



LMGTFY – WEB SCRAPING OP BASIS VAN STREEPJESCODE

MARC BOES – DIWUG – 16 FEBRUARI 2021

KORTE INTRODUCTIE

ZAKELIJK CIRKELTJE

- OOIIT BEGONNEN ALS CITIZEN DEVELOPER AVANT LA LETTRE (MET KLEITABLET, EXCEL 3.0, PYPYRUS, ACCESS 1.1 EN WORD 2.0)
- EIND 20^E EEUW MS OFFICE AFDELING GESTART BIJ BANK: DESKTOP SERVICES
- MID JAREN NUL EN TIEN MET SHAREPOINT AAN DE SLAG BIJ JSR, WORTELL, MOTION10
- LAATSTE JAREN M365 BIJ EEN ANDERE BANK ALS VP DESKTOP SERVICES

PRIVÉ

- WOONACHTIG IN ROTTERDAM MET VRIENDIN EN KAT
- GROTE KAT OP EERSTE SLIDE GEADOpteERD BIJ STICHTING LEEUW
- HOBBY'S: M365 EN WHISKY VERZAMELEN
- SOCIAL MEDIA: ZO WEINIG MOGELIJK



OVERZICHT VAN DE PRESENTATIE

- De aanleiding:
 - Een vrij uitgebreide whiskyverzameling,
 - een M365 Business Standard abonnement,
 - een Azure Pay-As-You-Go Dev/Test abonnement,
 - een prachtige whiskygegevensbron Whiskybase stelt geen data feeds beschikbaar.
- Het doel van de oplossing en het onderwerp van deze presentatie:
 - een zwaluwstaartje tussen de barcode scanner in Power Apps en pagina's van Whiskybase.
- De agenda:
 - De uitdaging, oftewel de handmatige acties die geautomatiseerd gaan worden.
 - De puzzelstukjes, oftewel de componentjes en de links hoe je dit inregelt.
 - Nog even concreet, oftewel de specifieke details van deze oplossing: 30'' demo, 5' Function App, 2' AAD protection, 7' connector, 2' Power Automate
 - Afsluiting.
- De presentatie met de links en de zelf gebrouwen code is als PDF hier te downloaden

DE UITDAGING

Mobile App met een knop die

- Barcode scant,
- de code opzoekt op Whiskybase,
- de gevonden pagina uitleest,
- de details in het invulscherm plaatst.



Save Save & Close + New Deactivate Delete Refresh Assign

Ardbeg Alligator

Whisky Collection

General Related

Name	Ardbeg Alligator	Bottling serie	---
Barcode	5010494904185	Bottled	13.07.2011
Bottle status	Open 2/4	Cask type	New American White ...
Category	Single Malt	Stated age	---
Distillery	Ardbeg	Bottle code	L11 194
Bottler	Distillery Bottling	Notes	---
Owner	Marc Boes (DTS)	Number of bottles	---
		Size	700 ml
		Strength	51.2 % Vol.
		Whiskybase URL	https://www.whi...

Whiskies Explore Contribute Market Shop 5010494904185

Search

5010494904185 Search

2 Whiskies

Bottler Brand Vinta... Bottled For sa... Rating

#	Name	Stated Age	Strength	Size	Bottled	Casknumber	Barcode	Rating	M
1	Ardbeg Alligator		51.2 % Vol.	700 ml	2011		5010494904185	88.46	
2	Ardbeg Alligator Untamed Release		51.2 % Vol.	700 ml	2011		5010494904185	88.53	

Whiskies Explore Contribute Market Shop Search the database...

Distilleries Brands Bottlers New releases Vintage Reviews & Notes Toplists

Ardbeg Alligator

OVERALL RATING 28 VOTES

88.46/100

My rating 0 Save

Category Single Malt

Distillery Ardbeg

Bottler Distillery Bottling

Bottled 13.07.2011


Casktype New American White Oak Barrels

Strength 51.2 % Vol.

Size 700 ml

Bottle code L11 194

Barcode 5010494904185



DE PUZZELSTUKJES

- Iets met een barcode scanner dat processen starten: Power Apps Mobile
- Iets dat pagina's kan scrapen: Azure Functions (in JavaScript, Python, C#, PowerShell etc.)
- Een koppeling tussen Azure Functions en Power Platform: Custom Connectors
- Lijm tussen custom connector en Power Apps: Power Automate

Hoe doe je dat en wat heb je nodig? Let me google that for you!

- Kosten: voor mijn type tenant en Azure abonnement ben je samen rond de EUR20/m kwijt, je kunt een gratis trial nemen.
- De Azure kant bestaat uit een resource groep, een azure function en een storage account. Zie daarvoor [Azure naamgevingsconventies](#).
- Een nieuwe Azure Function App werkt al. Extra zijn packages om web pagina's in te lezen en te parsen: axios en cheerio voor [JavaScript](#) (requests en BeautifulSoup voor Python). Python is gaaf, maar lastiger AAD te beveiligen.
- Web scrapen is legaal, maar houd rekening met het copyright op de informatie.
- Het maken van een AAD protected custom connector is stap voor stap beschreven in [dit artikel](#).
Tip 1: daarmee kun je een premium licentie (voor de REST connector) uitstellen of zelfs overbodig maken als je het alleen voor specifieke gevallen nodig hebt.
Tip 2: als een Power Apps Portal aanmaakt, dan krijg je een echte Dataverse cadeau. Geen idee of dat zo blijft, zo niet, dan werkt het natuurlijk ook met Lists.
- Power Apps werkt niet lekker met het ontvangen van data objecten, of eigenlijk alleen met strings. Met een flow kun je die logica buiten je App houden.
- Screen recording in Android is geblokkeerd, vandaar een [emulator](#) filmpje. De barcode is een foto om de camera te simuleren.

DEMO

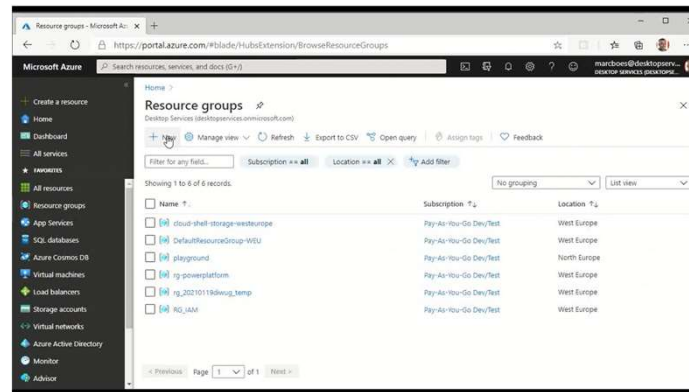
30 seconden demo van het
eindresultaat



FUNCTION APP

Vijf minuten:

- Azure Portal
 - Resource group
 - Function App
- VS Code
 - Function
 - Dependencies
- Validatie



<https://youtu.be/lpkoPyOEyIA>

WHY: Create resource group

WHERE:

<https://portal.azure.com/#blade/HubsExtension/BrowseResourceGroup>

WHAT: rg_20210216diwug

WHY: Create Function App

WHERE: <https://portal.azure.com/#create/hub>

WHAT: af-20210216diwug > rg_20210216diwug > Javascript > LTS14 > West Europe > stprowerplatform > Review and create

WHY: Create Folder

WHERE: D:\Data\2021

WHAT: 20210216DIWUG

WHY: Create Function

WHERE: VS Code

WHAT: Console > MKDIR D:\Data\2021\20210216DIWUG

WHAT: Azure > Create new Project > D:\Data\2021\20210216DIWUG

> Javascript > LTS > HTTP Trigger > diwug20210216 > Function

WHY: Insert Scraping code

WHERE: VS Code

WHAT: Copy / paste RAW code from presentation to index.js

WHY: Insert dependencies

WHERE: VS Code

WHAT: npm install --save axios cheerio

WHY: Deploy

WHERE: VS Code

WHAT: Deploy to function app > af-20210216diwug

WHY: Check

WHERE: Function App

WHAT: func-20210216diwug-temp > Copy function url > browse to url & barcode=080432402825&max=1

Resource groups - Microsoft Az...

https://portal.azure.com/#blade/HubsExtension/BrowseResourceGroups

Microsoft Azure

Search resources, services, and docs (G+/)

marcboes@desktopserv...
DESKTOP SERVICES (DESKTOPSE...

Create a resource

Home

Dashboard

All services

FAVORITES

All resources

Resource groups

App Services

SQL databases

Azure Cosmos DB

Virtual machines

Load balancers

Storage accounts

Virtual networks

Azure Active Directory

Monitor

Advisor

Home >

Resource groups

Desktop Services (desktopservices.onmicrosoft.com)

+

New

Manage view

Refresh

Export to CSV

Open query

Assign tags

Feedback

Filter for any field...

Subscription == all

Location == all

Add filter

Showing 1 to 6 of 6 records.

No grouping

List view

<input type="checkbox"/> Name	Subscription	Location	
<input type="checkbox"/> cloud-shell-storage-west europe	Pay-As-You-Go Dev/Test	West Europe	
<input type="checkbox"/> DefaultResourceGroup-WEU	Pay-As-You-Go Dev/Test	West Europe	
<input type="checkbox"/> playground	Pay-As-You-Go Dev/Test	North Europe	
<input type="checkbox"/> rg-powerplatform	Pay-As-You-Go Dev/Test	West Europe	
<input type="checkbox"/> rg_20210119diwug_temp	Pay-As-You-Go Dev/Test	West Europe	
<input type="checkbox"/> RG_IAM	Pay-As-You-Go Dev/Test	West Europe	

< Previous

Page

1

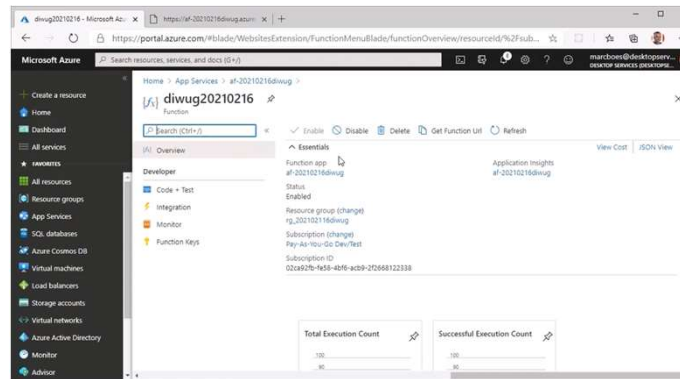
 of 1

Next >

AAD PROTECT APP

Twee minuten:

- Azure Portal
 - Protect calls to Azure Functions against anonymous access
 - Configure AAD authentication provider
- Validatie



<https://youtu.be/lpkoPyOEyIA>

WHY: Protect calls to Azure Functions against anonymous access

WHERE: Function in Azure Portal

WHAT: diwug20210216 > Authentication / Authorization > Turn on App Service Authentication > Login with AAD

WHY: Configure AAD authentication provider

WHERE: Function in Azure Portal, Authentication / Authorization

WHAT: Azure Active Directory > Express > OK > Save <refresh browser and see 'Express: Existing App'>

WHY: Checking the configuration

WHERE: anonymous browser session

WHAT: browse to url &barcode=080432402825&max=1

diwug20210216 - Microsoft Azure

https://af-20210216diwug.azure...

https://portal.azure.com/#blade/WebsitesExtension/FunctionMenuBlade/functionOverview/resourceId/%2Fsub...

Microsoft Azure

Search resources, services, and docs (G+)

marcboes@desktopserv...
DESKTOP SERVICES (DESKTOPSE...

Create a resource

Home

Dashboard

All services

FAVORITES

All resources

Resource groups

App Services

SQL databases

Azure Cosmos DB

Virtual machines

Load balancers

Storage accounts

Virtual networks

Azure Active Directory

Monitor

Advisor

Home > App Services > af-20210216diwug >

diwug20210216

Function

Search (Ctrl+)

Overview

Developer

Code + Test

Integration

Monitor

Function Keys

Enable

Disable

Delete

Get Function Url

Refresh

Essentials

Function app
af-20210216diwug

Status
Enabled

Resource group (change)
rg_20210216diwug

Subscription (change)
Pay-As-You-Go Dev/Test

Subscription ID
02ca92fb-fe58-4bf6-acb9-2f2668122338

Application Insights
af-20210216diwug

View Cost

JSON View

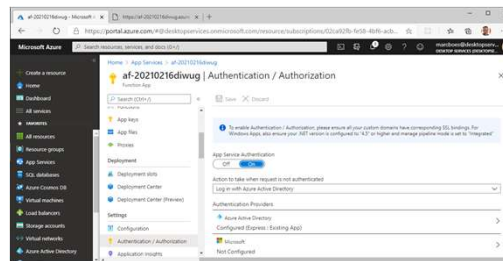
Total Execution Count

Successful Execution Count

CUSTOM CONNECTOR

Zeven minuten:

- Azure Portal
 - Create app registration for your Custom Connector in AAD
 - Get app registration secret for your Custom Connector in AAD
- Power Apps
 - Custom Connectors
- Validate



<https://youtu.be/hKkOKI3Ab34>

WHY: Create app registration for your Custom Connector in AAD

WHERE:

https://portal.azure.com/#blade/Microsoft_AAD_RegisteredApps/ApplicationsListBlade

WHAT: New Registration > af-20210216diwug-conn > Register > API permissions > Add APIs my organization uses > af-20210216diwug > Add user_impersonation

WHY: Get app registration secret for your Custom Connector in AAD

WHERE:

https://portal.azure.com/#blade/Microsoft_AAD_RegisteredApps/ApplicationsListBlade

WHAT: Certificates & secrets > New client secret > make note of value and client ID in overview.

Note: just to be sure, grant admin consent under API permissions

In Power App Studio, Data, Custom Connectors, create new

General

name: af-20210216diwug

logo: D:\Data\Desktop Services\Knowledge Base - Documents\Resources\Logo-Misc

description: DIWUG Demo

host: af-20210216diwug.azurewebsites.net

Security

OAuth2.0 > Azure Active Directory > Client ID > Client Secret >

<https://af-20210216diwug.azurewebsites.net>

Copy redirect url to connection app registration

<https://global.consent.azure-apim.net/redirect>

Definition

> New Action > DIWUG Demo > DIWUG learn more > DIWUG20210216 > Request > Import from <https://af-20210216diwug.azurewebsites.net/api/diwug20210216?code=...&barcode=080432402825&max=1>

> Request Query: modify the three parameters (default value code, barcode type and required, max required)

=> update before test

Test

WHY: Check that the connector works

af-20210216diwug - Microsoft A x | https://af-20210216diwug.azure... x | +

← → ↺ https://portal.azure.com/#@desktopservices.onmicrosoft.com/resource/subscriptions/02ca92fb-fe58-4bf6-acb... ☆ | ☆ | ☆ | [Profile] ...

Microsoft Azure Search resources, services, and docs (G+/)

marcboes@desktopserv...
DESKTOP SERVICES (DESKTOPSE...

Home > af-20210216diwug

af-20210216diwug | Authentication / Authorization

Function App

Search (Ctrl+/) Save Discard

Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Security
Events (preview)

Functions
Functions
App keys
App files
Proxies

Deployment

To enable Authentication / Authorization, please ensure all your custom domains have corresponding SSL bindings. For Windows Apps, also ensure your .NET version is configured to "4.5" or higher and manage pipeline mode is set to "Integrated"

App Service Authentication
Off On

Action to take when request is not authenticated
Log in with Azure Active Directory

Authentication Providers

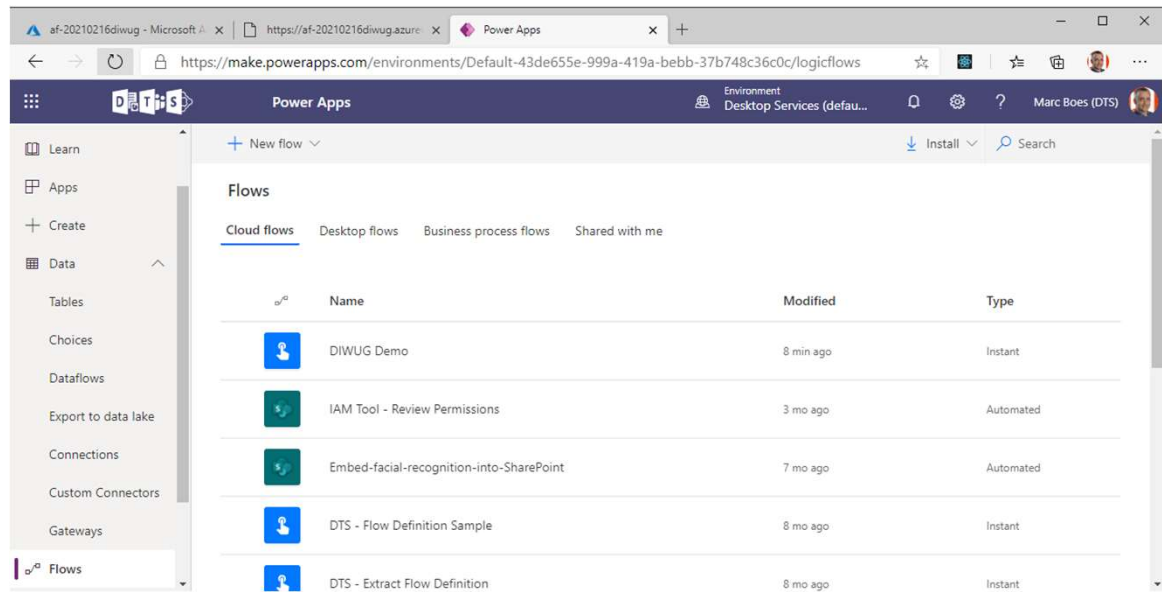
Azure Active Directory	>
Configured (Express : Existing App)	
Microsoft	>
Not Configured	
Facebook	>
Not Configured	

https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/reso...

POWER PLATFORM

Two minuten:

- Power Automate
- Github
 - Find this presentation and the VS code here:
[MJBoes/DIWUG20210216: Presentation Materials \(github.com\)](https://github.com/MJBoes/DIWUG20210216-Presentation-Materials)



<https://youtu.be/Gs2XNTQJTJs>

Power Automate

WHY: simple barcode > details validation

WHERE: <https://make.powerapps.com/>

WHAT: new manual flow

af-20210216diwug - Microsoft A x | https://af-20210216diwug.azure: x Power Apps x +

https://make.powerapps.com/environments/Default-43de655e-999a-419a-bebb-37b748c36c0c/logicflows






Power Apps Environment Desktop Services (defau... Marc Boes (DTS)

+ New flow

Learn Apps Create Data Tables Choices Dataflows Export to data lake Connections Custom Connectors Gateways Flows

Flows

Cloud flows Desktop flows Business process flows Shared with me

	Name	Modified	Type
	IAM Tool - Review Permissions	3 mo ago	Automated
	Embed-facial-recognition-into-SharePoint	7 mo ago	Automated
	DTS - Flow Definition Sample	8 mo ago	Instant
	DTS - Extract Flow Definition	8 mo ago	Instant
	SharePoint - Scheduled Document Publishing	8 mo ago	Scheduled

DIWUG 20210216 -.....mp4 ...

Open file

Show all



DANK VOOR DE AANDACHT

EN ALS ER VRAGEN ZIJN & TIJD OVER, DAN KAN DAT NU



JAVASCRIPT – INDEX.JS

```
const axios = require("axios");
const cheerio = require("cheerio");
module.exports = async function (context, req) {
  context.log('JavaScript HTTP trigger function processed a request.');
```

```
  if(!req.query.barcode||!req.query.max){
    context.res = { status:400,body:"Please pass a barcode and max returned"}
  }else{
    icount=1;
    lines=[];
    context.log('https://www.whiskybase.com/search?q='+req.query.barcode);
    const searchpage=await axios.get('https://www.whiskybase.com/search?q='+req.query.barcode
  ).then(result=>cheerio.load(result.data));
    while(icount<=Math.min(req.query.max,searchpage('.clickable').length)){
      line={url:searchpage('.clickable')[icount-1].attribs['href']};
      detailspage=await axios.get(line.url).then(result=>cheerio.load(result.data));
      line['name']=detailspage('header h1').text().toString().replace(/\t/g, ' ').replace(/\n/g, ' ').trim();
      detailspage('.block-desc dl').each((i,element)=>{
        dd='';dt=''
        detailspage(element.childNodes).each((i2,elem)=>{
          if(elem.name=='dt'){dt=detailspage(elem).text()}
          if(elem.name=='dd'){
            dd=detailspage(elem).text()
            line[dt]=dd;
          }
        })
      });
      icount+=1;
      lines.push(line);
    }
  }
  const responseMessage = JSON.stringify(lines);
  context.res = {
    // status: 200, /* Defaults to 200 */
    body: responseMessage
  };
}
```


JAVASCRIPT – INDEX.JS (RAW COPY / PASTE VERSION)

```
const axios = require("axios");
const cheerio = require("cheerio");

module.exports = async function (context, req) {
  context.log("JavaScript HTTP trigger function processed a request.");
  if(!req.query.barcode || !req.query.max){
    context.res = { status:400,body:"Please pass a barcode and max returned"}
  }else{
    icount=1;
    lines=[];
    context.log('https://www.whiskybase.com/search?q='+req.query.barcode);
    const searchpage=await
    axios.get('https://www.whiskybase.com/search?q='+req.query.barcode).then(result=>cheerio.load(result.data));
    while(icount<=Math.min(req.query.max,searchpage('.clickable').length)){
      line={url:searchpage('.clickable')[icount-1].attribs['href']};
      detailspage=await axios.get(line.url).then(result=>cheerio.load(result.data));
      line['name']=detailspage('header h1').text().toString().replace(/\t/g, "").replace(/\n/g, ' ').trim();
      detailspage('.block-desc dl').each((i,element)=>{
        dd="";dt=""
        detailspage(element.childNodes).each((i2,elem)=>{
          if(elem.name=='dt')(dt=detailspage(elem).text())
          if(elem.name=='dd'){
            dd=detailspage(elem).text()
            line[dt]=dd;
          }
        })
      });
      icount+=1;
      lines.push(line);
    };
    const responseMessage = JSON.stringify(lines);
    context.res = {
      // status: 200, /* Defaults to 200 */
      body: responseMessage
    };
  }
}
```

PYTHON - __INIT__.PY

```
import logging
import requests
from bs4 import BeautifulSoup

import azure.functions as func

def main(req: func.HttpRequest) -> func.HttpResponse:
    logging.info('Python HTTP trigger function processed a request.')

    barcode = req.params.get('barcode')
    if not barcode:
        try:
            req_body = req.get_json()
        except ValueError:
            pass
        else:
            barcode=req_body.get('barcode')
    if barcode:
        # Scrape the search result page for the details page url.
        wb_list_url=f"https://www.whiskybase.com/search?q={barcode}"
        wb_list = requests.get(wb_list_url, headers={'User-Agent': 'Mozilla/5.0'}).content
        ItemURL=BeautifulSoup(wb_list, 'html.parser').find_all('a',class_='clickable')[0]
        # Scrape the details page
        page = requests.get(ItemURL['href'], headers={'User-Agent': 'Mozilla/5.0'})
        itemPage=BeautifulSoup(page.content, 'html.parser')
        wbname=' '.join(itemPage.find('h1').text.split())
        keys=itemPage.find('div', class_='block-desc').find_all('dt')
        values=itemPage.find('div', class_='block-desc').find_all('dd')

        # Build the return object
        json_record={'Link':ItemURL['href'],'Name':wbname}
        i=0
        while i < len(keys):
            json_record[keys[i].text]=values[i].text
            i+=1
        return func.HttpResponse(str(json_record))
    else:
        return func.HttpResponse(
            "This HTTP triggered function executed successfully. Pass a parameter in the query string or in the request body.",
            status_code=200
        )
```

PYTHON - __INIT__.PY (RAW COPY / PASTE VERSION)

```
import logging
import requests
from bs4 import BeautifulSoup

import azure.functions as func

def main(req: func.HttpRequest) -> func.HttpResponse:
    logging.info('Python HTTP trigger function processed a request.')

    barcode = req.params.get('barcode')
    if not barcode:
        try:
            req_body = req.get_json()
        except ValueError:
            pass
        else:
            barcode=req_body.get('barcode')
    if barcode:
        # Scrape the search result page for the details page url.
        wb_list_url=f"https://www.whiskybase.com/search?q={barcode}"
        wb_list = requests.get(wb_list_url, headers={'User-Agent': 'Mozilla/5.0'}).content
        ItemURL=BeautifulSoup(wb_list, 'html.parser').find_all('a',class_='clickable')[0]
        # Scrape the details page
        page = requests.get(ItemURL['href'], headers={'User-Agent': 'Mozilla/5.0'})
        itemPage=BeautifulSoup(page.content, 'html.parser')
        wbname=' '.join(itemPage.find('h1').text.split())
        keys=itemPage.find('div', class_='block-desc').find_all('dt')
        values=itemPage.find('div', class_='block-desc').find_all('dd')

        # Build the return object
        json_record={'Link':ItemURL['href'],'Name':wbname}
        i=0
        while i < len(keys):
            json_record[keys[i].text]=values[i].text
            i+=1
        return func.HttpResponse(str(json_record))
    else:
        return func.HttpResponse(
            "This HTTP triggered function executed successfully. Pass a parameter in the query string or in the request body.",
            status_code=200
        )
```