

## Reference

- This example is from the following link
- <https://www.pluralsight.com/guides/web-scraping-with-beautiful-soup>  
(<https://www.pluralsight.com/guides/web-scraping-with-beautiful-soup>)
- Added some notes and made a few modifications

```
In [18]: ▶ 1  #!/usr/bin/env python3
          2  # -*- coding: utf-8 -*-
          3  #Created on Thu Sep  6 11:17:11 2018
          4
          5  #@author: kerry
          6
          7  #import libraries
          8
          9  from bs4 import BeautifulSoup
         10  import requests
         11  import csv
```

- Create a BeautifulSoup object and define the parser.

```
In [19]: ▶ 1  URL = 'https://en.wikipedia.org/wiki/List_of_game_engines'
          2  content = requests.get(URL)
          3  soup = BeautifulSoup(content.text, 'html.parser')
```

- BeautifulSoup can extract single or multiple occurrences of a specific tag and can also accept search criteria based on attributes such as:
- Find: This function takes the name of the tag as string input and returns the first found match of the particular tag from the webpage response as:

```
In [20]: 1 row = soup.find('tr') # Extract and return first occurrence of tr
2 print(row) # Print row with HTML formatting
3 print("=====Text Result=====")
4 print(row.get_text()) # Print row as text
```

```
<tr>
<th style="width: 12em">Name
</th>
<th>Primary <a href="/wiki/Programming_language" title="Programming language">programming language</a>
</th>
<th><a href="/wiki/Scripting_language" title="Scripting language">Scripting
</a>
</th>
<th><a class="mw-redirect" href="/wiki/Cross-platform" title="Cross-platform">Cross-platform</a>
</th>
<th>2D/3D oriented
</th>
<th>Target <a href="/wiki/Computing_platform" title="Computing platform">platform</a>
</th>
<th>Notable games
</th>
<th>License
</th>
<th class="unsortable">Notes and references
</th></tr>
=====Text Result=====
```

Name

Primary programming language

Scripting

Cross-platform

2D/3D oriented

Target platform

Notable games

License

Notes and references

In [21]:

```

1 #Findall: Use find_all to extract all the occurrences of a particular ta
2 rows = soup.find_all('tr')
3 for row in rows:          # Print all occurrences
4     print(row.get_text())

```

Cross-platform

2D/3D oriented

Target platform

Notable games

License

Notes and references

4A Engine

C++

Yes

- find\_all returns an object of ResultSet which offers index based access to the result of found occurrences and can be printed using a for loop.
- Pass List: find\_all can accept a list of tags as soup.find\_all(['th', 'td']) and parameters like id to find tags with unique id and href to process tags with href attribute as:

In [26]:

```

1 content = requests.get(URL)
2 soup = BeautifulSoup(content.text, 'html.parser')
3 tags = soup.find_all(id = True, href = True)

```

- Pass Function: A function can contain your customized logic to validate the tag and can be used as:

In [28]:

```

1 def isAnchorTagWithLargeText(tag):
2     """ Validate the anchor tag and should have text length greater than
3
4     return True if tag.name == 'a' and len(tag.get_text()) > 50 else Fal

```

```
In [29]: 1 content = requests.get(URL)
2 soup = BeautifulSoup(content.text, 'html.parser')
3 tags = soup.find_all(isAnchorTagWithLargeText, limit = 10)
4 for tag in tags:
5     print(tag.get_text())
```

advanced: lighting, shadows, interactive GUI surfaces

Star Wars: Knights of the Old Republic II: The Sith Lords

"A Gentle Introduction to Frogatto Formula Language"

"Artifact will use Source 2, bringing the engine to iOS and Android"

[https://en.wikipedia.org/w/index.php?title=List\\_of\\_game\\_engines&oldid=954338549](https://en.wikipedia.org/w/index.php?title=List_of_game_engines&oldid=954338549) ([https://en.wikipedia.org/w/index.php?title=List\\_of\\_game\\_engines&oldid=954338549](https://en.wikipedia.org/w/index.php?title=List_of_game_engines&oldid=954338549))

- Attribute Driven Search: The result of find\_all function can also contain Rows from other tables
- Unwanted values These are not desired most of the time. So, attributes like id, class, or value are used to further refine the search.
- Let's print the first found table (content table) to identify the attributes as:

```
In [30]: 1 table = soup.find_all('table')
2 print(table)
```

```
[<table class="wikitable sortable" style="text-align: center; font-size: 85%; width: auto; table-layout: fixed;">
<tbody><tr>
<th style="width: 12em">Name
</th>
<th>Primary <a href="/wiki/Programming_language" title="Programming language">programming language</a>
</th>
<th><a href="/wiki/Scripting_language" title="Scripting language">Scripting</a>
</th>
<th><a class="mw-redirect" href="/wiki/Cross-platform" title="Cross-platform">Cross-platform</a>
</th>
<th>2D/3D oriented
</th>
<th>Target <a href="/wiki/Computing_platform" title="Computing platform">platform</a>
</th>
<th>Notable games
```

- The content table has a unique CSS class attribute i.e. wikitable sortable which can be used to find the main content table as:

```
In [31]: 1 contentTable = soup.find('table', { "class" : "wikitable sortable"}) #
2 rows = contentTable.find_all('tr')
3 for row in rows:
4     print(row.get_text())
```

A-Frame (VR)

HTML, JavaScript

JavaScript

Yes

3D

Cross-platform

MIT

Open source Entity component system WebVR framework

- Here find is more suitable than find\_all, since only one table has wikitable sortable class property.
- Alternatively, the \_class (not available in old versions) attribute
- can be used as soup.find\_all('table', class\_="wikitable sortable").
- Nested Tags: Nested tags can be found using the select method as:

```
In [33]: 1 print(soup.select("html head title")[0].get_text()) # List of game engines
```

List of game engines - Wikipedia

- ```
1 - BeautifulSoup also allows you to mention tags as properties to find
first occurrence of the tag as:
```

```
In [35]: 1 content = requests.get(URL)
2 soup = BeautifulSoup(content.text, 'html.parser')
3 print(soup.head, soup.title)
4 print(soup.table.tr) # Print first row of the first table
```

```
<head>
<meta charset="utf-8"/>
<title>List of game engines - Wikipedia</title>
<script>document.documentElement.className="client-js";RLCONF={"wgBreakFrames":!1,"wgSeparatorTransformTable":["",""],"wgDigitTransformTable":["",""],"wgDefaultDateFormat":"dmy","wgMonthNames":["","January","February","March","April","May","June","July","August","September","October","November","December"],"wgRequestId":"Xqyi2gpAMNYAA7SsX9cAAAV","wgCSPNonce":!1,"wgCanonicalNamespace":"","wgCanonicalSpecialPageName":!1,"wgNamespaceNumber":0,"wgPageName":"List_of_game_engines","wgTitle":"List of game engines","wgCurRevisionId":954338549,"wgRevisionId":954338549,"wgArticleId":2323909,"wgIsArticle":!0,"wgIsRedirect":!1,"wgAction":"view","wgUserName":null,"wgUserGroups":["*"],"wgCategories":["Articles with short description","Use mdy dates from June 2018","All articles with unsourced statements","Articles with unsourced statements from July 2015","Video game engines","Technology-related lists"],"wgPageContentLanguage":"en","wgPageContentModel":"wikitext","wgRelevantPageName":"List_of_game_engines","wgRelevantArticleId":2323909,"wgIsProbablyEditable":!0,"wgRelevantPageIsProbablyEditable":!0,"wgRestrictionEdit":[],"wgRestrictionMove":[],"wgMediaViewerOnClick":!0,"wgMediaViewerEnabledByDefault":!0,"wgPopupsReferencePreviews":!1,"wgPopupsConflictsWithNavPopupGadget":!1,"wgVisualEditor":{"pageLanguageCode":"en","pageLanguageDir":"ltr","pageVariantFallbacks":"en"},"wgMFDisplayWikibaseDescriptions":{"search":!0,"nearby":!0,"watchlist":!0,"tagline":!1},"wgWMESchemaEditAttemptStepOversample":!1,"wgULSCurrentAutonym":"English","wgNoticeProject":"wikipedia","wgWikibaseItemId":"Q1809332","wgCentralAuthMobileDomain":!1,"wgEditSubmitButtonLabelPublish":!0};RLSTATE={"ext.globalCssJs.user.styles":"ready","site.styles":"ready","noscript":"ready","user.styles":"ready","ext.globalCssJs.user":"ready","user":"ready","user.options":"loading","ext.cite.styles":"ready","jquery.tablesorter.styles":"ready","jquery.makeCollapsible.styles":"ready","skins.vector.styles.legacy":"ready","wikibase.client.init":"ready","ext.visualEditor.desktopArticleTarget.noscript":"ready","ext.uls.interlanguage":"ready","ext.wikimediaBadges":"ready"};RLPAGEMODULES=["ext.cite.ux-enhancements","site","mediawiki.page.startup","skins.vector.js","mediawiki.page.ready","jquery.tablesorter","jquery.makeCollapsible","ext.gadget.ReferenceToolTips","ext.gadget.charinsert","ext.gadget.refToolbar","ext.gadget.extra-toolbar-buttons","ext.gadget.switcher","ext.centralauth.centralautologin","ext.popups","ext.visualEditor.desktopArticleTarget.init","ext.visualEditor.targetLoader","ext.eventlogging","ext.wikimediaEvents","ext.navigationTiming","ext.uls.compactlinks","ext.uls.interface","ext.cx.eventlogging.campaigns","ext.quicksurveys.init","ext.centralNotice.geoIP","ext.centralNotice.startUp"];</script>
<script>(RLQ=window.RLQ||[]).push(function(){mw.loader.implement("user.options@1hzgi",function($,jQuery,require,module){/*@nomin*/mw.user.tokens.set({"patrolToken":"+\\","watchToken":"+\\","csrfToken":"+\\"});});});</script>
<link href="/w/load.php?lang=en&modules=ext.cite.styles%7Cext.uls.interlanguage%7Cext.visualEditor.desktopArticleTarget.noscript%7Cext.wikimediaBadges%7Cjquery.makeCollapsible.styles%7Cjquery.tablesorter.styles%7Cskins.vector.styles.legacy%7Cwikibase.client.init&only=styles&skin=vector" rel="stylesheet"/>
```

```

<script async="" src="/w/load.php?lang=en&modules=startup&only=scri
pts&raw=1&skin=vector"></script>
<meta content="" name="ResourceLoaderDynamicStyles"/>
<link href="/w/load.php?lang=en&modules=site.styles&only=styles&
p;skin=vector" rel="stylesheet"/>
<meta content="MediaWiki 1.35.0-wmf.28" name="generator"/>
<meta content="origin" name="referrer"/>
<meta content="origin-when-crossorigin" name="referrer"/>
<meta content="origin-when-cross-origin" name="referrer"/>
<link href="/w/index.php?title=List_of_game_engines&action=edit" rel="a
lternate" title="Edit this page" type="application/x-wiki"/>
<link href="/w/index.php?title=List_of_game_engines&action=edit" rel="e
dit" title="Edit this page"/>
<link href="/static/apple-touch/wikipedia.png" rel="apple-touch-icon"/>
<link href="/static/favicon/wikipedia.ico" rel="shortcut icon"/>
<link href="/w/opensearch_desc.php" rel="search" title="Wikipedia (en)" typ
e="application/opensearchdescription+xml"/>
<link href="//en.wikipedia.org/w/api.php?action=rsd" rel="EditURI" type="ap
plication/rsd+xml"/>
<link href="//creativecommons.org/licenses/by-sa/3.0/" rel="license"/>
<link href="/w/index.php?title=Special:RecentChanges&feed=atom" rel="al
ternate" title="Wikipedia Atom feed" type="application/atom+xml"/>
<link href="https://en.wikipedia.org/wiki/List_of_game_engines" rel="canoni
cal"/>
<link href="//login.wikimedia.org" rel="dns-prefetch"/>
<link href="//meta.wikimedia.org" rel="dns-prefetch"/>
<!--[if lt IE 9]><script src="/w/resources/lib/html5shiv/html5shiv.js"></sc
ript><![endif]-->
</head> <title>List of game engines - Wikipedia</title>
<tr>
<th style="width: 12em">Name
</th>
<th>Primary <a href="/wiki/Programming_language" title="Programming languag
e">programming language</a>
</th>
<th><a href="/wiki/Scripting_language" title="Scripting language">Scripting
</a>
</th>
<th><a class="mw-redirect" href="/wiki/Cross-platform" title="Cross-platfor
m">Cross-platform</a>
</th>
<th>2D/3D oriented
</th>
<th>Target <a href="/wiki/Computing_platform" title="Computing platform">pl
atform</a>
</th>
<th>Notable games
</th>
<th>License
</th>
<th class="unsortable">Notes and references
</th></tr>

```

In [ ]: ▶

1

