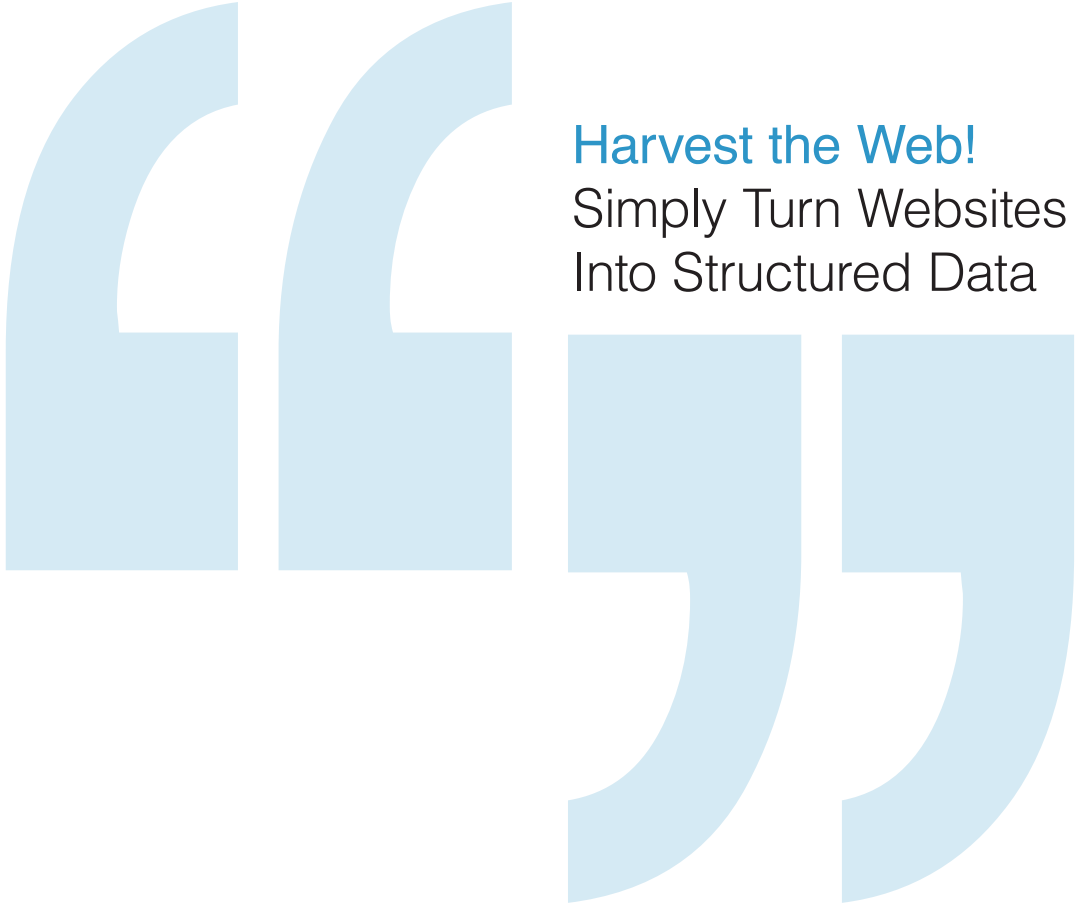




Handleiding **OutWit Hub**

Bjarn Saes, Melysha Val,
Tijanni van Groeningen & Tanja Visser

Datavisualisatie & Infographics - Tutorialdag 1



Harvest the Web!
Simply Turn Websites
Into Structured Data

Introductie

OutWit Hub is een handige tool om grote hoeveelheden informatie in je eigen databases of spreadsheets te krijgen. Met behulp van dit programma hoef je vrijwel niets handmatig te kopiëren, te plakken of over te schrijven. Mocht je dus bepaalde data willen analyseren, kan OutWit Hub je gigantisch veel tijd besparen.

OutWit Hub is een web scraper. Dat houdt in dat de software de achterliggende HTML-code van een website scant om informatie te achterhalen. Dergelijke software kan in meerdere situaties handig zijn. Stel dat je de prijzen van alle soorten Nespresso capsules van de Jumbo wil vergelijken met die van de Albert Heijn. Met OutWit Hub is dit slechts een kwestie van de juiste optie kiezen om de informatie te verzamelen en vervolgens te exporteren in het format dat jouw voorkeur geniet. Door de prijzen van al deze producten in (bijvoorbeeld) Excel te plaatsen, is prijsvergelijking een koud kunstje. Wanneer je alle prijzen handmatig zou moeten kopiëren en plakken zou het een veelvoud van je tijd kosten.

Het programma werd in 2010 ontwikkeld door OutWit Technologies. En alhoewel de laatste update dateert uit 2020, lijkt het programma aan vernieuwing toe te zijn. De user interface oogt alsof het 15 jaar geleden ontworpen is, maar ook de gebruiksvriendelijkheid laat op sommige momenten te wensen over. Zo zullen gebruikers met een Apple-product tegen verschillende problemen aanlopen. Zo is belangrijke informatie met regelmaat onzichtbaar en wil de applicatie voor sommigen helemaal niet opstarten. Mocht dit het geval zijn, probeer dan een eerdere versie te downloaden.

Desondanks heeft de digitale toepassing van OutWit Technologies een aantal zeer handige functies voor wie het programma wel naar behoren werkt. Ter herinnering, het doel van OutWit Hub is om zoveel mogelijk informatie in één keer van een informatiebron te 'scrapen'. Daar hebben de ontwikkelaars van het programma een aantal praktische functies voor in het leven geroepen. In deze handleiding zullen we een aantal van die opties uitvoerig uitlichten. Helaas wordt niet alles in detail besproken. Verschillende functies zoals de woordzoeker en de scraper voor documenten zijn namelijk alleen beschikbaar bij de betaalde versie die een slordige 700 euro kost.

Wat wel uitgebreid aan bod komt, zijn de functies voor links, afbeeldingen, contacten, tabellen, lijsten, tekst en meer! Zo leggen we natuurlijk uit hoe deze toepassingen werken, maar ook waar ze goed voor zijn en wanneer je ze kunt gebruiken. Lees deze handleiding aandachtig door en wordt een OutWit Hub expert.

Inhoud

6 Hoofdstuk 1 Leerdoelen

10 Hoofdstuk 2 Installeren

12 Hoofdstuk 3 Uitleg

12	3.1	Links
12	3.2	Images
14	3.3	Contacts
16	3.4.1	Data Tables
18	3.4.2	Data Lists
20	3.4.3	Data Guess
21	3.4.4	Data Scraped
22	3.5	Text
22	3.5.1	Text Words
23	3.5.2	Text News
24	3.5.3	Text Source
25	3.6	Automators
25	3.6.1	Automators Scrapers
26	3.7	History
28	3.8	Scraper

32 Hoofdstuk 4 Literatuurlijst

33 Hoofdstuk 5 Bijlagen

10	5.1	Bijlagen	Gasrijzen opdracht
12	5.2	Bijlagen	Gasrijzen (Digitale handout)
14	5.3	Bijlagen	Webpagina naar Excel (Digitale handout)

Hoofdstuk 1

Leerdoelen

In deze handleiding ga je alles leren over het gebruik van de gratis versie OutWit Hub. In het volgende hoofdstuk staat stapsgewijs uitgelegd hoe je de gratis versie van OutWit Hub kan installeren op Mac en Windows.

Een scraper zoals bijvoorbeeld OutWit Hub gebruik je voor het verzamelen van data van een webpagina. Maar wanneer gebruik je wel een scraper en wanneer niet?

Een scraper gebruik je zodra je informatie wilt verzamelen van een webpagina waarbij je direct inzicht krijgt in diverse gegevens. Dit doe je bijvoorbeeld zodra het te lastig is om dit handmatig uit te zoeken (bijvoorbeeld als er teveel gegevens zijn). Echter maak je alleen gebruik van een scraper zodra dit nodig is i.v.m. te veel gegevens en zodra je toestemming hebt om een webpagina te scrapen. Hieronder staat alles over het juridisch grijs gebied van scrapen.

Juridisch grijs gebied

Scrapen is echter soms grijs gebied. Per land kan het zelfs verschillen of het legaal of illegaal is. Als er scrapers bezig zijn met het ophalen van informatie van jouw website, dan kan het zomaar gebeuren dat de server overbelast raakt. Er wordt tijdens het scrapen namelijk gebruik gemaakt van automatische scripts. Dit betekent dat de website meerdere keren per seconde bezocht wordt. Dit is vele malen sneller dan een mens ooit kan klikken en dit is de server niet gewent. Mocht de host ontdekken dat dit vaak voorkomt, dan kan de host van de website de persoon opsporen (indien er gebruik wordt gemaakt van hetzelfde IP adres) en deze gebruiker als gevolg blokkeren.

Terugkomend op waarom scrapen grijs gebied kan zijn. Als je bezig bent met scrapen dan verzamel je data (in de meeste gevallen van anderen) en in sommige gevallen kan dit als eigendom gezien worden. Dit heeft het gevolg dat het beschouwd kan worden als onrechtmatig gebruik van informatie. Er bestaat zoiets als een databankenrecht en dit kan gaan over een verzameling van data. Zodra er voldoende is geïnvesteerd in de opbouw en onderhoud van een databank (informatie op een webpagina) kan dit gelden. Een webpagina met huizen die te koop staan is een voorbeeld van een beschermde databank, maar een website met nieuwsberichten is bijvoorbeeld niet beschermd.

Het advies dat je mee moet nemen in om van tevoren te controleren of de data die jij wilt scrapen vrij te gebruiken is. Dit kan je vaak vinden in de gebruikersvoorwaarden van een website en anders moet er eerst toestemming gevraagd worden aan de host voordat je kan beginnen (Breinstein, 2021).

Het leerdoel van deze handleiding over OutWit Hub in het algemeen is om kennis te krijgen over de basisfuncties binnen deze scraper tool.

3.1 Links

In het tabblad "links" vind je een overzicht van alle links binnen een webpagina, om deze vervolgens te filteren, ordenen en verzamelen om aan te werken bij het gebruik van de andere tools die OutWit Hub te bieden heeft.

Leerdoel: Het filteren en sorteren van de lijst met alle links binnen de betreffende webpagina.

3.2 Images

Images is een tool binnen OutWit Hub die je kan gebruiken om een overzicht te krijgen van alle afbeeldingen op een webpagina.

Leerdoel: Een overzicht van alle afbeeldingen op een webpagina krijgen, voor het makkelijk selecteren en downloaden van de gewenste afbeeldingen.

3.3 Contacts

De functie contacts kun je zeer goed gebruiken voor het verzamelen van contactgegevens van personen die geïnteresseerd zijn in een specifiek onderwerp/product. In het voorbeeld in hoofdstuk 3.3 wordt gebruik gemaakt van het verzamelen van contactgegevens van de concurrent met daarin geïnteresseerden.

Leerdoel: Een lijst ontwikkelen met e-mails zodat je gericht contact kan opnemen met gegarandeerd geïnteresseerden.

3.4 Data

De functie data bestaat uit vier deelfuncties. Met de functie data kan je data scrapen, filteren, sorteren en extraheren.

Tables

Tables gebruik je voor het ontwikkelen van een tabel van gegevens. In hoofdstuk 3.4.1 wordt uitgebreid uitgelegd hoe je data dat in tabelvorm staat kan binnenhalen en exporteren. Ook wordt er toegelicht hoe je de data kan filteren en sorteren.

Leerdoel: Gegevens ordelijk en inzichtelijk maken.

Lists

De functie 'Lists' is vrijwel hetzelfde is als de functie voor tabellen. Het gaat hier echter om data die weergegeven wordt in lijsten en niet in tabellen. Binnen HTML-code wordt onderscheid gemaakt tussen die twee.

Leerdoel: Gegevens ordelijk en inzichtelijk maken .

Guess

Deze functie geeft de gegevens weer uit de pagina die jij in OutWit Hub hebt gezet. Als je deze functie gebruikt zal de tool zelf een tabel maken met de geëxtraheerde data. In hoofdstuk 3.4.3 wordt toegelicht hoe je dit exact kan doen.

Leerdoel: Ontdekken wat de OutWit Hub tool kan zonder instructies van jou.

Scraped

De scraped functie is individuele functie, maar een weergave van de verzamelde gegevens in de functie 'scrapers'. In hoofdstuk 3.4.3 wordt toegelicht wat je te zien krijgt bij de functie scraped en in hoofdstuk 3.6.1 wordt toegelicht hoe je een scraper kan maken om dit resultaat te krijgen.

Leerdoel: De gegevens van de door jouw gemaakte scraper begrijpen.

3.5 Text

De text functie wordt gebruikt voor het tonen van eenvoudige teksten op een pagina waarbij overige inhoud wordt genegeerd. Je gebruikt deze functie voor verschillende doeleinden. Deze functie bestaat uit drie onderdelen:

Words (betaalde versie)

Deze functie laat alle woordenschat zien van de tekst. Met gebruik van de functie words kun je bijvoorbeeld ontdekken hoe vaak een specifiek woord in een tekst voorkomt.

News

De functie news is daarentegen heel handig om alle nieuwsartikelen op een website weer te geven in een makkelijk overzicht. Naast de lijst die je hieruit kan ontwikkelen, is er ook de mogelijkheid de data te filteren en te sorteren. Het uiteindelijke resultaat is ook te exporteren naar een Excel-bestand. In 3.5.2 leer je alles over het maken van een lijst met alle nieuwsartikelen van een voorbeeld webpagina.

Leerdoel: Alle nieuwsartikelen scrapen van een website.

Source

De functie source laat de gekleurde broncode van de pagina zien die jij in OutWit Hub hebt geplaatst. Deze codes zijn de belangrijkste onderdelen voor het scrapen (zoeken naar gegevens) van een website. De codes zijn zeer ingewikkeld als je er voor het eerst naar kijkt, maar daar is een oplossing voor. OutWit Hub maakt namelijk gebruik van kleurcodes. In hoofdstuk 3.5.3 leer je alles over de betekenis van kleuren in de broncode.

Leerdoel: Begrijpen van broncode met diverse kleurcodes.

In hoofdstuk 3.5 leer je alles over de mogelijkheden van deze functie en de verschillende onderdelen van deze functie.

3.6 Automators

De automators zijn de functies die Outwit Hub biedt om pagina's naar eigen wens te filteren. Outwit Hub stelt vijf automators beschikbaar in hun betaalde programma. Echter is alleen de automator 'scraper' te gebruiken in de gratis versie van het programma. Met deze functie kun je handmatig informatie uit een website filteren en vervolgens deze overzichtelijk in een tabel plaatsen.

Leerdoel: Bekend worden met de automators van Outwit Hub en leren filteren doormiddel van de scraper-functie.

3.7 History

Het laatste onderdeel wat aan bod komt is history. OutWit Hub slaat alle websites op die je hebt gebruikt tijdens het scrapen (zoeken naar gegevens). Dit wordt allemaal weergegeven in een tabelvorm die OutWit Hub voor jou maakt. In hoofdstuk 3.7 leer je alles over wanneer deze functie van pas kan komen en wat je er allemaal mee kan.

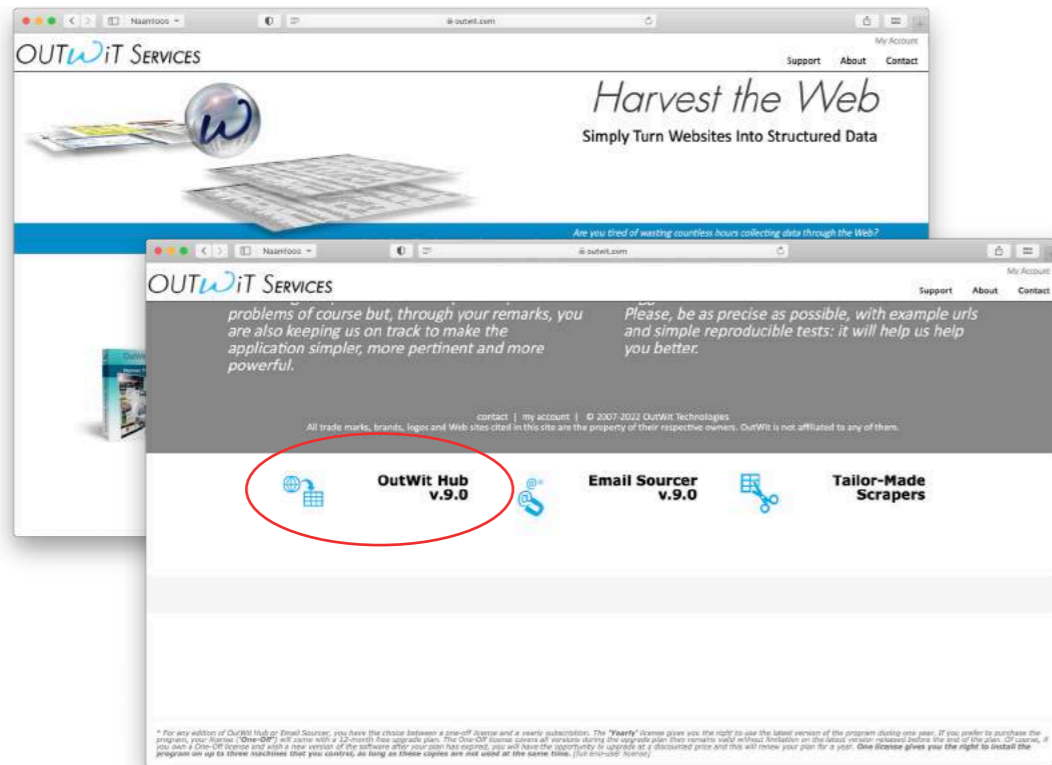
Leerdoel: Ontdekken waar je OutWit Hub voor hebt gebruikt en filteren van history resultaten.



“Get the Data the
Way You Want”

Hoofdstuk 2 Installeren

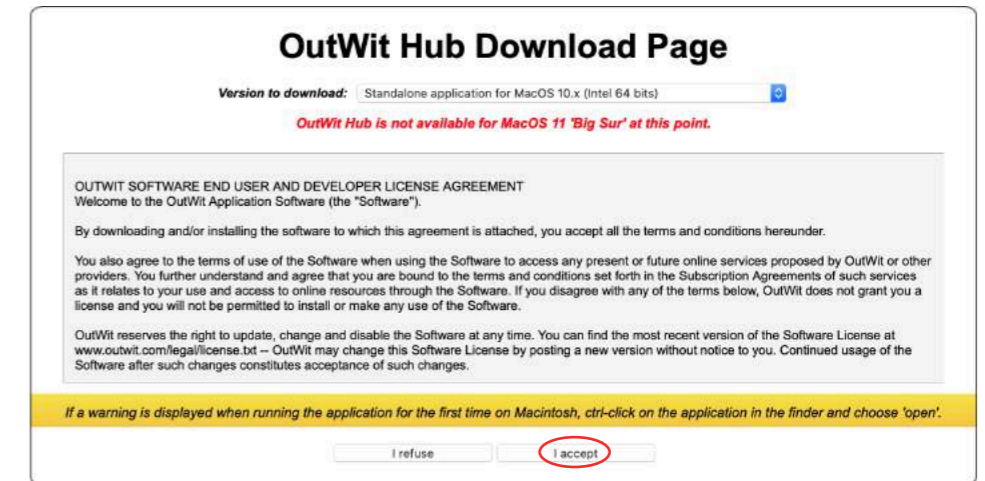
OutWit Hub is gemakkelijk te downloaden via het internet. Ga naar outwit.com en scrol helemaal naar beneden. Klik vervolgens op "OutWit Hub v.9.0".



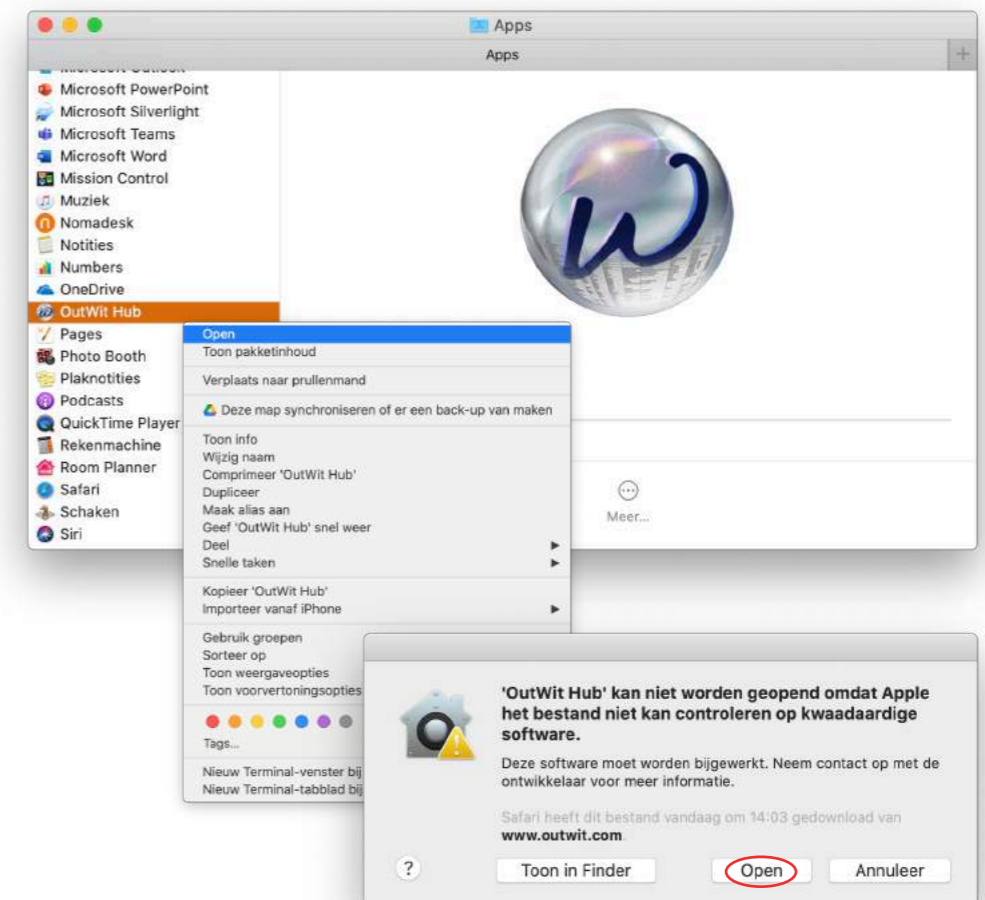
Hierdoor beland je op het volgende scherm. Klik op de blauwe "Free Download" knop om verder te gaan.



Nu hoef je alleen nog de voorwaarden te accepteren en de download zal van start gaan.



Mocht je een Apple-gebruiker zijn, vergeet dan niet de software te openen via rechtermuisknop en aan te geven dat je de software toch wil openen.



Hoofdstuk 3.1 Links

Het tabblad "links" laat een lijst zien met alle links die op de website van de, in de zoekbalk van de software, ingevoerde URL in zijn gevonden. De gevonden links kun je gebruiken om specifieke delen van de pagina uit elkaar te trekken en specifieker te kunnen bekijken.

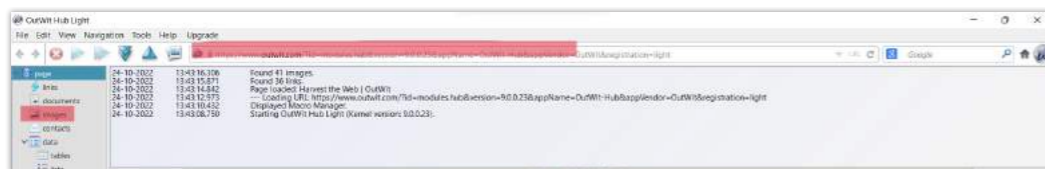
Om een beter overzicht te creëren kan de lijst met links door middel van klikken op de kolommen (bovenaan de lijst met alle links) worden gesorteerd en door middel de balk daaronder gefilterd en naar de Catch verplaatst worden.



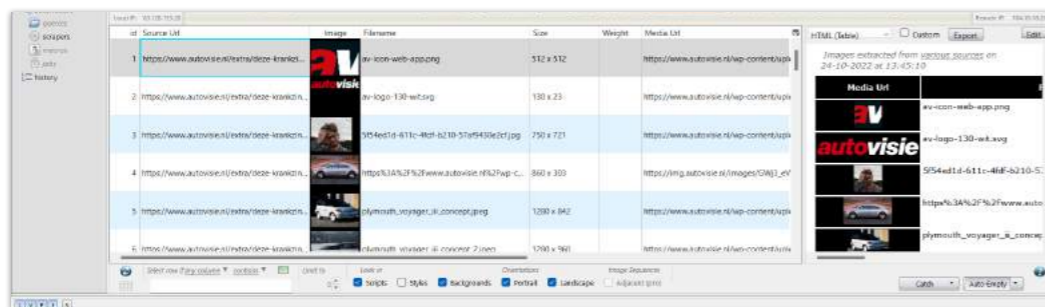
De Catch kun je zien als een soort (persoonlijk) archief waar je data en onderdelen in kunt verzamelen en opslaan om in andere tabbladen mee verder te kunnen werken. In hoofdstuk 3.7 History lees je hoe je links aan je "Catch" kan toevoegen.

Hoofdstuk 3.2 Images

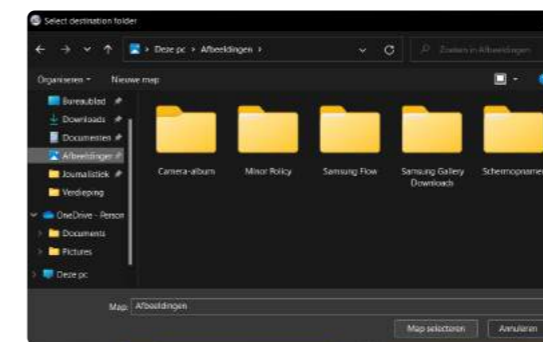
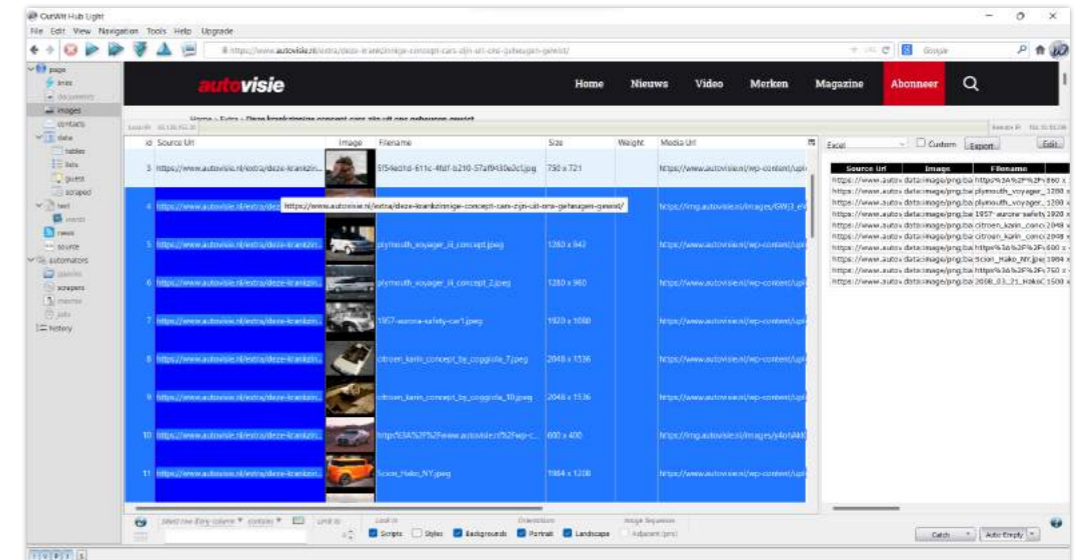
Nog een ingebouwde functie binnen Out WitHub is de 'image scrape'. Met deze tool kan je in een handomdraai allerlei foto's van een webpagina downloaden. Om daarmee aan de slag te gaan is het zoals gewoonlijk zaak dat je de URL van de desbetreffende webpagina invoert in de zoekbalk binnen OutWit Hub. Klik daarna op images, links in de toolbar.



In de onderstaande afbeelding zie je het scherm dat je vervolgens te zien krijgt. Zoals je ziet, heeft de software alle afbeeldingen gescraped. Dus ook het logo van het website, de afbeelding van de auteur etc.

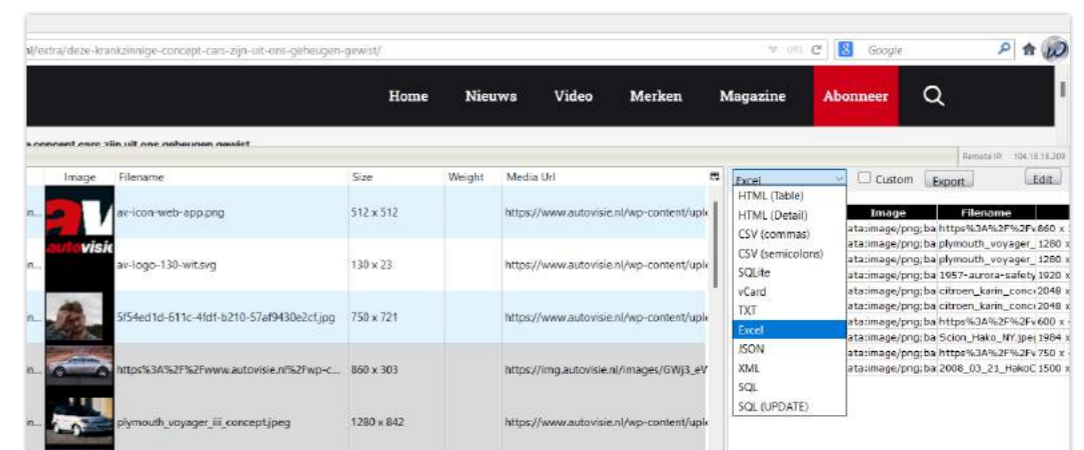


Om te voorkomen dat je harde schijf vol komt te staan met ongewenste afbeeldingen, kan je kiezen welke afbeeldingen je wil downloaden. Dat doe je door de afbeelding te selecteren met de rechtermuisknop terwijl je de control/commandtoets ingedrukt houdt. Daarna klik je linksboven op 'File' en vervolgens op 'Download selected files in'.



Daarna hoeft je enkel nog een map te kiezen waar je wil dat de afbeeldingen in belanden.

Je kan de afbeeldingen echter ook in tabelvorm weergeven. Dat doe je door aan de rechterzijde van het scherm een keuze te maken qua format (in dit geval Excel) en vervolgens op 'Export' te klikken.



Hoofdstuk 3.3

Contacts

De functie contacts kan je goed gebruiken voor het ontwikkelen van een lijst waarbij je (indien mogelijk) alle emails zichtbaar krijgt die gevonden kunnen worden.

Het ontwikkelen van deze lijst is makkelijk. Als eerste moet je ervoor zorgen dat je een pagina hebt die je wilt onderzoeken.

Voor dit voorbeeld wordt gebruik gemaakt van een betaalde backlink tracker (Majestic). Er zijn ook zeker gratis backlink trackers te vinden, maar die kunnen vaak minder links scrapen.

Als voorbeeld: Je wilt een eigen onderneming starten en wilt onderzoeken welke mensen praten over de branche waar jij in wilt werken. (In dit voorbeeld maak jij gebruik van Instagram).

- 1) Ga naar Instagram en typ de hashtag van de concurrent in of gebruik hun account pagina.
- 2) Kopieer de URL van de gehele pagina en plaats deze in Majestic (backlink site). Als je de historic index aan hebt staan krijg je nog meer resultaten.
- 3) Zorg dat de search op URL staat en klik op export as Excel.
- 4) Open het Excel-bestand en verwijder alle kolommen behalve kolom A (in dit geval is alleen kolom A van belang).
- 5) Zorg ervoor dat alles in hyperlinks staat dit doe je door in kolom B2: =HYPERLINK(A2) te zetten en dan trek je B2 naar beneden zodat dit bij de overige links ook gebeurt.
- 6) Sla het Excel op als webpagina: Save as webpage
- 7) Open OutWit Hub (zorg dat je op de 'page' staat) en klik op file -> open en open het bestand dat je zojuist hebt opgeslagen.
- 8) Klik rechtermuisklik op een link -> fast search for contacts -> in the links of this page
- 9) Vervolgens ga je automatisch naar de functie contacts en ontstaat er een lijst met gegevens van de backlinks.

Voor een visuele uitleg zie bron: Jesper Åström, (2018).

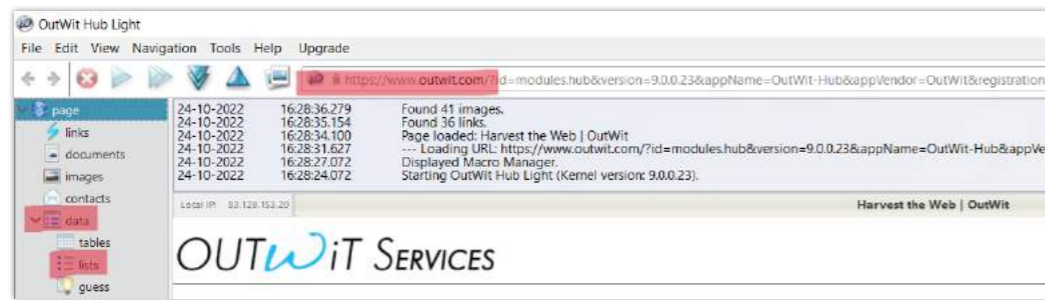
Vervolgens kan je de lijst met emails die hieruit komen gebruiken om contact te leggen met mensen die geïnteresseerd zijn in soortgelijke producten als die jij wilt verkopen.



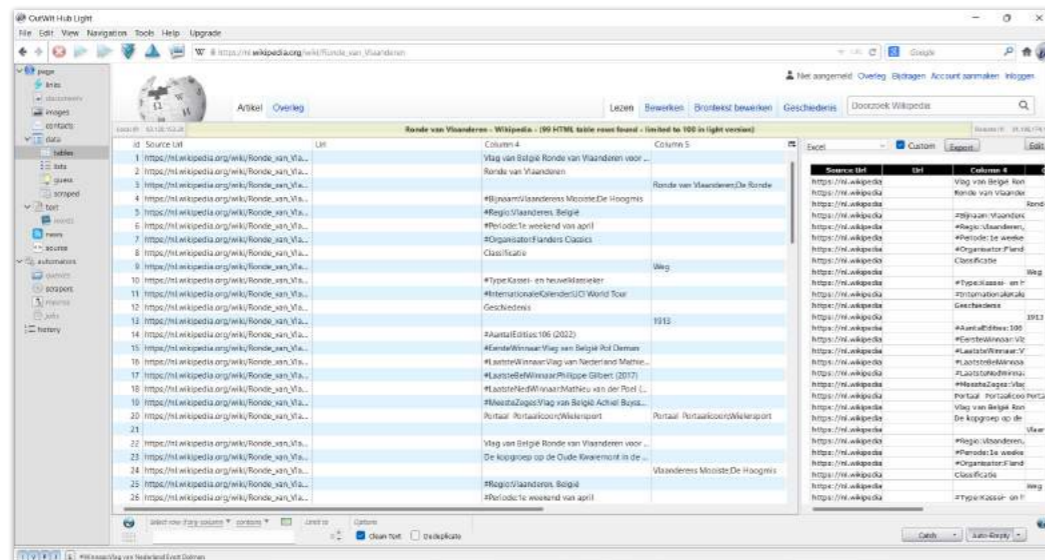
Find, extract and organize
all kinds of data and media
from online sources!

Hoofdstuk 3.4.1 Data Tables

Met de functie voor tabellen kun je binnen no-time een volledige tabel - zoals in de HTML-code wordt aangegeven met de scheidingstekens <table> en </table> - exporteren naar een format van jouw keuze. Dit kan om verschillende redenen handig zijn om onder de knie te krijgen. Stel nu dat je een artikel wil schrijven over de Ronde van Vlaanderen en je wil erachter komen hoeveel procent van de uiteindelijke winnaars uit België komen. Dit is in een programma als Excel heel makkelijk te berekenen. Op Wikipedia kan je een tabel vinden met alle winnaars, maar deze volledig kopiëren en plakken is geen optie. Om de tabel zo snel mogelijk in Excel te krijgen kun je de tabelfunctie van OutWit Hub gebruiken. Je begint zoals gewoonlijk met het invoeren van de link in de zoekbalk en klikt vervolgens op 'Tabel', onder het kopje 'Data', links in de toolbar.

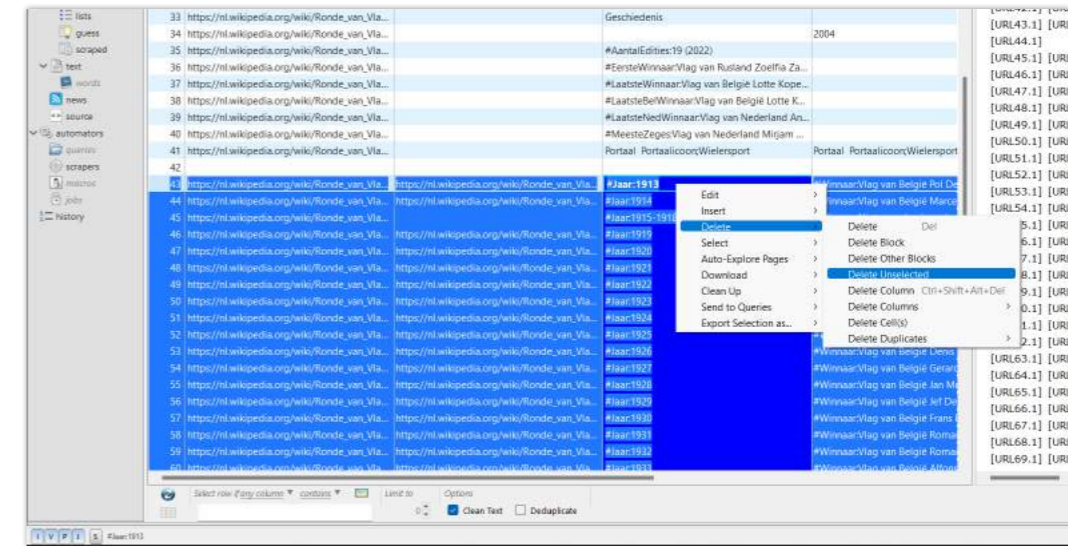


In de onderstaande afbeelding zie je het volgende scherm. Zoals je ziet, heeft de software alle tabellen die op de desbetreffende webpagina werden weergegeven, gescrept. Dus ook de inhoudsopgave en dergelijke.

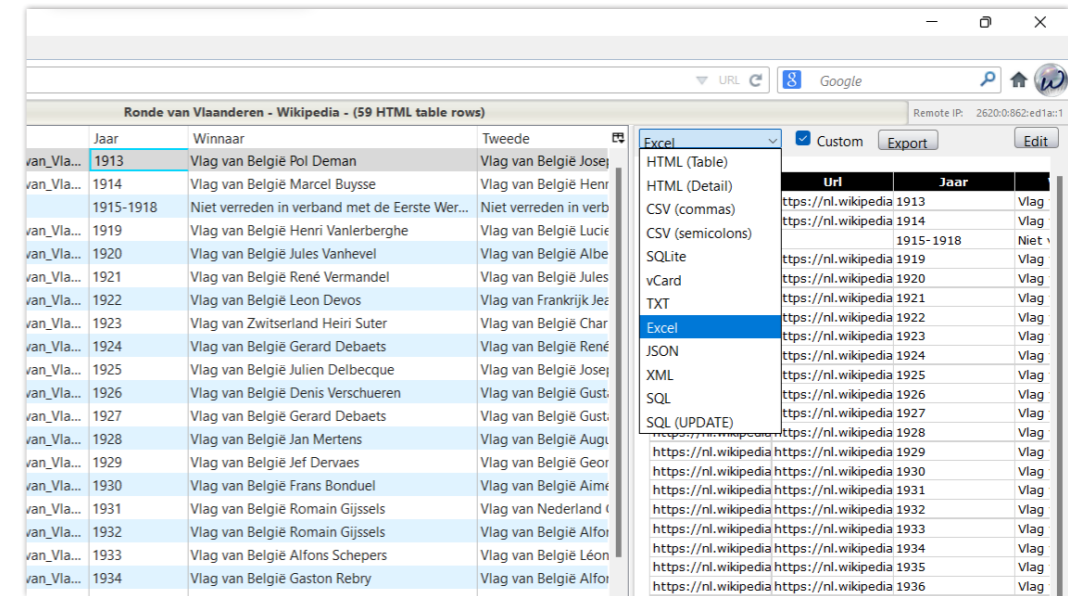


Vanzelfsprekend is alleen de tabel met winnaars nodig, dus kan de rest verwijderd worden om het wat overzichtelijker te maken. Om dat te doen klik je met de rechtermuisknop op de tabel die je nodig hebt (in dit geval de tabel met winnaars) en klik je vervolgens op 'Select' en daarna op 'Select block'. Nu is de gehele tabel met winnaars geselecteerd.

Om de overige tabellen te verwijderen klik je op de rechtermuisknop en vervolgens op 'Delete' en daarna op 'Delete unselected'.



In de onderstaande afbeelding kun je zien dat je enkel nog de tabel met winnaars overhoudt. Het enige dat daarna nog moet gebeuren is het kiezen van het juiste format. Wanneer je (in dit geval voor Excel) gekozen hebt, resteert alleen nog een klik op 'Export'.

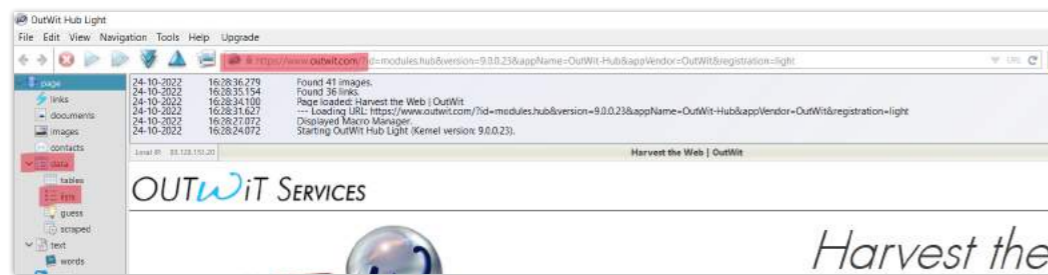


Het resultaat van deze paar handelingen is een overzichtelijke tabel met alle nodige informatie en kan Excel zijn werk gaan doen.

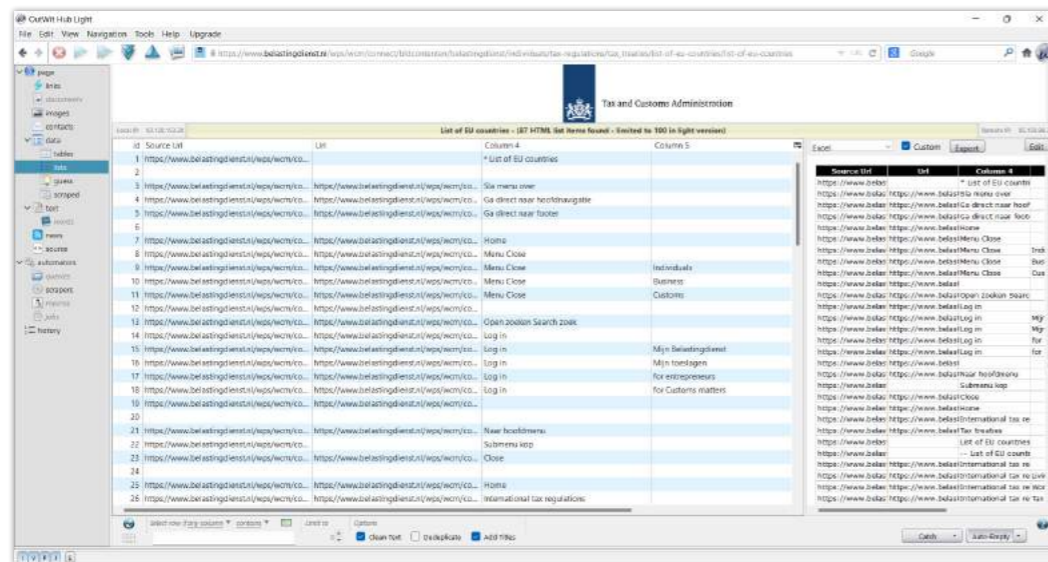
Hoofdstuk 3.4.2 Data Lists

De functie voor "lists" is een soortgelijke toepassing als die voor tabellen. Echter kijkt deze tool niet naar wat er in een tabel staat - zoals in de HTML-code wordt aangegeven met <table> en </table> - maar logischerwijs naar wat er in een lijst staat - aangeduid met en . De meeste websites gebruiken meerdere lijsten op een webpagina dus het zou kunnen dat er een grote hoeveelheid informatie gescreapt wordt die niet nodig is. Met de limiet van 100 elementen die je met de proefversie kunt scrapen is het in veel gevallen dus beter om je eigen scraper te maken. Hoe dat in zijn werk gaat, wordt later in deze handleiding uitgelegd.

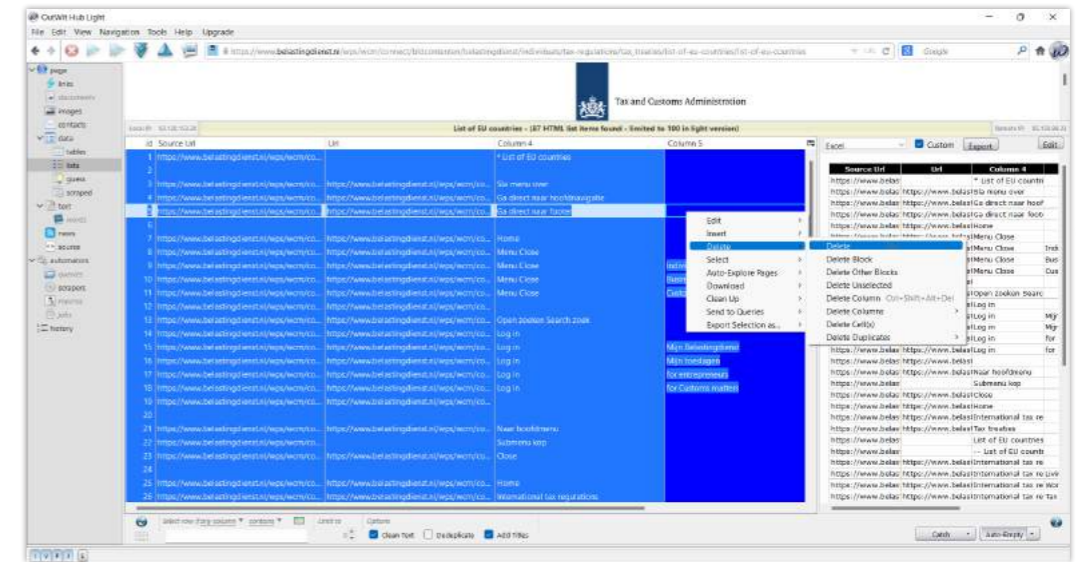
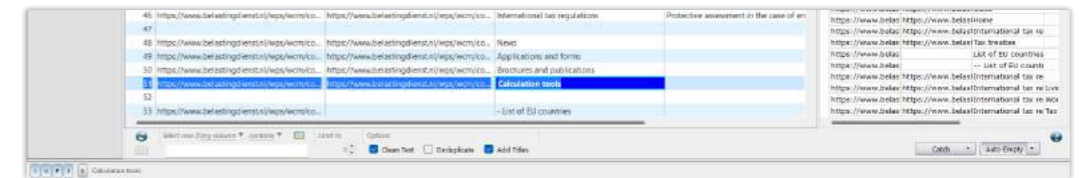
Zoals gezegd is de functie voor lijsten praktisch gezien hetzelfde als de tabelfunctie. Ook hier moet de link allereerst in de zoekbalk geplakt worden. Vervolgens klik je op lijsten. In dit voorbeeld wordt er een lijst gescreapt van alle EU-landen.



Wat meteen opvalt, is dat er dus meerdere lijsten gescreapt worden. Net als bij de tabel van hierboven gaan we de overige lijsten weghalen. De lijst met landen bestaat echter uit twee onderdelen. Het selecteren van één blok en de overige lijsten verwijderen zal hier dus niet werken. Dan zou het tweede deel van de lijst namelijk verloren gaan. In dit geval moeten de ongewenste lijsten op een andere manier verwijderd worden. Eén manier om dat te doen is door de eerste regel boven de lijst met EU-landen te selecteren.

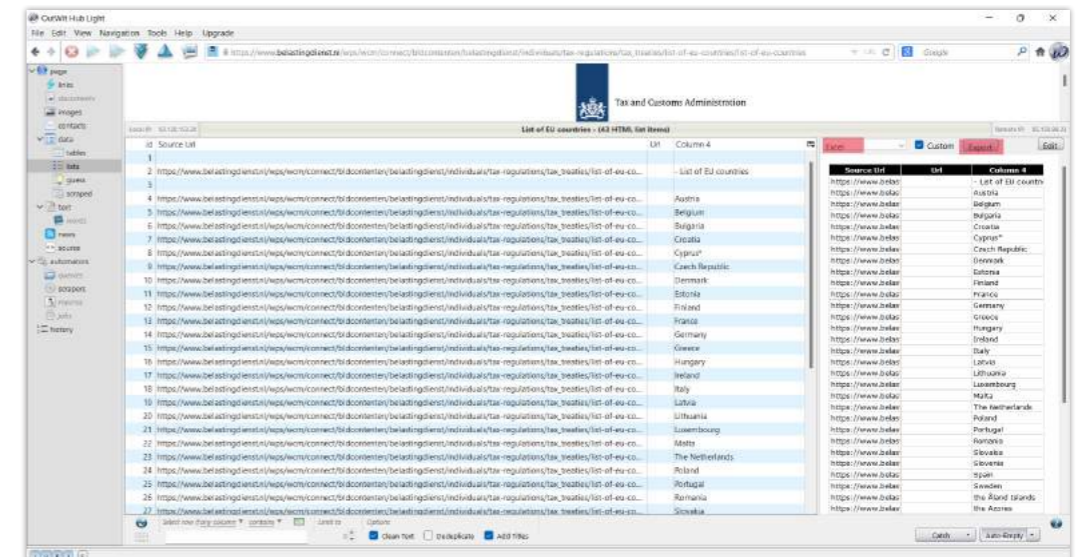


Selecteer het laatste onderdeel voor het benodigde deel van de lijst en selecteer vervolgens de allereerste regel met de linkermuisknop terwijl je de shifttoets ingedrukt houdt. Hierdoor wordt alle informatie die tussen de eerste en laatst aangeklikte regel geselecteerd. Om dit te verwijderen klik je via rechtermuisknop op 'Delete' en nogmaals 'Delete'.



Onderaan de lijst die we nodig hebben staat echter nog meer onnodige informatie. Hier moet hetzelfde principe toegepast worden.

Het resultaat is een overzichtelijke lijst van alle EU-landen. Nu is het zoals gewoonlijk een kwestie van het juiste format selecteren en exporteren.



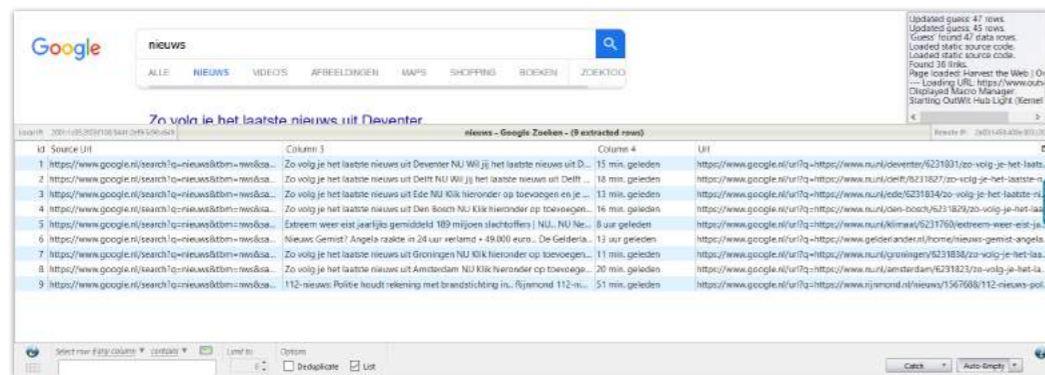
Hoofdstuk 3.4.3 Data Guess

De guess functie geeft de gegevens weer uit de pagina die in Outwit Hub is gezet. Outwit Hub maakt zelf een tabel met geëxtraheerde data. Dit doet Outwit Hub met behulp van automatische algoritmen. Deze algoritmen probeert de structuur van de pagina te begrijpen en gokt hier een gepaste scrape bij.

Outwit Hub analyseert de broncode van de pagina en gaat op zoek naar labels en scheidingstekens voor velden. Het programma maakt dus zelf een scraper aan zonder dat je zelf in de broncode op zoek moet gaan naar de gewenste gegevens.

Outwit Hub maakt bij deze functie zelf de scrape die wij in de tutorial samen gemaakt hebben. Het kan dus een functie zijn die heel veel tijd bespaart. Echter, zoals bij elk automatisch herkenningsalgoritme, zal dit niet in alle gevallen resulteren in het formaat dat gewenst is. Maar het is zeker een eenvoudig alternatief als je op zoek bent naar een simpele scraper.

Zo ziet bijvoorbeeld een automatische scrape van deze functie eruit voor de nieuwspagina op Google van vandaag (Nieuws-google zoeken, z.d.)



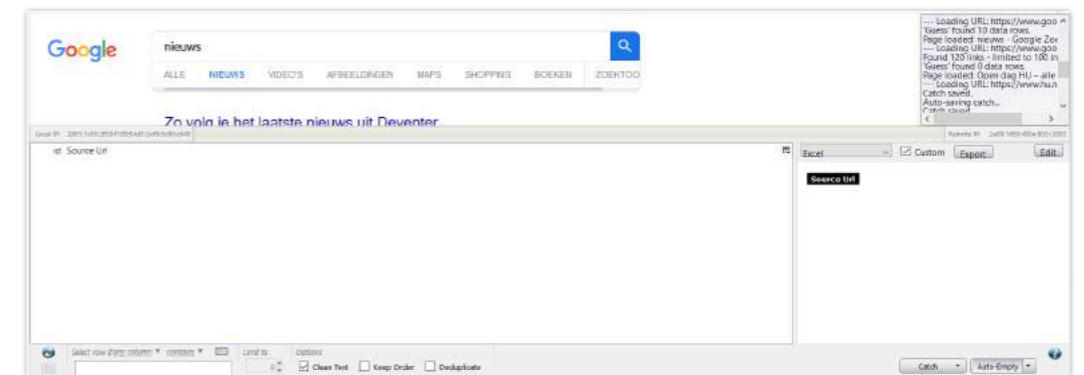
De enige stap die hiervoor gezet moet worden, is het toevoegen van een website in Outwit Hub. Vervolgens gokt Outwit Hub hoe de tabel eruit moet komen te zien.

Vervolgens is het net zoals bij bijna alle functies mogelijk om een catch (filter) te maken voor deze geautomatiseerde tabel. Om zo de informatie te filteren en mogelijk te exporteren in een Excel-bestand.

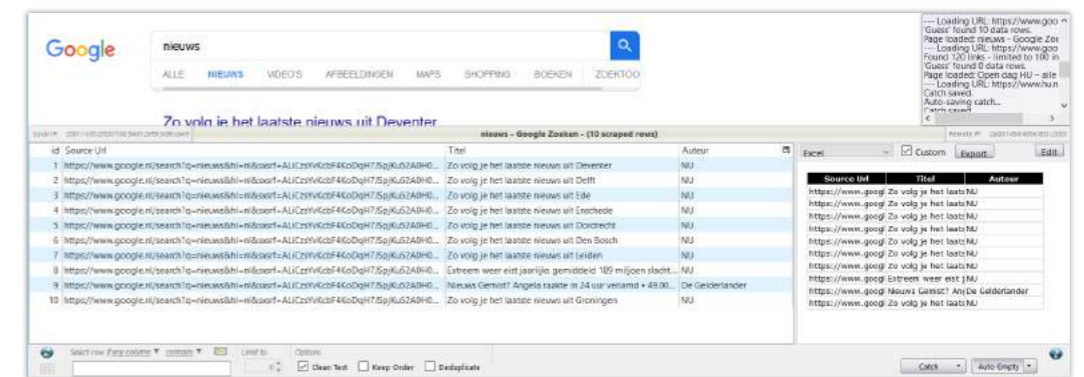
Hoofdstuk 3.4.4 Data Scraped

Het scraped venster heeft geen eigen functie, maar werkt als een weergave venster voor de functie 'scrapers' onder automaters (zie hoofdstuk 3.6.1 Scrapers).

Het venster geeft de resultaten weer van een scraper die handmatig gemaakt is door de Outwit Hub gebruiker. Zodra er geen scraper gemaakt of geactiveerd is zal dit venster ook leeg zijn.



Zodra er een aangepaste scraper geactiveerd wordt zal er een tabel verschijnen in dit venster. Als voorbeeld zie gebruiken wij de scraper die wij gezamenlijk tijdens de tutorial gemaakt hebben.



Zoals je ziet maakt Outwit Hub nu een tabel met informatie die wij graag willen scrapen in een tabel. Dit venster weergeeft de uitkomst van de scraper die wij gemaakt hebben.

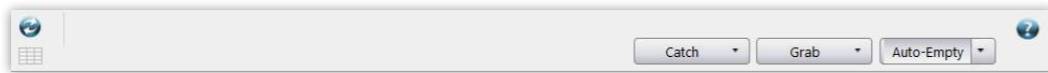
De tabel is te exporteren naar Excel, maar het is ook weer mogelijk om hier gebruik te maken van de catch (filter) om zo nog een verder gefilterde tabel te verkrijgen.

Hoofdstuk 3.5

Text

De text functie binnen Outwit Hub toont simpelweg alle eenvoudige tekst op een pagina. De tekstweergave geeft de tekstuele inhoud van de pagina die in Outwit Hub gezet is weer en negeert alle andere inhoud. De inhoud die wegvalt zijn bijvoorbeeld scripts, media, animaties, lay-outs, afbeeldingen, iconen, symbolen, etc.

Ook binnen de text functie is het mogelijk om een catch te maken om zo een gefilterde versie van de tekstweergave te maken. Deze gefilterde versie kan worden opgeslagen of worden geëxporteerd naar Excel.



Binnen de text functie zijn er nog drie verschillende functies mogelijk. Dit zijn de functies: words, news en source.

Hoofdstuk 3.5.1

Text Words

De Words functie is onderdeel van de betaalde versie van Outwit Hub en is niet beschikbaar voor gebruikers van de light versie.

In het kort geeft deze functie de woordenschat weer die op de pagina die in Outwit Hub gezet is. Hierin is de functie zo ingesteld dat elk woord apart gezien wordt. Hierin is in een oogopslag te zien welk woord met welke frequentie op de pagina te zien is.

id	Ordinal	Source URL	Freq	Word
1	1.0.1	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	upgrade plan
2	1.0.2	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	license gives
3	1.0.3	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	3/1469	one-off license
4	1.0.4	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	download application
5	1.0.5	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	email addresses
6	1.0.6	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	exploration button
7	1.0.7	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	sourcing tasks
8	1.0.8	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	vastly simplifies
9	1.0.9	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	email sourcer automates
10	1.0.10	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	found results
11	1.0.11	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	web page
12	1.0.12	https://www.outwit.com/?id=modales.hub&version=9.0.0.23&appName=OutWit-Hub&appVendor=OutWit®istration=light	2/1469	email sourcer

Hoofdstuk 3.5.2

Text News

De news functie toont een lijst met nieuwsartikelen (feed) op de huidige pagina of in het huidige domein. De nieuwsweergave wordt in een tabel gezet waarin diverse data over de pagina in te vinden is. De tabel bevat de volgende informatie:

Source-URL: de URL van de pagina waar de feed is gevonden

Feed Title: de naam van de feed

Title: de naam van het artikel

Article URL: de link naar het volledige artikel

Date: de datum en tijd van de uitgave van het artikel

Abstract: samenvatting van het artikel

Deze functie geeft snel en simpel een overzicht van alle nieuwsartikelen op een website. Om te laten zien hoe zo'n lijst met nieuwsartikelen of feed eruit ziet hebben wij een voorbeeld gemaakt van de AD nieuws pagina AD.nl, (z.d.).

id	Source URL	Feed Title	Title	Article URL	Date	Abstract
1	https://www.ad.nl/news/	ADnieuws	Kluis Kerkdankian reed account met boezemla...	https://www.ad.nl/show/police-kerkdankian-reed-...	23-10-2022 12:27:07	Korte updates van wereldsterren op sociale media: wij...
2	https://www.ad.nl/news/	ADnieuws	Een speciale libelle voor raaggetijden, met...	https://www.ad.nl/binnenland/een-speciale-libe...	23-10-2022 12:05:54	Tijdschrift libelle presenteert omhoog een adrie specia...
3	https://www.ad.nl/news/	ADnieuws	Waarom Buitenland 11 per meest wetters...	https://www.ad.nl/buitenland/waarom-buitenlan...	23-10-2022 11:00:52	Ekte week schrijft politiek wetshapper Hans van Soest...
4	https://www.ad.nl/news/	ADnieuws	Rishi Sunak stelt zich officieel kandidaat voor B...	https://www.ad.nl/buitenland/rishi-sunak-stelt-...	23-10-2022 09:46:53	Rishi Sunak heeft zich zondag officieel kandidaat gestel...
5	https://www.ad.nl/news/	ADnieuws	Zoektocht naar vermisten bootongeluk Terishe...	https://www.ad.nl/binnenland/zoektocht-naar-...	23-10-2022 09:34:59	De zoektocht naar de vermiste opvarenden van het boot...
6	https://www.ad.nl/news/	ADnieuws	Inzamelingsactie voor familie van Hebe stopt...	https://www.ad.nl/buitenland/inzamelingsactie-...	23-10-2022 08:07:15	De inzamelingsactie voor de familie van Hebe, het meis...
7	https://www.ad.nl/news/	ADnieuws	Tweede amber alert is één week, hoe vermiste...	https://www.ad.nl/binnenland/tweede-amber-al...	23-10-2022 05:29:44	De 11-jarige jongen die was vermist in Den Haag is van...
8	https://www.ad.nl/news/	ADnieuws	Oud president China plotseling afgevoerd tij...	https://www.ad.nl/buitenland/oud-president-chi...	23-10-2022 04:13:00	De Chinese leider Xi Jinping heeft het twintigste partij...
9	https://www.ad.nl/news/	ADnieuws	Fries dorp rouwt om bij bootongeluk omgek...	https://www.ad.nl/binnenland/fries-dorp-rouwt-...	23-10-2022 04:12:00	Wat een bijzondere vader-zoonrij had moesten worde...
10	https://www.ad.nl/news/	ADnieuws	Za wapenen adbezztars zich tegen skel: A...	https://www.ad.nl/autot/ze-wapenen-autobezitt...	23-10-2022 04:05:41	Voor wie het geluid van hagedisren op het dak van de...
11	https://www.ad.nl/news/	ADnieuws	Twee verpleegsters doodgegraven in zee...	https://www.ad.nl/buitenland/twee-verpleegste...	23-10-2022 00:24:33	Een man heeft opentagproeven (kake 199) twee verpl...
12	https://www.ad.nl/news/	ADnieuws	Duitse fitness tycoon veeggestort met klein vlie...	https://www.ad.nl/buitenland/duits-fitness-tyc...	23-10-2022 00:02:43	Een klein vliegtuigje, met daarin zes mensen onder wal...

Over het algemeen, net als voor de andere functies binnen Outwit Hub, kan de inhoud van deze weergave worden gefilterd en gesorteerd om specifieke gegevens te extraheren, met behulp van de filterbesturings-elementen in het gegevensblad onder de tabel. Je kunt het ook verplaatsen naar de Catch om zo een nieuwe gefilterde tabel te maken om deze vervolgens de exporteren naar een Excel-bestand.

Hoofdstuk 3.5.3 Text Source

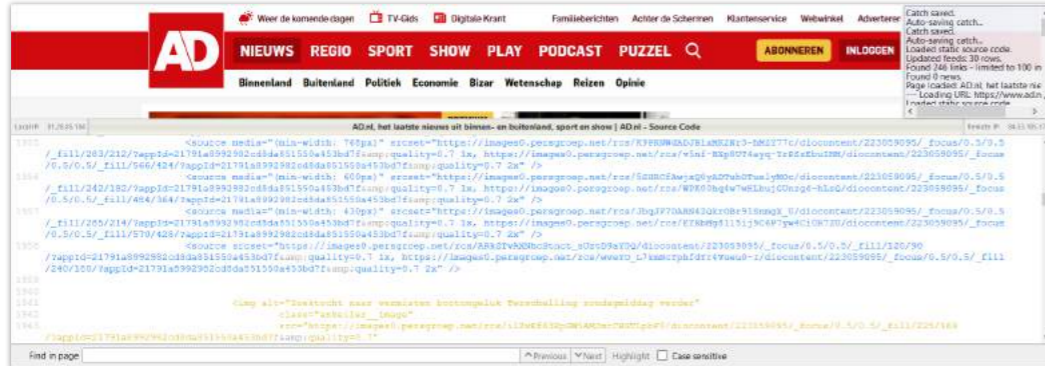
De source functie toont de gekleurde broncode van de pagina die in Outwit Hub geplaatst is. Dit is de code achter de website die ervoor zorgt dat de website eruit ziet zoals de website eruit ziet. Deze broncodes zijn te gebruiken bij het scrapen van de website zoals we in de tutorial in de les gedaan hebben.

Om de verschillende soorten broncodes te herkennen maakt Outwit Hub gebruik van een kleurensysteem. Deze is ontworpen voor het zoeken naar gegevens. Hierbij wordt de nadruk gelegd op de tekstuele inhoud die daadwerkelijk op de pagina wordt weergegeven. Er worden dus geen codes gebruikt voor programmeerdoeleinden.

De kleuren die gebruikt worden zijn:

- Weergegeven tekst (zwart):** namen van artikels
- HTML tags (blauw):** vaak tekst of namen van onderdelen op de website
- Links (roze):** linkjes van websites
- Comments (licht grijs):** reacties
- HTML Entities (donker grijs):** symbolen, bijvoorbeeld het copyright teken
- Styles (groen):** stijl van blokken op de pagina, bijvoorbeeld de vorm
- Scripts (paars):** blokken op de pagina, bijvoorbeeld ruimte voor advertenties
- Images (geel):** afbeeldingen

Dit ziet er zo uit bij bijvoorbeeld de AD pagina:



Binnen de broncodes is er ook een optie om te zoeken. Dit is erg handig want achter een kleine website kan een hele grote broncode zitten. Daarnaast zit er ook een highlight functie die ervoor zorgt dat het zoekwoord makkelijk in de code terug te vinden is wanneer dit nodig is.

Hoofdstuk 3.6 Automators

De automators zijn de functies die Outwit Hub biedt om een pagina te filteren op data. Er zijn vijf automaten binnen Outwit Hub: projecten, scrapers, macro's, jobs en queries. Van deze vijf is alleen de functie scraper gratis te gebruiken. Voor de andere vier functies moet je de pro (betaalde) versie van Outwit Hub aanschaffen.

Hoewel we vier van de vijf functies niet kunnen gebruiken is hier een korte beschrijving per automaat:

Scraper: een beschrijving van de gegevens op een pagina, die de markeringen definieert die in de codes achter de website gevonden kunnen worden. Door deze codes te 'scrapen' zijn de gegevens op de pagina te extraheren.

Macro: een momentopname van de volledige extractie van data op de pagina. Deze is gemaakt om een specifieke extractie met een klik op de knop uit te voeren.

Job: een vooraf ingestelde tijd waarop een actie of extractie uitgevoerd moet worden. Dit is handig om toe te passen wanneer er nieuws uitkomt en je bepaalde websites wil scrapen die op dat moment hierop reageren.

Queries: dit zijn sets van verschillende URL's waarop gelijktijdig verschillende andere automators uitgevoerd kunnen worden. Zo kun je een scraper combineren met een job om op een bepaalde tijd die bepaalde scraper uit te voeren.

Project: in een project komen alle vier de andere automaten bij elkaar in een complementaire scraper.

Hoofdstuk 3.6.1 Automators Scraper

Zoals eerder verteld, is een scraper een soort sjabloon die Outwit Hub vertelt hoe informatie uit een pagina moet worden gehaald. Wanneer tabellen of lijsten er niet in slagen om automatisch de structuur van een pagina te herkennen en de gegevens eruit te halen, heb je de mogelijkheid om een scraper te maken om zo Outwit Hub te vertellen hoe jij deze informatie overzichtelijk wilt scrapen.

Bij het maken van een scraper zijn de broncodes achter de website erg belangrijk. Deze broncodes zijn te herkennen aan de verschillende kleuren en staan vaak tussen dit soort <...> haakjes. Om te weten welke kleur wat betekend, zie hoofdstuk 3.5.3 Source.

Tijdens de tutorialdag hebben we uitgebreid gewerkt met het maken van zo'n soort scraper. Om te leren hoe je een scraper moet maken en te weten waar je hierbij precies rekening moet houden, zie de bijlagen "gasprijzen opdracht" en "gasprijzen (digitale handout)" (H. 5.1 en 5.2).

Hoofdstuk 3.7 History



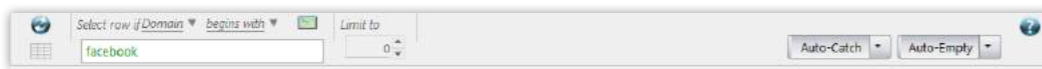
Het laatste onderdeel dat beschikbaar is op de light versie van Outwit Hub, is het onderdeel "History". Dit onderdeel is niets meer dan een onderdeel waarin je kunt kijken welke websites je gebruikt hebt tijdens het scrapen.

Outwit Hub creëert zelf een tabel met data over het gebruik van het programma. Hierin is exact te zien welke URL is gebruikt, wanneer deze URL voor het eerst is gebruikt, de laatste keer dat de URL gebruikt is en hoe frequent de URL gebruikt is.

Outwit Hub maakt dus voor de gebruiker een duidelijk overzicht van alle URL's die in het programma worden gebruikt. Maar dit kan voor een erg grote tabel zorgen wanneer het programma veel gebruikt wordt. Daarom zit bij deze tabel de "catch"-functie. Dit is simpelweg een filter om de tabel overzichtelijker te maken.

id	Domain	First Seen On	Last Seen On	Freq
4	www.google.nl	21-10-2022 16:23:07	21-10-2022 16:23:07	1
5	Blog.outwit.com	21-10-2022 16:23:07	21-10-2022 16:23:07	1
6	www.outwit.com	21-10-2022 16:23:07	21-10-2022 16:23:07	1
7	www.facebook.com	21-10-2022 16:24:18	21-10-2022 16:24:32	39
8	facebook.com	21-10-2022 16:24:19	21-10-2022 16:24:19	1
9	fbcdn.net	21-10-2022 16:24:26	21-10-2022 16:24:29	1
10	m.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1
11	nl.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1
12	fy-nl.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1
13	pl-pl.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1
14	tr-tr.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1
15	de-de.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1

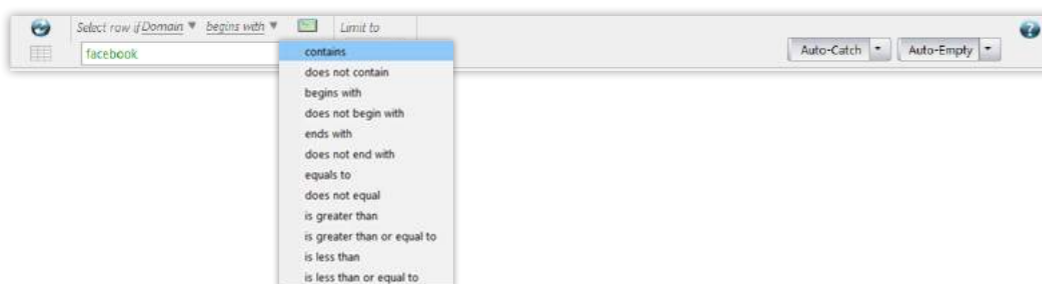
In de "catch" zitten nog verschillende functies. Zo heb je de "Empty"-functie. Hiermee maak je met een klik op de knop de hele tabel leeg. Naast deze functie heb je ook nog het zoekbalkje. Hierin kun je een zoekwoord van de naam van de website intypen om zo snel in de tabel alle URL's te vinden met dat zoekwoord.



Vervolgens heb je ook nog uitgebreide filter opties voor de filter die je wil gaan toepassen. Hiervoor zijn er twee uitklapvensters waarmee je de filter kunt specialiseren. Bij de het eerste venster kijkt de filter naar de rijen:



En bij de het tweede venster kijkt de filter naar hoe (of juist niet) het filterwoord in de tabel moet staan:



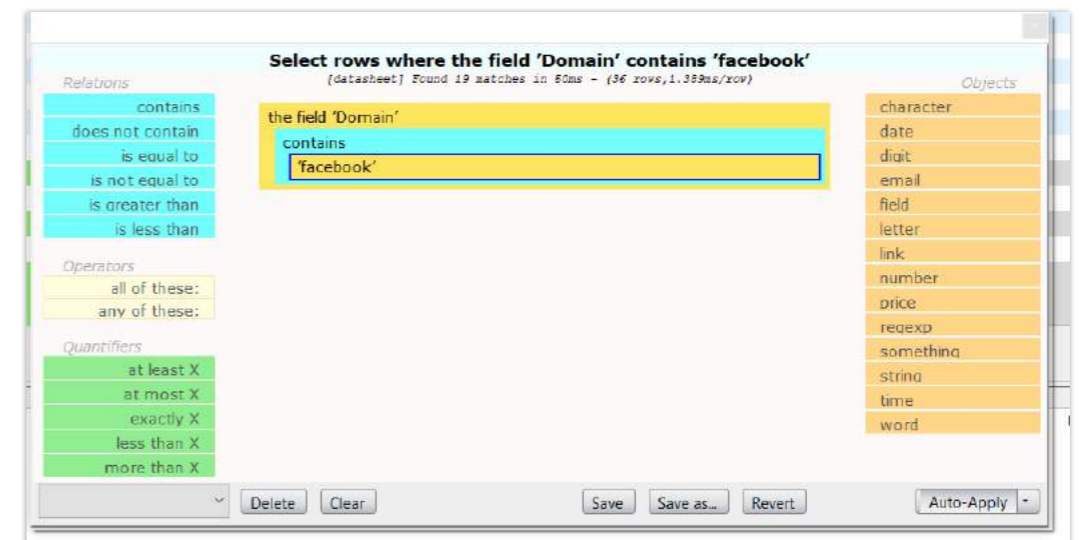
Zodra de hele filter is opgebouwd met de gewenste keuzes van de gebruiker, klikt de gebruiker op "auto-catch". Nu maakt Outwit Hub een gespecialiseerde tabel van de hele gebruiksgeschiedenis. Als voorbeeld is er gefilterd op Facebook:

id	Domain	First Seen	Last Seen	Frequency
1	www.facebook.com	21-10-2022 16:24:18	21-10-2022 16:24:32	39
2	facebook.com	21-10-2022 16:24:19	21-10-2022 16:24:19	1
3	nl.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1
4	fy-nl.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1
5	pl-pl.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1
6	tr-tr.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1
7	de-de.facebook.com	21-10-2022 16:24:32	21-10-2022 16:24:32	1

Deze gefilterde tabel is ook weer te exporteren naar Excel. Dit zijn de simpele en gebruiksvriendelijke filteropties die het History onderdeel van Outwit Hub te bieden heeft. Naast deze opties heeft Outwit Hub ook nog de selector editor optie. Deze is echter in beta en nog niet heel gebruiksvriendelijk. De selector editor is te vinden via het gekleurde icoontje in de catchfunctie:



Wanneer dit icoontje aangeklikt wordt, verschijnt de selector editor. Dit is een uitgebreide versie van de catch die al eerder aan bod kwam. Hiermee is het mogelijk om nog verder te filteren in de URL-geschiedenis. De selector editor en al zijn uitgebreide filteropties ziet er zo uit:



Met de light versie van Outwit Hub is het mogelijk om 100 URL's te filteren. Voor meer uitslagen is een betaalde versie nodig.

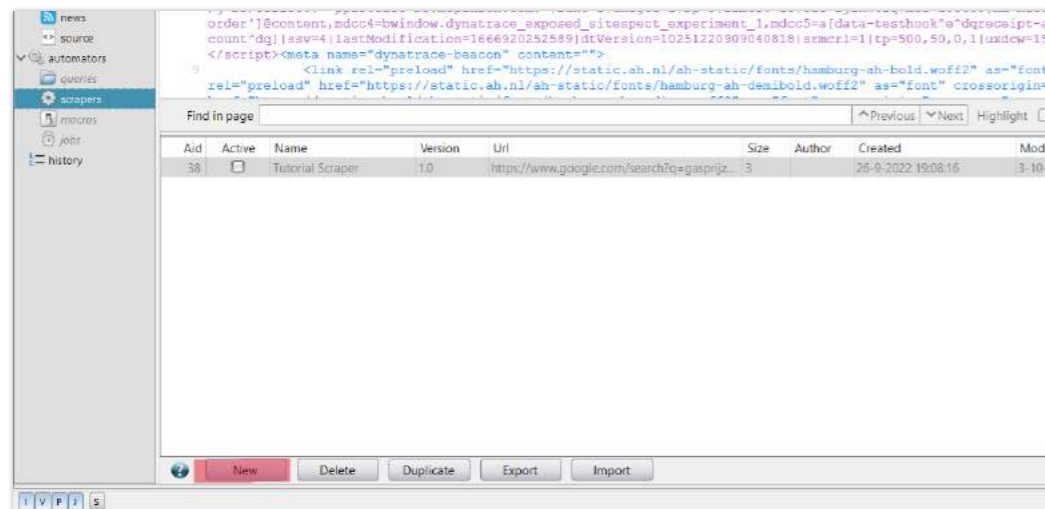
Hoofdstuk 3.8 Scraper

Logischerwijs is er niet voor iedere taak een ingebouwde tool. Het programma kan dus niet altijd met één druk op de knop jouw wensen inwilligen. In dit geval zal je je handjes moeten laten wapperen door een eigen scraper te bouwen. Met deze toepassing kun je zelf kraakhelder aangeven welke informatie je van een website af wil halen.

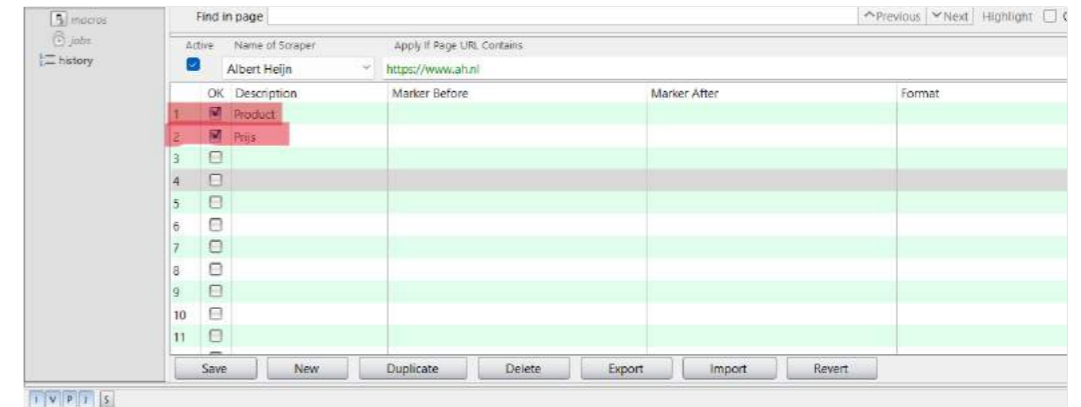
Zoals gewoonlijk plak je de link van de website in kwestie in de zoekbalk en klik je op 'Scraper' in de toolbar.



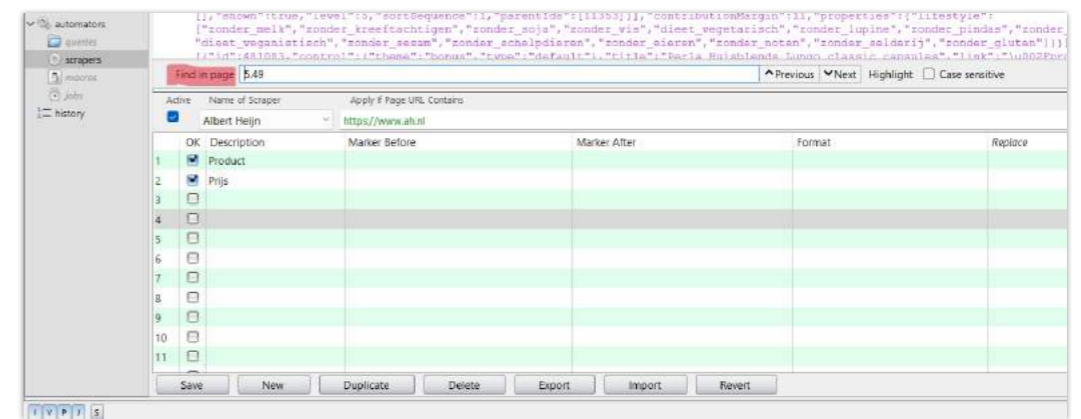
Vervolgens klik je op 'New'. Daarmee wordt er een nieuw werkveld gecreëerd waarin je de scraper gaat bouwen.



In het werkveld kan je de informatie die je wilt scrapen onderverdelen in kopjes. In dit geval willen we Albert Heijns prijzen voor alle koffiecapsules exporter. Daar hebben we de prijs en productnaam voor nodig, waar we twee verschillende regels voor aanmaken.



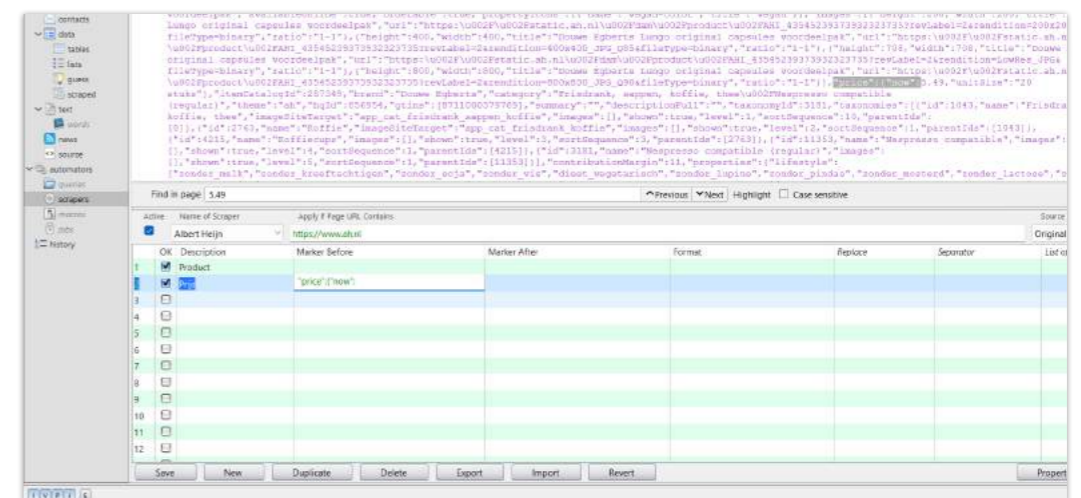
Om de scraper zijn werk te laten doen moeten we een zogeheten 'Marker before' en 'Marker after' invullen. Om te weten te komen wat we daar moeten opgeven moeten we allereerst naar de juiste informatie zoeken binnen de HTML-code. We moeten zoals gezegd de productnaam en de prijs hebben, dus wanneer je in de HTML-zoekbalk een willekeurige prijs of productnaam intikt, kom je meteen terecht waar je moet zijn.



Met de markers geef je aan waar de relevante informatie tussen staat

Stel nu dat in de HTML-code een productprijs wordt weergegeven met 123-€1.99-456. De marker before wordt dan '123-' en de marker after wordt '-456'. De software weet dan dat het de informatie moet scrapen die tussen die twee elementen staat.

In dit geval volgt de prijs na — "price":{ "now": — dat wordt dus de marker before.



Hoofdstuk 4

Literatuurlijst

Getting Started with OutWit Hub. (2008, juni). OutWit Hub Blog.
Geraadpleegd op 27 oktober 2022, van <https://blog.outwit.com/?p=52>

Breinstein. (2021, 19 mei). Web Scraping - een doorzettende trend. <https://www.breinstein.nl/de-digitale-wereld/web-scraping-een-doorzettende-trend/>

AD.nl, het laatste nieuws uit binnen- en buitenland, sport en show. (z.d.). AD.nl.
Geraadpleegd op 24 oktober 2022, van <https://www.ad.nl/nieuws/?referrer=https://www.google.com/>

Jesper Åström. (2018, 5 februari). How to use OutWit Hub to Scrape for Contacts [Video].
YouTube. Geraadpleegd op 24 oktober 2022, van <https://www.youtube.com/watch?v=XJAYn1IGY40>

nieuws - Google Zoeken. (z.d.). Geraadpleegd op 24 oktober 2022, van <https://www.google.nl/search?q=nieuws>

Hoofdstuk 5.1

Bijlagen

Gas Prijzen opdracht

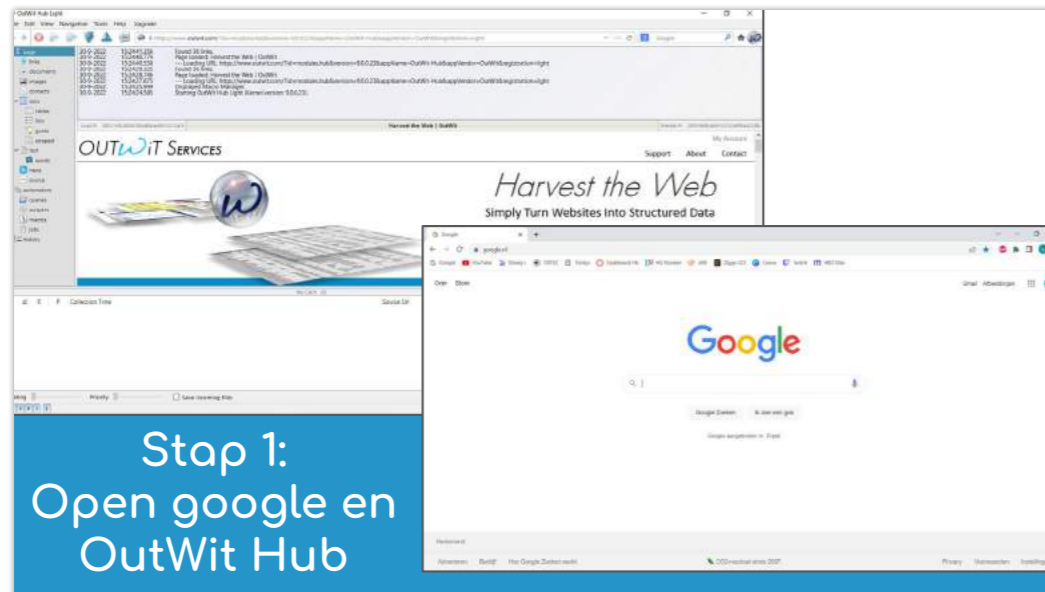
Stappen opdracht in de klas (gasprijzen)

1. Open OutWit Hub en open google zodat je op de startpagina staat
2. Zoek in google naar "gasprijzen"
3. Ga in google naar de nieuwspagina (zorg ervoor dat je zoekt naar recent en gesorteerd op relevantie)
4. Kopieer de link van de pagina (klik niet op een artikel)
5. Plak de link boven in de zoekbalk van OutWit Hub
6. De pagina van google verschijnt in Outwit Hub (het kan zijn dat je nog cookies etc. moet accepteren. Doe dit voordat je verder gaat)
7. Klik aan de linkerkant van het menu op het kopje "scrapers"
8. Creëer een nieuwe scraper door op 'new' te klikken
9. Geef de scraper een naam naar keuze, Kies bijvoorbeeld voor de naam "gasprijzen" omdat we hierop gaan zoeken. Klik op 'ok': Je ziet de nieuwe scraper ontstaan, te herkennen aan de lichtgroene blokjes.
10. Zoek naar het eerste artikel door de naam van het artikel in de zoekbalk. "find in page" de titel te typen.
11. Maak de drie onderdelen: titel, auteur en URL aan in de gecreëerde scraper. De volgorde is van belang. URL mag niet bovenaan staan anders kan dit gevolgen hebben voor de vormgeving van de tabel die we gaan scrapen.
12. Ga op zoek naar de marker before van de titel. Te herkennen aan de "<...>" in blauwe tekst. De titel is te herkennen aan de zwarte tekst en is de naam van het artikel.
13. Plak de geselecteerde tekst in de marker before kolom van de titel in de scraper.
14. Ga op zoek naar de marker after van de titel. Te herkennen aan de "<...>" in blauwe tekst. De titel is te herkennen aan de zwarte tekst en is de naam van het artikel.
15. Plak de geselecteerde tekst in de marker after kolom van de titel in de scraper.
16. Ga op zoek naar de marker before van de auteur. Te herkennen aan de "<...>" in blauwe tekst. De auteur is te herkennen aan de zwarte tekst en is de naam van de organisatie die het artikel geschreven heeft.
17. Plak de geselecteerde tekst in de marker before kolom van de auteur in de scraper.
18. Ga op zoek naar de marker after van de auteur. Te herkennen aan de "<...>" in blauwe tekst. De auteur is te herkennen aan de zwarte tekst en is de naam van de organisatie die het artikel geschreven heeft.
19. Plak de geselecteerde tekst in de marker after kolom van de auteur in de scraper.
20. Ga op zoek naar de marker before van de URL. Te herkennen aan de "<...>" in blauwe tekst. De URL is te herkennen aan de roze tekst en is de naam website.
21. Plak de geselecteerde tekst in de marker before kolom van de URL in de scraper.
22. Ga op zoek naar de marker after van de URL. Te herkennen aan de "<...>" in blauwe tekst. De URL is te herkennen aan de roze tekst en is de naam website.
23. Plak de geselecteerde tekst in de marker after kolom van de URL in de scraper.

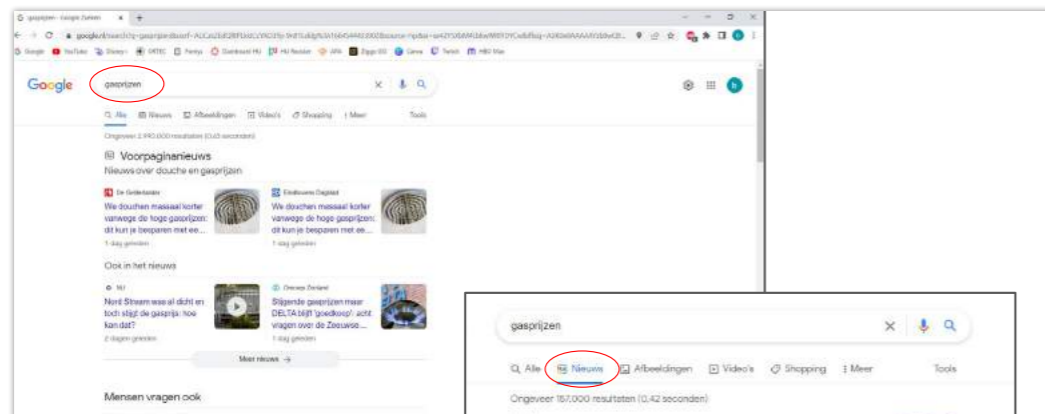
Hoofdstuk 5.2

Bijlagen

Gasrijzen (digitale handout)

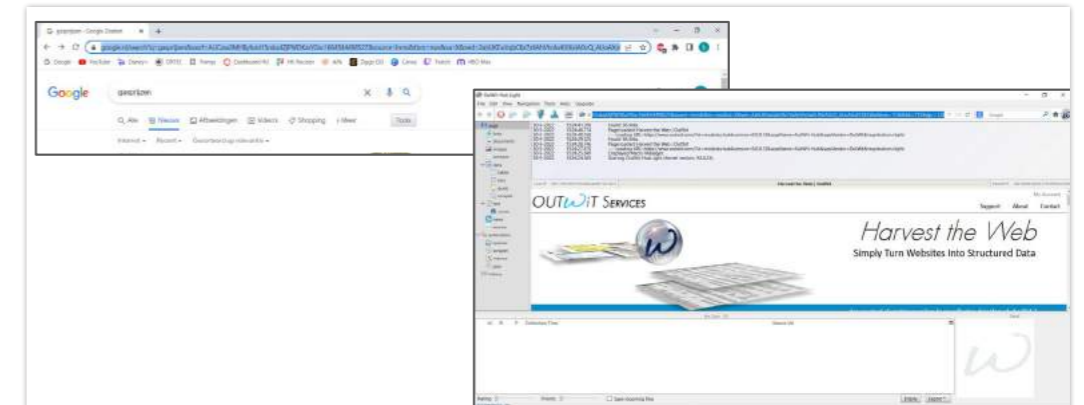


Step 1:
Open google en
OutWit Hub



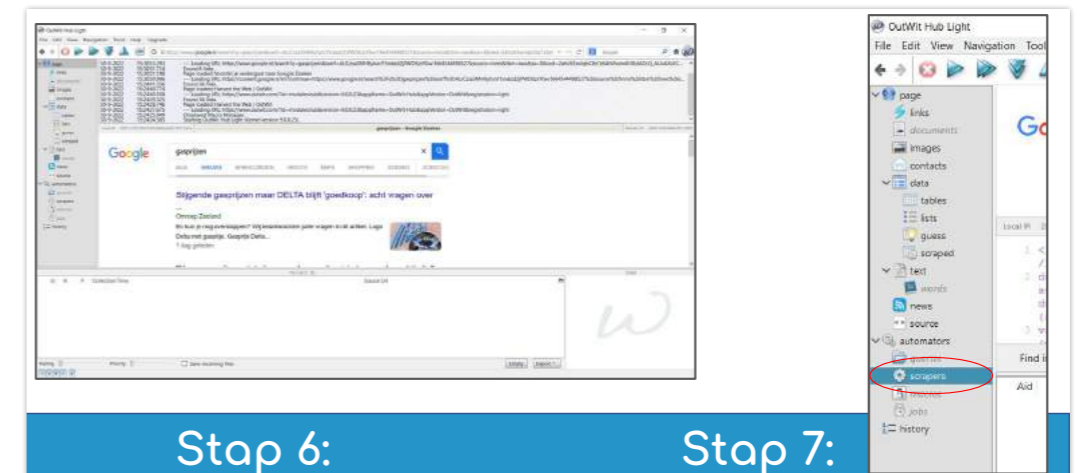
Step 2:
Zoek in google
naar gasrijzen

Step 3:
Ga naar nieuws, recent en
gesorteerd op relevantie



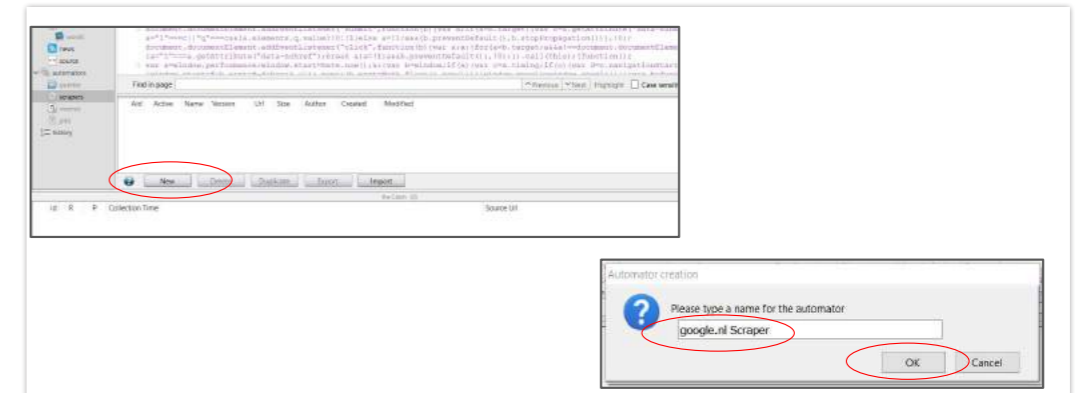
Step 4:
Kopieer de link

Step 5:
Plak de link in OutWit Hub



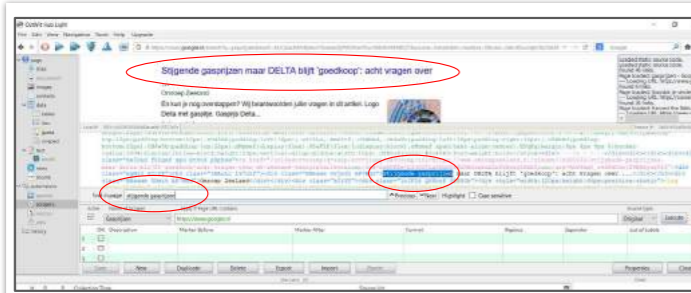
Step 6:
De google pagina
verschijnt in beeld

Step 7:
Klik op scrapers in het
linkermenu

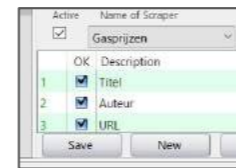


Step 8:
Creëer een 'new'
scraper

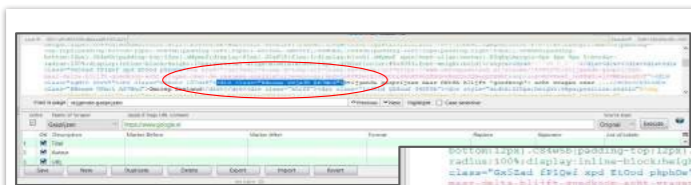
Step 9:
Geef de scraper een naam
en klik op 'ok'



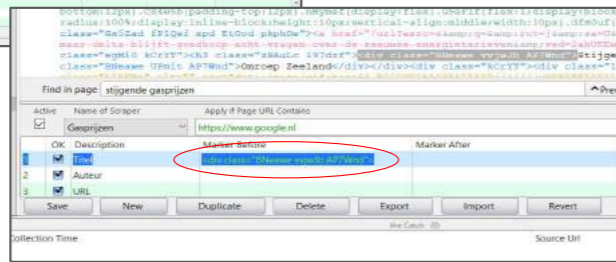
Step 10:
Zoek het eerste
artikel



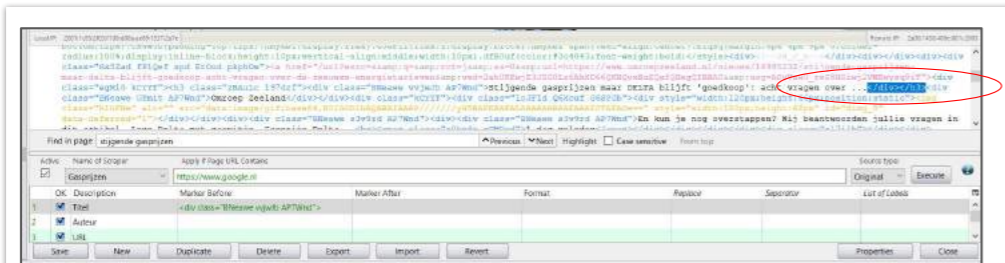
Step 11:
Maak drie onderdelen



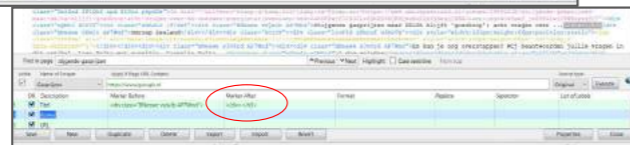
Step 12:
Zoek marker
before titel



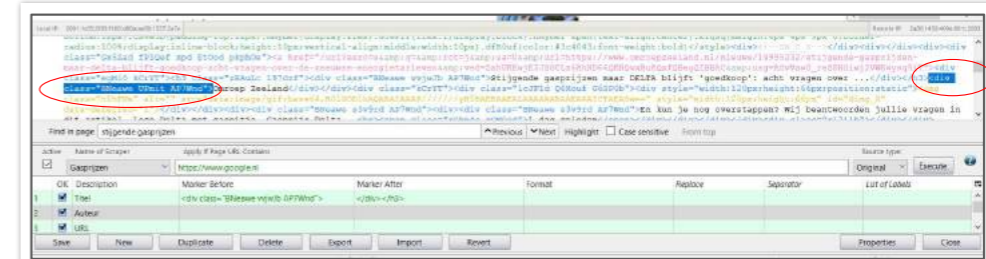
Step 13:
Kopieer en plak
before titel



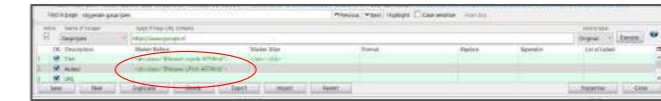
Step 14:
Zoek marker
after titel



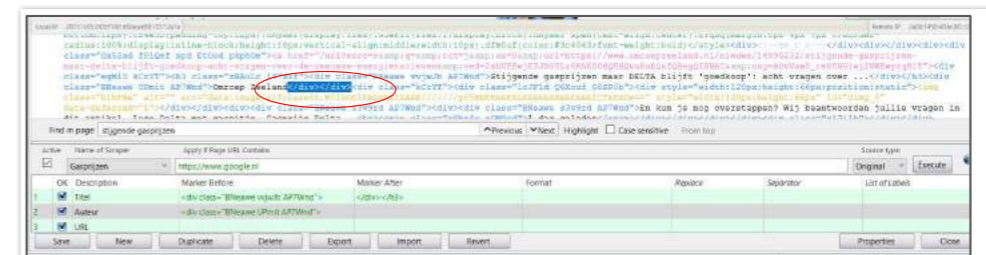
Step 15:
Kopieer en plak
after titel



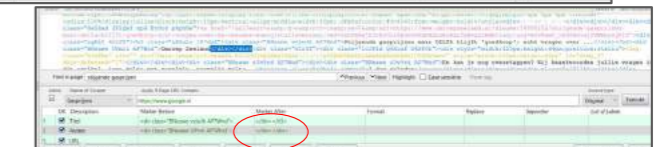
Step 16:
Zoek marker
before auteur



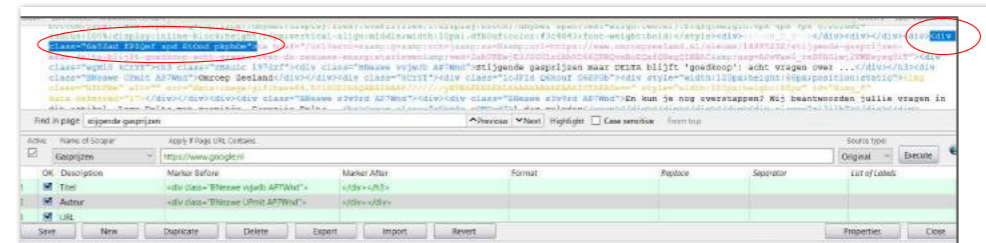
Step 17:
Kopieer en plak



Step 18:
Zoek marker
after auteur



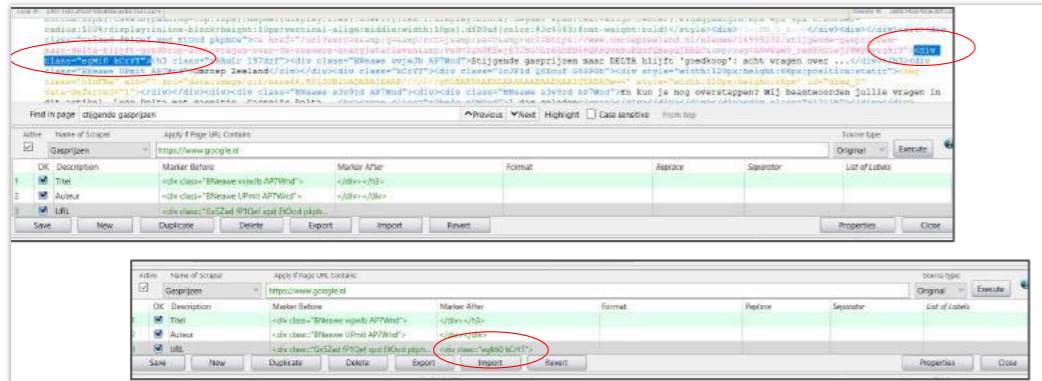
Step 19:
Kopieer en plak



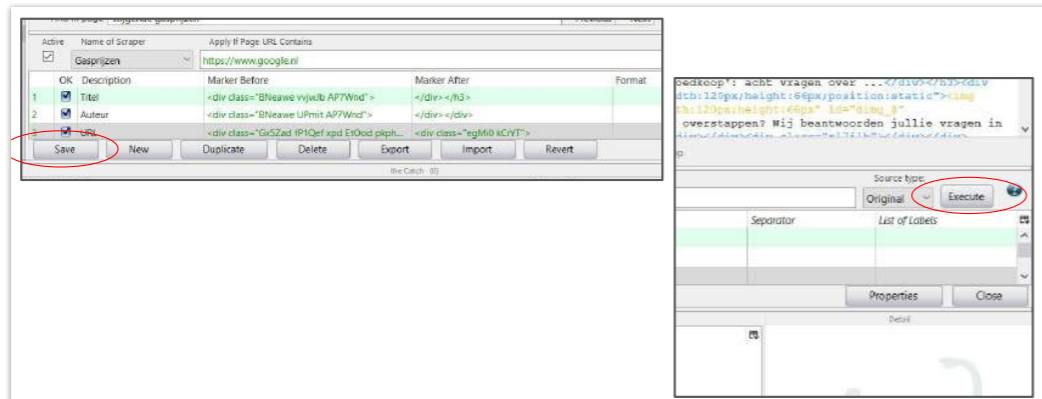
Step 20:
Zoek marker
before URL



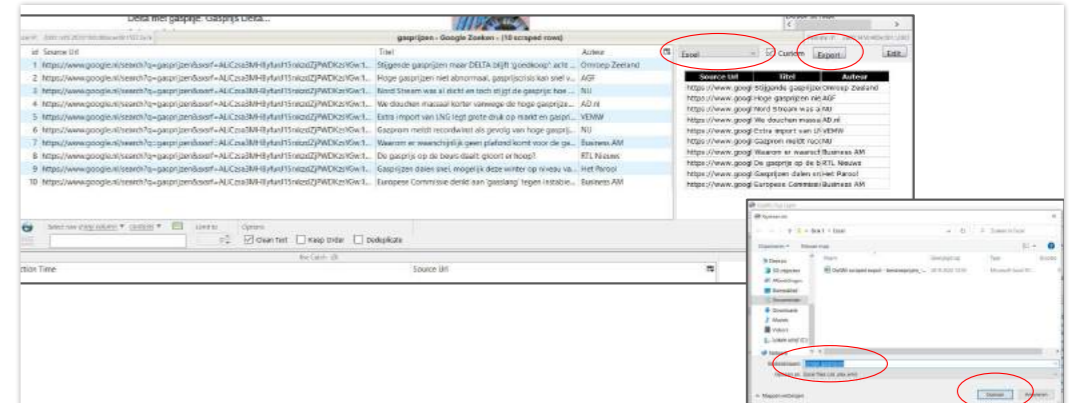
Step 21:
Kopieer en plak



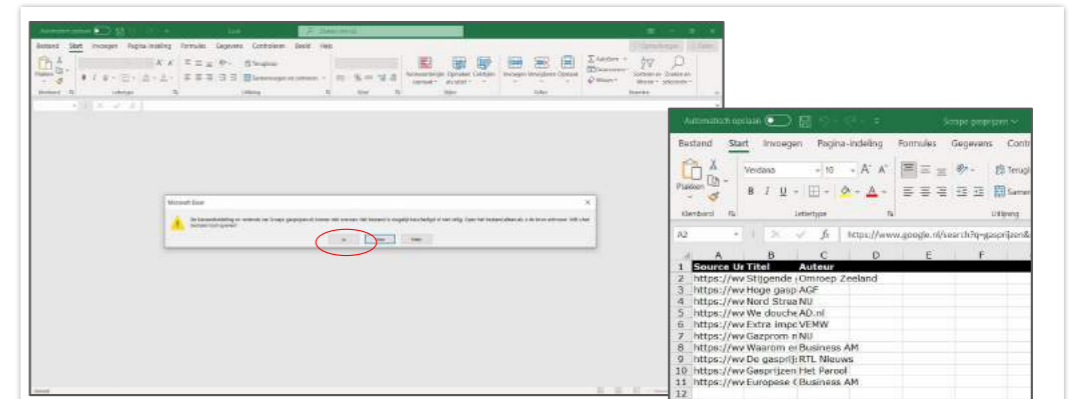
Step 22: Zoek marker after URL
Step 23: Kopieer en plak URL



Step 24: Sla je scraper op
Step 25: Klik op 'execute'



Step 26: Exporteer naar Excel
Step 27: Sla op onder eigen naam



Step 28: Melding Excel klik 'ja'
Resultaat

Hoofdstuk 5.3

Bijlagen

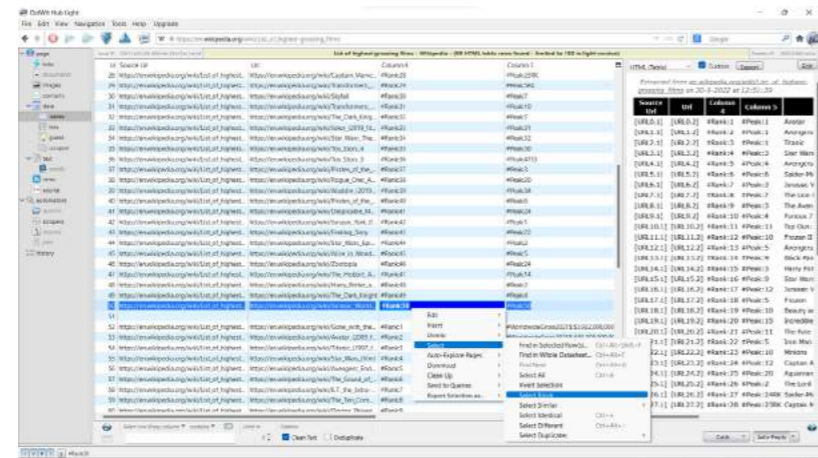
Webpagina naar Excel (Digitale handout)

Nu we hebben uitgelegd hoe het werkt mogen jullie het zelf proberen. In de chat stuur ik een link naar de Wikipedia Pagina met een tabel van de 50 succesvolste films ooit. Aan jullie om dit in een Excelsheet te krijgen zonder te knippen en te plakken.

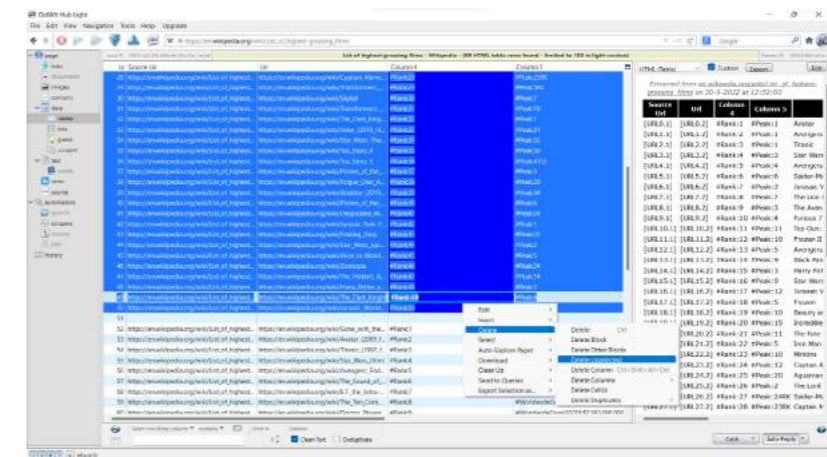
Om een tabel van een webpagina te scraperen moet je als eerste stap de URL van de website in de zoekbalk van het programma plakken. Je klikt op enter en vervolgens klik je links in de toolbalk op 'tables', onder het kopje data.



Vervolgens verschijnen de tabellen. Je ziet dat er nog een aantal andere tabellen op de pagina stonden die het systeem ook heeft meegenomen. Om het wat overzichtelijker te maken kunnen we die alvast weghalen. Dat doe je door met je cursor in het blok te staan dat je nodig hebt en vervolgens op je rechtermuisknop te drukken. Klik op 'select' en daarna 'select block'. Nu is de tabel die je nodig hebt geselecteerd.



Klik nu nog een keer op je rechtermuisknop en klik op 'delete' en vervolgens 'delete unselected'. Nu heb je alleen de juiste tabel nog over.



Nu is het alleen nog een kwestie van exporteren naar Excel. Klik rechtsboven het voorbeeld het dropdown menu aan, en kies voor Excel. Klik vervolgens op exporteren. Nu heb je een bestand met de 50 succesvolste films aller tijden en is het alleen nog een kwestie van de tabel opschonen.

