Table of Contents

1. Introduction
2. Data Extraction
3. Combining Data
4. Saving to CSV
5. Contributing
6. Dataset

Introduction

Chemical compounds play a crucial role in various scientific and industrial applications. This repository aims to centralize data related to different compounds, making it easier for researchers and enthusiasts to access and analyze relevant information.

Data Extraction

The provided Python code extracts data from a specific website related to chemical compounds. Here’s how it works:

* The code initializes a Chrome driver using Selenium and navigates to a target webpage.
* It retrieves information about chemical companies, including dates, product names, prices, and weights.
* The extracted data is then organized for further analysis.

Combining Data

In addition to extracting data, the code snippet demonstrates how to combine information from multiple DataFrames. By appending one DataFrame to another and resetting the index, you can create a comprehensive dataset.

Saving to CSV

Once you’ve collected and combined the data, you can save it to a CSV file. The provided code snippet shows how to specify the output path and export the DataFrame to a CSV format.

Dataset

1. ChemBook\_final.csv

This spreadsheet contains detailed information on various chemical products. Below is a breakdown of the columns and what type of information they contain:

* ID: A unique identifier for each entry.
* Name: The name of the chemical product or the source where the chemical data is listed.
* Tel: The telephone number associated with the chemical or source.
* Email: The email address associated with the chemical or source.
* CAS: Chemical Abstracts Service (CAS) registry number, a unique identifier for every chemical.
* Product N (Product Name): The name of the chemical product being sold.
* Stock We (Stock Weight): The weight of the stock being sold (presumably in grams or other relevant units).
* Price (in USD): The price of the stock in US dollars.
* Last Update: The date when the information was updated.
* Remarks: Additional notes or comments regarding the chemical product, often including purity percentage and URL for more details.
* Source\_URL: The url link to the page that the information was scraped from.

1. ChemBook\_final\_names:

This Sheet contains the unique names and id’s of the company names.

* ID: An id for the company name in the format CB\_Company\_#
* Company Name: The names of the companies.

1. ChemBook\_final\_CAS:

This Sheet contains all the unique CAS numbers and id’s.

* ID: An id for the company name in the format CB\_CAS\_#
* CAS: The CAS numbers of the chemicals

1. ChemBook\_final\_emails:

This Sheet contains all the unique emails and id’s

* ID: An id for the company name in the format CB\_Email\_#
* Email: The Email ID

1. ChemBook\_final\_names:

This Sheet contains the unique phone numbers in all formats and id’s.

* ID: An id for the company name in the format CB\_Company\_#
* Phone: The phone numbers of the companies, some records might have more than one number.