



AUTOMATION[®]
ANYWHERE

Go be great.

HTML Parser Package

Readme

Version 1.0.3

29/07/2020

Readme	1
1. Introduction	3
1.1 Overview	3
1.2 Use cases	3
2. Requirements & Prerequisites	4
2.1 System Requirements	4
2.2 Prerequisites	4
3. Getting Started	5
3.1 Quick Start	5
3.1.1 Setup	5
3.1.2 Configuration and Use	5
3.1.3 Samples	7
4. Support & FAQs	10
4.1 Support	10
4.2 FAQs	10

1. Introduction

This document contains all essential information for the user to make full use of this A2019 **HTML Parser Package**. It includes a description of the functions and capabilities and step-by-step procedures for setup & configuration of the **HTML Parser Package**.

1.1 Overview

The package provides actions to retrieve HTML data with various filters and selectors from a string, file or web page.

The package is based on the Java libraries

jsoup

<https://jsoup.org/>

and HtmlUnit

<https://htmlunit.sourceforge.io>

1.2 Use cases

The key use cases include:

- Retrieve elements from HTML documents such as emails or web pages
- Check for certain keywords or elements on a web page
- Convert HTML into valid XML

2. Requirements & Prerequisites

2.1 System Requirements

- Enterprise A2019.14+

2.2 Prerequisites

No prerequisites

3. Getting Started







3.1 Quick Start

3.1.1 Setup

Import the A2019 HTML Parser Package in the Control Room under Bots->Packages.

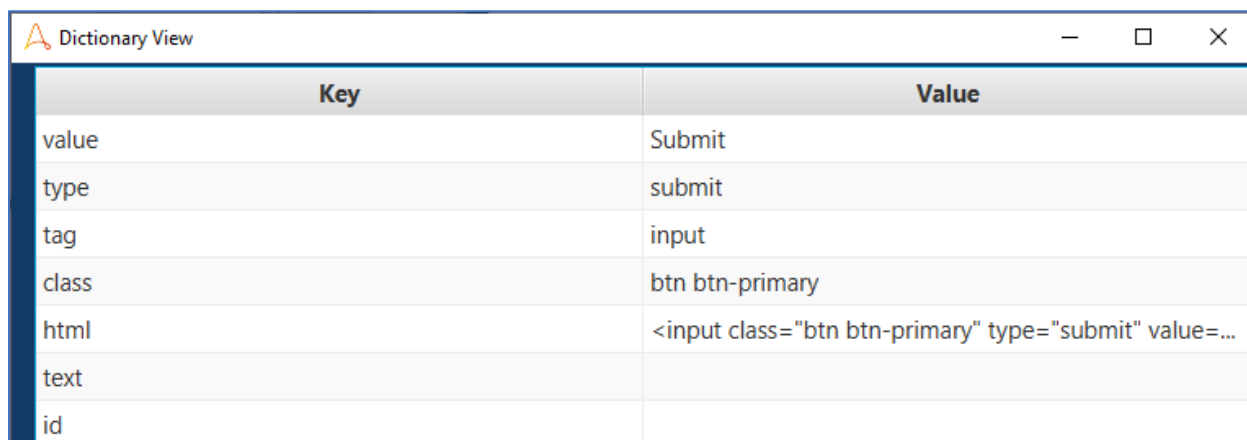
3.1.2 Configuration and Use

The package contains the following command actions

-  Advanced Search with RegEx
-  Get HTML Content
-  Convert HTML to XML
-  Search with Selector
-  Simple Search for Elements
-  Simulate Browser

If an action returns a list, the list elements are of type Dictionary and contain the following keys:
html, text, id, tag, value plus a key for every attribute name

Here a sample of a submit button



Key	Value
value	Submit
type	submit
tag	input
class	btn btn-primary
html	<input class="btn btn-primary" type="submit" value=...
text	
id	

Name	Description	Parameter	
Advanced Search with RegEx	RegEx Search by Attribute Value or Text Value	HTML	HTML string
		RegEx	Regular expression
		Attribute Name (optional)	If Query Type = Attribute, a Attribute Name is needed
		Query Type	By Attribute Value or Text Value
		List of Elements of type Dictionary	One dictionary for each HTML element
Get HTML Content	Get HTML text from a Web site or File, does not simulate a Browser	HTML Source	File name or Web URL
		Source Type	URL or File
		HTML	HTML Content of Web Site or File
Search with Selector	Advanced Search with Selector	HTML	HTML string
		Selector	Selector expression
		List of Elements of type Dictionary	One dictionary for each HTML element
Simple Search for Elements	Simple Search for Elements by Tag, ID, Class or Text	HTML	HTML string
		Query	Value
		Query Type	Type of element looking e.g. if Query is div and Query Type is Tag, it will find all <div> elements in the HTML string
		List of Elements of type Dictionary	One dictionary for each HTML element

Simulate Browser	Simulates a Browser to get HTML content.	URL	HTML string
		Chrome Driver Path	Chrome, Internet Explorer or Firefox
		HTML	HTML Content of Web Site
Convert HTML to XML	Converts HTML data to XML	HTML	HTML string
		XML	Converted XML string

A HTML element has the following, simplified schema

`<tag class="class" id="value" attribute="value"> text value</tag>`

3.1.3 Samples

1. Find all elements where the text value contains the word "Mercedes"

HTML Parser: Advanced Search with RegEx

RegEx Search by Attribute Value or Text Value

HTML

”” \$html\$ (x)

RegEx

”” (\W*(mercedes)\W) (x)

Attribute Name (optional)

”” (x)

Query Type

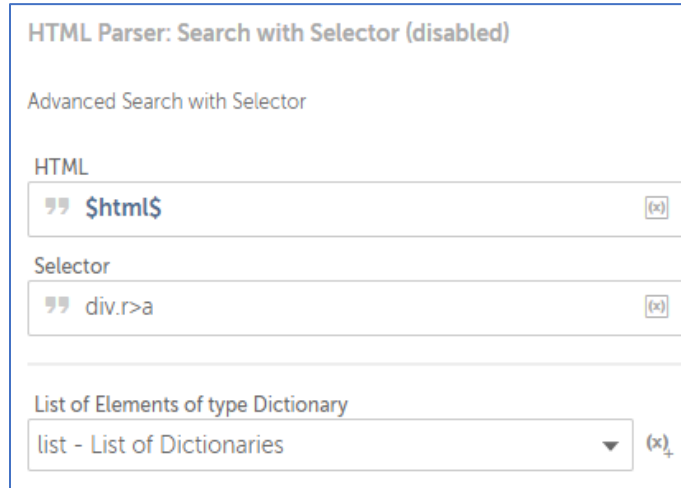
By Text Value ▼

List of Elements of type Dictionary

list - List of Dictionaries ▼ (x)

2. Find all elements that are `<a>` elements and have a `<div>` with `class="r"` element as a parent like

```
▼ <div class="r">  
  ▼ <a href="https://www.mercedes-amg.com/de.html"
```



The screenshot shows a web interface titled "HTML Parser: Search with Selector (disabled)". It features a section for "Advanced Search with Selector" with two input fields: "HTML" containing "\$html\$" and "Selector" containing "div.r>a". Below these is a "List of Elements of type Dictionary" dropdown menu showing "list - List of Dictionaries".

Some more selector expressions

- tagname: find elements by tag, e.g. `a`
- ns|tag: find elements by tag in a namespace, e.g. `fb|name` finds `<fb:name>` elements
- #id: find elements by ID, e.g. `#logo`
- .class: find elements by class name, e.g. `.masthead`
- [attribute]: elements with attribute, e.g. `[href]`
- [^attr]: elements with an attribute name prefix, e.g. `[^data-]` finds elements with HTML5 dataset attributes
- [attr=value]: elements with attribute value, e.g. `[width=500]` (also quotable, like `[data-name='launch sequence']`)
- [attr^=value], [attr\$=value], [attr*=value]: elements with attributes that start with, end with, or contain the value, e.g. `[href*/=/path/]`
- [attr~=regex]: elements with attribute values that match the regular expression; e.g. `img[src~=(?i)\.(png|jpe?g)]`
- *: all elements, e.g. `*`

You find a complete list of selectors here

<https://jsoup.org/cookbook/extracting-data/selector-syntax>

3. Find all elements that have a class=B6fmyf

```
▼ <div class="B6fmyf">
```

HTML Parser: Simple Search for Elements

Simple Search for Elements by ID, Tag, Text, Class or Attribute

HTML

 (x)

Query

 (x)

Query Type

By Class ▼

List of Elements of type Dictionary

list - List of Dictionaries ▼ (x)

4. Support & FAQs

4.1 Support

Free bots are not officially supported. You can get access to Community Support through the following channels:

- You can get access to Community Support, connecting with other Automation Anywhere customers and developers on [APeople](#) – the [Bot Building Forum](#), the [Bot Store Support Forum](#), or the [Developers Everywhere Group](#).
- Automation Anywhere also provides a [Product Documentation portal](#) which can be accessed for more information about our products and guidance on [Enterprise A2019](#).

4.2 FAQs

For questions relating to Enterprise A2019: See the [Enterprise A2019 FAQs](#).

Appendix A: Record of Changes

Instructions: Provide information on the version number, the date of the version, the author/owner of the version, and a brief description of the reason for creating the revised version.

No.	Version Number	Date of Change	Author	Notes
1	1.0.1	28/07/2020	Stefan Karsten	Initial version
2	1.0.3	29/07/2020	Stefan Karsten	Added HtmlUnit

Appendix B: References

No.	Topic	Reference Link
1	Overview of Enterprise A2019	Click here
2	Guidance: Building basic A2019 bots	Click here
3	Guidance: Building A2019 action packages	Click here
4	APeople Community Forum	Click here
5	Automation Anywhere University	Click here