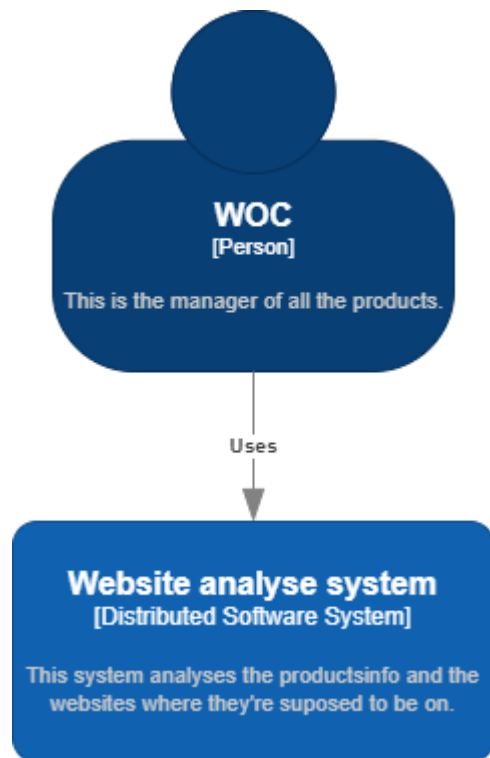


## Contents

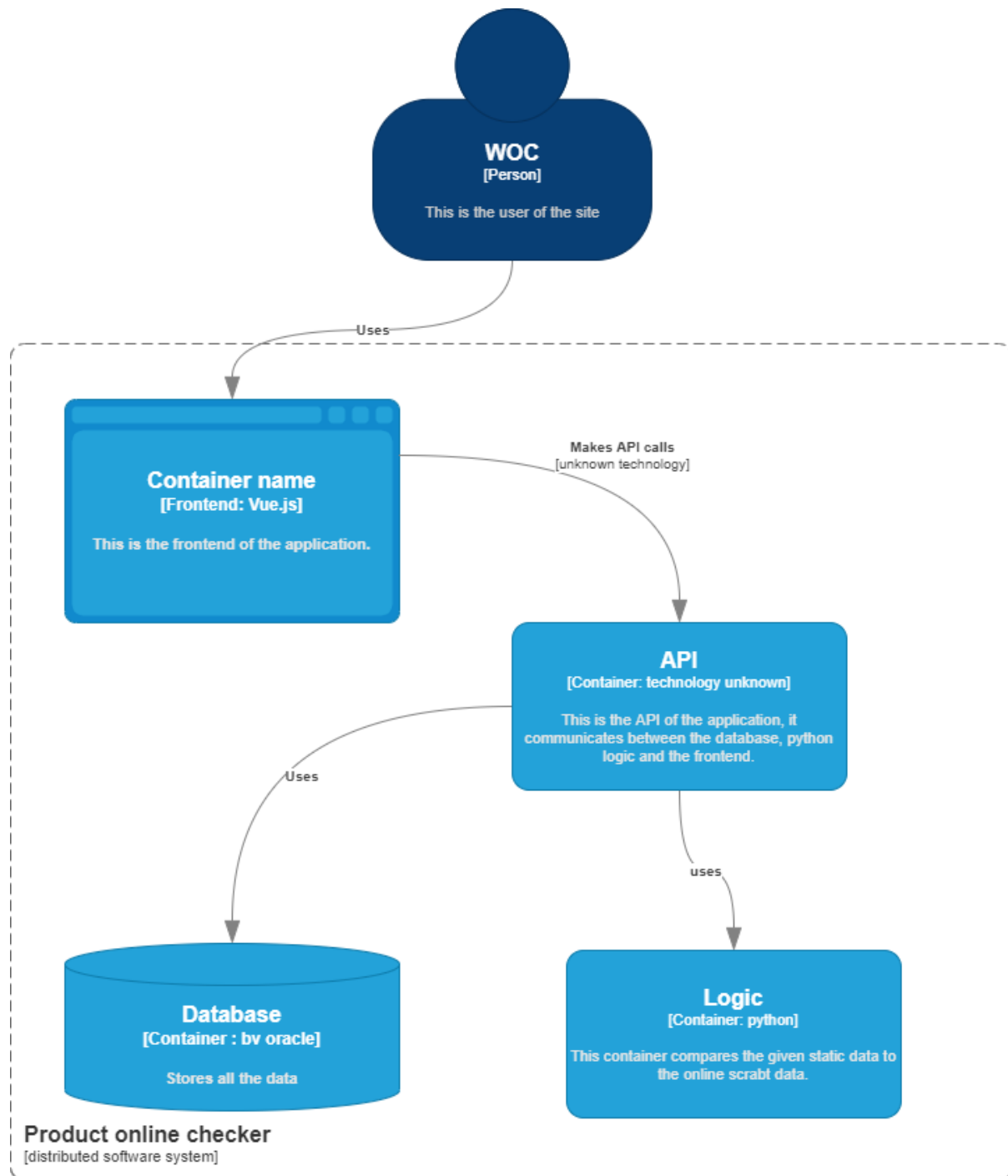
C4 model, layer 1:.....	2
C4 model, layer 2:.....	3
requirements.....	4
User stories.....	5
Story mapping.....	7
UI Analyse .....	8
Database entity relation diagram:.....	9
Business process.....	10
Documentatie voor werkmethode.....	12
Nieuwe IP-adres: IP ROTATION .....	13
Scraping.....	14



## C4 model, layer 1:



## C4 model, layer 2:



## requirements

FN-01 – The program gives a live score which indicates how similar the given data and the online data are.

FN-02 – The program can update the data weekly or on command.

FN-03 – The program checks if the data is live on the website.

FN-04 – The brand can upload an excel document with products that the program will check.

FN-05 – The brand can add retailers. Which the program will check in the weekly update.

NFN-01 – The technology used has to be still maintained and modern.

NFN-02 – In the frontend and backend there only can be one language.



WhiscoSoftware

# User stories

As a..

I want to..

So that...

Supplier	See my live scores overall	Supplier can see how many of my products are actually live
----------	----------------------------	--

Supplier	sort by retailer on overall products	I can see how much one retailer is showing my content live
----------	--------------------------------------	--

Supplier	See a dashboard with my products	I can see my live product score
----------	----------------------------------	---------------------------------

Supplier	search for a specific product	I can view one product
----------	-------------------------------	------------------------

Supplier	click on one product	I can see the score for one product
----------	----------------------	-------------------------------------

Supplier	Filter through my products ( percentages, high to low, low to high )	I can have an overall view of which of my products are mostly live and which are not
----------	--	--

Supplier	Sort my product scores by ( retailer )	I can see how much one retailer is showing my content live
----------	--	--



WhiscoSoftware

Supplier	See my live scores individually based on one product	Supplier can see how many of my advertisements are live on retailers websites
----------	--	---

Supplier	be able to add new products to my live score	my score is up to date with all my products score
----------	--	---

Supplier	Insert my excel sheet with the new product	the application can find the right information so the score is correct
----------	--	--

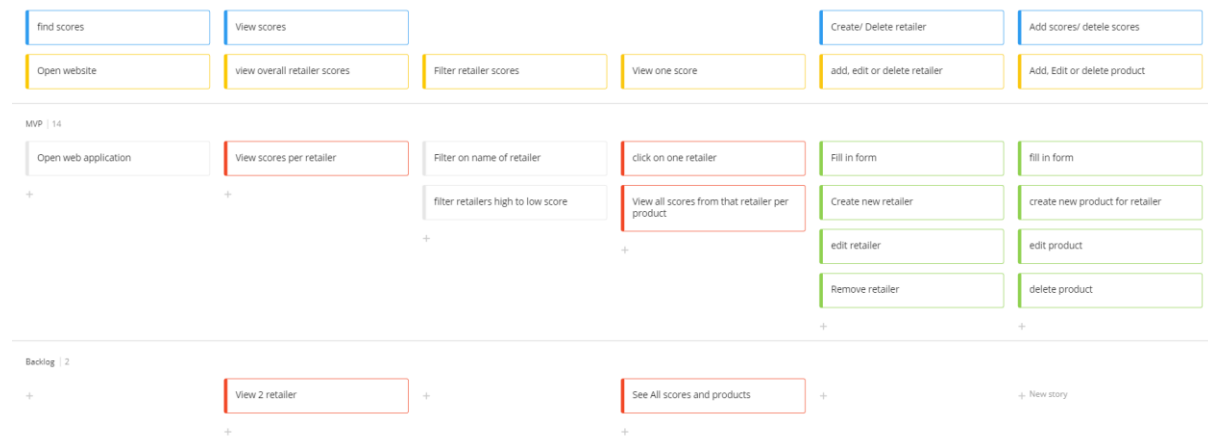
Supplier	Remove products	my score is up to date with all my products score
----------	-----------------	---

Supplier		
----------	--	--



WhiscoSoftware

# Story mapping




WhiscoSoftware


# UI Analyse

Voorbeeld:



search






Appel

Bulgaria


15%



Appel

Bulgaria


15%



Appel

Bulgaria


15%



Appel

Bulgaria


15%



Appel

Bulgaria

15%



Appel

Bulgaria

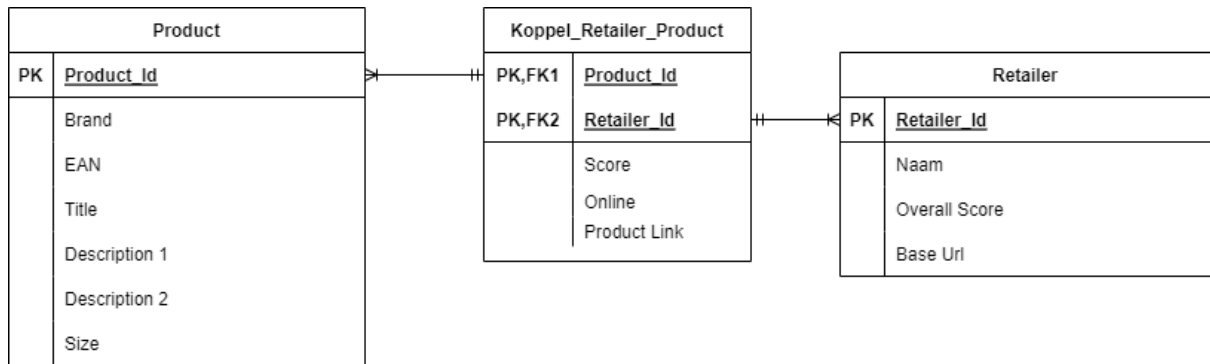
15%



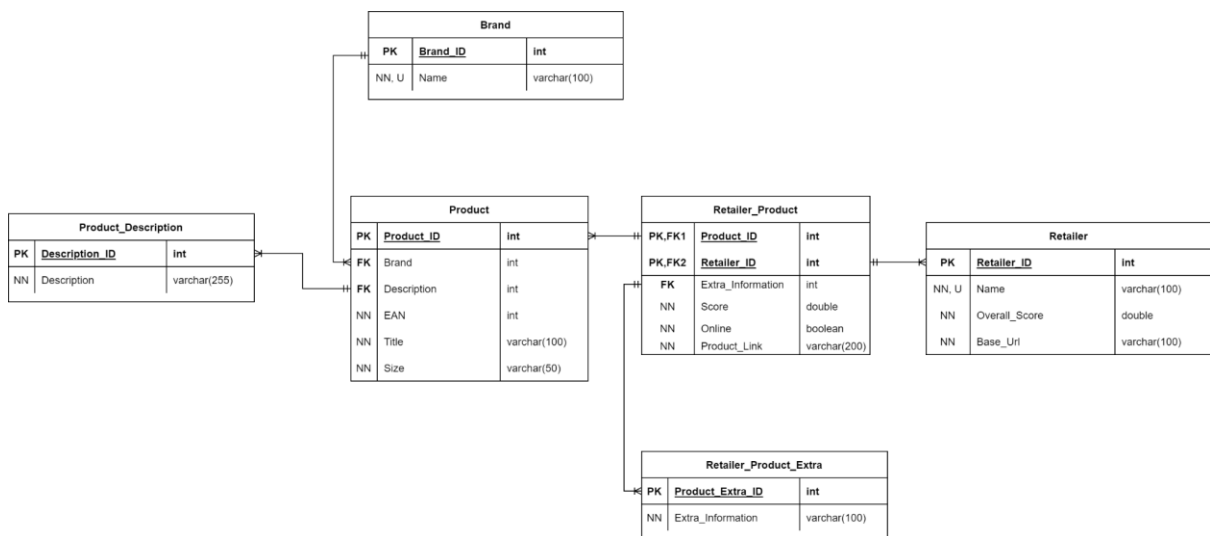
WhiscoSoftware



# Database entity relation diagram:



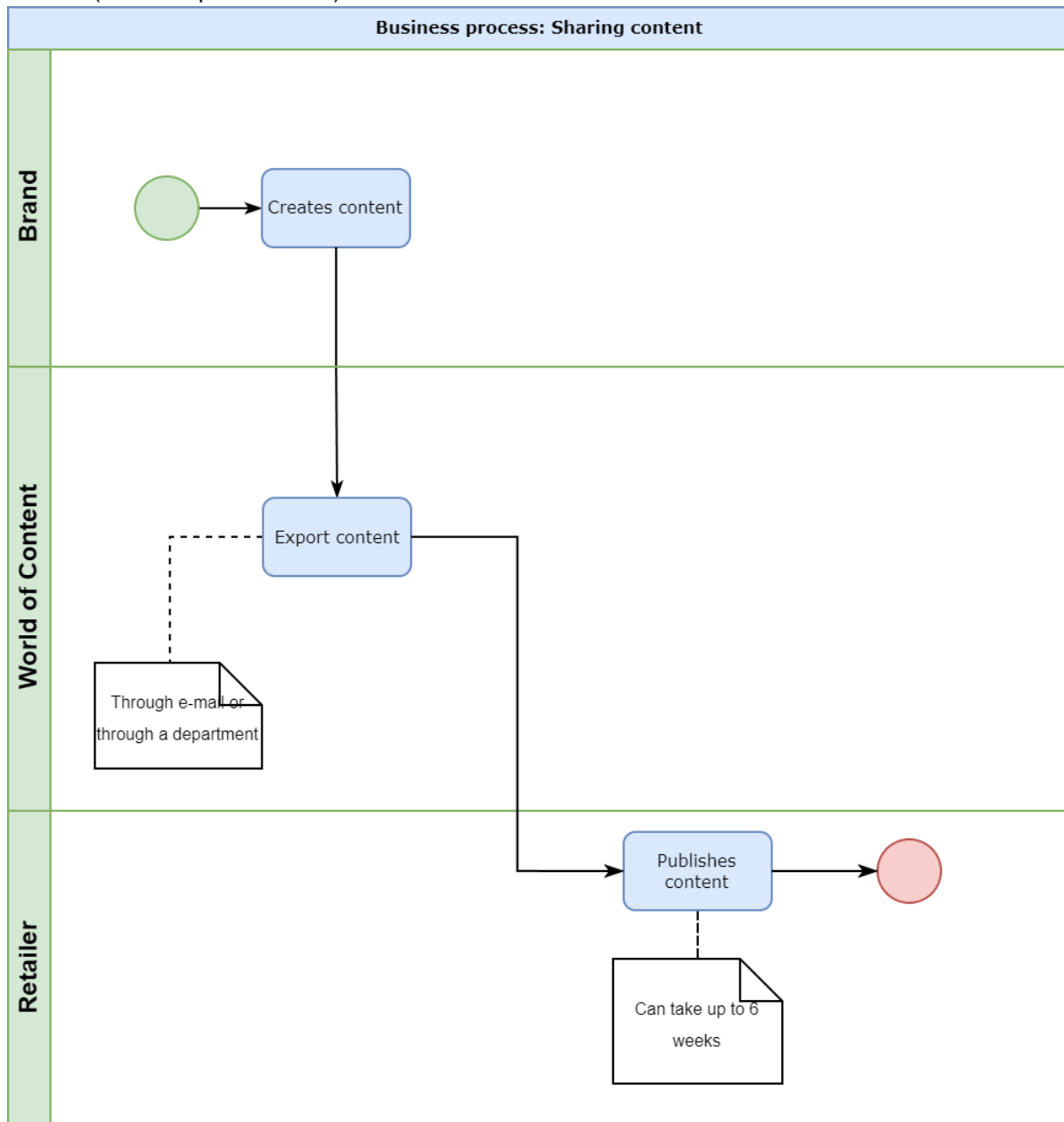
v1.1



WhiscoSoftware

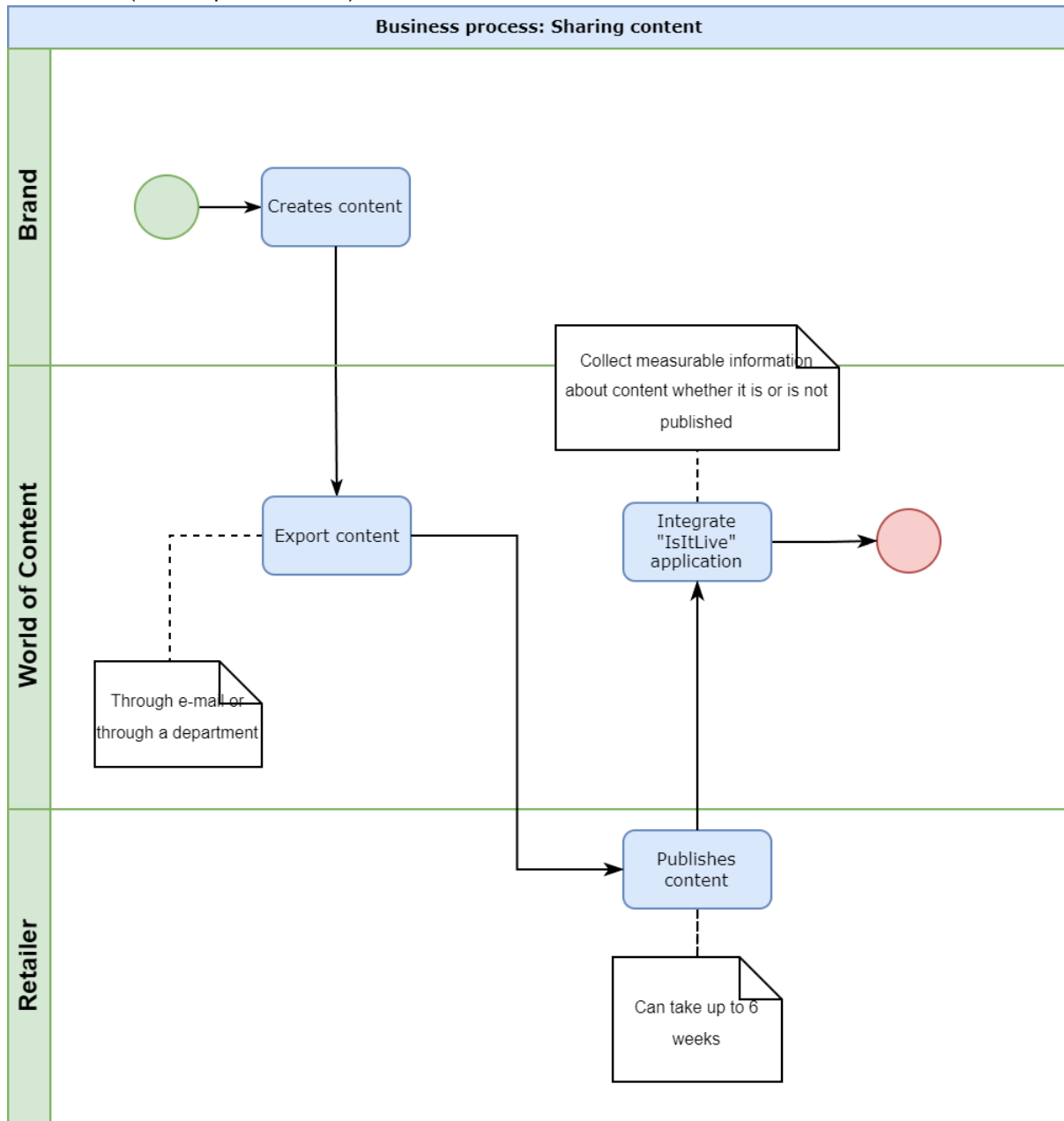
# Business process

Current (before implementation)



WhiscoSoftware

Automated (after implementation)



WhiscoSoftware

# Documentatie voor werkmethode

Er zijn verschillende manieren waarop dit project uitgewerkt kan worden. Daarom hebben wij vaak de vraag gehad waarom wij deze werkmethode hebben gekozen en hoe we die kunnen onderbouwen. Hiervoor hebben wij gedocumenteerd welke werkmethode wij hebben gekozen en welke voor en nadelen daarbij horen. Zo hebben wij een beter beeld bij de zwakke punten die onze applicatie heeft en kunnen we die versterken door er een oplossing bij te verzinnen.

## Gebruiken google zoekfunctie

- Een groot probleem bij het gebruiken van de google zoekfunctie was dat wij hierdoor niet een accuraat zoek patroon hadden. Omdat er een kans was dat wij de verkeerde website als resultaat zouden krijgen.
- Hiervoor hebben wij wat oplossingen bedacht om te zorgen dat de zoekmachine wat accuratere resultaten zou krijgen.
- Een van die oplossingen is bij het aanmaken van een nieuwe retailer wordt er gevraagd naar de basis URL van hun bedrijf. Om te kijken of de link wel overeenkomt met de website die onze scraper terug geeft.
- 

## Cookies



WhiscoSoftware

## Nieuwe IP-adres: IP ROTATION

- een probleem waar wij tegenaan liepen was dat het IP geblokkeerd werd doordat wij veel aan het scrapen waren op pagina's. Door wat onderzoek te doen hebben we wat oplossingen gevonden zodat we niet meer geblokkeerd worden door de website.

<https://www.scraperaapi.com/blog/5-tips-for-web-scraping/>

<https://www.scraperaapi.com/blog/free-shared-dedicated-datacenter-residential-rotating-proxies-for-web-scraping/>

<https://www.codementor.io/@scrapingdog/10-tips-to-avoid-getting-blocked-while-scraping-websites-16papipe62>

opties:

1. <https://www.scraperaapi.com/pricing/> -> deze scraper API zou het blokkeren van het IP tegengaan door een roterende IP. daarnaast bevat deze API ook andere hulpmiddelen. Het nadeel van deze optie is dat je op een account 1,000 gratis API credits per maand krijgt. Met ons project betekent dit dat het bedrijf (WoC) zeker moet betalen om deze oplossing te gebruiken. Aangezien World Of Content meer dan 1,000 klanten heeft en daarvoor meerdere API calls wil maken. Het voordeel is dat met deze API het een eenvoudige manier is om IP rotation staande te krijgen en te onderhouden.
2. <https://free-proxy-list.net/> -> Ons 2de idee was om een scraper te bouwen die een website met gratis proxies scraped. Om te zoeken naar een IP adres die niet geblokkeerd zou worden wanneer wij de calls willen uitvoeren voor het scrapen van producten. De website die we hebben uitgekozen ververs de Proxies elke 10 min. De nadelen hiervan zijn dat als de website niet werkt onze applicatie niet meer werkend is. Daarnaast als de scraper geen werkende IP adres zou vinden houdt het ook op. In het kort het is goedkoper maar minder betrouwbaar.
3. Eigen proxy server maken -> Een groot nadeel is is dat wij weinig kennis hebben over dit onderwerp en dat het daarom ook veel tijd gaat kosten. Het voordeel is is dat deze manier goedkoop en betrouwbaar



WhiscoSoftware

- [5 Tips For Web Scraping Without Getting Blocked/Blacklisted](#) 10: "Scrape Out of the Google Cache" → Dit is de laatste optie voor als alle andere opties niet werken. Dit is vooral omdat deze optie niet compleet betrouwbaar is. Via deze manier gebruikt onze applicatie niet rechtstreeks de website van de retailer, maar een zogenaamde 'snapshot' van een pagina, meestal een dag eerder gemaakt.

Voordelen:

- IP-adres zal niet worden geblokkeerd. We kunnen de pagina's zo vaak scrapen als we willen.

Nadelen:

- Informatie kan gedateerd zijn.
- Deze vorm werkt niet bij alle retailers (zie Jumbo).

wat is een proxy:

een proxy is een middenpersoon tussen jou en het internet. Deze computer stuur jij je data naartoe waarna de proxy het naar de eindbestemming. Een proxy server verbegit gedeeltelijk of volledig de persoonlijke informatie van de gebruiker. De p



## Scraping

Omdat wij een scraping tool gebruiken weten wij dat deze ook accurater gemaakt kan worden door dingen toe te voegen. Wij willen onze scraper specifieker laten zoeken zodat hij vaker het goede vergelijkt. hiervoor hebben wij wat stappen ondernomen waar de scraper doorheen gaat.

- zinnen
- lidwoorden weglaten
- woorden
- als het minder dan de helft is, maken we er 0% van.



WhiscoSoftware