

Onderzoeksrapport

Data van de landelijke bibliotheken verwerven en analyseren.

Onderzoeksrapport van Arnoud Aarnoudse
Studentnummer 0048940
In het kader van de opleiding HBO-ICT aan de
HZ University of Applied Sciences, Vlissingen

Datum: 12-02-2016

Onderzoeksrapport

Data van de landelijke bibliotheken verwerven en analyseren.

Auteur: Arnoud Aarnoudse
Studentnummer: 0048940
E-mailadres: justinaarnoudse@hotmail.com

Opleiding: HBO-ICT
Onderwijsinstelling: HZ University of Applied Sciences, Vlissingen
Eerste begeleider: W. Everse
Tweede begeleider: Nader te bepalen.
Stagebedrijf: Kenniscentrum Expertise- en Valorisatiemanagement
Stagebegeleider: H. de Bruin

Onderzoeksrapport
Inleverdatum: 12-02-2016
Stageplaats: Vlissingen

Versiebeheer

Versie	Auteur	Omschrijving
0.1	A.J.L	Eerste concept van het rapport.
0.2	A.J.L	Verbeteren inhoud van het rapport.
0.3	A.J.L	Verwerken feedback, verbeteren van elk hoofdstuk.
1.0	A.J.L	Eerste versie van het rapport.
1.1	A.J.L	Verbeteren rapport.
2.0	A.J.L	Tweede versie van het rapport.

Samenvatting

Het lectoraat DIO zocht naar een bron waar veel data uit gehaald kon worden voor hun zoekmethode, en kwam hierbij uit op de landelijke bibliotheken. De bibliotheken bevatten een grote hoeveelheid data die gebruikt kan worden om deze zoekmethode mee aan te vullen, en heeft ook een API beschikbaar. Deze API maakt het mogelijk om met het systeem van de landelijke bibliotheken te communiceren waardoor de bij hun beschikbare data opgevraagd kan worden. Het lectoraat zou de API willen gebruiken om er zoekopdrachten op uit te voeren, om zo meer data te verkrijgen voor hun eigen zoekmethode. Voor deze zoekopdrachten werden enkele zoektermen gebruikt, maar het was ook de bedoeling dat er een uitgebreide zoekopdracht uitgevoerd kan worden die meerdere zoektermen bevat. De volgende vraag stond centraal tijdens het onderzoek: “Welke kwalitatieve impact heeft het uitbreiden van de zoektermen met behulp van de grammatica der concepten op de ontsloten informatie van de API van de landelijke bibliotheken?”.

Dit onderzoek werd uitgevoerd met als doel om deze vraag te beantwoorden. Tijdens dit onderzoek werd de bestaande informatie over de API onderzocht. Zo werd de bestaande documentatie onderzocht, en werd de werking van de demo getest. De werking van de demo werd vergeleken met de beschrijvingen van de documentatie, waaruit bleek dat de documentatie klopte, al was deze nog niet volledig. Daarom werd er ook gecommuniceerd met de ontwikkelaar van de API om de onduidelijkheden te verhelderen. Vervolgens is de werking van de API ook daadwerkelijk getest door middel van verscheidene proof of concepts. Voor het ontwikkelen van deze proof of concepts werd er een library geselecteerd om de resultaten van de API te verwerken. Uiteindelijk bleek dat de functionaliteiten van de API en de geselecteerde library in de proof of concepts werkten, maar niet in samenwerking. Zo kon de library niet alle resultaten van de API verwerken.

Daarna werd er een experiment uitgevoerd met behulp van de demo van de API. Tijdens dit experiment werd er gezocht met een enkele zoekterm, en met een uitgebreide zoekopdracht met meerdere zoektermen. Om te kunnen bepalen welke kwalitatieve impact er plaatst vond moest er eerst bekend zijn welke eisen het lectoraat heeft voor het begrip kwaliteit. Daarom is er om deze definitie te achterhalen een open interview gehouden met de begeleider van het bedrijf.

Aan de hand van het open interview kon geconcludeerd worden dat de student en de begeleider het eens waren over wat de definitie van kwaliteit was, en dat de student de besproken definitie kon gebruiken voor het experiment. Uit het experiment was ook af te leiden dat er inderdaad een positieve kwalitatieve impact plaats vond door het uitbreiden van de zoekterm. De student heeft uiteindelijk ook met de ontwikkelaar van de API gesproken, en deze gaf een positief antwoord betreffende het ondersteunen van het lectoraat, mocht het lectoraat wensen de API te implementeren. Het advies van de student is om dit zeker te doen, aangezien er aangetoond is dat er een positieve kwalitatieve impact is, en het een goede ondersteuning zou zijn voor de zoekmethode van het lectoraat.

Summary

The lectureship DIO was in search of an extra data resource for their search method and ultimately decided that the national libraries would fit their requirements. The national libraries contain a large amount of data which can be used to supplement their search method and has an API available. This API communicates with the system of the national libraries which allows the lectureship DIO to obtain their data by sending their search requests. For the search query it needs to be possible to enter a single term while it should allow a possibility to expand it to multiple terms as well. The central question for this research was the following: "What qualitative impact will expanding the search terms, by using the grammar of concepts, have on the obtained information from the API of the national libraries?". The purpose of this research is to find out if extending the search terms on the available information of the lectureship will result in a qualitative impact on the result.

The research has been completed in order to answer this question. The current existing information has been researched during this study. The available documentation of the API has been read and the functionality of the demo has been tested. Thereafter the functionality of the API was assessed in a proof of concept to see if it corresponded with the actual documentation. The API developer was contacted after finding out that the documentation couldn't answer certain questions. The research of the current existing documentation proved to be correct even though it was not complete. The API was then used in multiple proof of concepts to research the usability. A library was selected for these proof of concepts to process the results of the API. The functionality of the API and the library were tested after this was proven, and proved to be working as documented. However there was a problem that occurred when these components were put to the test in cooperation with each other and they refused to function properly.

After having researched the functionality of the API an experiment was held by using the demo. During this experiment a search query was completed using a single search term and using an extended search term. To be able to determine the exact meaning of qualitative impact it was required to discover what this term meant for the lectureship DIO. An open interview with the mentor of the business was required to determine the definition. Following after, the definition was used to compare the results of the single search term with the results of the extended search term to determine if there actually was a qualitative impact.

On the basis of the open interview it was possible to conclude that the student and the mentor agreed on what the definition of a qualitative impact should be. The student was then allowed to use this definition during the experiment. As a conclusion to the experiment there had been proven to be a qualitative impact upon expanding the search terms. The student also spoke with the developer of the API who gave positive feedback with regards to supporting the lectureship. He has since reassured the lectureship that if they need any help with implementing the API that he would be more than happy to assist. The final advise of the student would be to certainly take up on this offer since there has been proven to be a qualitative impact. It certainly would be a great addition to the search method of the lectureship DIO.

Voorwoord

Tijdens het afstuderen heb ik veel kunnen leren over nieuwe methoden, die volgens mij in de toekomst zeer relevant zullen worden voor het gebruik van data, en het uitvoeren van een onderzoek. Gedurende het uitvoeren van het onderzoek ben ik veel obstakels tegen gekomen, maar gelukkig had ik genoeg doorzettingsvermogen. Hierdoor was het mogelijk om door te gaan met het onderzoek, en niet vast te lopen. Ook zou ik mijn begeleiders Mischa Beckers, Wouter Everse en Hans de Bruin willen bedanken voor hun begeleiding.

Vervolgens zou ik ook graag mijn collega's willen bedanken voor het verminderen van de druk van het afstuderen. Het hielp enorm dat ik in een prettige werksfeer kon werken, en waar ik mijn medewerkers ook om advies mocht vragen.

Uiteindelijk zou ik ook mijn vrienden en familie willen bedanken voor alle steun die ik ontvangen heb tijdens dit traject. Zo wil ik Leo Aarnoudse en Anja Aarnoudse, mijn ouders, bedanken voor de steun thuis die ik sinds het begin van deze opleiding, maar vooral in deze laatste fase van ze heb mogen ontvangen. Daarbij zou ik graag Ardjan Boeije en mijn overige klasgenoten willen bedanken voor al het advies dat ik gekregen heb m.b.t. het afstuderen. Tot slot wil ik ook mijn vrienden Sebastian Gevers en Charlotte Reiniers, en mijn oom Adri Aarnoudse bedanken voor hun medewerking en ondersteuning tijdens het schrijven van dit onderzoeksrapport.

Vlissingen, 12-02-2016

Inhoud

1	Inleiding.....	1
1.1	Aanleiding.....	1
1.2	Probleemstelling.....	2
1.2.1	Doelstelling	3
1.2.2	Vraagstelling	4
1.3	Theoretisch kader	5
2	Methoden en materialen	11
2.1	Deelvraag 1 – Onderzoeken API	11
2.1.1	Meetinstrumenten en operationalisatie.....	11
2.1.2	Analysemethoden	11
2.2	Deelvraag 2 – Gebruiken API	12
2.2.1	Meetinstrumenten en operationalisatie.....	12
2.2.2	Analysemethoden	13
2.3	Deelvraag 3 – Uitbreiden zoekopdracht	13
2.3.1	Meetinstrumenten en operationalisatie.....	13
2.3.2	Analysemethoden	14
2.4	Deelvraag 4 – Onderzoeken impact.....	15
2.4.1	Meetinstrumenten en operationalisatie.....	15
2.4.2	Analysemethoden	16
3	Resultaten	17
3.1	Deelvraag 1 – Onderzoeken API	17
3.1.1	Resultaten	17
3.1.2	Analyse.....	20
3.2	Deelvraag 2 – Gebruiken API	21
3.2.1	Resultaten	21
3.2.2	Analyse.....	23
3.3	Deelvraag 3 – Uitbreiden zoekopdracht	23
3.3.1	Resultaten	23
3.3.2	Analyse.....	24
3.4	Deelvraag 4 – Onderzoeken impact.....	24
3.4.1	Resultaten	24
3.4.2	Analyse.....	27
3.5	Totaalbeeld	27

4	Discussie	28
4.1	Deelvraag 1 – Onderzoeken API	28
4.1.1	Discussie	28
4.1.2	Deelconclusie	28
4.2	Deelvraag 2 – Gebruiken API	29
4.2.1	Discussie	29
4.2.2	Deelconclusie	30
4.3	Deelvraag 3 – Uitbreiden zoekopdracht	30
4.3.1	Discussie	30
4.3.2	Deelconclusie	30
4.4	Deelvraag 4 – Onderzoeken impact.....	31
4.4.1	Discussie	31
4.4.2	Deelconclusie	32
5	Conclusie	33
5.1.1	Vergelijking met ander onderzoek/theorie.....	33
5.1.2	Suggesties voor vervolgonderzoek	33
5.1.3	Tot besluit	34
	Literatuur	35
	BIJLAGEN	36
	Bijlage 1 - Requirement Specificatie.	37
	Bijlage 2 - Functioneel ontwerp.....	40
	Bijlage 3 – Technisch ontwerp Proof of Concepts.	47
	Bijlage 4 – Opzet Experimenten.	53
	Bijlage 5 – Conclusie experimenten.....	61

1 Inleiding

Dit onderzoeksrapport beschrijft het afstudeeronderzoek dat is uitgevoerd voor het lectoraat DIO¹(Duurzaam Innoveren en Ondernemen). De volgende beschrijving vat goed samen wat er belangrijk is voor het lectoraat:

“Valorisatie is de kern business van het kenniscentrum DIO. Of anders gezegd, de creatie van een dynamisch vermogen en duurzaamheidswaarde voor personen (hieronder vallen studenten, docenten, werkenden) en organisaties in de regio Zeeland door efficiënt (her)gebruik van kennis in een multidisciplinaire omgeving.” (Kenniscentra, 2014).

Oftewel, het omzetten van kennis naar commerciële producten of diensten, en er voor zorgen dat deze kennis en middelen ook toegepast kunnen worden. Het lectoraat is onderdeel van het kenniscentrum DIO, en voert niet alleen opdrachten uit voor klanten, maar ook voor zichzelf. Werknemers, docenten en studenten kunnen hier werken.

1.1 Aanleiding

Het lectoraat heeft momenteel een zoekmethode die werkt zoals de meeste andere zoekmethoden. In de zoekopdracht worden er woorden ingevuld, en vervolgens zoekt de zoekmethode waar deze woorden voorkomen in de tekst. Dit geeft alleen zoekresultaten waar dit woord in voorkomt, en niet resultaten die te maken hebben met dit woord zonder dat het woord er letterlijk in voor komt. Ook kan er niet gezocht worden naar relaties tussen begrippen. De zoekmethode herkent geen relaties, of bijbehorende woorden.

Het lectoraat werkt daarom aan een nieuwe zoekmethode die het mogelijk zal maken om bij het uitvoeren van een zoekvraag een zeer gepast resultaat te geven in de juiste context. Hiermee wordt bedoeld dat de zoekmethode zal zoeken naar wat voor relaties de gezochte begrippen hebben, en niet alleen waar de tekst hetzelfde woord bevat. Om dit te bereiken is het lectoraat de zoekmethode aan het uitbreiden en verbeteren. Dit zal gedaan worden door een nog te ontwikkelen zoekmethode. Deze zal databronnen doorlezen en automatisch relaties leggen tussen bepaalde begrippen met behulp van SKOS(Simple Knowledge Organization System) en andere methoden. De SKOS is een W3C standaard voor het vastleggen van een thesauri. (aisaac, 2012) Ook zal de zoekmethode naar deze al bestaande relaties zoeken. Deze manier van zoeken maakt het een semantische zoekmethode. Dit is omdat de zoekmethode zoekt met behulp van de semantiek tussen de zoektermen.

Op deze wijze kan er met behulp van meerdere bronnen een verband gelegd worden tussen meerdere begrippen, waardoor de context van de informatie verbeterd wordt. Momenteel heeft het lectoraat beschikking over een aantal thesauri waar al begrippen in beschikbaar zijn. Deze thesauri werken samen als een semantisch netwerk om de begrippen met elkaar te verbinden in een context.

¹ Tijdens het afronden van het onderzoek is deze naam veranderd, de nieuwe naam is nu “Kenniscentrum Expertise- en Valorisatiemanagement”. Voor het rapport wordt nog wel het lectoraat aangehouden.

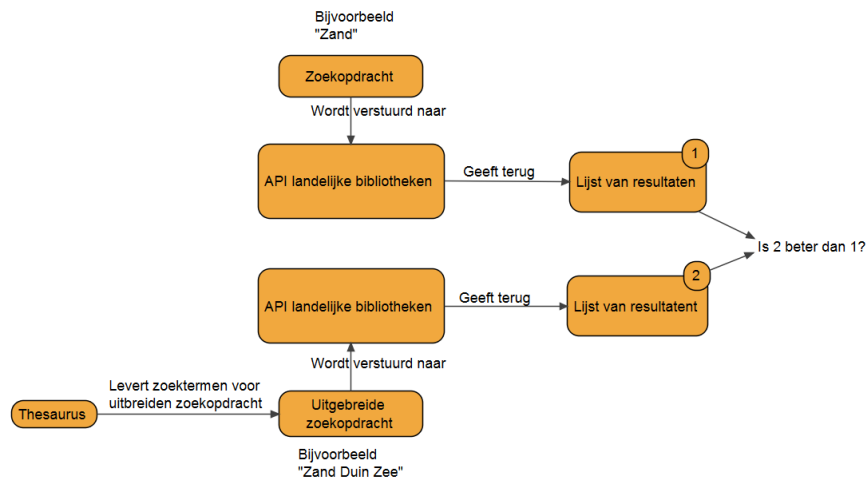
Door de zoekmethode te verbeteren zouden er bij deze thesauri nieuwe begrippen en nieuwe relaties tussen de begrippen ontdekt en gecreëerd kunnen worden. De te ontwikkelen methode zal gebruikt worden om de context van de informatie te verbeteren.

Het lectoraat zocht naar een nieuwe bron naast de thesauri waar veel data uit gehaald kan worden, omdat de te ontwikkelen zoekmethode relaties in kaart wil gaan brengen op basis van grote hoeveelheden data, en daarom dus veel databronnen moet hebben. Door het zoeken kwam het lectoraat uit op de landelijke bibliotheken van Nederland. De landelijke bibliotheken bevatten veel data die gebruikt kan worden om de bronnen van de zoekmethode van het lectoraat aan te vullen. Nu is het zo dat de landelijke bibliotheken een API(Application Program Interface) ontwikkeld hebben. Deze API maakt het mogelijk om met het systeem van de landelijke bibliotheken te communiceren waardoor er data opgehaald kan worden. Dit betekent dat hier heel veel data verkrijgbaar is. Nu is het aan de student om deze API en data te onderzoeken voor de zoekmethode. Om de API gemakkelijker bruikbaar te maken voor het lectoraat zou het handig zijn als de student uiteindelijk de bron via een prototype aan kan laten spreken, maar dit is niet kritiek voor het slagen van het onderzoek.

1.2 Probleemstelling

Het lectoraat wil de data van de bibliotheken beschikbaar maken voor de semantische zoekmethode. Hiermee wil het lectoraat meer data beschikbaar stellen voor hun zoekmethode, waardoor er meer informatie ontstaat. Door het verkrijgen van meer informatie kan het lectoraat deze nieuwe gegevens gebruiken met de zoekmethode om weer nieuwe informatie te vinden, en hier ook relaties in te leggen. Hierdoor ontstaat er een cyclus van verbetering en informatieverzameling. Momenteel is deze data nog niet aanwezig bij het lectoraat.

De bibliotheken hebben hun overstap naar de huidige digitale omgeving nog niet volledig voltooid, en daarom is mogelijk nog niet alles beschikbaar. De bibliotheken zullen veel data aanbieden, maar momenteel is er geen beeld van hoe deze data gebruikt kan worden. Hiermee wordt bedoeld dat het niet bekend is wat voor data de landelijke bibliotheken zullen leveren. Ook kunnen de zoekopdrachten verschillen. In de zoekopdracht kan er een enkele zoekterm ingevoerd worden, maar er kunnen ook meerdere zoektermen ingevoerd worden die samengesteld zijn uit de grammatica der concepten(De verzameling van de thesauri wordt binnen het lectoraat de grammatica der concepten genoemd). Dan is het ook maar de vraag of het uitbreiden van de zoektermen betere resultaten zal opleveren van de zoekopdracht bij de landelijke bibliotheken; welke kwalitatieve impact levert het uitbreiden van de zoekopdracht op de resultaten van de zoekopdracht? In figuur 1 is de probleemstelling kort samengevat.



Figuur 1: Concept map situatie van het concept. De API van de landelijke bibliotheken, en hoe dit aan zou sluiten bij het systeem van het lectoraat.

1.2.1 Doelstelling

Het doel van het onderzoek is om te achterhalen welke kwalitatieve impact het uitbreiden van de zoekopdracht met behulp van de grammatica der concepten levert op het resultaat van de zoekopdracht.

1.2.1.1 Relevantie

Dit hoofdstuk beschrijft de relevantie van het onderzoek vanuit meerdere perspectieven.

1.2.1.1.1 Bedrijfsrelevantie

De relevantie van het onderzoek voor het bedrijf is dat het bedrijf de resultaten hoopt te implementeren in het huidige systeem. Tijdens het onderzoek zal er geïnteracteerd en geëxperimenteerd worden met de API van de landelijke bibliotheken. De hierbij behaalde resultaten zullen worden vastgelegd en beschreven. Door het uitvoeren van dit onderzoek zal er ontdekt worden hoe de API te gebruiken is. Dit kan het bedrijf later gebruiken voor het implementeren van de API in hun eigen systeem. Ook zal er door het onderzoeken van de resultaten van de experimenten van de student te zien zijn of het uitbreiden van de zoektermen nut heeft of niet. Als blijkt uit het onderzoek dat dit niet het geval is, dan zou hier geen verdere moeite in gestopt moeten worden om dit te onderzoeken en mogelijk te maken.

1.2.1.1.2 Maatschappelijke relevantie

De maatschappelijke relevantie van het onderzoek is dat het uiteindelijke project van het lectoraat zal worden ingezet, voor bijvoorbeeld de zorgsector en andere belanghebbenden. Dit houdt in dat de resultaten van het onderzoek het mogelijk maken voor het lectoraat om te bepalen hoe de API van de bibliotheek gebruikt kan worden. Zo kan de meest kwalitatieve selectie aan data hieruit worden

verkregen, waardoor er veel informatie beschikbaar gesteld wordt voor maatschappelijke instanties om te gebruiken.

1.2.1.1.3 *Kennisgebied en theoretische relevantie*

De theoretische relevantie van het onderzoek is vooral merkbaar voor de API van de landelijke bibliotheken. Deze API heeft nog niet veel documentatie beschikbaar, en daarom ook geen voorbeelden. Dit houdt in dat alle experimenten van de student nieuw materiaal zijn voor deze API. Ook het experimenteren met de zoektermen en uitgebreide zoektermen is nog niet onderzocht en gedocumenteerd. Alle experimenten en resultaten zullen nieuwe toevoegingen zijn op de API van de landelijke bibliotheken, waar momenteel nog weinig over bekend is.

1.2.2 *Vraagstelling*

Dit zijn de vragen die beantwoord moeten worden tijdens het onderzoek om het doel te behalen.

1.2.2.1 *Centrale vraag*

De vraag die dit onderzoek probeert te beantwoorden luidt:

Welke kwalitatieve impact heeft het uitbreiden van de zoektermen met behulp van de grammatica der concepten op de ontsloten informatie van de API van de landelijke bibliotheken?

In deze context wordt er bedoeld met kwalitatieve impact, of de resultaten relevant zijn voor de zoekmethode van het lectoraat bij de ingevoerde zoektermen, en of hierdoor de kwaliteit van de context van de termen verbeterd.

1.2.2.2 *Deelvragen*

De hiervoor genoemde vraag valt uiteen in de volgende deelvragen:

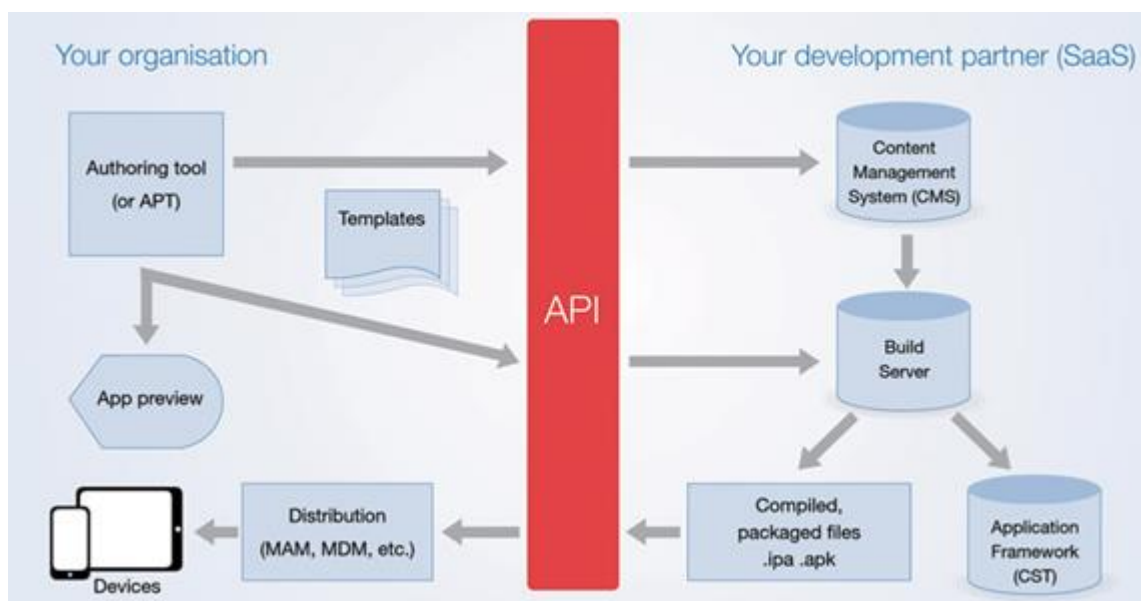
1. Hoe werkt de API van de landelijke bibliotheken?
2. Hoe kan de API gebruikt worden om informatie van de landelijke bibliotheken te ontsluiten op basis van een zoekopdracht?
3. Hoe kan de zoekopdracht uitgebreid worden met behulp van de grammatica der concepten?
4. Welke kwalitatieve impact heeft het uitbreiden van de zoektermen van de zoekopdracht op het resultaat van de zoekopdracht?

1.3 Theoretisch kader

Dit hoofdstuk beschrijft welke bronnen er gebruikt worden voor het opzoeken van informatie, en beschrijft bepaalde begrippen. Het onderzoek ligt in het kennisgebied van software engineering. Zo is het proof of concept ook nodig voor het onderzoek. Bij het proof of concept is het belangrijk dat deze uiteindelijk functioneel en kwalitatief optimaal wordt. De theorie gaat over termen uit dit gebied. Aangezien de focus van het onderzoek ligt op data integratie, en op het semantisch web, wordt hier ook over geschreven. Daarom is bijvoorbeeld het begrip data integratie belangrijk om te begrijpen. Het theoretisch kader is vooral gericht op die thema's die voor dit onderzoek belangrijk zijn, in dit geval zijn dit de begrippen die te maken hebben met semantiek. Zo zijn voor het begrijpen van semantiek begrippen zoals RDF belangrijk om de context van het onderzoek te kunnen begrijpen.

De toegang tot de data.

De landelijke bibliotheken hebben een API(Application Program Interface) beschikbaar die het lectoraat graag zouden gebruiken. API staat voor Application Program Interface, dit is een set routines, protocollen en functionaliteiten om een applicatie te kunnen ontwikkelen. De API maakt het mogelijk om verschillende onderdelen te gebruiken voor het ontwikkelen van een applicatie. Hierdoor is het mogelijk voor systemen om met elkaar te interacteren en de data op te halen van de landelijke bibliotheken. (Beal, n.d.) Er zijn er nog geen eerder uitgevoerde onderzoeken te vinden in context met de werking van de API van de landelijke bibliotheken. Het is een API die nog in ontwikkeling is, en ook vooral gebruikt word door andere bibliotheken. Hier waren geen onderzoeken of resultaten over te vinden. Anders zouden deze zeker bestudeerd geweest zijn voor het onderzoek. In figuur 2 is ook te zien hoe de API als laag zit tussen twee organisaties die beide een applicatie hebben, en via de API met elkaar communiceren.



Figuur 2: Een overzicht van hoe de API tussen meerdere systemen staat, (LEO Web Team, 2013)

De API werkt met behulp van SRU. SRU staat voor Search/Retrieve via URL. Dit is een standaard protocol dat op XML gebaseerd is, en dat gebruik wordt voor het uitvoeren van zoekopdracht. De syntax voor de zoekopdracht wordt toegepast door gebruik te maken van CQL, wat staat voor Contextual Query Language (SRU/CQL, 2013). CQL maakt het ook mogelijk om de zoekopdracht aan te passen, door middel van filters en operators. Zo kan er gezocht worden op alleen bepaalde onderdelen. Door te zoeken naar "dc.title any auto" wordt er gezocht naar alle zoekresultaten die het woord auto in de titel bevatten. Ook is het mogelijk om in de zoekopdracht gebruik te maken van zoekoperators. De zoekoperators maken het mogelijk om te zoeken met extra specificaties. Dus na het selecteren van welke onderdelen, zoals de titel, kan er ook gebruik gemaakt worden van "AND", "OR", "NOT" en "PROX". De "AND" zoekt naar "zoekterm 1" en "zoekterm 2", de resultaten moeten dus beide zoektermen bevatten. CQL heeft verder nog veel mogelijkheden voor het specificeren van de zoekopdracht.

Het verbinden van deze API en de zoekmethode zal waarschijnlijk gebeuren door middel van een interface. Een interface werkt als koppeling tussen verschillende onderdelen. Dit kan zijn tussen mens en machine, maar in dit geval gaat het over het verbinden van stukken code.

Door de API van de landelijke bibliotheken te verbinden met de zoekmethode is het mogelijk om de data te integreren. Dit wil zeggen, het samenvoegen van data uit meerdere bronnen, en deze in dezelfde wijze op te slaan en weergeven. Hierdoor kan er data van meerdere bronnen op dezelfde manier gebruikt worden aangezien de data gelijk is aan elkaar na de verwerkingen.

Het begrijpen en beschrijven van relaties tussen begrippen.

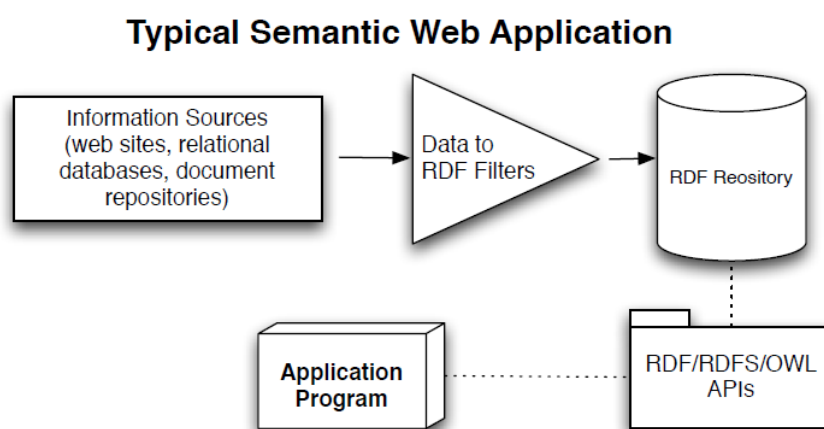
Met het gebruiken van de API is het lectoraat van plan de data van de landelijke bibliotheken toe te voegen als een nieuwe bron van data voor hun zoekmethode. Momenteel heeft het lectoraat wel bronnen zoals een thesaurus(mv. thesauri), maar nog geen bron die zoveel data bevat als de landelijke bibliotheken. Een thesaurus is een soort van woordenboek/naslagwerk/synoniemenlijst, waarin woorden kunnen worden opgezocht worden. De thesaurus kan inhouden dat bepaalde begrippen voor een onderwerp of categorie gegroepeerd worden, waar de relaties tussen deze begrippen ook worden bewaard. De thesaurus richt zich voornamelijk op een bepaalde categorie of onderwerp, en bevat hierdoor vervolgens alleen relevante definities voor.

Het aanwezig zijn van relaties tussen begrippen word ook wel semantiek genoemd, of in het Engels, semantic. Semantiek kan gezien worden als de betekenis van syntax en symbolen, maar ook als de relatie tussen tekens, en hun betekenis. (semantics, n.d.) (Houghton Mifflin Harcourt Publishing Company). Een korte beschrijving hiervoor is: Er is een dijk. Deze dijk is opgebouwd met zand. Dijk heeft dus een relatie met zand. Een dijk is vaak aan het water, dus heeft dijk een relatie met water. Wat is een relatie tussen zand en water? Dit is dijk, maar bijvoorbeeld ook strand. En zo kunnen meerdere begrippen aan elkaar verbonden worden met relaties. Dit is dus nodig om te begrijpen wat er gezocht word in de data die van de API terugkomt.

Deze semantiek worden tegenwoordig ook toegepast op het internet, en dit is ook de omgeving waar de API zich in bevindt. Dit wordt ook wel het semantic web genoemd. Semantic Web is een extensie(in de vorm van een framework) van het huidige web dat een eenvoudigere manier levert

voor het vinden, delen, hergebruiken en combineren van informatie. Het bouwt zich voort uit de mogelijkheden van taggen, om de informatie te gebruiken. (Semantic Web, n.d.). Het taggen gebeurt door middel van Open Annotations, dit zijn tags die gebruikt worden om de data en hun relaties te beschrijven. RDF en SKOS maken ook gebruik van deze annotations. (Van de Sompel, Ciccarese, & Sanderson).

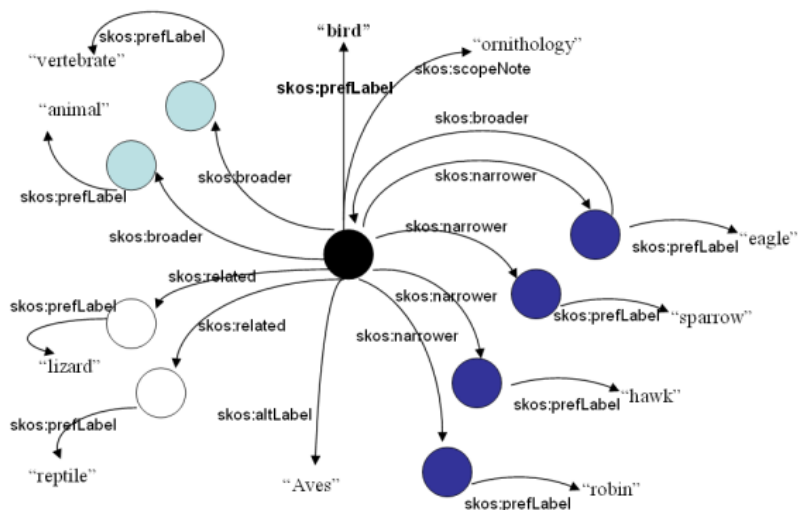
In figuur 3 volgt een overzicht van de algemene werking om beter te begrijpen hoe het semantisch web in elkaar steekt als het door een applicatie gebruikt zou worden. De data komt uit meerdere bronnen en wordt met behulp van RDF filters in een opslag geplaatst. De opslag wordt gebruikt om met behulp van technieken de gefilterde data uit te lezen in informatie die vervolgens in de applicatie gebruikt of getoond kan worden.



Figuur 3: Overzicht van een Semantic Web Applicatie, (Watson, 2010)

Om deze data overzichtelijk en duidelijk te beschrijven moet voor het aanduiden van de semantiek een structuur aanwezig zijn. Deze structuur moet vastgelegd en beschreven kunnen worden. Hiervoor wordt SKOS gebruikt. SKOS staat voor Simple Knowledge Organization System, en wordt gebruikt voor het beschrijven van thesauri of andere gestructureerde gecontroleerde vocabulaires, en is deel van het semantisch web. (aisaac, 2012) Op deze manier kunnen applicaties informatie met elkaar uitwisselen, of kan informatie georganiseerd worden. Daarom is het dus belangrijk voor het onderzoek om te weten wat dit is, zodat er begrepen kan worden hoe de data van de landelijke bibliotheken hierin past.

In figuur 4 is er een overzicht te zien van hoe SKOS er uit kan zien. In de figuur is te zien hoe er meerdere begrippen aan het begrip “bird” hangen. De “skos:narrower” bevatten termen die dieper ingaan op het begrip vogel, zo staan hier meerdere vogelsoorten. Aan de linkerkant staat er bijvoorbeeld “skos:broader”, met als begrip “animal”. Een vogel valt onder het begrip dier, dus de “broader” geeft aan dat er breder gekeken wordt. Op deze manier kunnen begrippen aan elkaar gerelateerd worden.



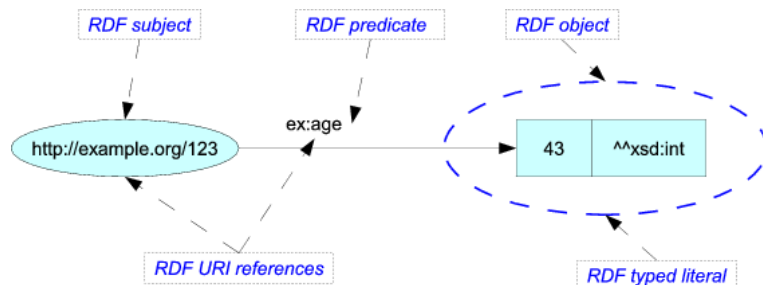
Figuur 4: Voorbeeld van SKOS, (Implementing a Government-wide Semantic Solution to Thesauri, 2005)

Het werken met semantische data.

Om te kunnen werken met semantische data wordt er gebruik gemaakt van meerdere onderdelen. Zo is er een structuur die vaak gebruikt wordt om de data te beschrijven. Dit is de Triple. De triple is een driedelige structuur, ook wel het subject-predicaat-object genoemd. Zo is een voorbeeld "De auto is blauw" de auto-is-blauw. Dit moet dus begrepen worden om te zien hoe de data in elkaar zit.

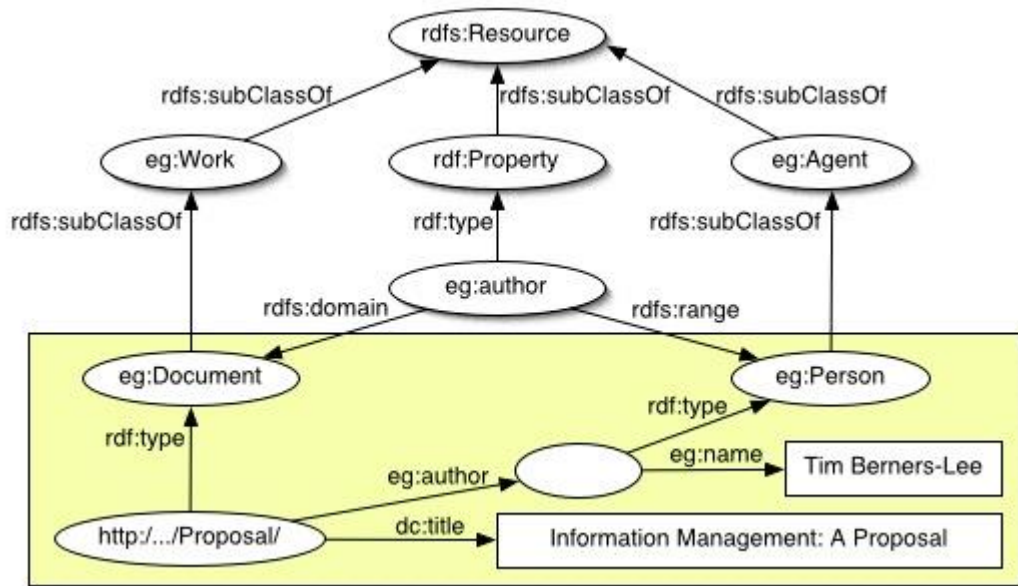
De data is vooral metadata. Metadata zijn gegevens over gegevens. Zo kan het over een ".pdf" bestand, de auteur, de datum, de grootte, en dergelijke beschrijven. Het gebruiken van metadata zorgt ervoor dat het makkelijker zou kunnen zijn om data te vinden, aangezien er al data over de data beschikbaar is.

Om deze data uit te wisselen wordt er gebruikt gemaakt van een standaard model. In dit geval is dit RDF. RDF staat voor Resource Description Framework, en is een standaard model voor het uitwisselen van data op het web. In figuur 5 is er een voorbeeld van RDF te zien. Er zijn meerdere onderdelen in RDF, die allemaal iets aangeven. Het onderwerp, het object, de referentie en dergelijke. RDF wordt gebruikt om uitspraken te maken over bronnen op het web. Dit gaat via een triple.



Figuur 5: Voorbeeld van RDF, (Nilsson, 2008)

Als uitbreiding van RDF is er het RDFS. RDFS staat voor RDF Schema, dit bevat meer complexe functies dan RDF. Door te weten hoe RDF werkt kan de data begrepen en gebruikt worden. In figuur 6 is er een beschrijving te zien over een artikel. De “dc: title” wijst naar “Information Management: A Proposal”, dit geeft aan dat dit de titel is. Net als SKOS worden de relaties op deze wijze aan elkaar verbonden. De SKOS en RDF worden samen gebruikt voor het beschrijven van semantische data.



Figuur 6: Voorbeeld van RDFS, (Brickley & Guha, 2002)

RDF werkt in combinatie met XML. XML staat voor Extensible Markup Language, en is een syntax om gestructureerde gegevens weer te geven als tekst. Op deze manier kan er metadata aan de informatie en tekst meegegeven worden, en is de tekst voor zowel mens als machine te lezen. Dit wordt vervolgens gebruikt om gegevens te versturen of op te slaan.

Om de beschikbare data te kunnen gebruiken moet er wel een taal beschikbaar zijn die hiertoe in staat is. In dit geval is dat SPARQL. SPARQL staat voor SPARQL Protocol and RDF Query Language is een RDF query taal die gebruikt wordt om RDF-gebaseerde data te bevragen of bewerken. Zoals de naam doet vermoeden, heeft de taal wat weg van SQL, en lijkt er met de query's wel op. Dit is dus nodig om de data eventueel te bewerken of te filteren.

2 Methoden en materialen

2.1 Deelvraag 1 – Onderzoeken API

1. Hoe werkt de API van de landelijke bibliotheken?

2.1.1 Meetinstrumenten en operationalisatie

Als eerste werd er besloten om de documentatie van de API (Groeneveld, Medema, Janssen, & Morée, 2015) te onderzoeken als wijze van deskresearch. Dit houdt in dat er gezocht werd naar informatie die in de documentatie aanwezig was. Er werd als eerst gezocht naar de algemene beschrijving van de API. Ook werd er gezocht naar de functionaliteiten van de API, en wat deze deden. Vervolgens werden de beschrijvingen van de resultaten onderzocht. Zo werd er gezocht in welk formaten deze beschikbaar waren, en wat hier in stond. De verwachting was dat al de beschrijvingen over de functionaliteiten en werkingen van de API hier aanwezig waren, en dat er voor de resultaten voorbeelden aanwezig waren. Bij de resultaten werd er ook verwacht dat hier aangegeven stond in welke formats deze beschikbaar waren, en wat de inhoud hier van was.

Na het lezen van de documentatie is er vervolgens gebruik gemaakt van de demo van de landelijke bibliotheken. Deze demo maakt gebruik van de API. Bij het gebruiken van de demo werd gekeken naar de werking van de demo, en welke resultaten hier uit kwamen. Zo werd er verwacht dat de demo de functionaliteiten van de API bevatte zoals deze beschreven werden in de documentatie, en dat hier inhoudelijk dezelfde resultaten uit kwamen. Na het onderzoeken van de aanwezige informatie werden onbeantwoorde en nieuwe vragen naar de ontwikkelaar van de API gemaild.

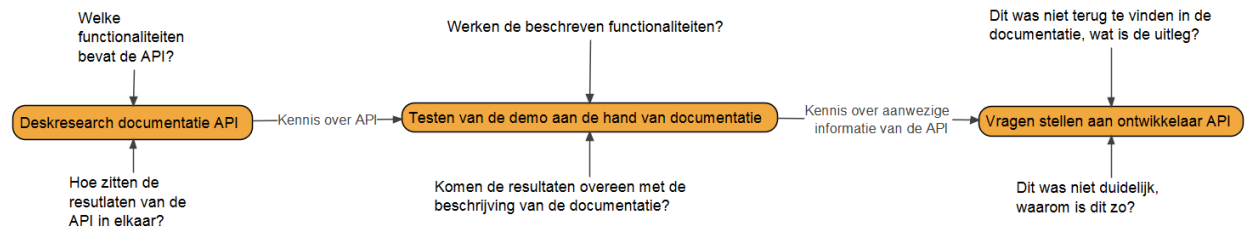
2.1.1.1 Onderbouwing keuze meetinstrument

De documentatie van de API werd doorzocht omdat hier de werking van API in beschreven stond. Om deelvraag 1 te kunnen beantwoorden was dit de beste optie aangezien de documentatie de API beschrijft en de antwoorden hier dus letterlijk in staan. Uit de deskresearch zal dan ook blijken of dit zo is. Het gebruiken van de demo is om de functionaliteiten in werking te zien, en om te kijken of de resultaten ook overeenkomen met de beschrijvingen van de documentatie. Deze kennis over de API was nodig om verder te kunnen met deelvraag 2, de kennis van de API vormde de basis voor deelvraag 2.

2.1.2 Analysemethoden

De inhoud van de documentatie, en de resultaten van het uitproberen van de demo werden door de student met elkaar vergeleken. De resultaten die uit de demo kwamen moesten voldoen aan de verwachtingen die getrokken waren uit het onderzoeken van de documentatie, dit werd met elkaar vergeleken door middel van een vergelijkingstabel. Het gebruiken van een tabel is duidelijk en overzichtelijk. Als de vergelijking overeenkwam dan klopte de documentatie ook daadwerkelijk met de werking van de API, en de demo die deze gebruikte. Vervolgens zijn de onbeantwoorde vragen en onduidelijkheden aan de ontwikkelaar gesteld. Op deze manier is er door de ontwikkelaar van de API zelf bevestigd of de resultaten klopten.

In figuur 8 in het overzicht te zien van deze stappen.



Figuur 8: Stappen meetinstrumenten deelvraag 1.

2.2 Deelvraag 2 – Gebruiken API

2. *Hoe kan de API gebruikt worden om informatie van de landelijke bibliotheken te ontsluiten op basis van een zoekopdracht?*

2.2.1 Meetinstrumenten en operationalisatie

Om dit te onderzoeken werden er proof of concepts ontwikkeld die gebruik maakten van de API door het gebruik van rapid prototyping. Er werden eisen en ontwerpen opgesteld voor de proof of concepts, namelijk het functioneel ontwerp, technisch ontwerp, en een requirement specificatie. In het functioneel ontwerp werden acceptatie ontwikkeld voor de proof of concepts. Deze werden gebaseerd op de documentatie. In de proof of concepts werd er geprobeerd de functionaliteiten van de API aan te roepen. Zo werd er verwacht dat de functionaliteiten van de API op dezelfde manier werkten als in de eerder bestudeerde demo, en zoals er beschreven stond in de documentatie. Het verwachte resultaat van het proof of concept was dat de resultaten overeenkwamen als de beschrijving in de documentatie. De proof of concepts werden ontwikkeld aan de hand van de documentatie van de API.

Om de API volledig te kunnen benutten moesten er ook libraries gekozen worden voor de proof of concepts. De library werd gekozen door middel van een pakketselectie. Zo werd er getest of de libraries al de benodigheden bevatten voor het verwerken van de resultaten van de API. Uiteindelijk werd er geprobeerd de library in werking te brengen in combinatie met de API van de landelijke bibliotheken in een proof of concept.

2.2.1.1 Onderbouwing keuze meetinstrument

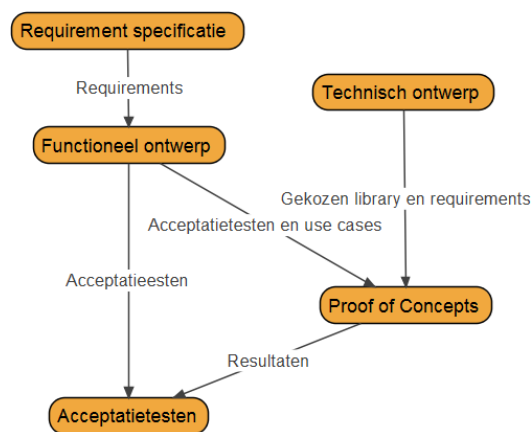
Er werd gekozen voor het maken van proof of concepts met behulp van rapid prototyping omdat hier de werking van de API mee getest kon worden zonder dat er teveel tijd geïnvesteerd werd aan het maken van documentatie en planningen. Het ging er namelijk om of de API werkt zoals beschreven staat in de documentatie, hier waren dus geen volledige ontwerpen voor nodig. Daarom zijn de proof of concepts ook gebaseerd op de documentatie, deze beschreven de functionaliteiten van de API. Het gebruiken van libraries was nodig omdat de resultaten die uit de API komen ook moesten worden

verwerkt. Deze kennis over het gebruiken van de API was nodig om met deelvraag 3 verder te kunnen gaan, aangezien de kennis over de werking en gebruik van de API daar nodig was.

2.2.2 Analysemethoden

Bij de proof of concepts werd er gekeken of de API gebruikt kon worden op dezelfde manier als in de demo, en of de externe libraries de resultaten konden verwerken en bewerken. Dit werd gedaan door de student, die observeerde of het gebruiken van de API in het proof of concept net zo werkte als in de demo. Dus de functionaliteiten van de demo werden vergeleken met de functionaliteiten van de proof of concept door middel van een vergelijkingstabel. Ook werden de acceptatietesten uitgevoerd om te kijken of de proof of concepts aan de eisen voldeden. Vervolgens werden de resultaten die binnen werden gehaald door de proof of concepts vergeleken met de resultaten van de demo om te zien of deze overeenkwamen. De resultaten werden doorgenomen om te kijken wat hier in stond, en of dat dit ook zo in de API beschreven stond.

In figuur 9 in het overzicht te zien van deze stappen.



Figuur 9: Stappen meetinstrumenten deelvraag 2.

2.3 Deelvraag 3 – Uitbreiden zoekopdracht

3. *Hoe kan de zoekopdracht uitgebreid worden met behulp van de grammatica der concepten?*

2.3.1 Meetinstrumenten en operationalisatie

Om deze deelvraag te beantwoorden moest er eerst onderzocht worden hoe een uitgebreide zoekopdracht er uit ziet. Om dit te achterhalen is er een open interview gehouden met de stagebegeleider en de bedrijfsbegeleider om te leren hoe dit te werk gaat met behulp van de grammatica der concepten. De open vraag voor het interview was “Hoe kan de zoekopdracht uitgebreid worden?”. Deze gesprekken zijn niet genoteerd, maar de resultaten van de open interview zijn wel direct opgenomen in de resultaten het rapport.

Vervolgens bedacht de student zelf ook nog een manier. Dit werd gedaan door te onderzoeken of het mogelijk was om de grammatica der concepten na te bootsen met de uit te voeren experimenten, en met de huidige kennis van de API.

2.3.1.1 *Onderbouwing keuze meetinstrument*

De onderbouwing hiervoor is dat de begeleiders weten hoe de uitgebreide zoekopdracht er uit zal komen te zien, en hoe deze wordt opgesteld. Deze kennis is bij hun aanwezig, en werd daarom ook daar gehaald. De keuze voor een open interview is omdat het antwoord geen ja of nee antwoord was, maar een antwoord waar een discussie over gehouden kon worden, hierdoor kon de student zelf ook eigen keuzes bespreken, en zo meteen vragen stellen waar nodig. Door het gebruiken van een open interview had de student de mogelijkheid om een kwalitatief antwoord te krijgen, en dit ook meteen te bevestigen. Deze informatie was ook nodig om verder te kunnen met deelvraag 4, aangezien de uitgebreide zoekopdrachten daar nodig waren.

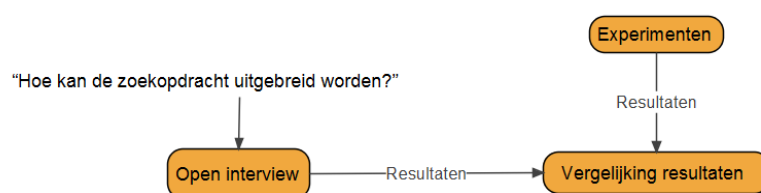
Het gebruiken van de resultaten om zelf de grammatica der concepten na te bouwen geeft extra input aan de werking en validiteit van het uitbreiden van de zoekopdracht. Ook zijn er op deze manier meer uitgebreide zoekopdrachten beschikbaar voor deelvraag 4.

2.3.2 *Analysemethodes*

De resultaten van het open interview werden geverifieerd tijdens het interview door het antwoord met de geïnterviewde te bespreken, en samen te bevestigen dat de resultaten van het interview klopten. Vervolgens is er met behulp van de demo van de landelijke bibliotheken geprobeerd om de uitgebreide zoekopdracht uit te voeren, om te kijken of de uitgebreide zoekopdracht bruikbaar was.

Het analyseren van de validiteit van het uitbreiden van de zoekopdrachten gebeurde tijdens de experimenten van deelvraag 4, waar de zoekopdracht uitgevoerd zal worden. Vervolgens worden de resultaten daar met elkaar vergeleken.

In figuur 10 in het overzicht te zien van deze stappen.



Figuur 10: Stappen meetinstrumenten deelvraag 2.

2.4 Deelvraag 4 – Onderzoeken impact

4. *Welke kwalitatieve impact heeft het uitbreiden van de zoektermen op het resultaat van de zoekopdracht?*

2.4.1 Meetinstrumenten en operationalisatie

Er werd een open interview gehouden met de opdrachtgever, met de startvraag “wat is kwaliteit?”. Als verwachting had de student dat de begeleider vooral de relevantie en hoeveelheid van de data belangrijk zou vinden, en niet de bron of betrouwbaarheid. Dit was ook wat de student zelf dacht van het begrip.

Vervolgens koos de student om naar het antwoord te zoeken op de deelvraag door middel van experimenten in een verkennend onderzoek. Er werd samen met de bedrijfsbegeleider besloten dat de experimenten uitgevoerd werden met behulp van de demo, en niet een voltooid prototype. Uiteindelijk is de conclusie van het experiment te gebruiken voor het beantwoorden van de hoofdvraag.

Voor de experimenten werd er een opzet geschreven die terug te vinden is in bijlage 4. De opzet volgt de volgende stappen:

1. **Er wordt een zoekopdracht ingevoerd bij de demo, deze wordt vervolgens uitgevoerd.**
2. **De student kiest vervolgens maximaal 10 artikelen uit van de zoekresultaten.**
3. **De omschrijvingen, onderwerpen en titels voor de gekozen artikelen uit de zoekresultaten wordt opgeslagen in een Word bestand.**
4. **De woorden in de omschrijving, onderwerpen en titels van de gekozen artikelen uit de zoekresultaten worden geteld en opgeslagen in een Word bestand.**
5. **Vervolgens breidt de student de originele zoekopdracht uit met de termen uit de top 50 meest voorkomende woorden van de inhoud van de omschrijving, en met begrippen uit de grammatica der concepten.**
6. **De stappen 1 tot en met 4 worden herhaalt, maar dan met de uitgebreide zoekopdracht in plaats van de standaard zoekopdracht.**

2.4.1.1 Onderbouwing keuze meetinstrument

Er werd gekozen voor een open interview omdat er geen vaste definitie was van het antwoord op de vraag “wat is kwaliteit”, en het dus waardevoller was om hierover met een open gesprek een antwoord over te krijgen. Het verkregen antwoord werd gebruikt omdat hiermee te bepalen welke kwalitatieve impact er was op de resultaten van de API. Het open interview was genoeg om te bepalen wat kwaliteit inhoud voor het lectoraat. Kwaliteit is namelijk een vaag begrip dat door de belanghebbenden gedefinieerd moest worden, en niet te bepalen valt uit bronnen of onderzoeken. Het is van belang wat er bij het Lectoraat zelf als kwaliteit wordt gezien, en niet wat er bij andere onderzoeken als kwaliteit gezien werd.

De reden voor het kiezen van het uitvoeren van de experimenten met de demo is dat het prototype alles zou moeten kunnen wat de demo ook kan, en dat het prototype niet volledig ontwikkeld is. Aangezien er geen uiteindelijk tijd was om het prototype af te krijgen was het gebruiken van de demo

het beste alternatief. De werking van de API was bekend, al functioneerde die nog niet met de externe libraries voor het lezen en bewerken van de resultaten uit de API. In plaats van de libraries werkende te krijgen werd er vooral gekeken naar het uitvoeren van het experiment. Door de demo te gebruiken kon er alsnog gezien worden hoe de API zou kunnen worden gebruikt. En omdat de gewenste informatie over het uitbreiden van een zoekopdracht met behulp van de grammatica der concepten op de ontsloten informatie van de landelijke bibliotheken nog niet aanwezig is. Er kon dus geen deskresearch uitgevoerd worden, daarom werd er gekozen voor een verkennend onderzoek. Door middel van de experimenten kan dit uitgetest en ontdekt worden.

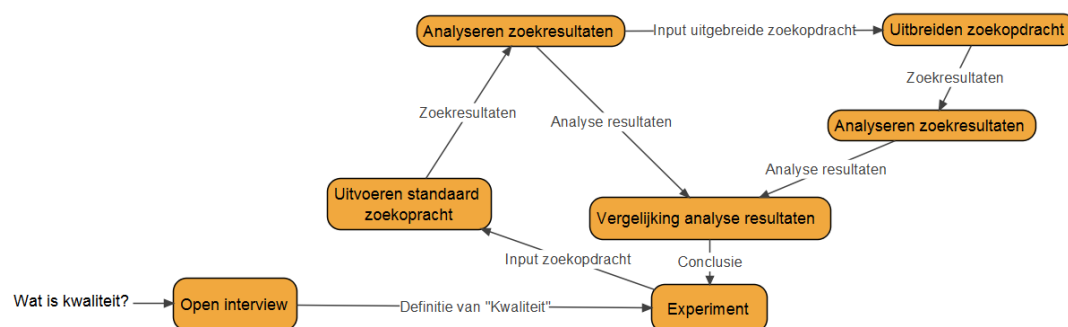
2.4.2 Analysemethoden

Het antwoord van het open interview werd tijdens het open interview met de begeleider besproken als analyse over de juistheid daarvan.

Het analyseren van de resultaten bij het gebruiken van de demo is gedaan door ze met elkaar te vergelijken tijdens de experimenten. Dit is ook terug te vinden in bijlage 3, het is de 7^e en laatste stap.

- 7. De student vergelijkt de top 50 meest voorkomende termen van de enkele zoekterm met die van de uitgebreide zoektermen, en de termen uit de grammatica der concepten om te kijken of deze overeenkomen, of juist nieuwe termen leveren.**

In figuur 11 is het overzicht te zien van deze stappen.



Figuur 11: Stappen meetinstrumenten deelvraag 4

3 Resultaten

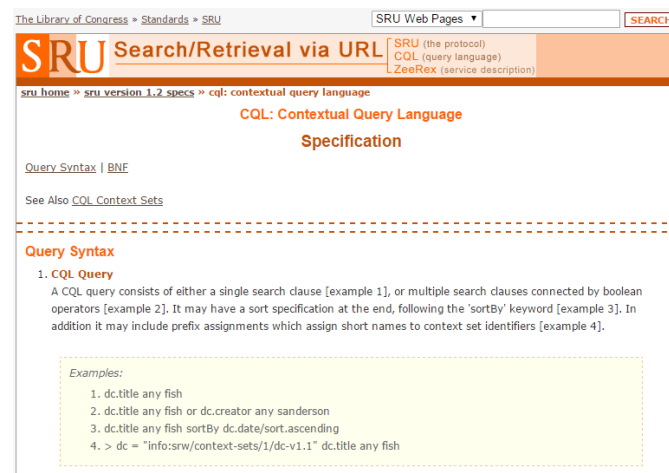
In dit hoofdstuk worden de resultaten van het onderzoek beschreven en geanalyseerd.

3.1 Deelvraag 1 – Onderzoeken API

1. Hoe werkt de API van de landelijke bibliotheken?

3.1.1 Resultaten

De API heeft documentatie die de werking van de API beschrijft, maar deze is niet volledig. Bepaalde onderdelen waren voor de student niet beschikbaar. De student had hiervoor contact gezocht met de ontwikkelaar van de API. Hierdoor kreeg de student toegang tot meer documentatie van de API. Hierin stond ook dat de resultaten in meerdere formats geleverd kunnen worden. Uit deze documentatie bleek dat de API ook gebruik maakt van SRU voor het uitvoeren van de zoekopdracht. Daarom heeft de student ook de documentatie hier van opgezocht. De volledige documentatie is terug te vinden in de bronnen (Groeneveld, Medema, Janssen, & Morée, 2015) en (SRU/CQL, 2013). In figuur 12 is er een gedeelte getoond van de documentatie over de SRU. Hier is te zien dat er beschrijvingen over de SRU aanwezig zijn.



Figuur 12: Voorbeeld documentatie SRU (SRU/CQL, 2013).

De documentatie van de landelijke bibliotheken beschreef ook de resultaten. Deze resultaten konden in een RDF/XML formaat opgevraagd worden. In figuur 13 is een voorbeeld te zien hoe de resultaten er uit zien. Ook stond er een overzicht over hoe de zoekresultaten in elkaar zaten, dit is terug te zien in figuur 14. De resultaten zijn beschreven als bestaand uit triples, die beschreven zijn in een RDF formaat, waarbij er gebruik is gemaakt van open annotations. De volledige documentatie is terug te vinden in (Groeneveld, Medema, Janssen, & Morée, 2015). In de resultaten bevonden zich ook GGC – OCLC(Gemeenschappelijk Geautomatiseerd Catalogiseersysteem voor Nederlandse bibliotheken) tags, deze bevatten meta data over het zoekresultaat.

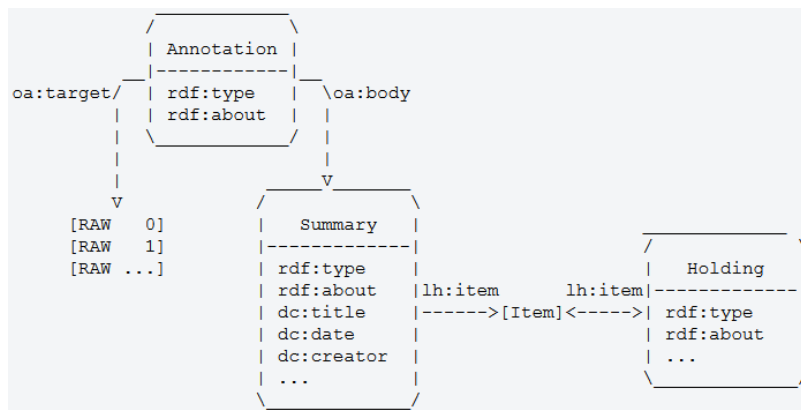
Regelnummers aan/uit doen

```

1 <Annotation xmlns="xmlns:oa="http://www.w3.org/ns/oa#"
2   rdf:about="http://data.bibliotheek.nl/nbc/summary#urn:GGC:oclc-ggc:780950577">
3   <hasTarget rdf:resource="urn:GGC:oclc-ggc:780950577"/>
4   <annotatedBy rdf:resource="http://bibliotheek.nl/id/bnl"/>
5   <motivatedBy rdf:resource="http://data.bibliotheek.nl/ns/nbc/oa#summarizing"/>
6   <hasBody>
7     <rdf:Description xmlns:dcterms="http://purl.org/dc/terms/"
8       <dcterms:type rdf:resource="http://dbpedia.org/ontology/Book"/>
9       <dcterms:title>High level title according to NBC</dcterms:title>
10      <dcterms:creator>High level creator is St Nicolaas</dcterms:creator>
11    </rdf:Description>
12  </hasBody>
13 </oac:Annotation>

```

Figuur 13: Voorbeeld van een resultaat. (Groeneveld, Medema, Janssen, & Morée, 2015).



Figuur 14: De opbouw van een resultaat. (Groeneveld, Medema, Janssen, & Morée, 2015).

Bij het onderzoeken van de demo werd er geprobeerd om inderdaad een zoekopdracht uit te voeren, en om de resultaten hier van te bekijken. De resultaten zijn gedeeltelijk te zien in figuur 15 en 16. In figuur 15 is er te zien hoe de zoekopdracht succesvol verwerkt is tot een lijst van resultaten, en in resultaat 16 is er te zien hoe de resultaten beschreven zijn in de achtergrond van de demo.

Dit formulier is bedoeld om te zien op welke manier een query vertaald wordt in CQL en hoe die als URL gerepresenteerd moet worden.

dijk

ApiKey filter: [- (default)]

Ranking

Leeftijd: [-] Datum: ☐ Aantal holdings: Bron: [-] Type: [-] Taal: [-]

Source

/snulsummary?operation=searchRetrieve&query=dijk&version=1.2&x-drilldown-format=rdxml&x-filter-common-keys=false&x-suggestionsQuery=dijk&x-term-drilldown-dcterms.type.uri,dcterms.language.uri,dcterms.creator,dcterms.date.year,chronological,dcterms.source.uri,dcterms.subject,dcterms.publisher,schema.genre,holding.lh.heldAt.uri,holding.ebook.levenvorm [link]

Toon technische informatie.

Resultaten

Query processing time (index): 0.801 s. Query processing time (api): 0.862. Including full SRU handling: 2.152 s.

Totaal aantal resultaten: 20526.

Filtering op werk: aanzetten

Groeperen: aanzetten

Clusteren: aanzetten

Sorteer op dcterms.date oplopend | dcterms.creator oplopend | dcterms.title oplopend

- Wild water**
Onderwerpen: romans en novellen ; oorspr. - Nederlands; Proza (romansnovellen); Literaire thriller
Auteur: Dijk, Ruben van
Publiek: Volwassenen
Genre: Fictie; Fictie; Thriller
Bron: Catalogus
Uitgegeven door: Bruna, Utrecht
Taal: Nederlands
Media: zonder medium; band; boek
Score: 72.5466156006
- Dijken van Nederland**
Onderwerpen: Dijken; landschaasarchitectuur; fotoboeken; diiken; Geschiedenis van de strijd teoen het water - Stormvloed; diikdoorbraken en overstromingen; Wetenschap en Techniek; Dijken

Facetten

Alle facetten

Type

- boek (10918)
- geluid (3490)
- afbeelding (2906)
- persoon of organisatie (2189)
- concept (2124)
- artikel (805)
- bladmuziek (241)
- brief (168)
- video (116)
- handschrift (92)

Meer...

Taal

- Nederlands (18772)
- Engels (1200)
- Westerlauwers Fries (496)
- Duits (110)
- Frans (65)
- Latijn (34)

Figuur 15: Het uitvoeren van een zoekopdracht in de demo. (Zoekplatform 1.8 - API (1.8.0.18), 2016)

← → ↻ apiozp.acc1.bibliotheek.nl/sru/summary?operation=searchRetrieve&query=dijk&version=1.2&x-drilldown-format=rdxml&x-filter-common-keys=false&x

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<srw:searchRetrieveResponse xmlns:srw="http://www.loc.gov/zing/srw/" xmlns:diag="http://www.loc.gov/zing/srw/diagnostic/" xmlns:xqcl="http://www.loc.gov/zing/cql/xqcl"
  xmlns:mesco_srw="http://mesco.org/namespace/srw#"
  <srw:version>1.2</srw:version>
  <srw:numberOfRecords>4907</srw:numberOfRecords>
  <srw:records>
    <srw:record>
      <srw:recordSchema>summary</srw:recordSchema>
      <srw:recordPacking>xml</srw:recordPacking>
      <srw:recordIdentifier>
        http://data.bibliotheek.nl/nbc/summary#aHR0cDovL2RhdGEuYm1ibGlvdGh1ZmSubmVvZ2djL3Bwb18zQ00TA0NjQ=
      </srw:recordIdentifier>
      <srw:recordData>
        <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-insta
          xmlns:nbcoa="http://data.bibliotheek.nl/ns/nbc/oa#" rdf:about="http://data.bibliotheek.nl/nbc/summary#aHR0cDovL2RhdGEuYm1ibGlvdGh1ZmSubmVvZ2djL3Bwb18zQ00TA0
            <oa:annotatedBy rdf:resource="http://data.bibliotheek.nl/id/bnl/">
            <oa:motivatedBy rdf:resource="http://data.bibliotheek.nl/ns/nbc/oa#summarizing"/>
            <oa:hasTarget rdf:resource="http://data.bibliotheek.nl/ggc/ppn/334490464"/>
            <oa:hasBody>
              <rdf:Description xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:lh="http://vocab.inf.ed.ac.uk/library/holdings#" xmlns:dcterms="http://purl.org
                xmlns:nbcoa="http://data.bibliotheek.nl/ns/nbc/oa#" xmlns:schema="http://schema.org/"
                <dcterms:source rdf:resource="urn:nbc:source:webggc.oclc.nl"/>
                <rdf:type rdf:resource="http://dbpedia.org/ontology/Book"/>
                <dcterms:type rdf:resource="http://dbpedia.org/ontology/Book"/>
                <dcterms:type rdf:resource="http://data.bibliotheek.nl/ns/nbc/type#Gedrukt_boek"/>
                <dcterms:title>Wild water</dcterms:title>
                <dcterms:creator>
                  <rdf:Description rdf:about="http://data.kb.nl/thesaurus/304967688">
                    <rdfs:label>Dijk, Ruben van</rdfs:label>
                  </rdf:Description>
                </dcterms:creator>
                <dcterms:date>2012</dcterms:date>
                <dcterms:publisher>Bruna, Utrecht</dcterms:publisher>
                <dcterms:language>
                  <rdf:Description rdf:about="urn:iso:std:iso:639:-2:dut">
                    <rdfs:label>Nederlands</rdfs:label>
                  </rdf:Description>
                </dcterms:language>
                <dcterms:description>
                  Tijdens de ontmoeting van een getrouwde vrouw met haar jonge minnaar in een vakantiehuisje bij Katwijk breekt een dijk door; ze lijken als ratten in c
                </dcterms:description>
                <dcterms:extent>280 p</dcterms:extent>
                <dcterms:subject>
                  <rdf:Description rdf:about="http://data.bibliotheek.nl/ns/nbc/WUV/nur#305">
                    <rdfs:label>Litteraire thriller</rdfs:label>
                  </rdf:Description>
                </dcterms:subject>
                <dcterms:subject>
                  <rdf:Description rdf:about="http://data.bibliotheek.nl/ns/nbc/Oclc/UNESCO/kmc112X#">
                    <rdfs:label>Proza (romans/novellen)</rdfs:label>
                  </rdf:Description>
                </dcterms:subject>
                <dcterms:subject>
                  <rdf:Description rdf:about="http://data.kb.nl/thesaurus/075629402">
                    <skos:altLabel>romans en novellen ; oorspr. - Nederlands</skos:altLabel>
                  </rdf:Description>
                </dcterms:subject>
                <dcterms:audience>
                  <rdf:Description rdf:about="http://data.bibliotheek.nl/ns/nbc/WB0_Biblion/audience#VLW">
                    <rdfs:label>Volwassenen</rdfs:label>
                  </rdf:Description>
                </dcterms:audience>
              </oa:hasBody>
            </oa:hasTarget>
          </oa:annotatedBy>
        </oa:hasTarget>
      </srw:recordData>
    </srw:record>
  </srw:records>
</srw:searchRetrieveResponse>
```

Figuur 16: De opbouw en inhoud van een resultaat van de demo. (Zoekplatform 1.8 - API (1.8.0.18), 2016)

Vervolgens is er met de ontwikkelaar van de API gesproken over de ontbrekende onderdelen, de uitwisseling van e-mails is opgeslagen, en gebruikt voor het verdere onderzoek. De vragen van de student zijn vervolgens ook in de FAQ van de API opgenomen. De FAQ was eerst ook nog niet bestaand in de documentatie. De FAQ en vragen zijn te zien in figuur 17.

Inhoud

1. Frequently Asked Questions

1. API

1.1. Search

1.1.1. The API returns results in different order between requests

1.1.1.1. The API returns results in different order between requests

1.1.1.2. Why can I get only 10 results at a time?

This may happen as a result of load balancing over two or more nodes giving different results. On a production environment this can only happen if one node failed and has been restored or if a node has been out of order for some time. These nodes do not follow the exact same pattern of adds and deletes, but might see only an add while other nodes see an add, delete and add. Unfortunately, the intermediate deletes do influence the ranking. On the Acceptation environment this happens more frequently as a result of more frequent releases of new nodes. If the differences become too big, the nodes must be recycled.

The search engine limits the number of results as a means to maintain response times under high load. Streaming much data to clients is no problem for the search engine, as it could do this efficiently by the thousands, but it is problematic nevertheless because of the following reasons:

- The amount of data to be retrieved by the system is quite large and this would take up many system resources which have to be shared by all libraries in the Netherlands.
- The amount of data streamed to the client will be very large and the latency of transmission of it would quickly dominate the response times seen by end-users. There is no way around this because it is the way the internet works.
- Many (if not most or all XML) parsers have trouble parsing the SRU responses, which are in XML+RDF format. Most parser cannot stream these responses and need to parse the whole response as one big document. This leads to even more latency (if the parser does not crash).
- The API is meant for implementing interactive apps while for harvesting there is a separate end-point available. The amount of information in one response is already huge (much much more than can be reasonably displayed on any device). This amount of data per response will likely be reduced in the future, not extended, and made available only when specifically requested.

Documentatie van de API (laatst bevestigd op 2015-04-28 09:35 door ENG@nsw.nl)

Figuur 17: Het resultaat van de gesprekken met de ontwikkelaar, nu ook aanwezig in de documentatie van de API. (Groeneveld, Medema, Janssen, & Morée, 2015).

Als laatste werden de resultaten naast elkaar gezet in een vergelijkingstabel.

	Documentatie	Demo
• Kan een zoekopdracht uitvoeren door middel van SRU.	✓	✓
• De resultaten komen in RDF/XML formaat terug.	✓	✓
• De resultaten zijn beschreven in RDF formaat.	✓	✓
• De resultaten zijn beschreven met open annotations.	✓	✓

3.1.2 Analyse

In de documentatie is te lezen dat er gebruikt gemaakt wordt van de SRU techniek om de informatie op te halen via de API. De documentatie geeft dit aan, maar dit is ook getest in de demo.

De SRU heeft een query waar de zoekterm in opgenomen wordt. Deze query bestaat uit meerdere onderdelen, zoals bijvoorbeeld in welke bestandssyntax de resultaten moeten worden opgehaald. Zo stond er documentatie dat het bestand in een XML/RDF formaat geleverd kan worden. Uit de documentatie blijkt er ook dat een SPARQL end-point aanwezig zou moeten zijn, maar deze werkt nog niet.

Uit de API komen XML/RDF bestanden, in deze bestanden staan meerdere open annotations tags. Zo staan er tags met de titel, onderwerpen, auteur, datum, categorie, omschrijving, doel en dergelijke. Deze tags zijn de Open Annotation tags. De documentatie beschrijft ook dat de resultaten als triples met Open Annotation tags geleverd worden. De GGC – OCLC tags waren ook aanwezig, maar deze zijn niet allemaal uitgezocht. Deze tags staan beschreven in hun documentatie, maar deze zullen niet allemaal relevant zijn. Het valt niet binnen de scope van het onderzoek om al deze tags te onderzoeken, de beschrijvingen zijn immers aanwezig in de documentatie. Na het bestuderen van de resultaten blijkt dat er slechts maximaal 10 zoekresultaten binnen komen. Na overleg met de ontwikkelaar van de API bleek dat dit is omdat de API gebruikt wordt door alle bibliotheken, en er daardoor veel data verstuurd wordt. Als al de zoekresultaten door iedereen opgehaald zouden worden door middel van deze SRU query, dan zou dit volgens de ontwikkelaar wel dagen kunnen duren. Ook bleek dat er alleen een lijst van 10 resultaten terug kwam, deze had nog wel meer informatie bij zich, maar niet de omschrijving. Ook is het mogelijk dat de zoekresultaten soms veranderen, dit komt door wijzingen in de nodes in de achtergrond van het systeem. Dit staat nu ook

in de FAQ van de landelijke bibliotheken. Het vinden van zoekresultaten ligt ook aan hoeveel data er beschikbaar is bij de landelijke bibliotheken betreffende de zoekopdracht. Er hoeft niet altijd een resultaat te zijn op de zoekopdracht, het kan best zijn dat de landelijke bibliotheken hier niks over beschikbaar hebben. Er zat wel bij de resultaten een link bij naar de pagina van het resultaat, waar alle informatie over het resultaat te vinden was.

De tabel laat zien dat de functionaliteiten bij de demo, en de documentatie beiden aanwezig zijn. Dit geeft aan dat de documentatie over de API klopt, en dat de API de functionaliteiten ook daadwerkelijk kan uitvoeren.

3.2 Deelvraag 2 – Gebruiken API

2. *Hoe kan de API gebruikt worden om informatie van de landelijke bibliotheken te ontsluiten op basis van een zoekopdracht?*

3.2.1 Resultaten

Met de kennis over de API zijn er ontwerpen voor proof of concepts gemaakt, deze zijn terug te vinden in bijlage 1,2 en 3. In deze ontwerpen zijn de eisen voor de proof of concepts beschreven. Ook de acceptatietesten voor de proof of concepts, en de pakketselectie van de libraries staan hier in beschreven.

De volgende functionaliteiten werden beschreven in het functioneel ontwerp:

- UC-01 – Verstuur zoekopdracht
- UC-02 – Ontvang resultaat
- UC-02 – Bevragen data

Vervolgens werd er getest of de API gebruikt kon worden met de externe library, maar hierna werd er al spoedig overgestapt op het gebruiken van de demo van de bibliotheken. Dit kwam onder andere doordat er problemen waren bij het werkende krijgen van de externe library met de resultaten van de API. Zo kon de externe library niet alle resultaten van de API gebruiken en verwerken. Daarom is er na het ontwikkelen van de individuele componenten al overgestapt naar het gebruiken van de demo. Dit werd ook aangeraden door de bedrijfsbegeleider.

Het testen van de proof of concepts werd uitgevoerd met behulp van de test cases uit het functioneel ontwerp in bijlage 2. Dit is een van de test cases.

Identificatienummer test case:	01
Test case voor de use case:	UC-01 – Verstuur zoekopdracht
Test specificatie:	De gebruiker kan het proof of concept uitvoeren. Hierdoor wordt de zoekopdracht verzonden naar de API.
Te testen requirements:	F-R1 - Uitvoeren zoekopdracht.

De actoren	De gebruiker van het proof of concept.
Pre Condities:	De gebruiker moet de query in de code ingevuld hebben. In deze query bevindt zich de zoekterm, of de uitgebreide zoekterm. Ook moet de gebruiker een API key hebben, en een internet verbinding.
Voorafgaande stappen of test cases:	Geen, dit is de eerste stap getest en genomen moet worden om te applicatie te kunnen gebruiken
Verwacht resultaat:	Het proof of concept verstuurt de zoekopdracht succesvol naar de API.
Uitvoerder test case:	Arnoud Aarnoudse
Resultaat:	Het Proof of Concept verstuurt inderdaad de zoekopdracht naar de API. Dit lukte met de enkele zoekterm, en de uitgebreide zoekterm.
Opmerkingen:	De zoekterm moet via de SRU manier uitgebreid worden, zoals "AND", "OR" en dergelijke. Deze zijn in de documentatie terug te vinden.

Voor de technische resultaten van de proof of concepts bevinden zich hier nog wat voorbeelden.

Een voorbeeld van de code van een proof of concept is te zien in figuur 18. Zo is er te zien hoe de SRU URL aangeroepen wordt, en hoe deze resultaten opgeslagen worden in een bestand.

```

10 function GetResults(){
11   $term = 'dijk';
12   $query = "http://api.ozp.accl.bibliotheek.nl/sru/summary?operation=searchRetrieve&query=$term&version=1.2&x-d";
13   $xml = file_get_contents($query);
14   $file = 'results.rdf';
15   file_put_contents($file, $xml);
16   echo($xml);
17 }
18
19 //http://www.w3.org/ns/oa#hasTarget of alle rdf:labels loopen
20 // coa:hasTarget rdf:resource="http://data.bibliotheek.nl/ggc/ppn/383798981"/>
21
22 //FOR LOOP HERE
23 function GetContent(){
24   $item = "334490464";
25   $query1 = "http://api.ozp.accl.bibliotheek.nl/about?profile=raw&uri=http://data.bibliotheek.nl/ggc/ppn/$item&";
26   $xml1 = file_get_contents($query1);
27   $file1 = 'content.rdf';
28   file_put_contents($file1, $xml1);
29   echo($xml1);
30 }
31
32 //echo GetBetween();
33 GetResults();
34 echo "-----<br>";

```

Figuur 18: Voorbeeld code proof of concept

De resultaten van dit proof of concept zijn te zien in figuur 19. Zo is er te zien hoe hier veel verschillende triples staan met hun eigen tags.

```

-----<br>
dcterms:audience.uri.byOntologyMidden Groep 42 dcterms:audience.uri.byOntologyMidden Groep 31 dcterms:audience.uri.byOntologyAngst5 dcterms:audience.uri.byOntologyHbo en universiteit4 dcterms:audience.uri.byOntologyLet op met kinderen tot 12 jaar6 dcterms:audience.fictie: 9-12 jaar27 dcterms:audience.uri.byOntologyJeugd fictie: 7-8 jaar28 dcterms:audience.uri.byOntologyJeugd fictie: aanvankelijk lezen6 dcterms:audience.uri.byOntologyOntspanning voor volwassenen42 dcterms:audience.uri.byOntologyDiscriminatie1 dcterms:audience.uri.byOntologySchadelijk/alle leeftijden1 dcterms:audience.uri.byOntologySeks1 dcterms:audience.uri.byOntologyJeugd100 schema:genre.uri.byOntologyNon-Fictie dcterms:creator.uri.byOntologyDijk. Louis van647 dcterms:creator.uri.byOntologyDorrestijn. Hans90 dcterms:creator.uri.byOntologyDort. Wietke dcterms:creator.uri.byOntologyDijk. Martin van94 dcterms:creator.uri.byOntologyDijk. Hanneke van86 dcterms:creator.uri.byOntologyDijk. Peter holding.lh.heldAt.uriBibliotheek Vlissingen199 holding.lh.heldAt.uriBibliotheek Terneuzen147 holding.lh.heldAt.uriBibliotheek Goes147 holding.lh.heldAt.uriBibliotheek Zierikzee84 holding.lh.heldAt.uriBibliotheek Oostburg73 holding.lh.heldAt.uriBibliotheek Hulst73 holding.lh.heldAt.uri.byOntologyStichting Zeeuwse Bibliotheken (POI)3813 holding.lh.heldAt.uri.byOntologyBibliotheek Terneuzen147 holding.lh.heldAt.uri.byOntologyBibliotheek Oosterschelde515 holding.lh.heldAt.uri.byOntologyBibliotheek Oost-Souburg111 holding.lh.heldAt.uri
-----<br>
Wild water2012Tijdens de ontmoeting van een getrouwde vrouw met haar jonge minnaar in een vakantiehuisje bij Katwijk breekt een dijk door: ze cmThriller2.13.16.19.25.27.28.33.34.40.41.62.201.283.286.331.332.333.334.336.337.338.339.340.342.343.344.345.346.348.350.351.353.354.355.06-111999-21-09-1417-59-12Aavrdaccontent duttekstxtxdamedia/dutzonder mediumnrdacarrer/dutbandncobdragerboekA334491 2012EbtbEbtbFCTVLWnl@Wild waterRuben van Dijk304967688DijkDijkvanvanRubenRubenpseud. van Thomas van SlobbeBrunaUtrecht280 p2 een getrouwde vrouw met haar jonge minnaar in een vakantiehuisje bij Katwijk breekt een dijk door: ze lijken als ratten in de val te zitten.Literaire van2012Bruna. UtrechtNederlandsTijdens de ontmoeting van een getrouwde vrouw met haar jonge minnaar in een vakantiehuisje bij Katwijk breekt

```

Figuur 19: Voorbeeld resultaten proof of concept.

Tijdens het experimenteren kwamen er uit de proof of concepts inderdaad resultaten uit in een RDF format, met triples en open annotation tags.

3.2.2 Analyse

De acceptatietesten uit het functioneel ontwerp werden uitgevoerd. Hieruit bleek dat de student en de begeleider beiden tevreden waren over de werking van het proof of concepts, en de resultaten van de API. De testen toonden aan dat de API gebruikt kon worden, en dat de functionaliteiten van de API het ook deden door middel van de proof of concepts.

Bij het ontwikkelen van de proof of concepts werd er eerst gekozen voor de RAP library, maar deze bleek uiteindelijk niet in staat de resultaten van de zoekopdracht uit te lezen. Daarom werd er vervolgens overgestapt op de 2^e keuze van het pakketselectie, de EasyRDF library. Maar ook deze library was niet in staat om de zoekresultaten uit te lezen. Zo konden beide libraries wel het bestand lezen waar alle informatie van 1 item in staat, maar niet het bestand waar de lijst van 10 items in staat. Vanwege tijdsgebrek kon er uiteindelijk geen werkend prototype gemaakt worden van de proof of concepts. Het was verstandiger geweest om collega's om hulp te vragen voor de verdere ontwikkeling van het prototype, maar uiteindelijk werd er samen met de bedrijfsbegeleider en stagebegeleider besloten dat het prototype niet afgerond hoefde te worden.

3.3 Deelvraag 3 – Uitbreiden zoekopdracht

3. *Hoe kan de zoekopdracht uitgebreid worden met behulp van de grammatica der concepten?*

3.3.1 Resultaten

Uit het gesprek bleek dat de uitgebreide zoekopdracht bestaat uit de originele zoekopdracht plus de bijbehorende zoektermen. Deze bijbehorende zoektermen kunnen op twee manieren gevonden worden. Een van deze methoden is door de bijbehorende zoektermen op te zoeken in de grammatica der concepten van het lectoraat. De thesaurus is beschikbaar op de volgende locatie:

<http://195.93.238.49/dementie/index.php/Home>.

De tweede manier is om de zoekresultaten van de originele zoekopdracht te gebruiken om hier relevante termen uit te halen, en deze vervolgens toe te voegen aan de originele zoekopdracht. Deze resultaten zijn terug te vinden in de resultaten van de experimenten in bijlage 5.

Bij het uitvoeren van de zoekopdracht kan er ook gebruik gemaakt worden van filters en zoekoperators van de CQL. Deze toevoegingen kunnen gebruikt worden om de zoekopdracht nog verder uit te breiden als dit nodig zou zijn. In het volgende voorbeeld zijn alle filters en operators in een SRU query ge-highlight. Voor al de mogelijkheden voor de SRU is de documentatie beschikbaar. (SRU/CQL, 2013)

```
/sru/summary?operation=searchRetrieve&query=Dementie AND Neurologie&version=1.2&x-drilldown-format=rdxml&x-filter-common-keys=false&x-suggestionsQuery=Dementie Neurologie&x-term-drilldown=dterms:subject
```

3.3.2 Analyse

Bij de documentatie stond ook dat er gebruik gemaakt wordt van SRU, wat weer gebruik maakt van CQL. Bij CQL zijn er meerdere mogelijkheden beschikbaar voor het manipuleren van de zoekopdracht. De CQL in de SRU geeft veel mogelijkheden voor de zoekopdrachten. Maar een voorbeeld hiervan is: ***dc.title all "vis kikker"***.

Hier wordt er gezocht naar alle titels van de artikelen waar de woorden vis en kikker in voorkomen

Door het gebruiken van de mogelijkheden van CQL kan er op verschillende onderdelen gezocht worden binnen de landelijke bibliotheken. Maar een ander voorbeeld is dat er ook gebruik gemaakt kan worden van zoekoperators. Deze kunnen gebruikt worden met: 'AND', 'OR', 'NOT' en 'PROX'. De functionaliteiten hiervan zijn ook terug te vinden in de documentatie van de CQL.

Voor het experiment is er gebruik gemaakt van de AND, aangezien dit is wat de demo gebruikt, en omdat dit is hoe de begeleider de uitgebreide zoekopdracht beschreef. (SRU/CQL, 2013) De AND zoekt dus naar alle termen die in de zoekopdracht staan. AND is Engels voor "en", er hiermee wordt er gezocht naar zoekterm1 en zoekterm2. Het zoeken met de door de thesaurus uitgebreide zoekterm, en met de door de zoekresultaten uitgebreide zoekopdracht gaven dezelfde soorten resultaten terug. Ze konden dus beide als uitgebreide zoekopdracht gebruikt worden.

3.4 Deelvraag 4 – Onderzoeken impact

4. *Welke kwalitatieve impact heeft het uitbreiden van de zoektermen van de zoekopdracht op het resultaat van de zoekopdracht?*

3.4.1 Resultaten

Uit het open interview is gebleken dat de opdrachtgever kwaliteit ziet als het aantal relevante data dat verkrijgbaar is. De bronnen, lengte of betrouwbaarheid van de data heeft hier geen hoge prioriteit. De student zelf was het hier ook wel mee eens, aangezien er uiteindelijk nieuwe data toegevoegd moet worden aan de bron. Dit is omdat het over een grote hoeveelheid data gaat, met ook een grote hoeveelheid meta data. Zelfs als er weinig of korte bronnen tussen zitten, dan nog is het, door het grote aantal beschikbare meta-data en artikelen, een grote hoeveelheid data.

Vervolgens is het experiment uitgevoerd. De opzet en conclusie van de experimenten zijn terug te vinden in bijlage 4 en 5. In deze bijlagen is er te lezen wat de stappen waren voor het uitgevoerde experiment, en wat de antwoorden waren op deze stappen.

In de volgende tabel worden de begrippen die overeenkomen met de begrippen van de grammatica der concepten oranje gekleurd. De nieuwe begrippen worden groen gekleurd, en de missende begrippen worden rood gekleurd.

Grammatica der Concepten	Dementie	Dementie Neurologie	Dementie Apathie
1. ontmoetingscentra dementie	1. dementie - 10 times (4.9%)	1. dementie - 9 times (8.7%)	1. dementie - 3 times (3.9%)
2. vasculaire dementie	2. en - 10 times (4.9%)	2. neurologie - 8 times (7.7%)	2. apathie - 3 times (3.9%)
3. frontotemporale dementie	3. voor - 8 times (3.9%)	3. geneeskunde - 8 times (7.7%)	3. ouderen - 3 times (3.9%)
4. body-dementie	4. met - 7 times (3.4%)	4. van - 5 times (4.8%)	4. het - 3 times (3.9%)
5. psychogeriatrische zorg	5. van - 7 times (3.4%)	5. proefschrift - 4 times (3.8%)	5. van - 3 times (3.9%)
6. apathie	6. gezondheid - 5 times (2.4%)	6. universiteit - 4 times (3.8%)	6. met - 3 times (3.9%)
7. MMSE	7. praktische - 5 times (2.4%)	7. en - 3 times (2.9%)	7. zijn - 3 times (3.9%)
8. progressieve supranucleaire paralyse	8. hun - 5 times (2.4%)	8. een - 3 times (2.9%)	8. bij - 2 times (2.6%)
9. Alzheimer telefoon	9. lichaam - 5 times (2.4%)	9. het - 3 times (2.9%)	9. psychiatrie - 2 times (2.6%)
10. Afleidende met iets vertrouwd	10. neurologie - 5 times (2.4%)	10. amsterdam - 3 times (2.9%)	10. begeleiden - 2 times (2.6%)
11. vakantieopvang	11. algemeen - 5 times (2.4%)	11. dementia - 3 times (2.9%)	11. mensen - 2 times (2.6%)
12. Langzaam wennen aan nieuwe situaties.	12. op - 5 times (2.4%)	12. ziekte - 2 times (1.9%)	12. die - 2 times (2.6%)
13. Reisberg	13. de - 5 times (2.4%)	13. de - 2 times (1.9%)	13. apathisch - 2 times (2.6%)
14. Op vakantie gaan in een kleinschalige omgeving	14. informatie - 4 times (1.9%)	14. tot - 2 times (1.9%)	14. dementerende - 2 times (2.6%)
15. milde cognitieve stoornissen	15. zijn - 4 times (1.9%)	15. neuropsychiatrie - 2 times (1.9%)	15. psychogeriatrische - 2 times (2.6%)
16. diagnostiek	16. het - 4 times (1.9%)	16. cognitieve - 2 times (1.9%)	16. zorg - 2 times (2.6%)
17. Wet zorg en dwang kortdurende opvang	17. over - 4 times (1.9%)	17. diagnostiek - 2 times (1.9%)	17. verzorgende - 2 times (2.6%)
	18. haar - 4 times (1.9%)	18. kan - 2 times (1.9%)	18. beroepen - 2 times (2.6%)
	19. een - 4 times (1.9%)	19. parkinson - 1 time (1.0%)	19. richtlijnen - 2 times (2.6%)
	20. medische - 3 times (1.5%)	20. depressie - 1 time (1.0%)	20. en - 2 times (2.6%)
	21. mensen - 3 times (1.5%)	21. depressiviteit - 1 time (1.0%)	21. geneeskunde - 1 time (1.3%)
	22. bij - 3 times (1.5%)	22. intelligentie - 1 time (1.0%)	22. begeleiding - 1 time (1.3%)
	23. familie - 3 times (1.5%)	23. schets - 1 time (1.0%)	23. geriatrische - 1 time (1.3%)
	24. die - 3 times (1.5%)	24. tevens - 1 time (1.0%)	24. neurologie - 1 time (1.3%)
	25. ze - 3 times (1.5%)	25. bijdrage - 1 time (1.0%)	25. depressief - 1 time (1.3%)
	26. leven - 3 times (1.5%)	26. existentieel - 1 time (1.0%)	26. achtergronden - 1 time (1.3%)
	27. wat - 3 times (1.5%)	27. antropologisch - 1 time (1.0%)	27. doel - 1 time (1.3%)
	28. boek - 3 times (1.5%)	28. georiënteerde - 1 time (1.0%)	28. verantwoording - 1 time (1.3%)
	29. leeftijd - 2 times (1.0%)	29. vergeten - 1 time (1.0%)	29. depressieve - 1 time (1.3%)
	30. patiënten - 2 times (1.0%)	30. specialisme - 1 time (1.0%)	30. depressies - 1 time (1.3%)
	31. jaar - 2 times (1.0%)	31. bij - 1 time (1.0%)	31. depressiviteit - 1 time (1.3%)
	32. veel - 2 times (1.0%)	32. dementiesyndroom - 1 time (1.0%)	32. neurosen - 1 time (1.3%)
	33. gids - 2 times (1.0%)	33. consensusbijeenkomst - 1 time (1.0%)	33. de - 1 time (1.3%)
	34. hoe - 2 times (1.0%)	34. die - 1 time (1.0%)	34. te - 1 time (1.3%)
	35. kan - 2 times (1.0%)	35. met - 1 time (1.0%)	35. gedragsproblemen - 1 time (1.3%)
	36. dementerenden - 2 times (1.0%)	36. nog - 1 time (1.0%)	36. voor - 1 time (1.3%)
	37. als - 2 times (1.0%)	37. cognitie - 1 time (1.0%)	37. ook - 1 time (1.3%)
	38. hoofd - 2 times (1.0%)	38. gedrag - 1 time (1.0%)	38. als - 1 time (1.3%)
	39. heeft - 2 times (1.0%)	39. hoofddlijnen - 1 time (1.0%)	39. deze - 1 time (1.3%)
	40. doet - 2 times (1.0%)	40. anatomische - 1 time (1.0%)	40. op - 1 time (1.3%)
	41. er - 2 times (1.0%)	41. boek - 1 time (1.0%)	41. ziektebeelden - 1 time (1.3%)
	42. te - 2 times (1.0%)	42. dat - 1 time (1.0%)	42. probleemgedrag - 1 time (1.3%)
	43. niet - 2 times (1.0%)	43. voor - 1 time (1.0%)	43. verzorgenden - 1 time (1.3%)
	44. door - 2 times (1.0%)	44. patiënten - 1 time (1.0%)	44. helpenden - 1 time (1.3%)
	45. maar - 2 times (1.0%)	45. als - 1 time (1.0%)	45. snel - 1 time (1.3%)
	46. aan - 2 times (1.0%)	46. auteur - 1 time (1.0%)	46. psychische - 1 time (1.3%)
	47. naar - 2 times (1.0%)	47. werd - 1 time (1.0%)	47. om - 1 time (1.3%)
	48. moet - 2 times (1.0%)	48. jaren - 1 time (1.0%)	48. een - 1 time (1.3%)
	49. samen - 2 times (1.0%)	49. niet - 1 time (1.0%)	49. omgang - 1 time (1.3%)
	50. alzheimer - 2 times (1.0%)	50. kennis - 1 time (1.0%)	50. cliënt - 1 time (1.3%)

Als de resultaten van de zoekopdrachten met elkaar vergeleken worden, in plaats van met de begrippen van de grammatica der concepten, dan is er ook te zien dat ook hier weer begrippen terug naar voren komen. Dit is aangetoond in de volgende tabel.

Dementie	Dementie Neurologie	Dementie Apathie
1. dementie - 10 times (4.9%)	1. dementie - 9 times (8.7%)	1. dementie - 3 times (3.9%)
2. en - 10 times (4.9%)	2. neurologie - 8 times (7.7%)	2. apathie - 3 times (3.9%)
3. voor - 8 times (3.9%)	3. geneeskunde - 8 times (7.7%)	3. ouderen - 3 times (3.9%)
4. met - 7 times (3.4%)	4. van - 5 times (4.8%)	4. het - 3 times (3.9%)
5. van - 7 times (3.4%)	5. proefschrift - 4 times (3.8%)	5. van - 3 times (3.9%)
6. gezondheid - 5 times (2.4%)	6. universiteit - 4 times (3.8%)	6. met - 3 times (3.9%)
7. praktische - 5 times (2.4%)	7. en - 3 times (2.9%)	7. zijn - 3 times (3.9%)
8. hun - 5 times (2.4%)	8. een - 3 times (2.9%)	8. bij - 2 times (2.6%)
9. lichaam - 5 times (2.4%)	9. het - 3 times (2.9%)	9. psychiatrie - 2 times (2.6%)
10. neurologie - 5 times (2.4%)	10. amsterdam - 3 times (2.9%)	10. begeleiden - 2 times (2.6%)
11. algemeen - 5 times (2.4%)	11. dementia - 3 times (2.9%)	11. mensen - 2 times (2.6%)
12. op - 5 times (2.4%)	12. ziekte - 2 times (1.9%)	12. die - 2 times (2.6%)
13. de - 5 times (2.4%)	13. de - 2 times (1.9%)	13. apathisch - 2 times (2.6%)
14. informatie - 4 times (1.9%)	14. tot - 2 times (1.9%)	14. dementerende - 2 times (2.6%)
15. zijn - 4 times (1.9%)	15. neuropsychiatrie - 2 times (1.9%)	15. psychogeriatrische - 2 times (2.6%)
16. het - 4 times (1.9%)	16. cognitieve - 2 times (1.9%)	16. zorg - 2 times (2.6%)
17. over - 4 times (1.9%)	17. diagnostiek - 2 times (1.9%)	17. verzorgende - 2 times (2.6%)
18. haar - 4 times (1.9%)	18. kan - 2 times (1.9%)	18. beroepen - 2 times (2.6%)
19. een - 4 times (1.9%)	19. parkinson - 1 time (1.0%)	19. richtlijnen - 2 times (2.6%)
20. medische - 3 times (1.5%)	20. depressie - 1 time (1.0%)	20. en - 2 times (2.6%)
21. mensen - 3 times (1.5%)	21. depressiviteit - 1 time (1.0%)	21. geneeskunde - 1 time (1.3%)
22. bij - 3 times (1.5%)	22. intelligentie - 1 time (1.0%)	22. begeleiding - 1 time (1.3%)
23. familie - 3 times (1.5%)	23. schets - 1 time (1.0%)	23. geriatrische - 1 time (1.3%)
24. die - 3 times (1.5%)	24. tevens - 1 time (1.0%)	24. neurologie - 1 time (1.3%)
25. ze - 3 times (1.5%)	25. bijdrage - 1 time (1.0%)	25. depressief - 1 time (1.3%)
26. leven - 3 times (1.5%)	26. existentieel - 1 time (1.0%)	26. achtergronden - 1 time (1.3%)
27. wat - 3 times (1.5%)	27. antropologisch - 1 time (1.0%)	27. doel - 1 time (1.3%)
28. boek - 3 times (1.5%)	28. georiënteerde - 1 time (1.0%)	28. verantwoording - 1 time (1.3%)
29. leeftijd - 2 times (1.0%)	29. vergeten - 1 time (1.0%)	29. depressieve - 1 time (1.3%)
30. patiënten - 2 times (1.0%)	30. specialisme - 1 time (1.0%)	30. depressies - 1 time (1.3%)
31. jaar - 2 times (1.0%)	31. bij - 1 time (1.0%)	31. depressiviteit - 1 time (1.3%)
32. veel - 2 times (1.0%)	32. dementiesyndroom - 1 time (1.0%)	32. neurosen - 1 time (1.3%)
33. gids - 2 times (1.0%)	33. consensusbijeenkomst - 1 time(1.0%)	33. de - 1 time (1.3%)
34. hoe - 2 times (1.0%)	34. die - 1 time (1.0%)	34. te - 1 time (1.3%)
35. kan - 2 times (1.0%)	35. met - 1 time (1.0%)	35. gedragsproblemen - 1 time (1.3%)
36. dementerenden - 2 times (1.0%)	36. nog - 1 time (1.0%)	36. voor - 1 time (1.3%)
37. als - 2 times (1.0%)	37. cognitie - 1 time (1.0%)	37. ook - 1 time (1.3%)
38. hoofd - 2 times (1.0%)	38. gedrag - 1 time (1.0%)	38. als - 1 time (1.3%)
39. heeft - 2 times (1.0%)	39. hoofdlijnen - 1 time (1.0%)	39. deze - 1 time (1.3%)
40. doet - 2 times (1.0%)	40. anatomische - 1 time (1.0%)	40. op - 1 time (1.3%)
41. er - 2 times (1.0%)	41. boek - 1 time (1.0%)	41. ziektebeelden - 1 time (1.3%)
42. te - 2 times (1.0%)	42. dat - 1 time (1.0%)	42. probleemgedrag - 1 time (1.3%)
43. niet - 2 times (1.0%)	43. voor - 1 time (1.0%)	43. verzorgenden - 1 time (1.3%)
44. door - 2 times (1.0%)	44. patiënten - 1 time (1.0%)	44. helpenden - 1 time (1.3%)
45. maar - 2 times (1.0%)	45. als - 1 time (1.0%)	45. snel - 1 time (1.3%)
46. aan - 2 times (1.0%)	46. auteur - 1 time (1.0%)	46. psychische - 1 time (1.3%)
47. naar - 2 times (1.0%)	47. werd - 1 time (1.0%)	47. om - 1 time (1.3%)
48. moet - 2 times (1.0%)	48. jaren - 1 time (1.0%)	48. een - 1 time (1.3%)
49. samen - 2 times (1.0%)	49. niet - 1 time (1.0%)	49. omgang - 1 time (1.3%)
50. alzheimer - 2 times (1.0%)	50. kennis - 1 time (1.0%)	50. cliënt - 1 time (1.3%)

3.4.2 Analyse

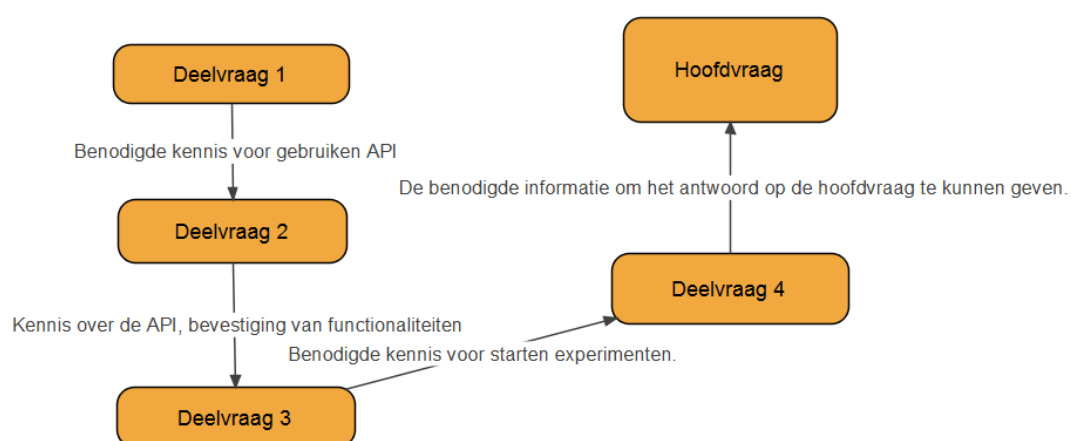
De zoekresultaten van de enkele en de uitgebreidere zoektermen verschillen van elkaar. Zo komen er inderdaad nieuwe artikelen naar voren aangezien er met een nieuwe zoekterm gezocht wordt. Dit levert dus inderdaad nieuwe artikelen, en dus nieuwe data die gebruikt kan worden voor het vinden van relaties tussen de begrippen. Maar soms komen er ook al eerder aanwezige zoekresultaten naar voor. Deze leveren dan geen nieuwe informatie, maar bevestigen de al aanwezige informatie. Ook ontbreekt er vaak data bij verschillende velden zoals de omschrijving, of de onderwerpen. Zo zijn velden soms niet volledig of gewoon niet ingevuld. Uiteindelijk heeft dit geen negatieve invloed op de zoekresultaten, aangezien er altijd velden ingevuld zullen zijn, en deze dus data leveren die vervolgens weer gebruikt kan worden.

In de eerste tabel van de resultaten is te zien dat er begrippen zijn die overeenkomen met de begrippen van de grammatica der concepten. Het begrip apathie is teruggevonden omdat dit gebruikt is voor de uitgebreide zoekopdracht, maar het begrip diagnostiek werd ook gevonden. Hier was niet naar gezocht. Dit betekent dus dat de zoekresultaten inderdaad gerelateerde begrippen leveren.

Er zijn ook veel begrippen die nog niet aanwezig zijn in de grammatica der concepten, dit komt ook grotendeels omdat er nog niet veel begrippen aanwezig zijn in de grammatica der concepten. Maar het toont wel aan dat er inderdaad nieuwe begrippen gevonden kunnen worden. In de tweede tabel is te zien dat de uitgebreide zoekopdrachten, in vergelijking met de standaard zoekopdracht, veel overeenkomende, maar ook nieuwe begrippen bevatten. Dus ook hiermee is aan te tonen TODO

Aan de resultaten is dus te zien dat ook bij het uitbreiden van de originele zoekterm er relevante nieuwe begrippen naar voren komen. Het is mogelijk dat het aantal resultaten kleiner is, maar dit is niet van belang, aangezien dit niet onder de gedefinieerde term “kwaliteit” valt.

3.5 Totaalbeeld



Figuur 20: Concept map totaalbeeld deelvragen.

4 Discussie

In dit hoofdstuk worden de deelvragen bediscussieerd, waarna er vervolgens een conclusie wordt getrokken.

4.1 Deelvraag 1 – Onderzoeken API

1. Hoe werkt de API van de landelijke bibliotheken?

4.1.1 Discussie

Het beoogde doel van de deelvraag was om kennis op te doen over de API, zodat deze vervolgens gebruikt kan worden.

De documentatie van de API werd onderzocht, dit zorgde ervoor dat de werking van de API bekend werd. Dit werd vervolgens nogmaals onderzocht door het uitvoeren van de beschreven functionaliteiten op de demo. Dit betrof het kunnen zoeken door middel van een zoekopdracht, en het filteren van de zoekresultaten. Dit was de beste methode omdat hiermee de werking, en de validiteit van de documentatie van de API getest kon worden.

De resultaten die hier uitkwamen klopten ook. Zo werd er voldaan aan de verwachtingen van de student met betrekking tot de documentatie. De documentatie beschreef daadwerkelijk ook de werking van de API, en de API kon inderdaad zoekopdrachten uitvoeren zoals in de documentatie beschreven stond. Daarbij kwamen de resultaten in het correcte formaat en met de juiste inhoud uit de test. De resultaten bevatten ook zoals er beschreven stond meerdere tags van het GGC – OCLC. Verder stonden er nog wel functionaliteiten beschreven in de documentatie die niet bruikbaar waren, maar dat werd daar ook gemeld.

De resultaten zijn betrouwbaar aangezien deze nog eens besproken zijn met de ontwikkelaar van de API, en vervolgens zelf nog zijn gecontroleerd met behulp van de demo. De volledigheid van deze resultaten is dus ook naar behoren, het zou kunnen dat er nog kleine puntjes niet beschreven zijn, maar deze lijken momenteel geen invloed te hebben op het gebruiken van de functionaliteiten.

4.1.2 Deelconclusie

De API maakt gebruik van SRU. De functionaliteit van de SRU is te zien in de bron (Groeneveld, Medema, Janssen, & Morée, 2015). Zo is er te zien in de documentatie dat de zoektermen in de query worden gezet als een normaal woord. De API is aan te spreken met behulp van een URL waar vervolgens een APIkey achter geplaatst moet worden. De API is te gebruiken door elke methode of programmeertaal die een URL kan bezoeken, en die hier resultaten van kan ontvangen.

De zoektermen komen in de URL van de SRU query binnen, dit is terug te zien in de documentatie van de bron (Groeneveld, Medema, Janssen, & Morée, 2015). De zoektermen zien er normaal uit, het is gewoon een woord. Dit is hoe ze ingetypt worden, er hoeft niets aan gedaan te worden. Het kan ook als een normaal woord neergezet worden in de query.

Het resultaat komt er uit als een XML/RDF formaat, dat gevuld is met meerdere triples. Bij het beschrijven van de informatie in het RDF bestand wordt er gebruik gemaakt van open annotations. De triples bevatten ook tags, deze staan beschreven in de documentatie in de bron (Groeneveld, Medema, Janssen, & Morée, 2015). Maar niet alle tags zijn beschreven, sommigen van deze tags moeten zelf worden opgezocht bij het GGC – OCLC, hier staat per tag beschreven wat deze data betekend. (Richtlijnen voor het catalogiseren in het GGC). Deze tags bevatten meta data.

4.2 Deelvraag 2 – Gebruiken API

2. *Hoe kan de API gebruikt worden om informatie van de landelijke bibliotheken te ontsluiten op basis van een zoekopdracht?*

4.2.1 Discussie

Het doel van de deelvraag was om te weten hoe de API gebruikt kon worden. Hiermee kon gevalideerd worden dat de API ook echt bruikbaar was, en hierdoor kwam er meer kennis over de API en de resultaten beschikbaar.

Er werd gekozen voor proof of concepts omdat alleen de zoekmethoden van de API gebruikt moesten worden om te bewijzen dat deze het deden. Dit was de snelste en meest geschikte methode om dit te doen. Er moesten vooraf geen uitgebreide ontwerpen en diagrammen gemaakt worden omdat de student met een agile methode het proof of concept ontwikkelde, en dus aan de slag kon met de API om naar resultaten te streven op een kleine schaal. Vervolgens werd er om dezelfde reden ook gekozen om de te gebruiken libraries te testen in proof of concepts. Deze libraries waren nodig om de informatie van de API te kunnen verwerken.

Omdat het niet gelukt is de externe libraries te laten werken met de API, en omdat het laten functioneren van de API niet het belangrijkste onderdeel van het onderzoek was, werd er uiteindelijk gekozen om de antwoorden op de deelvragen te vinden door het gebruiken van de demo. De libraries werkten niet volledig met de API, zo kon de lijst van zoekresultaten niet bevraagd worden. De artikel details pagina werkte wel. Daarom is er vervolgens overgestapt op de demo. Bij de demo kon de student meteen zoeken met de zoekterm, en kon de resultaten hiervan zien en doornemen. Dit hielp met het beantwoorden van de deelvragen, die vooral hierover gingen. Vervolgens werd de verkregen kennis van de proof of concepts gecombineerd met de verkregen kennis van het experimenteren met de demo, en samen leidde dit tot het antwoord op de deelvraag.

De resultaten zijn gedeeltelijk betrouwbaar. Het is gelukt om de meeste functionaliteiten werkend te krijgen, maar het uitlezen van het bestand met de lijst van resultaten is niet gelukt. Bij de demo van de landelijke bibliotheken werkt dit wel, maar dit is de student niet gelukt om te repliceren. Bij het ontwikkelen van de proof of concepts had er eerder overlegd moeten worden met de collega's om proberen dit probleem op te lossen. Ook had er gebruik gemaakt kunnen worden van een andere ontwikkeltaal, hier waren mogelijk wel libraries voor beschikbaar die alle bestanden succesvol uit konden lezen. Maar dit waren problemen aan de kant van het proof of concept, de API werkte zoals er beschreven stond, en dat maakt de resultaten betrouwbaar.

4.2.2 Deelconclusie

De API is aan te roepen door middel van een URL, en maakt gebruik van SRU. Het aanroepen van de API kan via meerdere programmeertalen, zo is voor het proof of concept PHP gebruikt. De API kan gebruikt worden om een zoekquery op te stellen in een SRU URL, met meerdere operators en filters, om zo een lijst aan zoekresultaten te ontvangen. Het formaat van de zoekresultaten staat ook beschreven in de documentatie van de API (Groeneveld, Medema, Janssen, & Morée, 2015), maar bij de proof of concepts is er gebruik gemaakt van het aanroepen van de XML.RDF. Aangezien semantische data in een RDF formaat geleverd wordt, en dit dus uniform is met de al aanwezige techniek bij het lectoraat. Deze resultaten bevatten triples en zouden vervolgens met een library voor het bewerken en lezen van RDF data bevestigd kunnen worden om zo de gewenste resultaten hier uit te ontvangen.

4.3 Deelvraag 3 – Uitbreiden zoekopdracht

3. *Hoe kan de zoekopdracht uitgebreid worden met behulp van de grammatica der concepten?*

4.3.1 Discussie

Het doel van deze deelvraag was om te weten hoe de zoekopdracht uitgebreid kan worden, anders kan er immers geen uitgebreide zoekopdracht uitgevoerd worden.

Eerst werd er gepraat met de begeleiders over hoe de uitgebreide zoekopdracht van de grammatica der concepten er uit zou zien. Hier bleek uit dat dit een zoekopdracht is, met meer woorden. Dus het toevoegen van meerdere zoektermen aan het originele zoekterm. Vervolgens werd er besloten door de student om de resultaten van de standaard zoekopdracht te onderzoeken, om hier vervolgens zelf nieuwe begrippen uit te halen om toe te voegen aan de standaard zoekopdracht. Zo werd het gebruiken van de grammatica der concepten nagebootst, en zo kwamen er meer resultaten beschikbaar. Uit de documentatie van de API bleek ook dat er bij het gebruiken van CQL er ook meerdere operators en filters beschikbaar zijn voor de zoekopdracht.

De resultaten van het interview zijn betrouwbaar, aangezien deze direct van de begeleiders komen. De begeleiders weten immers zelf hoe hun de zoekopdracht uitbreiden. De resultaten van het nabootsen van de grammatica der concepten door het gebruiken van de resultaten van de experiment is minder betrouwbaar. Het experiment is eenmaal uitgevoerd, en gaf positieve resultaten. Het zou goed kunnen dat de volgende keer het andere resultaten geeft. Ook kwamen niet met alle zoekopdracht zoekresultaten naar boven. Het experiment zou vaker uitgevoerd moeten worden om betrouwbaardere resultaten te krijgen. Ook zou het toepassen van de verschillende soorten filters kunnen leiden tot nieuwe resultaten. Het uitproberen van alle mogelijkheden van de filters en operators van de CQL had de resultaten betrouwbaarder gemaakt, hier was alleen helaas geen tijd voor.

4.3.2 Deelconclusie

De zoekopdracht kan uitgebreid worden door hier meerdere zoektermen aan toe te voegen. Deze zoektermen komen uit de grammatica der concepten, die meerdere begrippen bevat. Door het

selecteren van begrippen uit deze lijst, en deze toe te voegen aan de originele zoekopdracht kan de zoekopdracht uitgebreid worden. Vervolgens kunnen er meerdere zoekoperators aan de zoekopdracht toegevoegd worden. Dit is mogelijk door de CQL technieken binnen de SRU, deze bevatten veel verschillende mogelijkheden voor het gebruiken van filters om de zoekopdracht te wijzigen. Zoektermen kunnen ook uit de zoekresultaten gehaald worden, en op deze wijze recursief gebruikt worden.

4.4 Deelvraag 4 – Onderzoeken impact

4. *Welke kwalitatieve impact heeft het uitbreiden van de zoektermen op het resultaat van de zoekopdracht?*

4.4.1 Discussie

Het doel van deze deelvraag was om de kennis op te halen die benodigd is voor het beantwoorden van de deelvraag, door het onderzoeken van de impact met behulp van experimenten kan de hoofdvraag beantwoord worden.

Als eerst moest de term kwaliteit verduidelijkt worden. Hiervoor vormde de student zijn eigen mening over de definitie van kwaliteit. Vervolgens voerde de student een gesprek met de begeleider, in de vorm van een open interview. Er werd hiervoor gekozen omdat er geen standaard antwoord op te vinden was, en omdat er samen gepraat moest worden over wat voor de begeleider de eisen zijn aan begrip kwaliteit. Vervolgens na het vaststellen van de definitie van kwaliteit werden de experimenten uitgevoerd. Bij de experimenten werden de resultaten van de enkele zoekterm en de uitgebreide zoekterm met elkaar vergeleken. De details van het experiment zijn terug te vinden in bijlage 5. Er werd gekozen voor een vergelijkend onderzoek omdat de impact gevonden moest worden. De beste keuze om dit te bereiken was om de resultaten met elkaar te vergelijken zodat in de resultaten, die ook terug te vinden zijn in bijlage 5, het antwoord op de deelvraag te gevonden kon worden.

Wel werden er minder resultaten gevonden door het uitbreiden van de zoekopdracht, maar zolang het nieuwe informatie gaf dan voldoet het nog aan de definitie van kwaliteit. Als er meer data beschikbaar word later, of er worden meerdere zoekopdracht met verschillende CQL operators uitgevoerd, dan zullen er ook meer artikelen beschikbaar zijn. Ook hadden sommige artikelen geen omschrijving. Bij deze artikelen konden er alsnog begrippen uit de titel en omschrijving gehaald worden, maar het geeft wel aan dat niet alle artikelen volledig zijn. Ook kwamen er af en toe enkele spellingsfouten in voor. Maar aangezien er gezocht wordt naar hoe vaak een begrip voorkomt, of wat de relevantie is, zullen de fout gespelde woorden vanzelf afvallen, tenzij deze in meerdere artikelen herhalend fout geschreven zijn.

De resultaten zijn gedeeltelijk betrouwbaar. Het is duidelijk dat, door wat het lectoraat verstaat onder het begrip “kwaliteit”, zolang er resultaten zijn, dit bijdraagt aan een positieve kwalitatieve impact. Maar het experiment is slechts eenmaal uitgevoerd, hierdoor kan het zijn dat de volgende experimenten andere resultaten zouden kunnen geven, al zou dit niet veel invloed hebben op de conclusie. Om de resultaten betrouwbaarder te maken zouden de experimenten meerdere keren uitgevoerd moeten worden. Ook zouden de experimenten aangepast kunnen worden als er meerdere

soorten uitgebreide zoekopdrachten waren, met meerdere zoekfilters en operators. Hiervan is bij dit experiment nog geen gebruik van gemaakt.

4.4.2 Deelconclusie

Na het uitbreiden van de zoekopdracht werden sommige begrippen, die bij de enkele zoekvraag werden gevonden, opnieuw weergegeven, waarbij er een ook aantal nieuwe begrippen die in eerste instantie niet door de enkele zoekterm naar voren kwamen, nu wel naar voren kwamen.

Door het vergelijken van de gevonden begrippen en de begrippen die aanwezig waren in de grammatica der concepten, bleek ook dat er veel nieuwe begrippen naar voren kwamen, en dat de grammatica momenteel weinig begrippen bevat betreffende “dementie”.

Door het uitbreiden van de zoekopdracht komen er nieuwe en relevante begrippen naar voor, daarom heeft het uitbreiden van de zoekopdracht een positieve kwalitatieve impact op het resultaat van de zoekopdracht.

5 Conclusie

Welke kwalitatieve impact heeft het uitbreiden van de zoektermen met behulp van de grammatica der concepten op de ontsloten informatie van de API van de landelijke bibliotheken?

Het uitbreiden van de zoektermen met behulp van de grammatica der concepten zorgt er voor dat er minder zoekresultaten tevoorschijn komen. Maar deze zoekresultaten bevatten wel nieuwe begrippen, die vervolgens gebruikt kunnen worden als nieuwe informatie. En als de zoekresultaten dezelfde begrippen bevatten, dan kunnen deze gebruikt worden om al aanwezige relaties tussen begrippen te bevestigen. Het uitbreiden van de zoektermen geeft dus een positieve kwalitatieve impact. Door het wijzigen van de zoekfilters zou dit nog kunnen wijzigen, maar dit is niet volledig getest tijdens het onderzoek.

Het doel van het onderzoek is om te achterhalen welke kwalitatieve impact het uitbreiden van de zoekopdracht met behulp van de grammatica der concepten levert op het resultaat van de zoekopdracht.

Het doel van het onderzoek is bereikt. Zoals in het hierboven gegeven antwoord op de hoofdvraag te lezen is kan er geconcludeerd worden dat er een positieve kwalitatieve impact aanwezig is. Het uitbreiden van de zoekopdracht levert nieuwe relevante begrippen, die vervolgens ook weer gebruikt kunnen worden voor nieuwe zoekopdrachten. En in het geval dat de resultaten al bestaande begrippen opleveren, dan levert dit een bevestiging voor aanwezige relaties tussen de begrippen.

5.1.1 Vergelijking met ander onderzoek/theorie

Op de data en documentatie van de landelijke bibliotheken na is er nog geen eerder onderzoek uitgevoerd, en valt het hier ook niet mee te vergelijken. Voor meer informatie over de werking van de SRU zijn er al wel eerder onderzoeken uitgevoerd die gebruikt zouden kunnen worden. Dit zou kunnen helpen met het implementeren van de API bij het lectoraat. Over semantiek zijn er al meerdere onderzoeken uitgevoerd, ook bij het lectoraat is er zeer veel informatie aanwezig over dit gebied.

5.1.2 Suggesties voor vervolgonderzoek

Er is gebleken uit het huidige onderzoek dat:

- De API van de landelijke bibliotheken werkt zoals in de documentatie staat, maar nog wel uitgebreid kan worden.
- Er veel documentatie beschikbaar is over de tags die gebruikt worden voor het beschrijven van de meta data, en voor het gebruiken van de filters en operators van de CQL in SRU.
- Er heel veel data beschikbaar is bij de landelijke bibliotheken.
- Het uitbreiden van de zoektermen een positieve kwalitatieve impact heeft.

- De ontwikkelaar van de API van de landelijke bibliotheken bereid is om met het lectoraat samen te werken aan een implementatie van de API.

Voor een vervolgonderzoek zijn er meer dan genoeg mogelijkheden. Hiervoor zou het lectoraat samen met de ontwikkelaar van de API in bespreking kunnen gaan. Momenteel kan er zonder de hulp van de ontwikkelaar per zoekopdracht een lijst van maximaal 10 zoekresultaten opgehaald worden. Dit is voor het gebruik niet optimaal. Zo is op deze wijze ook de hoeveelheid data niet optimaal. Door met de ontwikkelaar in bespreking te gaan kan er samen gewerkt worden aan het ophalen van alle gewenste data, dat betekend dus onder andere dat er meer dan 10 resultaten per zoekopdracht opgevraagd kunnen worden. Zo zou er afgesproken kunnen worden tussen de landelijke bibliotheken en het lectoraat welke data ze zouden willen binnenkrijgen. Het zal waarschijnlijk niet nodig zijn om bijvoorbeeld al de GPS gegevens per artikel binnen te krijgen. Dit heeft voor de landelijke bibliotheken ook weer een kleinere impact op hun systeem, en zo kan er samen gewerkt worden aan het afstemmen van het gebruik van de API.

Het is ook zeker mogelijk om de API te gebruiken voor de huidige methoden die ontwikkeld worden bij het lectoraat, de API levert namelijk ook RDF bestanden met triples er in, en kan hier dus op aansluiten. Maar tijdens het testen ging er wel iets mis bij het gebruiken van een library om het RDF bestand te doorzoeken. Dit kan op meerdere manieren opgelost worden, maar het wordt het lectoraat aangeraden om eerst met de landelijke bibliotheken in discussie te gaan. Het is mogelijk dat de data dan misschien uiteindelijk op een andere wijze aangeleverd zou worden.

Het is ook een goed idee om de tags van het GGC – OCLC te onderzoeken, om te kijken wat voor hen mogelijk relevante informatie zou kunnen zijn. Verder is het ook goed om de mogelijkheden van de CQL te onderzoeken, om te kijken welke filters en operators er gebruikt zullen worden bij het gebruiken van de API.

5.1.3 Tot besluit

Het wordt het lectoraat zeker aangeraden om door te gaan met een vervolgonderzoek. Het gebruiken van de API zal zeker meer data geven, waardoor er meer informatie gevonden kan worden. Dit zal een goede toevoeging zijn voor de zoekmethode van het lectoraat. Ook is er bewezen dat het uitbreiden van de zoektermen wel degelijk een positieve kwalitatieve impact levert op de resultaten bij de API, en dat het mogelijk is de API te implementeren. Dit betekend dat het ook bruikbaar zal zijn voor het lectoraat en de grammatica der concepten. Daarom raadt de student aan om toch zeker met de ontwikkelaar van de API verder te werken om de data van de API bij het lectoraat te implementeren.

Literatuur

- aisaac. (2012, 12 13). *SKOS Simple Knowledge Organization System - Home Page*. Retrieved 02 06, 2016, from w3: <https://www.w3.org/2004/02/skos/>
- Beal, V. (n.d.). *API - application program interface*. Retrieved 2 2, 2015, from webopedia: <http://www.webopedia.com/TERM/A/API.html>
- Brickley, D., & Guha, R. (2002, 11 12). *RDF Vocabulary Description Language 1.0: RDF Schema*. Retrieved 2 9, 2015, from w3: <http://www.w3.org/TR/2002/WD-rdf-schema-20021112/>
- Groeneveld, E., Medema, J. P., Janssen, T., & Morée, H. (2015, 06 11). *NBC+ Search Platform Documentation*. Retrieved 6 11, 2015, from wiki.zp.seecr: <https://wiki.zp.seecr.nl/Documentation>
- Houghton Mifflin Harcourt Publishing Company. (n.d.). *semantics*. Retrieved 2 2, 2015, from thefreedictionary: <http://www.thefreedictionary.com/semantics>
- Implementing a Government-wide Semantic Solution to Thesauri*. (2005, 11). Retrieved 2 9, 2015, from pdfpower: <http://www.pdfpower.com/XML2005Proceedings/ship/71/sall-71.HTML>
- Kenniscentra. (2014). Retrieved 2 2, 2015, from hz: <http://hz.nl/nl/werkenleren/Kennisdeling%20en%20samenwerking/Pages/Kenniscentra.aspx>
- LEO Web Team. (2013, 6 28). *Creating, publishing and distributing apps at enterprise scale*. Retrieved 2 9, 2015, from leolearning: <http://leolearning.com/blog/creating-publishing-and-distributing-apps-at-enterprise-scale/>
- Nilsson, M. (2008, 1 14). *Expressing Dublin Core metadata using the Resource Description Framework (RDF)*. Retrieved 2 9, 2015, from dublincore: <http://dublincore.org/documents/dc-rdf/>
- Richtlijnen voor het catalogiseren in het GGC*. (n.d.). Retrieved 02 08, 2016, from oclc: <http://support.oclc.org/ggc/richtlijnen/?id=13&ln=nl>
- Semantic Web*. (n.d.). Retrieved 2 2, 2015, from webopedia: http://www.webopedia.com/TERM/S/Semantic_Web.html
- semantics*. (n.d.). Retrieved 2 2, 2015, from webopedia: <http://www.webopedia.com/TERM/S/semantics.html>
- SRU/CQL. (2013, 12 26). *SRU/CQL*. Retrieved 5 26, 2015, from loc: <http://www.loc.gov/standards/sru/>
- Van de Sompel, H., Ciccarese, P., & Sanderson, R. (n.d.). Retrieved 2 7, 2016, from <https://www.w3.org/ns/oa>
- Watson, M. (2010). *Practical Semantic Web and Linked Data Applications*.
- Zoekplatform 1.8 - API (1.8.0.18)*. (2016, 2 7). Retrieved 2 7, 2016, from apiozp: <http://apiozp.acc1.bibliotheek.nl/>

BIJLAGEN

Bijlage 1 - Requirement Specificatie.

Requirement Specification

Project Naam: Proof of Concept API
Bedrijf: Lectoraat DIO
Informatie Systeem: PHP
Opgesteld door: Arnoud Aarnoudse

Document Eigenaren	Project/Organisatie Rol
Arnoud Aarnoudse	Documentbeheer, Programmeur, Stagiair

Gerefereerde documenten:

Document	Versie
-	-

1 | Requirements specificatie

Hier staan de requirements beschreven, deze zijn onderverdeeld in Functionele requirements, die functionaliteiten beschrijven die in het proof of concept zelf voorkomen, en niet Functionele requirements die system prestaties beschrijven. Deze requirements zijn gebaseerd op de functionaliteiten van wat een zoekmachine moet doen, en op de opdracht van de afstudeerder.

1.1 | FUNCTIONELE REQUIREMENT

Functioneel requirement	MoSCoW	Requirement	Uitleg requirement
F-R1	M	Uitvoeren zoekopdracht.	De gebruiker* moet de zoekopdracht kunnen laten uitvoeren.
F-R2	M	Ontvangen resultaten.	De gebruiker moet de resultaten kunnen ontvangen.
F-R3	S	Bevragen data.	De gebruiker moet de data kunnen bevragen.

*De gebruiker = De persoon die met het Proof of Concept werkt.

1.2 | NIET-FUNCTIONELE REQUIREMENT

Niet functionele requirement	MoSCoW	Requirement	Uitleg requirement	KPI voor Niet functionele requirements
NF-R1	M	Ontvangen resultaten in RDF formaat.	Het proof of concept moet het resultaat van de zoekopdracht ontvangen in de vorm van het RDF formaat, en deze moet triples en open annotations bevatten.	Deze handeling moet binnen 5 seconden voltooid zijn.

Bijlage 2 - Functioneel ontwerp.

Functioneel Ontwerp

Project Naam: Proof of Concept API
Bedrijf: Lectoraat DIO
Informatie Systeem: PHP
Opgesteld door: Arnoud Aarnoudse

Document Eigenaren	Project/Organisatie Rol
Arnoud Aarnoudse	Documentbeheer, Programmeur, Stagiair

Gerefereerde documenten:

Document	Versie
Requirement Specificatie	1.0

Use Cases

In dit hoofdstuk worden de use cases beschreven.

1.1 USE CASES

1.1.1 UC-01 – Verstuur zoekopdracht

Pre Conditie:

De gebruiker moet de query in de code ingevuld hebben. In deze query bevindt zich de zoekterm, of de uitgebreide zoekterm. Ook moet de gebruiker een API key hebben, en een internet verbinding.

Requirement(s):

F-R1 - Uitvoeren zoekopdracht..

Actoren:

De gebruiker van het proof of concept.

Beschrijving:

De gebruiker is in staat een zoekopdracht te versturen naar de API.

1.1.1 UC-02 – Ontvang resultaat

Pre Conditie:

De gebruiker moet een internetverbinding hebben.

De gebruiker moet al een zoekopdracht verstuurd hebben naar de API.

Requirement(s):

F-R2 - Ontvangen resultaten.

NF-R1 - Ontvangen resultaten in RDF formaat.

Actoren:

De gebruiker van het proof of concept.

Beschrijving:

De gebruiker van het proof of concept is in staat de resultaten te ontvangen in RDF formaat.

1.1.2 UC-02 – Bevragen data

Pre Conditie:

De gebruiker moet een internetverbinding hebben.

De gebruiker moet al een zoekopdracht verstuurd hebben naar de API.

Requirement(s):

F-R3 – Bevragen data.

Actoren:

De gebruiker van het proof of concept.

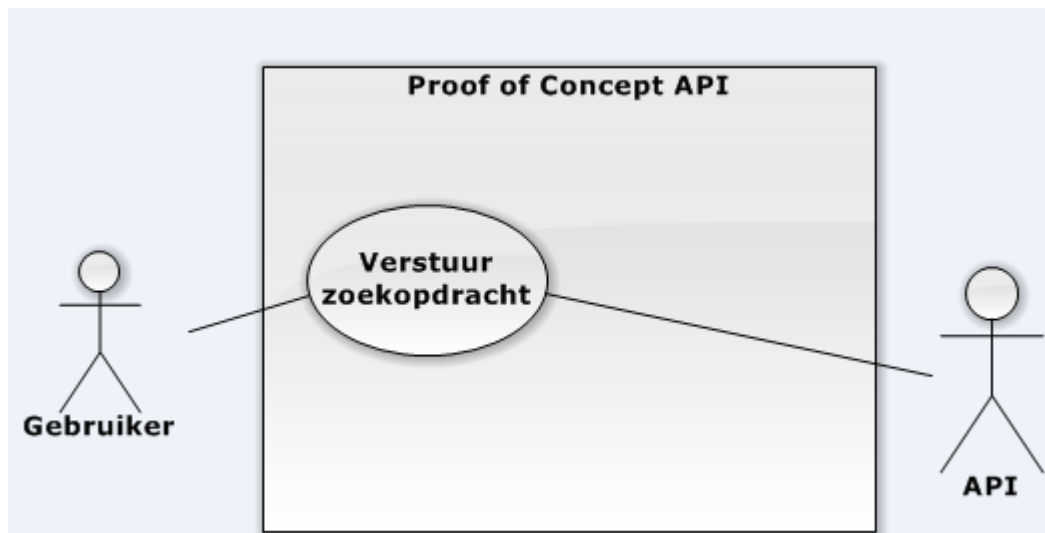
Beschrijving:

De gebruiker van het proof of concept is in staat de resultaten te ontvangen in RDF formaat.

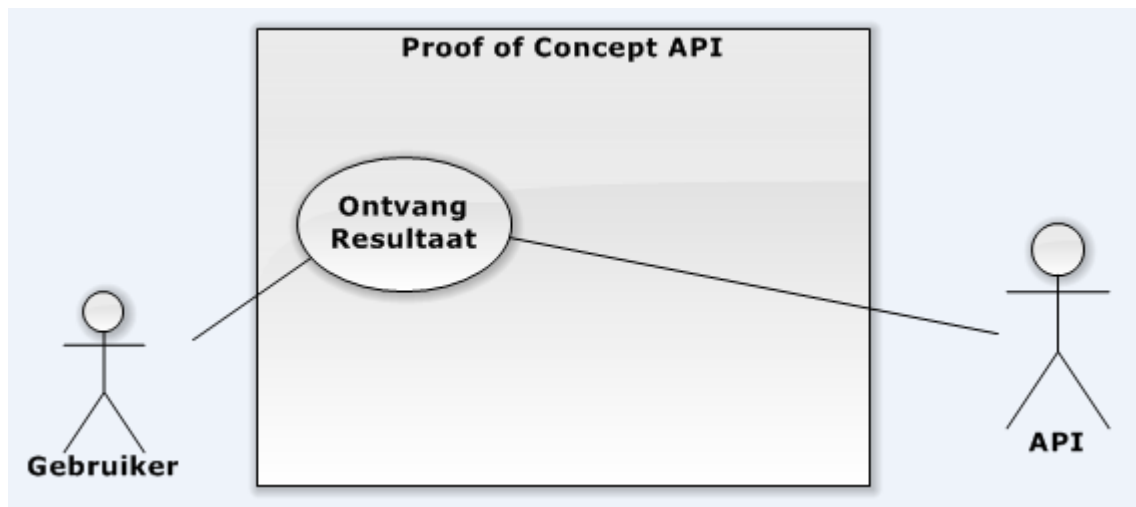
Use Case Diagram

In dit hoofdstuk staat per besproken use case een use case diagram.

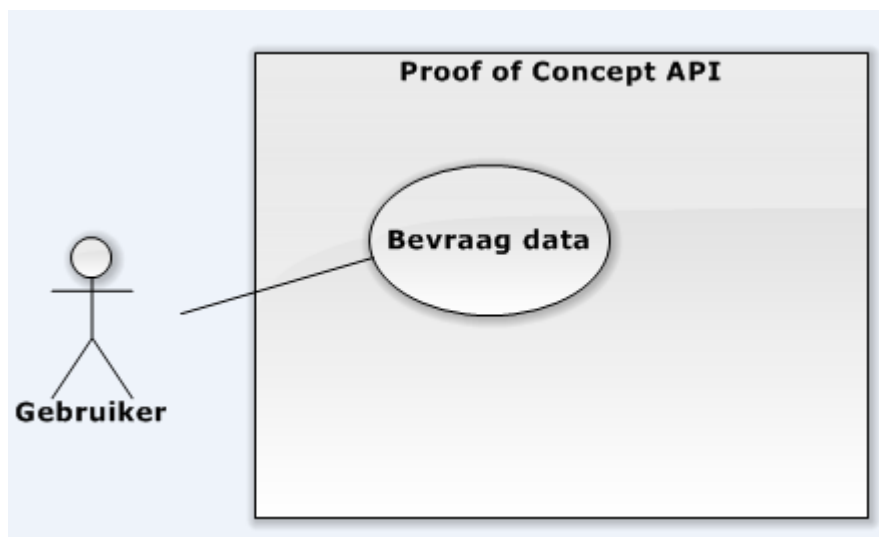
1.2 UC-01 VERSTUUR ZOEKOPDRACHT USE CASE DIAGRAM



1.3 UC-02 ONTVANG RESULTAAT USE CASE DIAGRAM



1.4 UC-03 BEVRAAG DATA USE CASE DIAGRAM



2. Testen

In dit hoofdstuk zullen de test cases voor het proof of concept beschreven worden.

2.1 TEST CASES

Dit zijn de test cases voor het proof of concept.

2.1.1 Test Case 1

Identificatienummer test case:	01
Test case voor de use case:	UC-01 – Verstuur zoekopdracht
Test specificatie:	De gebruiker kan het proof of concept uitvoeren. Hierdoor wordt de zoekopdracht verzonden naar de API.
Te testen requirements:	F-R1 - Uitvoeren zoekopdracht.
De actoren	De gebruiker van het proof of concept.
Pre Condities:	De gebruiker moet de query in de code ingevuld hebben. In deze query bevindt zich de zoekterm, of de uitgebreide zoekterm. Ook moet de gebruiker een API key hebben, en een internet verbinding.
Voorafgaande stappen of test cases:	Geen, dit is de eerste stap getest en genomen moet worden om te applicatie te kunnen gebruiken
Verwacht resultaat:	Het proof of concept verstuurt de zoekopdracht succesvol naar de API.
Uitvoerder test case:	Arnoud Aarnoudse
Resultaat:	Het Proof of Concept verstuurde inderdaad de zoekopdracht naar de API. Dit lukte met de enkele zoekterm, en de uitgebreide zoekterm.
Opmerkingen:	De zoekterm moet via de SRU manier uitgebreid worden, zoals "AND", "OR" en dergelijke. Deze zijn in de documentatie terug te vinden.

2.1.2 Test Case 2

Identificatienummer test case:	02
Test case voor de use case:	UC-02 – Ontvang resultaat.
Test specificatie:	De gebruiker ontvangt het resultaat van de API na dat de zoekopdracht is verzonden, en is verwerkt door de API.
Te testen requirements:	F-R2 – Ontvangen resultaten. NF-R1 - Ontvangen resultaten in RDF formaat.
De actoren	De gebruiker van het proof of concept.
Pre Condities:	De gebruiker moet een internetverbinding hebben. De gebruiker moet al een zoekopdracht verstuurd hebben naar de API.
Voorafgaande stappen of test cases:	Test Case 1
Verwacht resultaat:	Er komen resultaten terug in een RDF formaat die gevonden zijn door het uitvoeren van de zoekopdracht. Het RDF bestand moet triples en open annotations bevatten
Uitvoerder test case:	Arnoud Aarnoudse

Resultaat:	Als resultaat kwam er ook daadwerkelijk een RDF bestand binnen. In dit bestand bevonden zich de resultaten voor de zoekopdracht. In het bestand bevonden zich triples, met open annotations.
Opmerkingen:	Er bevind zich heel veel meta data bij de resultaten.

2.1.3 Test Case 3

Identificatienummer test case:	03
Test case voor de use case:	UC-03 – Bevraag data.
Test specificatie:	De gebruiker bevestigt de data die verkregen is van de API. Zo zullen de doel paden van de artikelen opgevraagd worden. Ook de omschrijvingen zullen opgevraagd worden.
Te testen requirements:	F-R3 – Bevestigen data.
De actoren	De gebruiker van het proof of concept.
Pre Condities:	De gebruiker moet al resultaten beschikbaar hebben.
Voorafgaande stappen of test cases:	Test Case 2
Verwacht resultaat:	Uit de bevestiging van de data komen alleen de bevestigde gegevens naar voren.
Uitvoerder test case:	Arnoud Aarnoudse
Resultaat:	Het resultaat van de lijst met 10 artikelen kon niet bevestigd worden. Het resultaat met de gegevens over 1 artikel was wel te bevestigen, en gaf de gewenste gegevens weer.
Opmerkingen:	Het is onduidelijk of het niet werken van de bevestiging komt door de gekozen library, of door een fout in het resultaat van de API.

2.2 ACCEPTATIE TEST

De acceptatietest zal uitgevoerd worden door het proof of concept uit te voeren waar de ontwikkelaar en begeleider bij zitten. Tijdens deze test zal er door beide personen gekeken worden naar de werking van het proof of concept, en of de werking resultaten acceptabel zijn. Dit wordt gebaseerd op de requirements, en op persoonlijke input.

Bijlage 3 – Technisch ontwerp Proof of Concepts.

Technisch Ontwerp

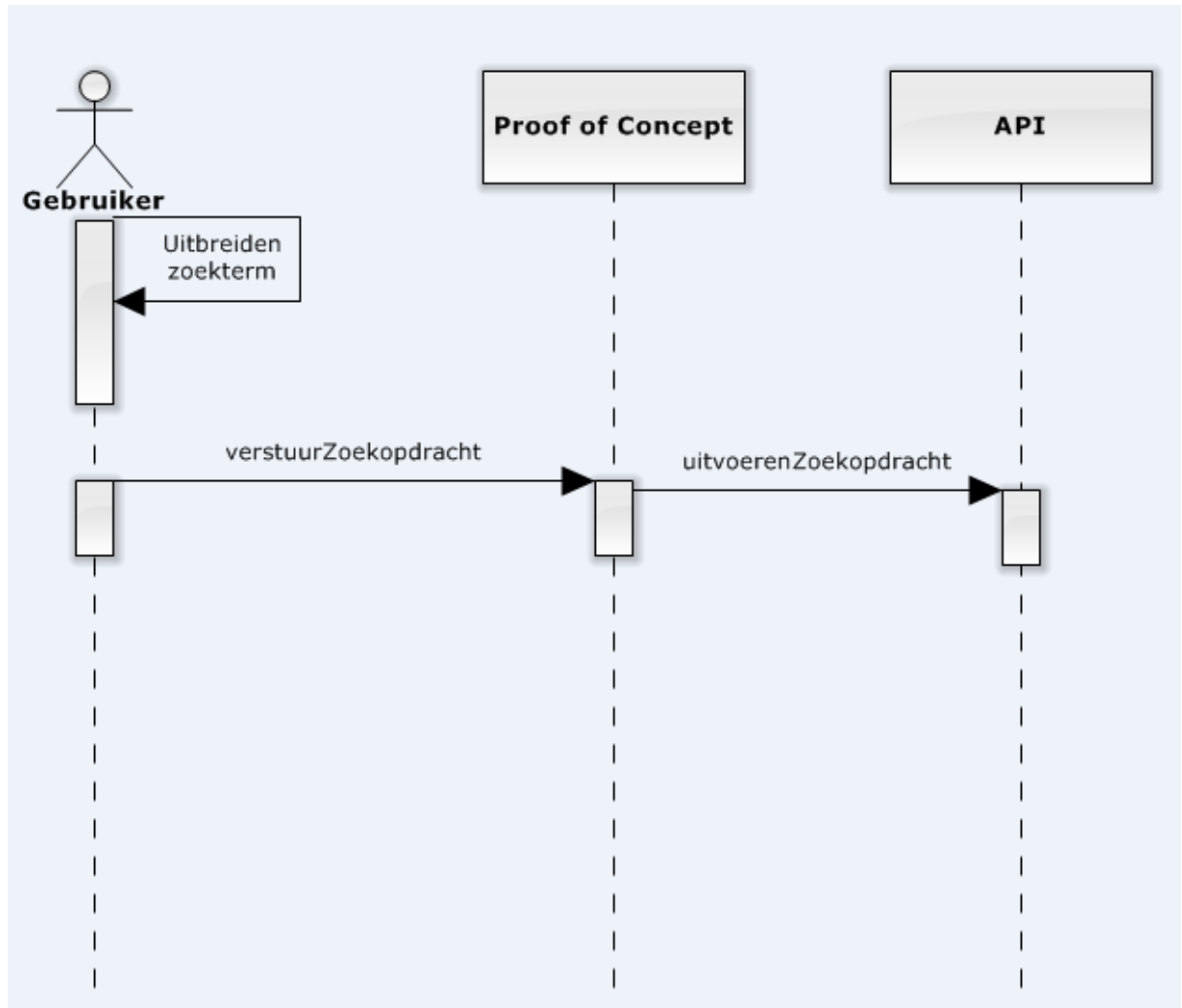
Project Naam: Proof of Concept API
Bedrijf: Lectoraat DIO
Informatie Systeem: PHP
Opgesteld door: Arnoud Aarnoudse

Document Eigenaren	Project/Organisatie Rol
Arnoud Aarnoudse	Documentbeheer, Programmeur, Stagiair

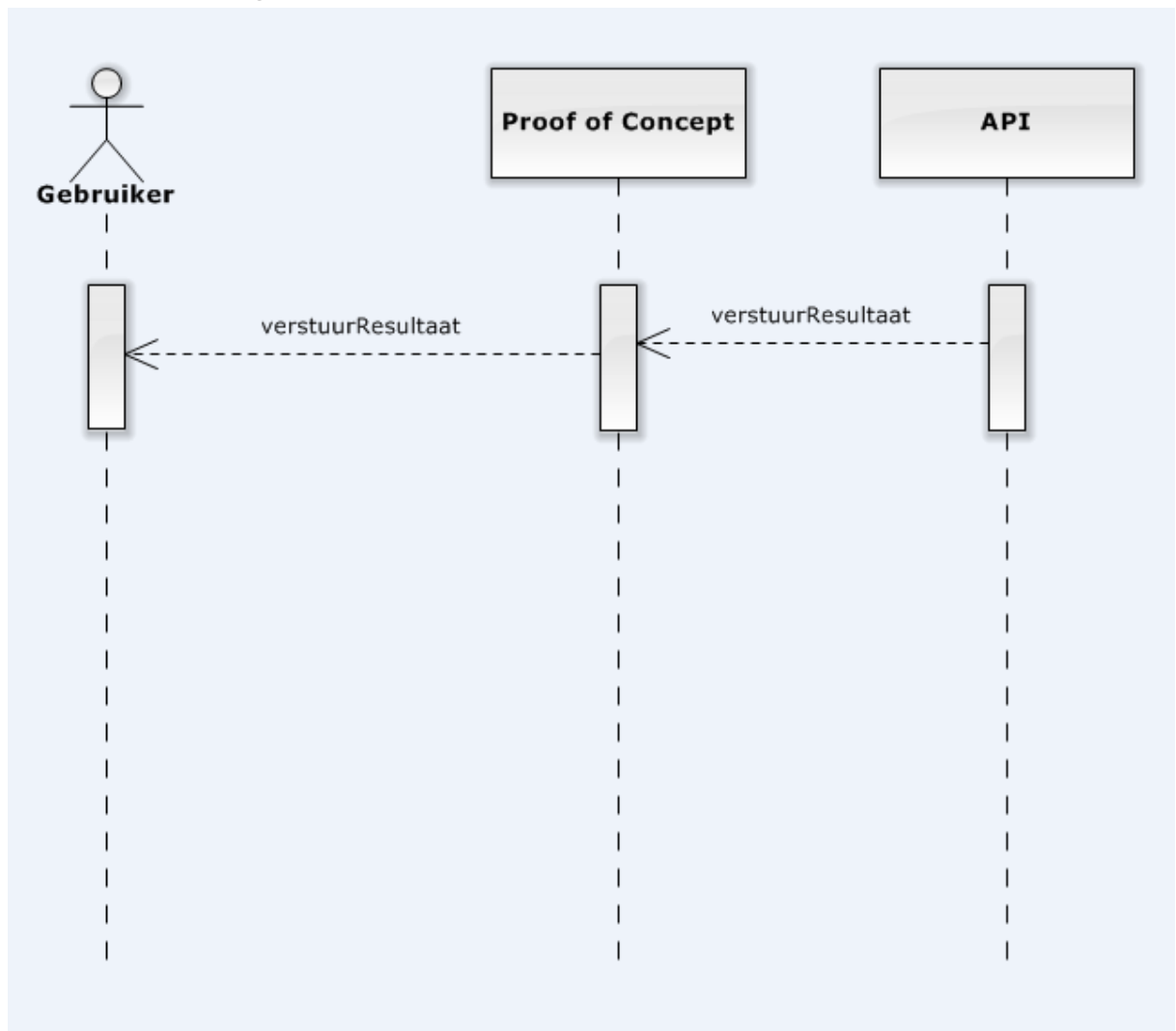
1. System Sequence Diagrams

In dit hoofdstuk zullen de use cases uitgewerkt worden tot SSD 's voor het proof of concept.

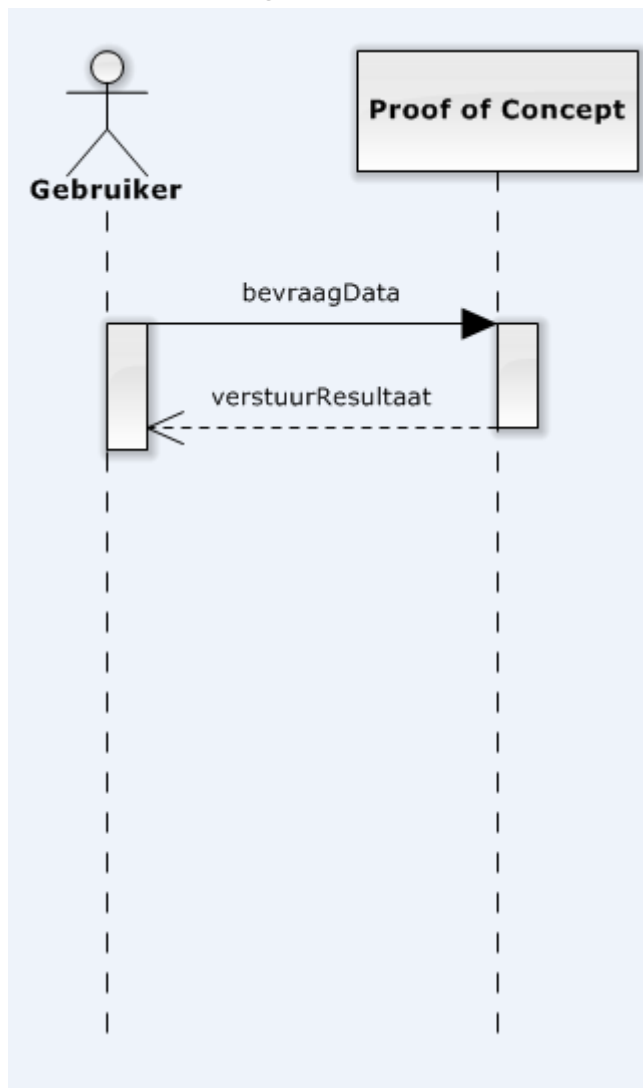
1.1 SYSTEM SEQUENCE DIAGRAM 1: VERSTUUR ZOEKOPDRACHT



1.2 SYSTEM SEQUENCE DIAGRAM 2: ONTVANG RESULTAAT



1.3 SYSTEM SEQUENCE DIAGRAM 3: BEVRAAG DATA



2 Ontwikkelmethode

Voor het ontwikkelen van het proof of concept wordt er gekozen voor een agile rapid prototyping ontwikkelmethode. De redenering hier voor is dat er niet, tot weinig bekend is over de te testen API. Door hier met een flexibele manier aan te werken zullen er sneller resultaten gehaald worden.

2.1 TOOLS

IDE: Voor IDE zal Aptana gebruikt worden. Aptana is een bekende tool voor het ontwikkelen van code. De programmeur heeft hier ook de meeste ervaring mee en kan hier ook makkelijk support voor vinden.

2.2 LIBRARIES

Voor het kiezen van de libraries is er gezocht naar welke libraries er beschikbaar zijn voor het werken met RDF bestanden en PHP. Vervolgens is er een pakketselectie uitgevoerd op deze libraries.

	EasyRDF	RAP	Graphite
Kan lokaal RDF uitlezen	✓	✓	✓
Kan SPARQL uitvoeren op RDF bestanden	✓	✓	✓
Heeft documentatie	✓	✓	✓
Is simplistisch, bevat niet te veel overige onderdelen, en maakt ook niet gebruik van andere libraries of programma's.	X	✓	X
Werkt in PHP	✓	✓	✓
Online terug te vinden in andere voorbeelden, blogs en dergelijke.	✓	✓	X

Door de pakketselectie is er gekozen voor de RAP, aangezien deze aan alle wensen voldoet.

Bijlage 4 – Opzet Experimenten.

Doel van het experiment:

Het doel van het experiment is om te onderzoeken wat voor effect het uitbreiden van de zoekopdracht geeft op de bijbehorende zoekresultaten. Het gaat hierbij om de zoekopdracht die wordt uitgevoerd met behulp van de demo van de landelijke bibliotheken. Deze demo maakt gebruik van de API die gebruikt kan worden door het lectoraat DIO, en die gebruikt wordt voor de proof of concepts van de student.

De zoekopdracht kan bestaan uit een standaard zoekopdracht, of een uitgebreide zoekopdracht. De standaard zoekopdracht is een zoekopdracht met één zoekterm, terwijl de uitgebreide zoekopdracht bestaat uit meerdere zoektermen. Deze meerdere zoektermen komen uit de grammatica der concepten. De reden dat deze zoektermen uit de zoekresultaten moeten komen is om te kijken of de uitgebreide zoekopdrachten elkaars zoekresultaten kunnen gebruiken om nieuwe uitgebreide zoekopdrachten uit te voeren. Zo wordt de grammatica der concepten nagebootst. Uit deze uitgebreide zoekopdrachten worden vervolgens relevante, maar nieuwe zoekresultaten verwacht, die opnieuw gebruikt kunnen worden voor het opstellen van uitgebreide zoekopdrachten. Door deze stappen te herhalen kunnen er mogelijk nieuwe relevante begrippen gevonden worden voor het oorspronkelijke begrip. Dit experiment moet aantonen of dit ook echt het geval is.

Materialen

De demo:

De demo is van de landelijke bibliotheken, en maakt gebruik van de API van de landelijke bibliotheken. De demo is opgezet om de werking en functionaliteit van de API aan te tonen. Aangezien het de student niet gelukt is om zelf een prototype te ontwikkelen met deze API is er door de student en de begeleider gekozen om de demo te gebruiken. Deze demo kan immers wat de student in het proof of concept probeerde werkend te krijgen, en bevat daarom alle benodigde functionaliteiten.

De demo wordt gebruikt om de zoekopdracht bij in te voeren, en vervolgens uit te laten voeren. Uit deze demo komen zoekresultaten, deze zullen onderzocht worden tijdens het experiment om het effect van het uitbreiden van de zoekopdracht vast te stellen.

Deze demo is momenteel te bereiken op: <http://apiozp.acc1.bibliotheek.nl/demo>

De word counter:

Het aantal woorden in de zoekresultaten zal tijdens het experiment geteld moeten worden. Om deze handeling te kunnen uitvoeren is er gekozen voor een website die dit kan doen. De gekozen website is in staat om het aantal woorden in een tekst te

tellen, en kan ook aangeven hoe vaak een bepaald woord voorkomt in de gegeven tekst. Dit is ook de reden dat de student voor deze website gekozen heeft.

Deze website is momenteel te bereiken op de volgende URL:

<http://wordcounttools.com/>

Inputs en outputs

Inputs:

De zoekopdracht voor de demo:

De zoekopdracht wordt ingevoerd als normale tekst in de zoekbalk van de demo.

Deze bestaat uit een enkele zoekterm in het geval van een standaard zoekopdracht, of uit meerdere zoektermen in het geval van een uitgebreidere zoekopdracht.

Voorbeeld standaard zoekopdracht: "Strand"

Voorbeeld uitgebreide zoekopdracht: "Strand Zee Zand" of "Zee Zand"

Daadwerkelijke initiële zoekopdracht: "Dementie"

De tekst voor de woorden teller:

De tekst voor de woorden teller zal ingevoerd worden als tekst in het invul scherm, zodat het te lezen is. Als de tekst te groot is om te gebruiken in het invul scherm van de site, dan zal het als '.txt' of als '.pdf' bestand geüpload worden.

Outputs:

De output van de demo:

Als output bij de demo wordt er een lijst van 10 relevante items verwacht, en een lijst met meerdere bijbehorende facetten waar op gefilterd kan worden. Deze lijst komt normaal als '.rdf' bestand uit de API, maar de demo verwerkt de data al tot een inzichtelijke en bruikbare lijst. Bij deze 10 items wordt er een titel verwacht, een beschrijven, een auteur, en een categorie. Bij elk item uit deze lijst zullen woorden en zinnen staan die het item beschrijven. Dit gaat dan bijvoorbeeld over het onderwerp, het genre, de auteur en overige. Deze data zal vervolgens gebruikt worden bij de woorden teller.

De output van de woorden teller:

Als output van de woorden teller wordt er een lijst verwacht waar de woorden staan die in de ingevoerde tekst voorkwamen, met hierbij het aantal in nummers en procenten van hoe vaak deze voorkwamen in de tekst. Dit zal gebruikt worden om te kijken of er woorden vaak voorkomen, dit kan mogelijk een verband aantonen met de zoektermen.

Methoden

Het experiment zal als volgt uitgevoerd worden:

1. **Er wordt een zoekopdracht ingevoerd bij de demo, deze wordt vervolgens uitgevoerd.**
2. **De student kiest vervolgens maximaal 10 artikelen uit van de zoekresultaten.**
3. **De omschrijvingen, onderwerpen en titels voor de gekozen artikelen uit de zoekresultaten wordt opgeslagen in een Word bestand.**
4. **De woorden in de omschrijving, onderwerpen en titels van de gekozen artikelen uit de zoekresultaten worden geteld en opgeslagen in een Word bestand.**
5. **Vervolgens breidt de student de originele zoekopdracht uit met de termen uit de top 50 meest voorkomende woorden van de inhoud van de omschrijving, en met begrippen uit de grammatica der concepten.**
6. **De stappen 1 tot en met 4 worden herhaalt, maar dan met de uitgebreide zoekopdracht in plaats van de standaard zoekopdracht.**
7. **De student vergelijkt de top 50 meest voorkomende termen van de enkele zoekterm met die van de uitgebreide zoektermen, en de termen uit de grammatica der concepten om te kijken of deze overeenkomen, of juist nieuwe termen leveren.**

De uitleg van de stappen is als volgt:

1. **Er wordt een zoekopdracht ingevoerd bij de demo, deze wordt vervolgens uitgevoerd.**

De zoekopdracht zal in de zoekbalk van de demo worden getypt door de student. Vervolgens drukt de student op de enter toets, of klikt hij op de zoekknop om de zoekopdracht uit te voeren.

Het verwachte resultaat is dat de demo de zoekopdracht verwerkt, en vervolgens een lijst van een onbekend aantal(maximaal 10) relevante vondsten weergeeft.

2. **De student kiest vervolgens maximaal 10 artikelen uit van de zoekresultaten.**

De reden dat er zelfstandig gekozen wordt is omdat niet alle zoekresultaten een omschrijving of pagina hebben. Dit maakt uiteindelijk niet uit als alle artikelen uit de zoekresultaten gebruikt worden, maar bij het selecteren van een paar zoekresultaten is het beter om er voor te zorgen dat deze wel aanwezig zijn. Anders kunnen er ook geen woorden geteld worden.

De reden dat de student deze artikelen uit de zoekresultaten kiest is om de student de persoon is die dit experiment uitvoert. Er zijn namelijk geen andere personen bij de uitvoering van het experiment betrokken. Daarom is het logisch dat de student zelf de 10 zoekresultaten uitkiest. Ook kan de student bij het selecteren van de 10 artikelen kijken of er wel omschrijvingen aanwezig zijn. Anders zijn deze zoekresultaten voor het experiment niet bruikbaar.

Er worden 10 zoekresultaten gekozen om de API zelf standaard 10 zoekresultaten terug geeft. Het is mogelijk om nog verder door te klikken voor nieuwe zoekresultaten op de demo, maar in de eerste instantie zijn het 10 zoekresultaten. Er zouden meerdere zoekopdrachten moeten worden uitgevoerd om met de API meer dan 10 zoekresultaten te krijgen, waar er dan bijvoorbeeld gebruik gemaakt wordt van filters om op deze manier nieuwe resultaten te krijgen. In dat opzicht zijn dit ook nieuwe zoekopdrachten. Daarom dat er voor 10 zoekresultaten gekozen wordt, want deze zijn per zoekopdracht beschikbaar.

Het verwachte resultaat is dat er zoekresultaten, van een onbekend aantal(maximaal 10) aanwezig zijn. De zoekresultaten worden verwacht weergegeven te worden in de standaard vorm van de demo. Deze vorm is een lijst van maximaal 10 items, waar er op de items geklikt kan worden voor een detail pagina. Er wordt verwacht dat de items in deze lijst minimaal een titel, onderwerp en omschrijving hebben.

3. De omschrijvingen, onderwerpen en titels voor de gekozen artikelen uit de zoekresultaten wordt opgeslagen in een Word bestand.

Er wordt voor de omschrijving gekozen om deze door de API geleverd wordt, en daarom later ook binnengehaald kan worden door de proof of concepts/applicatie die gebruikt zal maken van de API. Ook wordt er voor de omschrijving gekozen om hier waarschijnlijk een beschrijvende tekst in staat, met relevante begrippen. De reden dat de omschrijving wordt opgeslagen is om de zoekresultaten kunnen wijzigen, en de omschrijvingen ook. Dit houdt in dat als er later opnieuw gezocht word, de omschrijving dan anders kan zijn. Dit zou inhouden dat deze zoekresultaten niet volledig herleidbaar zouden zijn. Dit is belangrijk als het experiment later gevalideerd moet worden. Daarom worden de omschrijvingen opgeslagen.

Het wordt opgeslagen in een Word bestand om dit een standaard tekstbewerker is die het goed doet. Zo kan de opmaak verbeterd worden voor een betere leesbaarheid van de data. Dit is handig aangezien de data door de student zelf gelezen moet kunnen worden, en mogelijk moet het ook gepresenteerd worden aan de opdrachtgever. In beide gevallen is het dan van belang dat de data overzichtelijk is.

De verwachting is dat de student in staat is om de omschrijvingen, onderwerpen en titels op te slaan in een word document, dat later nog te gebruiken is

4. De woorden in de omschrijving, onderwerpen en titels van de gekozen artikelen uit de zoekresultaten worden geteld en opgeslagen in een Word bestand.

De reden dat de woorden geteld worden van de omschrijving, is om te zien welke woorden het meeste voorkomen. De relevantie van een begrip kan opvallen als het meerdere keren voorkomt. Bij de begrippen staat ook hoe vaak het voorkomt in de tekst, en hoeveel procent dit woord uitmaakt van de gehele tekst. Deze waardes worden hier voorlopig niet in meegenomen. De redenering hiervoor is dat de student momenteel nieuwe termen uit deze 10 zoekresultaten hoopt te vinden. Het is dan mogelijk dat een begrip duizend maal voorkomt in één resultaat, maar dat is dan mogelijk om dit resultaat specifiek over dat begrip gaat. Als er een begrip is dat ook vaak voorkomt in de andere 9 zoekresultaten, dan komt het begrip dus vaker voor, en is relevanter aan de initiële zoekopdracht. Daarom kijkt de student niet specifiek naar het aantal gevonden woorden in elk resultaat, maar naar welke woorden er in meerdere zoekresultaten voorkomen.

Voor het tellen van de woorden verwacht de student dat de woorden in de tool geplakt kunnen worden om de woorden te tellen. Vervolgens verwacht de student daar als resultaat een top 50 woorden van. In deze top 50 woorden worden veel woorden zoals “het”, “de” en “een” verwacht, maar ook begrippen. Er wordt gekozen voor een top 50 om er veel tussenvoegsels, lidwoorden, en dergelijke bij zullen zitten. Door het kiezen van de top 50 kan de student hier voorbij lezen, en zelf de begrippen hier uit halen.

5. Vervolgens breidt de student de originele zoekopdracht uit met de termen uit de top 50 meest voorkomende woorden van de inhoud van de omschrijving, en met begrippen uit de grammatica der concepten.

Dit zullen tussen 2-5 termen zijn om dit de uitgebreide zoektermen nog op een redelijke grootte houdt, is de mening van de student. Als de zoekterm te ver uitgebreid wordt dan is de kans groter dat hier minder zoekresultaten uit terug komen. De reden dat dit gedaan wordt is zodat er naar nieuwe zoektermen gezocht kan worden. Op deze manier kunnen er meerdere zoekresultaten uitgevoerd worden, die als het goed is nieuwe zoektermen geven, die opnieuw gebruikt kunnen worden. Dit is efficiënter dan één hele lange zoekopdracht, die minder zoekresultaten weergeeft.

Er wordt gekozen voor de top 50 van de omschrijvingen om deze termen uiteindelijk ook de informatie is die binnen zal komen van de gevonden artikelen van de API, en niet de top 50 van de ingelezen artikelen/boeken.

De verwachting is dat de student termen vind die in verschillende top 50's voorkomen, en deze termen gebruikt voor het maken van een uitgebreide zoekterm.

6. De stappen 1 tot en met 7 worden herhaalt, maar dan met de uitgebreide zoekopdracht in plaats van de standaard zoekopdracht.

De stappen 1 tot en met 7 worden herhaalt om het dezelfde manier van zoeken is, alleen met een andere input. De stappen om de resultaten te verwerven en verwerken zijn identiek.

De verwachting is dat de stappen succesvol doorlopen worden voor de uitgebreide zoekopdracht, en dat er vervolgens nieuwe zoekresultaten uitkomen die niet bij de standaard zoekopdracht beschikbaar waren. Ook worden er zoekresultaten verwacht die wel aanwezig waren bij de standaard zoekopdracht, aangezien de zoektermen nog wel relevant moeten zijn, alleen niet alle zoekresultaten zullen terug te vinden zijn.

7. De student vergelijkt de top 50 meest voorkomende termen van de enkele zoekterm met die van de uitgebreide zoektermen, en de termen uit de grammatica der concepten om te kijken of deze overeenkomen, of juist nieuwe termen leveren.

De reden dat er gekeken wordt of ze overeenkomen is om te kijken of er door het uitbreiden van de zoekresultaten nog steeds overeenkomende termen naar voren komen. Dit geeft aan dat deze term een kans heeft belangrijk te zijn, en een verbindingspunt is. Op deze manier kunnen de belangrijke termen er op den duur uitgehaald worden als ze heel vaak bij veel verschillende artikelen tevoorschijn komen, met verschillende zoekopdrachten.

De reden dat er gekeken wordt of er nieuwe termen gevonden kunnen worden is om te zien of het uitbreiden van de zoektermen voor de zoekopdracht een effect heeft in het vinden van nieuwe gerelateerde termen. Het vinden van deze termen is belangrijk om op den duur weer nieuwe uitgebreide zoekopdrachten te maken, om hier uitgebreidere zoektermen in te kunnen vinden.

De verwachting is dat er overeenkomende begrippen zijn, dit zou aangeven dat door het uitbreiden van de zoektermen er nog steeds overeenkomende begrippen gevonden worden. Er wordt ook verwacht dat er nieuwe begrippen

zijn, dit betekend dat door het uitbreiden van de zoektermen er nieuwe informatie gevonden kan worden die hoogst waarschijnlijk gerelateerd is aan de initiële zoekterm.

Bijlage 5 – Conclusie experimenten.

Resultaten

Het experiment zal als volgt uitgevoerd worden:

1. **Er wordt een zoekopdracht ingevoerd bij de demo, deze wordt vervolgens uitgevoerd.**
2. **De student kiest vervolgens maximaal 10 artikelen uit van de zoekresultaten.**
3. **De omschrijvingen, onderwerpen en titels voor de gekozen artikelen uit de zoekresultaten wordt opgeslagen in een Word bestand.**
4. **De woorden in de omschrijving, onderwerpen en titels van de gekozen artikelen uit de zoekresultaten worden geteld en opgeslagen in een Word bestand.**
5. **Vervolgens breidt de student de originele zoekopdracht uit met de termen uit de top 50 meest voorkomende woorden van de inhoud van de omschrijving, en met begrippen uit de grammatica der concepten.**
6. **De stappen 1 tot en met 4 worden herhaalt, maar dan met de uitgebreide zoekopdracht in plaats van de standaard zoekopdracht.**
7. **De student vergelijkt de top 50 meest voorkomende termen van de enkele zoekterm met die van de uitgebreide zoektermen, en de termen uit de grammatica der concepten om te kijken of deze overeenkomen, of juist nieuwe termen leveren.**

De uitleg van de stappen is als volgt:

1. **Er wordt een zoekopdracht ingevoerd bij de demo, deze wordt vervolgens uitgevoerd.**

“Het verwachte resultaat is dat de demo de zoekopdracht verwerkt, en vervolgens een lijst van een onbekend aantal(maximaal 10) relevante vondsten weergeeft.”

Stap 1 ging niet zoals verwacht. De zoekopdracht kon worden ingevuld, en vervolgens kwam er een lijst. Deze lijst bevatte in sommige gevallen geen, of minder dan 10 resultaten. Af en toe bevatte de lijst wel 10 resultaten. Het ligt er aan wat er gezocht wordt, en hoeveel artikelen er beschikbaar zijn bij de landelijke bibliotheken die voldoen aan de zoekopdracht.

2. **De student kiest vervolgens maximaal 10 artikelen uit van de zoekresultaten.**

“Het verwachte resultaat is dat er zoekresultaten, van een onbekend aantal(maximaal 10) aanwezig zijn. De zoekresultaten worden verwacht weergegeven te worden in de standaard vorm van de demo. Deze vorm is een lijst van maximaal 10 items, waar er op de items geklikt kan worden voor een

detail pagina. Er wordt verwacht dat de items in deze lijst minimaal een titel, onderwerp en omschrijving hebben.”

Stap 2 ging bijna zoals verwacht. Er kwam een lijst met een aantal resultaten, en op deze resultaten kon geklikt worden voor een detail pagina. Maar deze pagina bevatte niet altijd een omschrijving of een onderwerp. Wel zat er nog veel informatie bij, die ook gebruikt zou kunnen worden door het lectoraat DIO. Er zat zelfs al semantische data in, waar de connecties tussen sommige onderwerpen werden uitgelegd.

3. De omschrijvingen, onderwerpen en titels voor de gekozen artikelen uit de zoekresultaten wordt opgeslagen in een Word bestand.

De verwachting is dat de student in staat is om de omschrijvingen, onderwerpen en titels op te slaan in een word document, dat later nog te gebruiken is.

Stap 3 is gelukt. De student kon de omschrijvingen, onderwerpen en titels opslaan in het word bestand, mits deze aanwezig waren. Deze kunnen nu gebruikt worden om het experiment te controleren. Wel viel op dat er soms spelfouten in de resultaten zitten.

4. De woorden in de omschrijving, onderwerpen en titels van de gekozen artikelen uit de zoekresultaten worden geteld en opgeslagen in een Word bestand.

“Voor het tellen van de woorden verwacht de student dat de woorden in de tool geplakt kunnen worden om de woorden te tellen. Vervolgens verwacht de student daar als resultaat een top 50 woorden van. In deze top 50 woorden worden veel woorden zoals “het”, “de” en “een” verwacht, maar ook begrippen. Er word gekozen voor een top 50 om er veel tussenvoegsels, lidwoorden, en dergelijke bij zullen zitten. Door het kiezen van de top 50 kan de student hier voorbij lezen, en zelf de begrippen hier uit halen.”

Stap 4 is gelukt. De student kon succesvol de woorden in de woorden teller site plakken, en deze laten tellen. Hier kwam inderdaad een lijst van woorden uit, met de hoeveelheid van hoe vaak het woord voorkwam.

5. Vervolgens breidt de student de originele zoekopdracht uit met de termen uit de top 50 meest voorkomende woorden van de inhoud van de omschrijving, en met begrippen uit de grammatica der concepten.

“De verwachting is dat de student termen vind die in verschillende top 50's voorkomen, en deze termen gebruikt voor het maken van een uitgebreide zoekterm.”

Stap 5 is gelukt. Er kon een uitgebreide zoekopdracht gemaakt worden. Het lukte om een uitgebreide zoekopdracht te maken voor de grammatica der concepten, en om een uitgebreide zoekopdracht te maken voor de zoekresultaten van de originele zoekterm.

6. De stappen 1 tot en met 7 worden herhaalt, maar dan met de uitgebreide zoekopdracht in plaats van de standaard zoekopdracht.

“De verwachting is dat de stappen succesvol doorlopen worden voor de uitgebreide zoekopdracht, en dat er vervolgens nieuwe zoekresultaten uitkomen die niet bij de standaard zoekopdracht beschikbaar waren. Ook worden er zoekresultaten verwacht die wel aanwezig waren bij de standaard zoekopdracht, aangezien de zoektermen nog wel relevant moeten zijn, alleen niet alle zoekresultaten zullen terug te vinden zijn.”

Stap 6 ging niet zoals verwacht, bij de uitgebreide zoekopdracht werden er geen zoekresultaten gevonden. De uitgebreide zoekopdrachten moesten ingekort worden om weer aan zoekresultaten te komen. Dit zou mogelijk verholpen kunnen worden door het gebruik van zoek operators, moment wordt er een “AND” gebruikt, waar er dus gezocht wordt naar beide termen, dit zou later met andere operators kunnen. Maar door het inkorten van de zoekopdracht konden de stappen alsnog uitgevoerd worden.

7. De student vergelijkt de top 50 meest voorkomende termen van de enkele zoekterm met die van de uitgebreide zoektermen, en de termen uit de grammatica der concepten om te kijken of deze overeenkomen, of juist nieuwe termen leveren.

“De verwachting is dat er overeenkomende begrippen zijn, dit zou aangeven dat door het uitbreiden van de zoektermen er nog steeds overeenkomende begrippen gevonden worden. Er wordt ook verwacht dat er nieuwe begrippen zijn, dit betekend dat door het uitbreiden van de zoektermen er nieuwe informatie gevonden kan worden die hoogst waarschijnlijk gerelateerd is aan de initiële zoekterm.”

Stap 7 is gelukt. Eerst werden alle gevonden begrippen uit de zoekresultaten vergeleken met de begrippen uit de grammatica der concepten. Hier bleek dat er begrippen waren die al aanwezig waren in de grammatica, maar ook veel begrippen die nog niet aanwezig waren. Dit betekend dus dat er inderdaad nieuwe begrippen naar voorkomen voor de grammatica der concepten. Vervolgens zijn de begrippen van de zoekresultaten van de uitgebreide zoekopdrachten vergeleken met de originele zoekopdracht. Hier bleek uit dat er begrippen waren die bij alle zoekresultaten voorkwamen. Dit toont aan dat bij het uitbreiden van de zoekopdracht er nog steeds overeenkomende begrippen zijn. Maar er werden ook meerdere begrippen gevonden die nog niet aanwezig waren bij de originele zoekopdracht.

Conclusie experiment

"Het doel van het experiment is om te onderzoeken wat voor effect het uitbreiden van de zoekopdracht geeft op de bijbehorende zoekresultaten. Het gaat hierbij om de zoekopdracht die wordt uitgevoerd met behulp van de demo van de landelijke bibliotheken. Deze demo maakt gebruik van de API die gebruikt kan worden door het lectoraat DIO, en die gebruikt word voor de proof of concepts van de student.

De zoekopdracht kan bestaan uit een standaard zoekopdracht, of een uitgebreide zoekopdracht. De standaard zoekopdracht is een zoekopdracht met één zoekterm, terwijl de uitgebreide zoekopdracht bestaat uit meerdere zoektermen. Deze meerdere zoektermen komen uit de grammatica der concepten. Maar ook zullen de zoektermen uit de zoekresultaten gehaald worden.

De reden dat deze zoektermen uit de zoekresultaten moeten komen is om te kijken of de uitgebreide zoekopdrachten elkaars zoekresultaten kunnen gebruiken om nieuwe uitgebreide zoekopdrachten uit te voeren. Zo word de grammatica der concepten nagedaan. Uit deze uitgebreide zoekopdrachten komen vervolgens relevante, maar nieuwe zoekresultaten, die opnieuw gebruikt kunnen worden voor het opstellen van uitgebreide zoekopdrachten. Door deze stappen te herhalen kunnen er mogelijk nieuwe relevante begrippen gevonden worden voor het oorspronkelijke begrip. Dit experiment moet aantonen of dit ook echt het geval is."

Het doel van het experiment is behaald. Het effect van het uitbreiden op de zoekopdracht geeft een positief effect. Bij de standaard zoekopdracht worden er begrippen gevonden. Door het tellen van deze begrippen, kon er een uitgebreide zoekopdracht opgezet worden met behulp van de nieuwe begrippen. Na het uitvoeren en vergelijken van deze uitgebreide zoekopdracht bleek dat er wel minder zoekresultaten naar voren kwamen, maar wel met nieuwe begrippen. Ook kwamen er overeenkomende begrippen naar voren. Dit toont dus aan dat de zoekresultaten nog steeds relevant zijn aan de standaard zoekopdracht, en daarom dus een positief effect hebben.

Na het uitbreiden van de originele standaard zoekopdracht met de begrippen van de grammatica der concepten werden er ook weer overeenkomende en nieuwe begrippen gevonden. Maar ook dit keer was het aantal resultaten lager. Het uitbreiden van de zoekopdracht leid in beide gevallen dus tot overeenkomende, maar vooral nieuwe begrippen. Ook bevatten de zoekresultaten nog veel metadata die niet bij de experimenten betrokken zijn, maar wel benut kunnen worden. Als het lectoraat DIO gebruik zou willen maken van de API dan zou alle informatie dus opgevraagd kunnen worden.

Vervolgens kan er met de zoekmethode waar hun aan werken naar alle informatie gekeken worden, zo kan er nog meer informatie uit de data getrokken worden. Dus vanwege de hoeveelheid nieuwe data die beschikbaar komt door het uitbreiden van de zoekopdracht is te concluderen dat het uitbreiden van de originele zoekterm een positief effect heeft.