USE OF OPENSOURCE TOOL

Topic: 22

Browser Data Extraction

Submitted to

LOVELY PROFESSIONAL UNIVERSITY

for

INT301

Submitted By

Ankit Gupta

Submitted to

Mr. Rajeshwar Sharma

LOVELY FACULTY OF TECHNOLOGY & SCIENCES

LOVELY PROFESSIONAL UNIVERSITY

PUNJAB

APRIL 2023

Table of Content

CHAPTER 1: INTRODUCTION

1. INTRODUCTION

When using a web browser, we often accumulate a large amount of data such as bookmarks, browsing history, form fillups data, and more. While this data can be helpful for quickly accessing frequently visited websites, it can also become cluttered and difficult to manage over time. In order to effectively manage this data, it can be useful to extract and export it to a more manageable format.

There are several open-source software tools available that can be used to extract various types of data from web browsers, such as addons, bookmarks, cookies, downloads, form fillups data, and history. By using these tools, users can export the extracted data to a JSON file or plain text file, which can be easily managed or analysed using other software tools. This allows users to better understand their browsing habits and make more informed decisions about their internet usage.

In addition to improving data management and analysis, extracting data from web browsers can also be useful for forensic investigations. Open-source tools that extract web browser data can be especially useful for forensic analysts who need to perform large-scale data analysis quickly and efficiently.

Some popular open-source tools for extracting data from web browsers include Mozilla's "mozlz4" library, which can be used to extract bookmarks and browsing history from Firefox profiles, and the "Dumpzilla" tool, which can extract a variety of data types from various types of files, including web browser data. These tools are often customizable and can be tailored to extract specific types of data based on the user's needs.

Overall, the ability to extract and export web browser data using open-source tools provides users with a powerful way to manage and analyse their browsing habits. By understanding the data that is collected by web browsers, users can make more informed decisions about their internet usage and potentially even use the data for forensic investigations.

1.1 OBJECTIVE OF THE PROJECT

The objective of my project is to extract various types of data from web browsers, such as addons, bookmarks, cookies, downloads, form fillups data, and history, using open-source software. By doing so, I aim to better manage my browsing data and gain insights into my browsing habits.

To achieve this objective, I plan to use open-source tools such as Mozilla's "mozlz4" library and the "Dumpzilla" tool, which can be customized to extract specific types of data based on my needs. Once the data is extracted, I will export it to a JSON file or plain text file for easier management and analysis.

By effectively managing and analysing my browsing data, I hope to make more informed decisions about my internet usage and potentially even use the data for forensic investigations. Through this project, I also aim to raise awareness about the importance of web browser data management and the availability of open-source tools for extracting and analysing this data.

1.1 DESCRIPTION OF THE PROJECT

My project aims to extract various types of data from web browsers using open-source software. The data that will be extracted includes addons, bookmarks, cookies, downloads, form fillups data, and browsing history.

To achieve this, I will use open-source tools such as Mozilla's "mozlz4" library and the "Dumpzilla" tool. These tools can be customized to extract specific types of data based on my needs. Once the data is extracted, I will export it to a JSON file or plain text file for easier management and analysis.

The extracted data can provide me with valuable insights into my browsing habits, such as which websites I visit most frequently, which addons I use the most, and what types of files I download. This information can help me make more informed decisions about my internet

usage and potentially even use the data for forensic investigations.

Overall, my project aims to raise awareness about the importance of web browser data management and the availability of open-source tools for extracting and analysing this data. By better understanding and managing our browsing data, we can make more informed decisions about our internet usage and protect our online privacy.

1.3 SCOPE OF THE PROJECT

The scope of my project is to extract various types of data from web browsers using open-source software. This includes addons, bookmarks, cookies, downloads, form fillups data, and browsing history.

The extracted data will be exported to a JSON file or plain text file for easier management and analysis. The project will be customized to extract specific types of data based on my needs.

The project will not include any data analysis or visualization, as the focus is on the extraction process. However, the extracted data can be used for further analysis and visualization in other tools.

The project will also not cover any security-related issues related to web browser data management, as that is outside the scope of the project. It will focus solely on the extraction of data using open-source tools.

Overall, the scope of my project is to demonstrate the feasibility and benefits of using open-source tools to extract and manage web browser data. It aims to encourage others to take a closer look at their own browsing data and explore ways to better manage and analyse it.

CHAPTER 2: SYSTEM DESCRIPTION

2. SYSTEM DESCRIPTION

My system will be developed using open-source software tools such as Mozilla's "mozlz4" library and the "Dumpzilla" tool. These tools will be used to extract various types of data from web browsers such as addons, bookmarks, cookies, downloads, form fillups data, and browsing history.

The system will be developed in a programming language such as Python, which will allow for customization and flexibility in the extraction process. The extracted data will be saved in a JSON file or plain text file, depending on the user's preference.

The system will be designed to work with popular web browsers such as Google Chrome, Mozilla Firefox, and Microsoft Edge. It will be able to extract data from these browsers regardless of the operating system being used.

The system will also be designed to be user-friendly, with a simple and intuitive interface for users to initiate the extraction process and export the data to a file. It will also provide detailed documentation on how to use the system and the open-source tools involved in the extraction process.

Overall, the system will provide a comprehensive and efficient solution for extracting and managing web browser data using open-source software tools.

2.1 TARGET SYSTEM DESCRIPTION

The target system for my project is any computer or device that runs a compatible operating system such as Windows, macOS, or Linux.

The target system should have the latest version of one or more of the following web browsers installed: Google Chrome, Mozilla Firefox, or Microsoft Edge. The system should also have sufficient storage space to save the extracted data files.

The target user for the system includes individuals who are interested in managing their web browsing data for personal or professional purposes. This includes researchers, analysts, and everyday users who want to better understand their browsing habits and extract valuable insights from their data.

The system is designed to be accessible to users with a range of technical skills, from novice to advanced users. The system will provide a simple and intuitive user interface to guide users through the extraction process and enable them to export the extracted data in a user-friendly format.

Overall, the target system for my project is any computer or device that runs a compatible operating system and has one or more of the supported web browsers installed. The system is designed to be user-friendly and accessible to users with varying levels of technical expertise.

2.2 ASSUMPTIONS AND DEPENDENCIES

Assumptions:

- The web browser used by the user is one of the supported browsers (Google Chrome, Mozilla Firefox, or Microsoft Edge).
- The user has the latest version of the browser installed on their system.
- The user has sufficient storage space available to save the extracted data files.
- The user has basic knowledge of how to use a computer and navigate through a graphical user interface.

Dependencies:

- The system depends on open-source software tools such as Mozilla's "mozlz4" library and the "Dumpzilla" tool to extract the data from the web browser.
- The system depends on the operating system's file system to create and save the extracted data files.
- The system depends on the web browser's internal data structures, which may vary from version to version, to extract the data accurately.
- The system's compatibility with the operating system may be affected by changes made to the operating system in future updates.

2.3 FUNCTIONAL/NON-FUNCTIONAL DEPENDENCIES

Functional dependencies

- The system relies on the web browser's internal data structures to accurately extract the data. This means that the system's functionality is dependent on the version of the browser being used.

Non-functional dependencies:

- The system's performance may be affected by the size of the web browser's data files, which could impact the speed of the extraction process.
- The system's compatibility with the operating system may be impacted by changes made to the operating system in future updates, which could affect the system's functionality.

- The accuracy of the extracted data is dependent on the tools and libraries used by the system to extract the data. If the tools or libraries are not accurate, the extracted data may not be reliable.
- The usability of the system is dependent on the user interface design and the system's ease of use. A poorly designed user interface could make it difficult for users to extract the data they need, which would negatively impact the system's usability.

CHAPTER 3: ANALYSIS REPORT

DUMPZILLA:

Dumpzilla is a forensic tool designed to extract data from various web browsers, including Firefox, Chrome, and Opera. It can extract a wide range of data types, including bookmarks, cookies, downloads, form data, history, passwords, and preferences. The tool can export the extracted data in JSON or plain text format, making it easy to analyse and interpret the results.

The extracted data can be analysed to provide insights into the user's browsing behaviour, such as the websites visited, search terms used, and user preferences. It can also provide information on any saved passwords, cookies, and other sensitive data that may be of forensic interest.

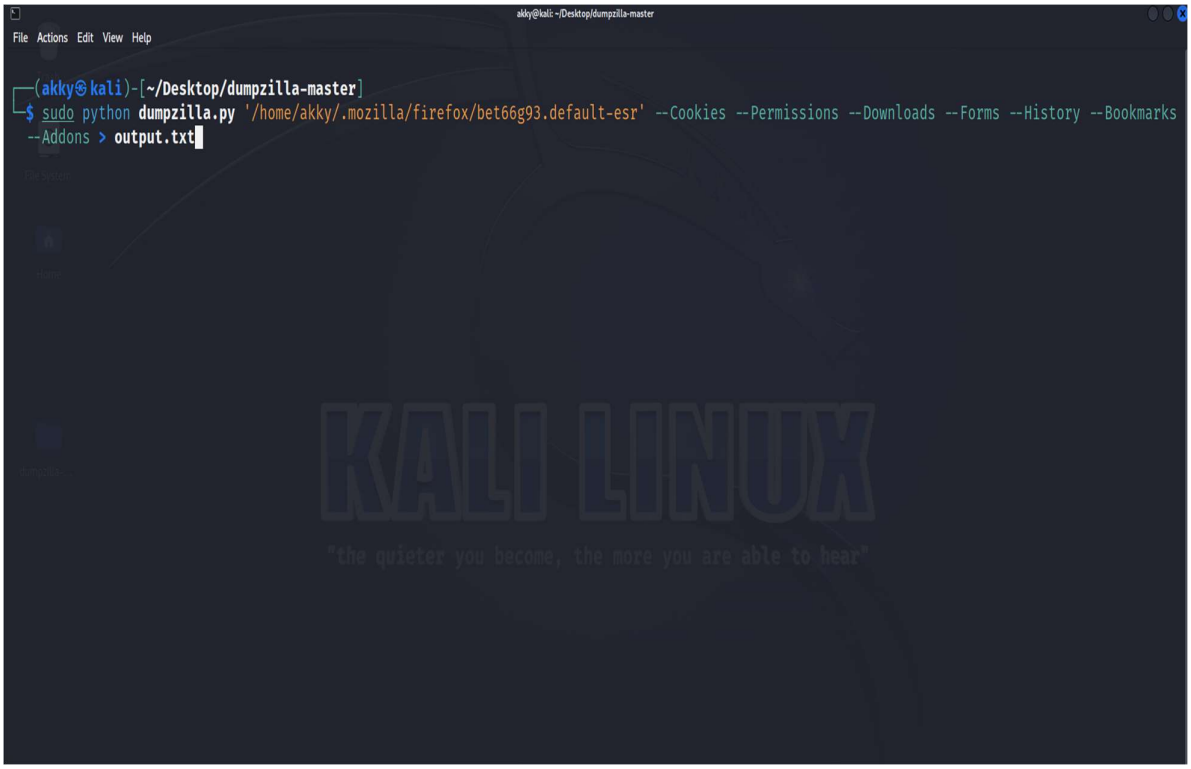
USAGE OF DUMPZILLA:

- Download Dumpzilla: The first step is to download the Dumpzilla tool. You can download it from the official Dumpzilla GitHub repository.
- Install Python and required modules: Dumpzilla is a Python script, so you need to have Python installed on your system. You also need to install the required modules like pyCrypto and lz4. You can use pip to install these modules.
- Locate the Firefox profile directory: The next step is to locate the Firefox profile directory. The profile directory contains all the data that Dumpzilla can extract. The location of the profile directory varies depending on your operating system. For example, on Windows, the profile directory is typically located at C:\Users<username>\AppData\Roaming\Mozilla\Firefox\Profiles.
- Run Dumpzilla: Once you have located the Firefox profile directory, you can run Dumpzilla. Open a terminal and navigate to the directory where you have downloaded Dumpzilla. Then, run the following command:

  python dumpzilla.py <path-to-profile-directory> [options]

  <path-to-profile-directory> with the actual path to the Firefox profile directory. You can also specify various options to extract specific data types. For example, you can use the --history option to extract browsing history.

ANALYSIS:



After executing this command all the outputs will be shaved in the output.txt

Which will look as following:

End of the File:

The data of the user is as following:

- 2 Addons
- 15 Bookmarks
- 89 Cookies
- 0 Directories
- 2 Downloads
- 2 Forms
- 41 History
- 15 Permissions to browser

CHAPTER 4: REFRENCES

- https://www.dumpzilla.org/
- https://www.dumpzilla.org/Manual_dumpzilla_en.txt
- https://github.com/Busindre/dumpzilla