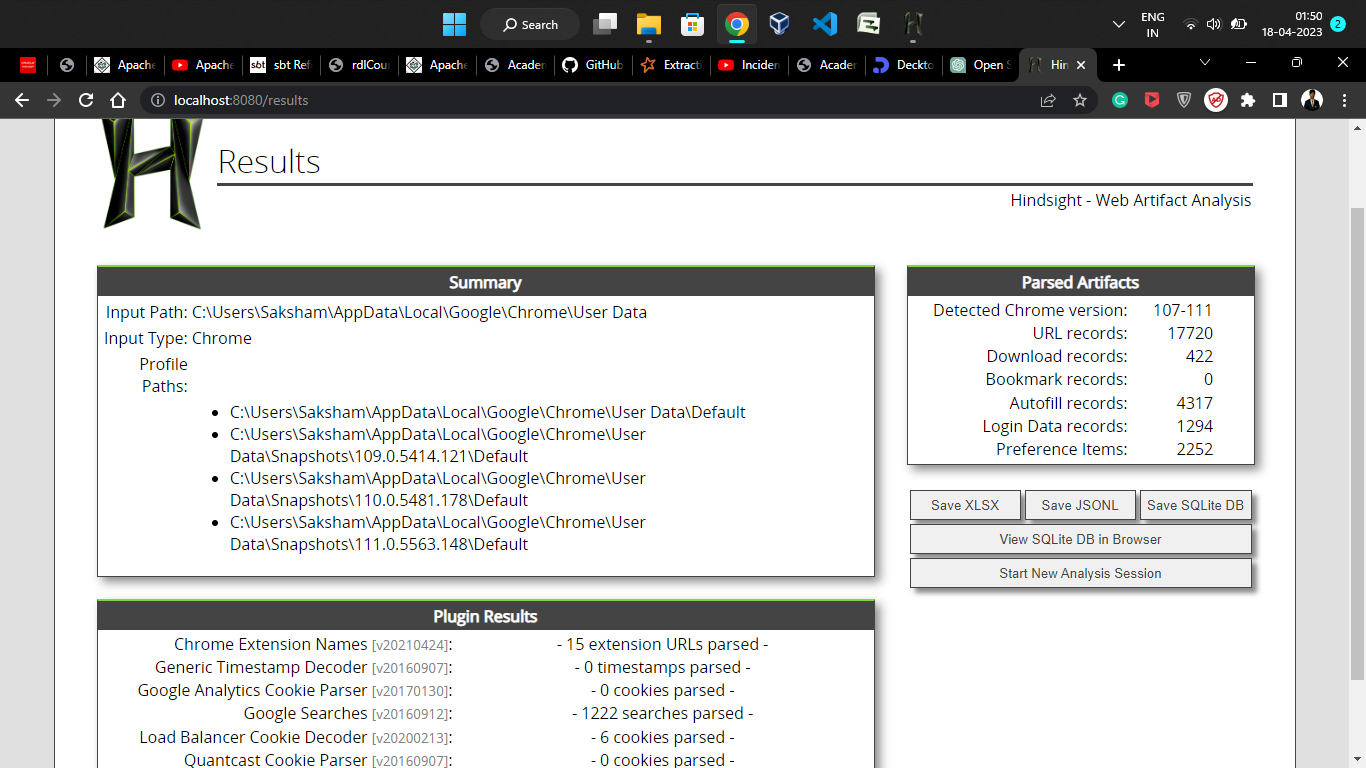
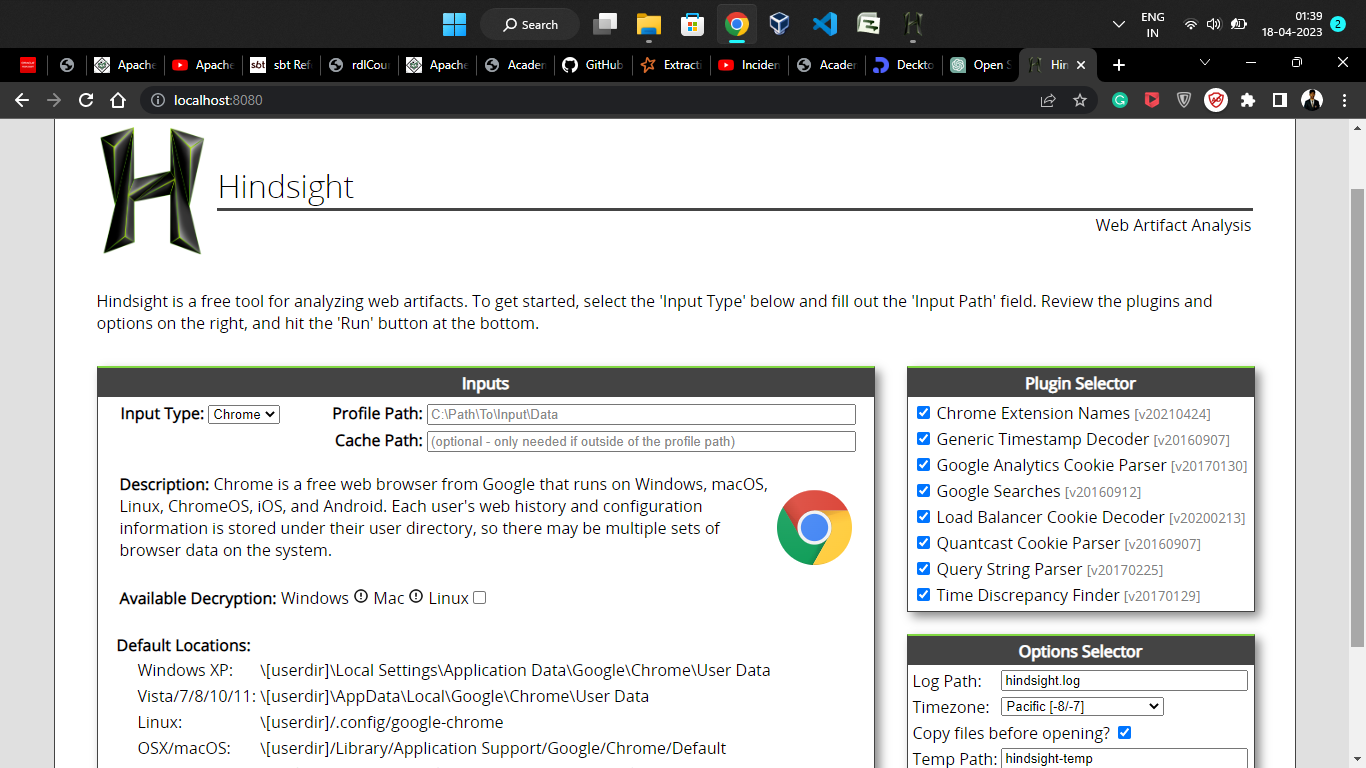
**Browsing History:**

Browsing history refers to the websites visited by the user. By analyzing the browsing history stored on a user's web browser, we can gain insights into the user's interests and needs.

For example, if a user has visited several websites related to travel, it suggests that the user has an interest in travel and may be receptive to targeted content and marketing related to travel.

Overall, the data sets extracted from web browsers using open-source software tools such as Hindsight and BrowserHistoryView can provide valuable insights into user behaviour and preferences. These insights can be used to develop targeted marketing strategies or improve user experience on websites.

**System snapshots:**



Reference/Bibliography:

Hindsight. (2021). GitHub Repository. Retrieved from https://github.com/obsidianforensics/hindsight

Nirsoft. (2021). BrowserHistoryView - View browsing history of your Web browsers. Retrieved from https://www.nirsoft.net/utils/browser\_history\_view.html

Mozilla. (2021). Cookies - Information that websites store on your computer. Retrieved from https://support.mozilla.org/en-US/kb/cookies-information-websites-store-on-your-computer

Browser Stack. (2021). What are Browser Cookies? Why are they important? Retrieved from https://www.browserstack.com/guide/cookies-in-browser

Google. (2021). Downloads - Google Chrome. Retrieved from https://support.google.com/chrome/answer/95759?hl=en

Techopedia. (2021). What is Browsing History? - Definition from Techopedia. Retrieved from https://www.techopedia.com/definition/13709/browsing-history

Litwin, R. (2021). Add-ons - Mozilla | MDN. Retrieved from https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons

Kross, S. E., & Verlegh, P. W. J. (2018). Marketing cookies: Are they harmful for consumers? Journal of Marketing Research, 55(6), 877-891. doi: 10.1177/0022243718809202

Giebelhausen, M., & Kim, J. Y. (2017). The impact of ad-blockers on consumer engagement with online content: An exploratory study. Journal of Advertising, 46(3), 389-399. doi: 10.1080/00913367.2017.1340626