

Today I applied these alteryx tools to perform web scraping. Day 03 #60daysofUdacity



Text
Input



Download



Browse



Select



Output
Data

Lesson 01, section 16
Web-scraping solution

Rosana Santos



Alteryx Designer x64 - New Workflow1*

File Edit View Options Help

Search view

★ Favorites < In/Out Preparation Join Parse Transform In-Database Reporting Documentation Spatial Interface Data Investigation Predictive AB Testing

Browse Input Data Output Data Text Input Data Cleansing Filter Formula Sample Select Sort Join Union Text To Columns Summarize Comment

Text Input (1) - Configuration

Field1

1 https://community.alteryx.com/t5/ACE-Program/bd-p/ace

Scrape a page on the Alteryx site containing the list of Alteryx ACEs. The ACE program highlights the most influential and supportive experts of Alteryx Analytics.

Start Here.yxmd x New Workflow1* x +

Alteryx Designer x64

Finished running in 2.7 seconds

[Learn about AMP Engine](#)

☐ Don't show this message again OK

Results - Text Input

1 of 1 Fields Cell viewer 1 record displayed

Record	Field1
1	https://community.alteryx.com/t5/ACE-Program/...

Search Data Metadata Actions

Windows Taskbar: Digite aqui para pesquisar, R, File Explorer, Edge, Alteryx Designer, Mail, Task View, Chrome, LinkedIn, Alteryx, Chrome, PowerPoint, Network, Signal, POR PTB2, 21:33 12/12/2020

Alteryx Designer x64 - New Workflow1*

File Edit View Options Help

Search view

★ Favorites < In/Out Preparation Join Parse Transform In-Database Reporting Documentation Spatial Interface Data Investigation Predictive AB Testing

Browse Input Data Output Data Text Input Data Cleansing Filter Formula Sample Select Sort Join Union Text To Columns Summarize Comment

Browse (3) - Configuration

Start Here.yxmd x New Workflow1* x +

Run

Profile

DownloadData x

Summary

Type	Records	Data Type	Size
V_WString	1		1,073,741,823

● Ok 1 100.00%

Unique 1 100.00%

● Null 0 0.00%

● Not Ok 0 0.00%

● Empty 0 0.00%

Length Statistics

Min	Max	Average	Shortest Value	Longest Value	First Alphanumeric Value	Last Alphanumeric Value
219,4...	219,4...	219,4...	<IDO...	<IDO...	<IDO...	<IDO...

Results - Browse (3) - Input

3 of 3 Fields Cell Viewer 1 record displayed, 94 KB Search Data Metadata Actions

Record #1, Field DownloadData (V_WString)

```
<!DOCTYPE html> <html prefix="og: http://ogp.me/ns#" dir="ltr" lang="en" class="no-js">
<head>
<script type="text/javascript"> (window.NREUM||(NREUM={})).loader_config={licenseKey:"90ec53e80f",applicationID:"290187372";window.NREUM||(NREUM={}).__nr_require=function(e,t,r){var n=e.getEntries();n.length>0&&d("lcp",[n.length-1])}function o(e){e.getEntries().forEach(function(e){e.hadRecentInput||d("cls",[e])});function a(e){if(e instanceof m&&!g){var t=Matl hstart";y.forEach(function(e){document.addEventListener(e,a,1)}))}({d}),0,4:[function(e,t,n){function r(e,t){if(!i)return!1;if(e!==i)return!1;if(t)return!0;if(!o)return!1;for(var n=o.split(","),r=t.sp "",o=0;for(r in e).call(e,r)&&(n[o]=t(r,e[r]),o+=1);return n}var i=Object.prototype.hasOwnProperty;t.exports=r},{}],7:[function(e,t,n){function r(e,t,n){t||(t=0),"undefined"!==typeof n&&(n=e moveEventListener:m,emit:n,get:g,listeners:v,context:t,buffery:abort:a,aborted:!1);return b}function o(){return new r(function a(){(s.api||s.feature)&&(p.aborted=!0,s=p.backlog={})}var c= ort;f(h,function(t,n){e[t]||(e[t]=n)});var n=a();c("mark",["onload",n+b.offset,null,"api"],c("timing",["load",n]);var r=p.createElement("script");r.src="https://" + e.agent.t.parentNode.insertBefore(r,"wrap-function":function(e,t,n){function r(e){return!(e&&e instanceof Function&&e.apply&&!e[a])}var i=e("ee"),o=e(7),a="nr@original",c=Object.prototype.hasOwnProperty,f=!1,t.ex try{e.emit("internal-error",t)}catch(n){return e||(e=i),n.inPlace=u,n.flag=a,n},{}],0,["loader"]);</script>
</head>
<body>
</body>
</html>
```

Record	Field1	DownloadData	DownloadHeaders
1	https://community.alteryx.com/t5/ACE-Program/...	<!DOCTYPE html> <html prefix="og: http://ogp....	HTTP/1.1 200 OK

Contains the entire webpage

Windows Taskbar: Digite aqui para pesquisar, R, File Explorer, Edge, Task Manager, Mail, Chrome, LinkedIn, Alteryx, Chrome, PowerPoint, Network, Time: 21:37 12/12/2020

Alteryx Designer x64 - New Workflow1*

view

FileEditViewOptionsHelp

FavoritesIn/OutPreparationJoinParseTransformIn-DatabaseReportingDocumentationSpatialInterfaceData InvestigationPredictiveAB Testing

BrowseInput DataOutput DataText InputData CleansingFilterFormulaSampleSelectSortJoinUnionText To ColumnsSummarizeComment

Workflow - Configuration

CanvasWorkflowRuntimeEventsMeta Info

Canvas Options

Layout Direction:Horizontal

Annotations:Show

Connection Progress:Show Only When Running

Start Here.yxmdNew Workflow1*

»↺▶Run

Alteryx Designer x64

Finished running in 3.1 seconds with 1 field conversion error

[Learn about AMP Engine](#)

☐ Don't show this message againOK

Results - Workflow - Messages

All0 Errors1 Conv Errors0 Warnings2 Messages2 FilesLast RunConfiguration

Text Input (1)

Download (2)

Browse (3)

Output Data (5)

Output Data (5)

Designer x64

1 record was output

Records transferred: 1; active connections: 1; average records per minute: 28

1 records

DownloadData: "<!DOCTYPE html><html prefix="og: http://ogp.me/ns#" dir="ltr" lang="en" class="no-js"><head><script type="text/javascript">(windo

1 records were written to "C:\Users\Rosana\Desktop\Udacity\ Creating an Analytical Dataset\01 Web scraping, WOW moment.csv"

Finished running in 3.1 seconds with 1 field conversion error

Windows Taskbar

Windows Start ButtonSearch Bar (Digite aqui para pesquisar)Taskbar Icons (Alteryx, File Explorer, Edge, etc.)System Tray (Network, Volume, Date/Time: 22:01 12/12/2020)

Salvamento Automático 01 Web scraping, WOW moment.csv Pesquisar Rosana Ferreira Soares dos Santos

Arquivo Página Inicial Inserir Desenhar Layout da Página Fórmulas Dados Revisão Exibir Desenvolvedor Ajuda Compartilhar Comentários

Colar Área de Transferência Fonte Alinhamento Número Estilos Células Edição

A1 <!DOCTYPE html><html prefix="og: http://ogp.me/ns#" dir="ltr" lang="en" class="no-js">

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
193	"contest.entries_per_page_num": 20,																				
194	"layout.messages_per_page_linear": 30,																				
195	"integratedprofile.cta_manage_topics_dismissal_timestamp": -1,																				
196	"profile.shared_profile_test_group": false,																				
197	"integratedprofile.cta_personalized_feed_dismissal_timestamp": -1,																				
198	"integratedprofile.curated_feed_size": 5,																				
199	"contest.one_kudo_per_contest": false,																				
200	"integratedprofile.enable_social_networks": false,																				
201	"integratedprofile.my_interests_dismissal_timestamp": -1,																				
202	"profile.language": "en",																				
203	"layout.friendly_dates_max_age_days": 31,																				
204	"layout.threading_order": "thread_descending",																				
205	"blog.toc_heading_list_style": "disc",																				
206	"useRecService": false,																				
207	"layout.module_welcome": "",																				
208	"imageupload.max_uploaded_images_per_upload": 100,																				
209	"imageupload.max_uploaded_images_per_user": 2000,																				
210	"integratedprofile.connect_mode": "",																				
211	"tkb.toc_maximum_heading_level": "2",																				
212	"tkb.toc_heading_list_style": "disc",																				
213	"sharedprofile.show_hoverscard_score": true,																				

01 Web scraping, WOW moment

Pronto Digite aqui para pesquisar

POR PTB2 22:04 12/12/2020

That was my first attempt.
the result is not similar to that presented in
the solution. But I'll try again tomorrow!!!

Learning is

Powerful

8:-)