Scraping the AAS (Asset Administration Shell) using Python and Selenium for the retrieval and search of updated manuals and files.

Table of Contents

[Function Flow Chart 1](#_Toc134087963)

[Function Descriptions 3](#_Toc134087964)

[Parse product page: 3](#_Toc134087965)

[Check if modal exists 3](#_Toc134087966)

[Check if cookies exists 3](#_Toc134087967)

[If modal and cookies exist click close 3](#_Toc134087968)

[Get product series name 3](#_Toc134087969)

[Click to destination 4](#_Toc134087970)

[Clean data 4](#_Toc134087971)

[Parse links 4](#_Toc134087972)

[Get doc text 4](#_Toc134087973)

[Combine data 4](#_Toc134087974)

# Function Flow Chart

Function Flow Chart

A screenshot of a computer

Description automatically generated with medium confidence

# Function Descriptions

## Parse product page:

Accepts as arguments: url, is\_initial

url: accepts a string of the web page to parse

is\_initial: accepts a bool type of true or false

This function is the main function that is executed to get the proper information. It calls these following functions:

1. Check if modal exists -> bool
2. Check if cookies exists -> bool
3. If modal and cookies exists click and close
4. Get product series name -> dictionary
5. Click to destination
6. Clean data -> dictionary:
   1. Parse links -> Generator of type string
      1. Returns an array of strings
   2. Get doc text -> list
      1. Return an array of strings
   3. Combine data -> dictionary
      1. Return a dictionary Object

## Check if modal exists

This function is used in the initial run of the program. It runs a check whether on the AAS website a pop up modal is visible or not and returns a bool of True or False.

## Check if cookies exists

This function is also used in the initial run of the program. It also runs a check but looks for a cookies dialogue and return a bool of True or False

## If modal and cookies exist click close

This function is also used in the initial run of the program and receives two arguments. This first argument is the return value of the Check if modal exists function and the second argument is the return value of the check if cookies exist function. This runs a check that if any of these are true, to then click the close button if the modal exists and to click accept if there is a cookies dialogue available.

## Get product series name

This functions returns a dictionary Object and takes no arguments. It grabs the product name, product series, and product type from the information tab of the product page and return the foundational data structure needed to append the document information.

## Click to destination

Takes no arguments and returns nothing.

This function states what it does. It is a clicker function to reveal HTML that is hidden within the single page website by Java Script.

## Clean data

Takes as an argument the foundational data that is returned from the function Get product series name and returns a dictionary Object

### Parse links

Used in the Clean data function

Take as an argument a list and returns a Generator of type string

### Get doc text

Takes two arguments: data of a dictionary type, and a web element of type list

Used in the Clean data function. We pass down the 1st argument from the clean data function and use this as our foundational data structure. The second argument is a list of web elements that text needs to be extracted from. It returns the foundational data structure with the appended text in the documents section of the dictionary Object.

### Combine data

Used in the Clean data function.

This function combines the text data and links to complete the data structure in filling in the empty links section in the foundational data structure. It takes 2 arguments. The first argument is the return statement from the Get doc text function and the second argument is the return statement from the Parse links functions. It then returns a completed data structure of type dictionary needed to write to the JSON file.