# Functionele analyse

“Van Havermaet heeft enorm veel (meerdere terabytes) ongestructureerde data. Een deel van deze data is manueel gestructureerd gemaakt, het grootste deel is dat echter nog niet. De opdracht van jullie bestaat er uit om een (Proof of Concept) PoC te ontwikkelen om op een visuele manier de juiste documenten terug te vinden. In een eerste fase is het de bedoeling om een tussentaal te definiëren zodat er efficiënt gezocht kan worden naar de juiste documenten en dat links tussen verschillende documenten gedefinieerd kunnen worden. In een tweede fase is het vervolgens de bedoeling om deze links te visualiseren en ervoor te zorgen dat er in deze visuele setting eenvoudig naar documenten gezocht kan worden. Bij onderstaand voorbeeld FOODPF.html is er een visualisatie, maar het zoeken naar de juiste informatie is niet gemakkelijk door o.a. leesbaarheid en hoeveelheid van links. Het is van belang om gedurende de gehele opdracht rekening te houden met de performantie aangezien het gaat om een enorme hoeveelheid aan documenten gaat.”

**Problemen:**

* Ongestructureerde data, waarvan een deel manueel ingevoerd
* Moeilijk om documenten terug te vinden
* Opvraging data
* JSON – String

**Oplossing:**

* Visualisatie op uniforme manier, met informatie over entiteiten en hun relaties
* Het visualiseren van de toegangsrechten uit de azure active directory en SQL.
* Inhoud van de JSON bestanden omzetten naar entiteiten en deze met samen met de originele sql databank objecten in een nieuwe databank plaatsen
* Informatie over bestanden ophalen via sharepoint API calls en in DB opslaan
* Het gebruiken van een key-woord dat niet in een tekst voorkomt als zoekterm. (topic modelling)
* Bij het zoeken rekening houden met de afgelegde weg en op basis hiervan documenten voorstellen.

**Technologieën:**

* Microsoft Azure: Verwerking van documenten
* Azure Active Directory
* Sharepoint: Opslag van de documenten
* Sharepoint API
* SQL & Json (en ETL)
* D3js voor visualisatie

**Groepsrollen:**

* Projectleider: Arin Er
* Contactpersoon: Arin Er
* Hoofdprogrammeur: Stef Michiels
* Database verantwoordelijke(n) Pieter-Jan Van Der Schueren, Stijn Saenen
* Sharepoint en AAD verantwoordelijke: Pieter-Jan Van Der Schueren
* Scrum-master: Stijn Saenen

