Open data webservice van Overheid.nl

Handleiding voor het bevragen van de zoekdienst via SRU

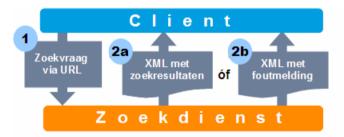
Samenvatting & inleiding

Via Overheid.nl kunnen burgers en professionals zoeken in regelgeving, bekendmakingen en andere datacollecties van de overheid. Om het hergebruik van deze data mogelijk te maken, is er een technische koppeling (ook wel webservice of API genoemd) beschikbaar waarmee eenvoudig kan worden gezocht in de data van Overheid.nl. Deze webservice is onderdeel van de zoekfunctionaliteit die bekend is als 'de zoekdienst'. Dit document is een technische handleiding bedoeld voor ontwikkelaars van partijen die de data op Overheid.nl willen ontsluiten.

De collecties lokale regelingen, lokale bekendmakingen, lokale vergunningen, overheidsorganisaties en producten en diensten kunnen op deze wijze worden uitgevraagd. Er zijn met de webservice complexe zoekvragen mogelijk als: "zoek alle regelingen van 'Amsterdam' met het woord 'leges' in de *titel* EN met het woord 'huwelijk' in de *tekst*".

Mogelijke **toepassingen** zijn bijvoorbeeld het tonen van regelgeving op gemeentelijke websites, het tonen van bekendmakingen op een kaart en het ontsluiten van data op mobiele telefoons of tablets.

Het onderstaande schema geeft in hoofdlijnen weer hoe de webservice werkt.



De webservice is gebaseerd op de **internationale open standaard** Search & Retrieval by URL (SRU). Het is mogelijk om **real-time** te zoeken met de webservice.

Op het gebruik van deze webservice is een fair use policy van kracht.

Dit document beschrijft niet hoe documenten moeten worden *aangeleverd* bij de zoekdienst; alleen het *uitvragen* van documenten wordt beschreven.

Gebruikers van de webservice wordt verzocht zich te registreren via servicecentrum@logius.nl, om een contactadres (e-mailadres) te hebben bij wijzigingen aan de webservice.

Inhoud

Samenvattir	ng & inleiding	2
1	Collecties waarin via de webservice kan worden gezocht	4
2	Gebruikte termen in dit document	5
3 3.1 3.2	Overzicht van de werking van de webservice Zoekvragen, resultaten en foutmeldingen Real-time bevragen of synchroniseren	6 6 6
4 4.1 4.2 4.3	Zoekvraag: de opbouw van webservice-requests Opbouw van de URL CQL-query: zoeken in collecties Uitvragen via HTTP GET of HTTP POST	7 7 8 10
5 5.1 5.2 5.3 5.4	Het zoekresultaat: de opbouw van een XML-response Overzicht: <searchretrieveresponse> op hoofdlijnen Het zoekresultaat: <record> Extra informatie op resultaatsetniveau: <extraresponsedata> Advies: ontwikkel robuuste foutafhandeling</extraresponsedata></record></searchretrieveresponse>	11 11 12 14 18
6	Foutmeldingen: de diagnostics-functie	18
7 7.1	Collectiespecifieke informatie Op welke velden kan worden gezocht? De explain-functie	20 20
8	Vragen of opmerkingen over dit document	21
9	Bijlage: Voorbeelden van SRU-queries	22
10 10.1 10.2	Bijlage: Naamgevingsconventies van veldnamen Categorie 1: Veldnamen met een 1-op-1 relatie met metagegevens Categorie 2: Verrijkte veldnamen	22 22 23
11	Bijlage: Ondersteuning van de SRU-standaard	23
12	Bijlage: voorbeeld XML-zoekresultaat	24
13	Bijlage: Doorontwikkeling van de SRU-webservice	26

Collecties waarin via de webservice kan worden gezocht

In de volgende collecties kan worden gezocht via de webservice:

Collectie	Afkorting	Omschrijving
Lokale regelingen	CVDR	Actuele en historische teksten van regelingen die gelden voor gemeenten, provincies en waterschappen en voor Bonaire, Saba en de voormalige Nederlandse Antillen. De bron van deze collectie is de Centrale Voorziening Decentrale Regelgeving (CVDR).
Lokale bekendmakingen	вм	Lokale bekendmakingen van gemeenten, provincies en waterschappen. Deze worden tevens in plaatselijke kranten gepubliceerd.
Lokale vergunningen	VG	Vergunningen die zijn verleend of aangevraagd en worden gehandhaafd door een gemeente, provincie of waterschap.
Overheids- organisaties	00	De adresgegevens, telefoonnummers en webadressen van alle Nederlandse overheidsorganisaties.
Producten en diensten	SC	De verzameling van producten en diensten (collectienaam: Samenwerkende Catalogi) die op grond van landelijke wetgeving worden aangeboden of gevraagd door de Nederlandse overheid. Ook steeds meer producten van de lokale overheden zijn hier te vinden.
Overheid.nl	LNK	Zoekingang om te zoeken diverse plaatsgebonden collecties. Zie "Bij u in de buurt" op de homepage van Overheid.nl. De getoonde resultaten zijn uit de collecties links, Samenwerkende Catalogi en landing pages.

Met de onderstaande URL kunnen bijvoorbeeld alle CVDR-documenten worden opgevraagd (met voor het voorbeeld maximaal 10 resultaten):

http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrieve&x-connection=cvdr&startRecord=1&maximumRecords=10&query=keyword="""

Door in het voorbeeld "CVDR" te vervangen door een van de andere afkortingen uit de tabel hierboven, kan in de andere collecties worden gezocht.

2 Gebruikte termen in dit document

Voordat we overgaan op een toelichting op de werking van de webservice, is het noodzakelijk om een aantal termen te definiëren.

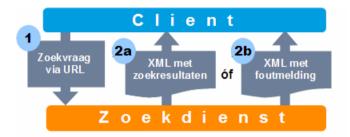
→ Leeswijzer: Lees deze termen eerst door voor het verdere document te lezen.

Webservice	Technisch mechanisme om gegevens uit te wisselen. Ook wel koppeling, koppelvlak, interface of API (Application Programming Interface) genoemd (hoewel deze termen niet geheel hetzelfde zijn, is dit voor dit document niet van belang).
OWMS	Overheid Web Metadata Standaarden. Via XML gestandaardiseerde waardenlijsten en termen bedoeld om uitwisselbaarheid van data te verbeteren. Zie www.standaarden.overheid.nl. De collecties op Overheid.nl voldoen vrijwel allemaal aan deze standaarden.
Zoekdienst	De zoekmachine achter Overheid.nl die wordt gebruikt door zowel de user interface van Overheid.nl als door de in dit document beschreven webservice.
SRU	Seach and Retrieval by URL. Internationale open standaard voor webservices waarmee kan worden gezocht in gestructureerde data. De relevante delen van deze standaard worden beschreven in dit document; zie http://www.loc.gov/standards/sru/ voor extra achtergrondinformatie.
CQL	Contextual Query Language. Taal waarmee zoektermen kunnen worden gedefinieerd (vergelijkbaar met SQL). De relevante delen van deze taal worden beschreven in dit document; zie http://www.loc.gov/standards/sru/specs/cql.html voor extra achtergrondinformatie.
Extension	Uitbreidingsmechanisme van SRU waarmee extra informatie kan worden uitgevraagd (in het SRU-request) en geretourneerd (in de SRU-response). Zie http://www.loc.gov/standards/sru/specs/common.html#extraData
Veldnaam	Een inhoudelijk veld waarop gezocht kan worden, bijvoorbeeld "dcterms:title" voor de titel van een document. Nota bene: in SRU wordt dit een 'index' genoemd, maar omdat dit in de databasewereld vaak een andere betekenis heeft, wordt in dit document alleen 'veldnaam' gebruikt.
Zie ook	Overige termen worden als bekend verondersteld voor de doelgroep van dit document. Zie de volgende pagina's op wikipedia voor meer informatie: XML; URL; client; record; query;

3 Overzicht van de werking van de webservice

De webservice volgt de internationale open standaard SRU en werkt in hoofdlijnen zoals in dit hoofdstuk beschreven.

3.1 Zoekvragen, resultaten en foutmeldingen



- De client stelt een zoekvraag in de vorm van een URL met parameters. De volgende zoekvraag zoekt bijvoorbeeld in de CVDR-collectie naar documenten met het woord "algemene" in de titel, met een maximum van 10 resultaten: http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrieve&x-connection=cvdr&startRecord=1&maximumRecords=10&query=title=algemene Deze stap wordt toegelicht in hoofdstuk 4.
- De webservice retourneert een XML-response met daarin één van de twee mogelijke berichten:
 - **2a. XML met zoekresultaten**. Dit zoekresultaat kan worden gebruikt om te presenteren aan een gebruiker of het kan worden gebruikt voor verdere verwerking. Dit wordt verder beschreven in hoofdstuk 5.
- **2b. XML met een foutmelding.** Dit wordt verder beschreven in hoofdstuk 6. Hiernaast is het mogelijk om per collectie uit te vragen op welke inhoudelijke veldnamen kan worden gezocht. Dit wordt beschreven in hoofdstuk 7.

3.2 Real-time bevragen of synchroniseren

Er zijn verschillende mogelijkheden om de webservice te gebruiken.

Het is mogelijk om **real-time** de webservice te bevragen, waarbij een zoekvraag van een website, mobiele applicatie, etcetera direct wordt doorgestuurd naar de zoekdienst. Het resultaat in XML kan dan worden getransformeerd zodat het kan worden getoond aan de eindgebruiker.

Een andere strategie is om de collecties dagelijks te **synchroniseren**. Dit wordt aangeraden voor grootverbruikers, om een overschrijding van de fair use policy te voorkomen. Het is vooralsnog helaas niet mogelijk om een lijst van *verwijderde* documenten in een bepaalde periode op te vragen.

Zoekvraag: de opbouw van webservice-requests



4.1 Opbouw van de URL

Om een zoekvraag te stellen, dient er een URL te worden opgebouwd die alle zoekcriteria bevat. Een voorbeeld hiervan is:

http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrieve&x-connection=cvdr&startRecord=1&maximumRecords=10&query=title=algemene

Door op de bovenstaande URL te klikken, kan worden bekeken hoe een XML response eruit ziet. De URL voor het stellen van een zoekvraag bestaat uit de volgende onderdelen:

Onderdeel / parameter	Ver- plicht?	Toelichting
http://zoekdienst.overh eid.nl/sru/Search	Ja	Het adres van de webservice. Voor testdoeleinden is er ook een acceptatieomgeving.
version=1.2	Ja	SRU-versie, moet altijd waarde 1.2 hebben.
operation=searchRe trieve	Ja	Operatie die aangeeft dat er wordt gezocht in een collectie.
x-connection=	Ja	De afkorting van de collectienaam. Mogelijke waarden: cvdr; bm; vg; oo; sc; lnk. (zie de tabel in hoofdstuk 1Fout! Verwijzingsbron niet gevonden.)
startRecord= &maximumRecords=	Ja	Deze parameters worden doorgaans in combinatie gebruikt om een beperkte set resultaten op te vragen, bijvoorbeeld resultaat 11-20 van de in totaal 600 zoekresultaten. Met startRecord wordt aangegeven vanaf welk record resultaten moeten worden teruggegeven (in geval van het voorbeeld: 11) en met maximumRecords hoeveel er totaal moeten terugkomen (in het voorbeeld: 10). Het maximale aantal zoekresultaten voor een query ligt op

		4020. Voor het synchroniseren van collecties moeten er dus meerdere zoekvragen worden gesteld die onder deze grens blijven.
x-info-1- accept=any	Nee	Als gewenst is dat de webservice <extraresponsedata> en <extrarecorddata> parameters retourneert, moet dit aan alle requests worden toegevoegd. Zie voor een toelichting http://www.loc.gov/standards/sru/resources/accept.html</extrarecorddata></extraresponsedata>
query=	Ja	De daadwerkelijke zoekvraag in de vorm van een 'CQL query', bijvoorbeeld "creator=Amsterdam and title=legesverordening". Zie de volgende paragraaf voor een toelichting.

4.2 CQL-query: zoeken in collecties

Voor het uitvragen van collecties in de zoekdienst kunnen zoekvragen worden gesteld in CQL (Contextual Query Language). De CQL query staat in de URL na query=.

Het onderstaande voorbeeld is een query voor de lokale regelingen (CVDR) collectie: http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrieve&x-connection=cvdr&startRecord=1&maximumRecords=10&query=title=algemene

Voor de bondigheid worden in de voorbeelden in dit hoofdstuk het gedeelte vóór de CQL query weggelaten:

...&query=dcterms.title=algemene

Tip: de afgekorte URL's in dit hoofdstuk zijn hyperlinks. Klik erop om het zoekresultaat te tonen.

4.2.1 Zoeken in veldnamen

De zoekcriteria hebben betrekking op de veldnamen in een collectie. In de CQL query kan volstaan worden met de naam zonder namespace prefix, dus bijvoorbeeld title=Paspoort is gelijk aan dcterms.title=Paspoort.

Opmerking: in dit hoofdstuk worden voorbeelden gebruikt waarin veldnamen voorkomen zoals dcterms.title (voor de titel van een document) en dcterms.creator (voor de auteur van een document). In hoofdstuk 7 wordt toegelicht welke veldnamen er per collectie beschikbaar zijn.

4.2.2 Niet gevoelig voor hoofdletters, accenten, etc.

Het maakt niet uit of een zoekvraag met of zonder accenten, koppeltekens, diakritische tekens, hoofdletters of kleine letters gesteld wordt. Ofwel: Een zoekopdracht voor 'cafe' vindt ook 'café', 'CAFÉ' en 'Café'.

4.2.3 Operatoren

De volgende operatoren etc. kunnen worden gebruikt in CQL:

Relations

= gedeeltelijke match, tekst komt voor in het betreffende veld

== volledige match, tekst komt overeen met de waarde in het betreffende veld

<, >, <=, >= vergelijking, van toepassing op datum-velden

adi resultaten bevatten de opgegeven zoektermen in exact de

volgorde als opgegeven: een zogenaamde "exact phrase search"

all resultaten bevatten de opgegeven zoektermen in willekeurige

volgorde

any resultaten bevatten in ieder geval een van de opgegeven zoektermen

Booleans

and resultaten moeten aan beide criteria wordt voldaan

or resultaten moeten aan minstens één van beide criteria voldoen not resultaten moeten niét aan het criterium achter 'not' voldoen

Wildcards

* wildcard

? wildcard voor 1 willekeurig karakter in een string

Let op: het gebruik van aanhalingstekens betekent hier dus geen *exact search phrase* zoals dat voor andere zoekmachines vaak wel het geval is. Hiervoor wordt de 'adj' (adjacent) relatie gebruikt.

Zie http://www.loc.gov/standards/sru/resources/cql-context-set-v1-2.html voor een uitvoeriger beschrijving.

4.2.4 Complexere zoekvragen

Full-text zoeken

Om full-text in een collectie te zoeken, kan de volgende zoekopdracht worden gebruikt: ...&query=keyword=fiets

Het hangt van de collectie af in welke velden precies wordt gezocht bij een full-text search.

Een hele collectie uitvragen

Voor een lege zoekopdracht worden lege aanhalingstekens gebruikt. Onderstaande zoekopdracht resulteert in het uitvragen van de gehele collectie lokale regelingen: ...&query=keyword="""

Zoekopdracht op meerdere veldnamen

Om verkeerde interpretatie van queries te voorkomen wordt aanbevolen criteria te groeperen met behulp van haakjes. Waarden kunnen het best tussen dubbele quotes ("") in de query worden opgenomen.

... &query=(title="Algemene Plaatselijke Verordening") and (dcterms.creator=Utrecht)

4.2.5 Sorteren

De zoekresultaten worden standaard gesorteerd op relevantie. Deze wordt bepaald op basis van het voorkomen van zoektermen in metadata en/of de bodytekst.

Voor het sorteren op een specifieke veldnaam kan de parameter sortBy gebruikt worden:

...&query=keyword=fiets sortBy dcterms.modified/sort.descending

Sortby kan niet op alle veldnamen gebruikt worden. Gebruik de 'explain'-operatie (zie hoofdstuk 7) om te achterhalen voor welke veldnamen sortBy ondersteund wordt.

4.2.6 Zoeken op postcode, organisatietype en organisatie

Voor veel collecties is het mogelijk om op postcode te zoeken. Vaak wordt in een zoekvraag een postcode gecombineerd met een organisatienaam of organisatietype. Hier volgt een toelichting op dit soort zoekvragen.

Wanneer in de zoekvraag een (4-cijfer) postcode of gemeente als locatie wordt gespecificeerd, bepaalt de zoekdienst zelf welke organisaties relevant zijn. Dit wordt bepaald op basis van:

- de (commerciële) postcodetabel met relaties tussen gemeenten, (4-cijfer) postcodes en provincies. Deze postcodetabel wordt periodiek aangepast;
- het OWMS-framework met daarin geografische relaties tussen diverse soorten organisaties.
 Momenteel zijn dat de relaties gemeente-waterschap en gemeente-GGD. Deze relaties worden aangepast indien nodig.

Bij aanduiding van locatie met een postcode bepaalt de zoekdienst dus welke gemeente het betreft, in welke provincie die ligt, welke waterschappen in die gemeente actief zijn, en welke regionale organisaties op die postcode of in die gemeente actief zijn (afhankelijk van hoe de organisatie de informatie aanlevert).

De (4-cijfer) postcode kan als volgt worden opgenomen in een CQL query:

...&query=postcodeCijfers=3514.

4.3 Uitvragen via HTTP GET of HTTP POST

Het is mogelijk om de webservice via HTTP GET of HTTP POST uit te vragen. SOAP wordt niet ondersteund. Zie http://www.loc.gov/standards/sru/specs/transport.html voor meer informatie.

5 Het zoekresultaat: de opbouw van een XML-response

De zoekdienst geeft een zoekresultaat terug in XML conform de SRU-specificaties. De opbouw van de



XML wordt in dit hoofdstuk beschreven.

Indien de SRU-query niet wordt begrepen door de zoekdienst, of als er een fout optreedt, dan volgt een standaard SRU-foutbericht. Dit wordt beschreven in hoofdstuk 6. Zie de bijlage voor een voorbeeld van een volledig XML-response.

5.1 Overzicht: <searchRetrieveResponse> op hoofdlijnen

Het XML-zoekresultaat bevat ieder afzonderlijk gevonden record (resultaat), extra informatie per record en extra informatie over de hele resultaatset.

Ten behoeve van de leesbaarheid zijn de namespaces in de voorbeelden weggelaten. De generieke SRU-elementen vallen in de namespace "http://www.loc.gov/zing/srw/".

In hoofdlijnen ziet een zoekresultaat er als volgt uit (merk op: de regelnummers staan niet in het resultaat: deze worden gebruikt om naar te verwijzen):

```
1. <searchRetrieveResponse>
2.
       <version>1.2
       <numberOfRecords>14</numberOfRecords>
3.
       <records>
4.
5.
               <record>...</record>
6.
              <record>...</record>
7.
8.
       </records>
9.
       \verb|<nextRecordPosition>| 11</nextRecordPosition>|
10.
       <extraResponseData>
11.
       </extraResponseData>
12. </searchRetrieveResponse>
```

Alle elementen in bovenstaand overzicht zijn standaard SRU elementen.

Regel	Omschrijving
1	<searchretrieveresponse> is het root-element.</searchretrieveresponse>

2	Het element <version> heeft betrekking op de versie van de SRU-standaard.</version>
3 & 9	Het element <numberofrecords> geeft het totaal aantal gevonden resultaten weer. Dit staat los van het aantal records in het XML-bericht dat wordt bepaald door gebruik van 'maximumRecords' in de query. Wordt in de zoekvraag bijvoorbeeld maximumRecords=10 gespecificeerd, en zijn er 14 resultaten, dan bevat <numberofrecords> de waarde 14. Het aantal records in het XML-bericht wordt niet expliciet opgenomen, maar kan worden bepaald door het aantal <record> elementen te tellen. Indien van toepassing wordt het element <nextrecordposition> (regel 9) opgenomen, waarmee wordt verwezen naar het eerstvolgende record in de resultatenset. In het voorbeeld zou dat 11 zijn.</nextrecordposition></record></numberofrecords></numberofrecords>
4	Het element <records> is het moederelement van alle <record> elementen.</record></records>
5-7	leder <record> bevat 1 zoekresultaat. Dit element wordt in de volgende paragraaf toegelicht.</record>
10	Het element <extraresponsedata> bevat extra informatie op het niveau van de hele resultaatset. Hieronder vallen onder meer facetten. Dit element wordt in paragraaf 5.3 toegelicht.</extraresponsedata>

5.2 Het zoekresultaat: <record>

Elk record element bevat informatie over het gevonden resultaat (bijvoorbeeld een document of een product). Deze bestaat uit een aantal generieke SRU-elementen (in het overzicht hieronder dikgedrukt) en een aantal specifiek voor de zoekdienst.

```
1. <record>
      <recordSchema>http://standaarden.overheid.nl/sru/
3.
       <recordPacking>xml</recordPacking>
       <recordData>
5.
              <gzd>
6.
                     <originalData>
8.
                     </originalData>
9.
                     <enrichedData>
10.
11.
                     </enrichedData>
12.
              </gzd>
13.
       </recordData>
14.
       <recordPosition>1</recordPosition>
15.
       <extraRecordData>
16.
17.
       </extraRecordData>
18. </record>
```

Regel	Omschrijving			
1	leder <record> bevat een gevonden resultaat, bijvoorbeeld een document of een product.</record>			
2	leder <record> voldoet aan een <recordschema> dat gelijk is voor alle collecties in de zoekdienst.</recordschema></record>			
3	De <recordpacking> is voor de zoekdienst altijd 'XML' en niet 'string'.</recordpacking>			
4	Onder het element <recorddata> staat de metadata van het record.</recorddata>			
5	Het <gzd> element staat voor Gemeenschappelijke Zoekdienst. Dit is het root element voor de metadata.</gzd>			
6 – 8	Onder <originaldata> staat de data in OWMS-formaat zoals deze ook in de broncollectie staat. Bij sommige collecties zal dit alleen de data tussen het <meta/>-element zijn. Bij andere collecties zal dit zowel de informatie in de <meta/> als in de <body> moeten worden meegestuurd. In dat geval wordt ook het containerbegrip meegestuurd. Het element <originaldata> bestaat uit de elementen <owmskern> (met daarin verplichte metadatavelden volgens OWMS), <owmsmantel> (met optionele metadatavelden volgens OWMS) en een collectiespecifiek element. Deze opbouw is bedoeld om te standaardiseren tussen collecties: elke collectie zal bijvoorbeeld een element <dcterms:title> hebben waarin de titel van een document of product is te vinden.</dcterms:title></owmsmantel></owmskern></originaldata></body></originaldata>			
9–11	In het element <enricheddata> staan veldnamen die niet als element in de broncollectie staan, maar daar wel uit kunnen worden afgeleid. Voorbeelden hiervan zijn <organisatietype> bij alle collecties en <publicatieurl> bij de lokale regelingen (CVDR)-collectie.</publicatieurl></organisatietype></enricheddata>			
14	In het element <recordposition> wordt de positie van het record binnen de totale resultatenset getoond.</recordposition>			
15-17	 Onder <extrarecorddata> staan enkele extra kenmerken van het <record>.</record></extrarecorddata> <version>: De zoekdienst retourneert de eenvoudige versie of de volledige versie van de metadata. Eenvoudige versie zijn alleen de veldnamen, die nodig zijn voor de resultatenlijst op Overheid.nl. Hier staat altijd version = full. De version = simple is niet geïmplementeerd.</version> <ti> <ti></ti></ti>			

gemeentenaam), dan is er ook geen snippet en wordt deze niet getoond.

5.3 Extra informatie op resultaatsetniveau: <extraResponseData>

→ **leeswijzer**: De informatie in <extraResponseData> zal niet voor iedere ontwikkelaar relevant zijn. Het betreft geavanceerdere functionaliteit die niet noodzakelijk is voor een eenvoudige uitvraag van informatie.

Onder het element <extraResponseData> staat extra informatie op het niveau van de hele resultaatset:

- Facetten: <facetedResults>. Facetten zijn een soort filters waardoor een resultaat verder kan worden verfijnd.
- Sparql-verrijkingsinformatie: <owmsOntology>. Als een gemeente fuseert met een andere gemeente, dan wil de eindgebruiker vaak graag ook de resultaten van de 'oude' gemeente zien als hij op de 'nieuwe' gemeente zoekt. Voor dit soort verrijking is de Sparql-verrijkingsinformatie bedoeld.
- Facet- en spellingssuggesties: <suggestions>. Soms spelt een eindgebruiker een naam niet correct. Hiervoor zijn facet- en spellingssuggesties bedoeld.
- Sponsored links <sponsoredLinks>. Bijvoorbeeld de landingpages van Overheid.nl die aan de gebruiker worden getoond als deze zoekt op vooraf gedefinieerde zoektermen.

In dit hoofdstuk worden de bovenstaande elementen verder beschreven.

5.3.1 Facetten: <facetedResults>

5.3.1.1. Wat zijn facetten?

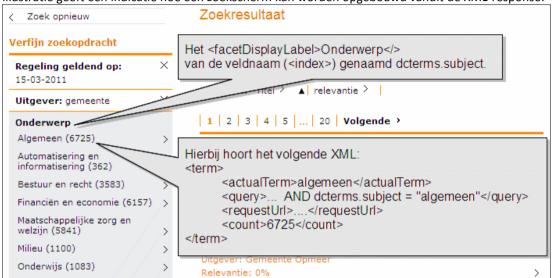
Facetten kunnen door eindgebruikers worden gebruikt om het zoekresultaat verder te verfijnen. In het screenshot hieronder staat een detail van een zoekresultaat op Overheid.nl:



In het voorbeeld zijn er 32.566 zoekresultaten. Als de gebruiker zou klikken op 'Algemeen', dan wordt het zoekresultaat *beperkt* tot alle documenten die als onderwerp 'Algemeen' hebben. Merk op dat in het voorbeeld al is geklikt op het facet uitgever, waardoor er alleen nog maar uitgevers van het organisatietype 'gemeente ' in het zoekresultaat zitten (dit is te zien aan 'Uitgever: gemeente' in het schermvoorbeeld).

5.3.1.2. Werking van facetten

Met de SRU-response is het mogelijk om vergelijkbare functionaliteit te maken. De onderstaande illustratie geeft een indicatie hoe een zoekscherm kan worden opgebouwd vanuit de XML-response:



Facetten worden in hoofdlijnen als volgt opgebouwd in de response-XML:



Regel	Omschrijving		
1	<facetedresults> is het moederelement voor alle facetten</facetedresults>		
2-5	De <datasource> heeft als kenmerken een <datasourcedisplaylabel> (een korte omschrijving van de databron voor de eindgebruiker), een <datasourcedescription> (een toelichting op de databron voor de eindgebruiker) en een <baseurl> (de URL waar de databron te benaderen is).</baseurl></datasourcedescription></datasourcedisplaylabel></datasource>		
6-9	Hier staan de daadwerkelijke facetten, zie hieronder voor een toelichting.		

Een <facet> is als volgt opgebouwd:

```
1. <facet>
```

```
2.
       <facetDisplayLabel>Uitgever</facetDisplayLabel>
3.
       <facetDescription>De naam van de organisatie...</facetDescription>
4.
       <index>dcterms.creator</index>
       <relation>=</relation>
5.
6.
       <terms>
7.
               <term>
8.
                      <actualTerm> Duiven </actualTerm>
9.
                      <query>... AND dcterms.creator = "Duiven"</query>
10.
                      <requestUrl>...</requestUrl>
11.
                      <count>3</count>
12.
               </t.erm>
13.
               <term>
14.
15.
               </term>
16.
       <terms>
17. </facet>
```

Regel	Omschrijving
2	Het <facetdisplaylabel> kan worden getoond aan de gebruiker. Het is de naam van het facet.</facetdisplaylabel>
3	De <facetdesciption> is een toelichting op het facet.</facetdesciption>
4	De <index> is de veldnaam van het facet (ofwel: hierop kan worden gefilterd).</index>
5	<relation> is de gebruikte operator voor het facet.</relation>
7-11	Een <term> is de waarde waarop kan worden gefilterd. Deze heeft een <actualterm> die als link kan worden getoond aan de gebruiker, een <query> die kan worden uitgevoerd als de gebruiker op de link klikt (merk op dat deze hetzelfde is als de originele query PLUS het extra filtercriterium), de bijbehorende <requesturl> en een <count> van het aantal keren dat een resultaat met deze <term> voorkomt in het zoekresultaat.</term></count></requesturl></query></actualterm></term>

5.3.1.3. Type facetten

Er zijn drie type facetten:

- String facet. Dit is een simpel facet op een veldnaam. Dit type facet kan bijvoorbeeld worden gebruikt als een gebruiker alleen documenten wil zien van de gemeente Stadskanaal.
- Range facet (ook wel periode facet genoemd als het om datums gaat). De zoekdienst definieert op welke periode er kan worden gefilterd. Mogelijke waarden zijn bijvoorbeeld 'Afgelopen week' of 'Afgelopen maand'. Zie de Bekendmakingencollectie op Overheid.nl voor een voorbeeld van een range facet.
- Taxonomie facet (of hiërarchisch facet). Op een taxonomie of hiërarchisch facet kun je doordrillen (= inzoomen). Een voorbeeld: als er 10 provincies (organisatieType = provincie) en 100 gemeentes ((organisatieType = provincie) in je

zoekresultaat zitten, dan wil je niet het facet van creator al zien (dus een opsomming van al deze provincies en gemeentes). Pas als je doorklikt op provincies, wil je al de onderliggende provincies zien. In dit voorbeeld zou de definitie van het hiërarchische facet zijn: organisatieType -> creator. In deze definitie wordt dus het 'parent' facet met het 'child' facet verbonden.

5.3.1.4. Facetzoeken volgens SRU 2.0 als extensie in SRU 1.2

SRU versie 1.2 (de gebruikte webservicestandaard voor de zoekdienst) bevat geen facetzoeken. Facetzoeken is wel mogelijk in de conceptversie van SRU 2.0. Daarom is besloten facetzoeken uit SRU 2.0 te gebruiken in SRU 1.2 als extension.

Er is geen aparte requestparameter voor het zoeken op facetten. Welke facetten er in de response-XML worden teruggegeven, hangt af van de collectie en wordt bepaald door de server.

5.3.2 Sparql-verrijkingsinformatie: <owmsOntology>

Als een gemeente fuseert met een andere gemeente, dan wil de eindgebruiker vaak graag ook de resultaten van de 'oude' gemeente zien als hij op de 'nieuwe' gemeente zoekt. Of als een gebruiker op een organisatie zoekt, wil hij graag ook documenten van organisatieonderdelen van die organisatie zien. Voor dit soort verrijking is de Sparql-verrijkingsinformatie op basis van de OWMS-ontologie bedoeld.

Informatie hierover wordt eenmalig opgenomen op het niveau van de XML-response in het aan de zoekdienst aangeleverde Sparql-result XML-formaat. Een XML-fragment daarvan ziet er als volgt uit:

```
<owmsOntology:rdfEnrichment</pre>
xmlns:owmsOntology="http://standaarden.overheid.nl/sru/ontology">
      <result>
     <binding name="oudeGemeente">
     <uri>http://standaarden.overheid.nl/owms/terms/Venhuizen (gemeente)</uri>
     </binding>
    <binding name="preflabelOudeGemeente">
     <literal xml:lang="nl">Venhuizen</literal>
    </binding>
    <binding name="property">
     <uri>http://standaarden.overheid.nl/owms/terms/successor</uri>
    </binding>
    <binding name="nieuweGemeente">
<uri>http://standaarden.overheid.nl/owms/terms/Drechterland (gemeente)</uri>
    </binding>
    <binding name="preflabelNieuweGemeente">
     teral
datatype="http://www.w3.org/2001/XMLSchema#string">Drechterland</literal>
    </binding>
   </result>
</owmsOntology:rdfEnrichment>
```

Van belang is in ieder geval om te weten op basis van welke relatie uit de OWMS-ontologie additionele records zijn toegevoegd. Indien gewenst kan in de user interface melding gemaakt worden van deze relatie. Bovenaan een resultatenlijst kan dan staan: "In de zoekresultaten kunnen organisaties voorkomen die later opgegaan zijn in de organisatie(s) waarop u zocht". Voor collecties met bijvoorbeeld een parent-child relaties kan dat zinnetje worden "In de

zoekresultaten kunnen organisaties voorkomen die onderdeel uitmaken van de organisatie(s) waarop u zocht".

In het geval van de relatie Bekendmakingtype - ThemaIndelingOverheid wordt ook Sparql XML teruggegeven. Wanneer je Bekendmakingen opvraagt zonder specifiek Bekendmakingtype zouden alle results van de Sparql result-XML moeten worden teruggeleverd.

5.3.3 Facet- en spellingsuggesties: < suggestions >

Suggestions worden opgenomen in de extraResponseData als in de specificaties van de collectie is aangegeven dat gebruik wordt gemaakt van spellingsuggesties of facetsuggesties. De suggestie met label en requestUrl wordt als volgt gestructureerd:

5.3.4 Sponsored links: <sponsoredLinks>

Gesponsorde links zijn extra resultaten die voorkomen als de gebruiker op een bepaalde voorgedefinieerde term zoekt, zoals "Onderwijs".

5.4 Advies: ontwikkel robuuste foutafhandeling

De structuur van een XML-response is relatief vrij. Dit heeft als voordeel dat er afhankelijk van de zoekvraag meer of minder elementen kunnen worden teruggegeven. Ook is het denkbaar dat er in de toekomst meer elementen zullen worden teruggegeven onder bijvoorbeeld <extraRecordData> of <extraResponseData> (voor een dergelijke flexibiliteit zijn deze elementen ook bedoeld, zie het extension mechanisme van SRU).

Dit heeft wel tot gevolg dat de client robuust geprogrammeerd dient te zijn om rekening te houden met het al dan niet aanwezig zijn van bepaalde elementen. Programmeer daarom altijd een robuuste foutafhandeling in de code die de SRU-response inleest.

6 Foutmeldingen: de diagnostics-functie

Indien het SRU-request incorrect is of als er een andere fout is opgetreden, dan retourneert de zoekdienst een of meer foutmeldingen in XML (genaamd diagnostics).



Een dergelijke foutmelding ziet er als volgt uit:

Regel	Omschrijving	
1	Onder het element <diagnostics> hangen 1 of meer foutmeldingen.</diagnostics>	
3	<uri>: In het element URI wordt de officiële URI opgenomen van de SRU-standaard. Deze begint met info:srw/diagnostic/1/ gevolgd door het relevante nummer van de foutmelding.</uri>	
4	In <details> wordt de SRU-beschrijving opgenomen uit de tweede kolom in de hierna volgende tabel.</details>	
5	In <message> wordt de foutmelding opgenomen die de achterliggende software (Fast) genereert.</message>	

SRU kent een breed scala aan standaardfoutmeldingen, waarvan de volgende zijn geïmplementeerd:

Nr.	<details></details>	Details Format
1	General system error	Debugging information (traceback) Bij een andere foutmelding die hieronder niet wordt genoemd, wordt deze foutmelding gegeven + details van de onderliggende zoektechnologie.
2	System temporarily unavailable	
4	Unsupported operation	
6	Unsupported parameter value	

7	Mandatory parameter not supplied	
8	Unsupported Parameter	
10	Query syntax error	Wanneer een fout optreedt die is gerelateerd aan het zoeken.
50	Result sets not supported	Wanneer een fout optreedt die is gerelateerd aan de response-XML. Bijvoorbeeld: — De result set is niet compleet. Dit kan komen doordat het genereren van een result set is onderbroken. — Er werden teveel records opgevraagd om een result set te genereren.
80	Sort not supported	Wanneer een fout optreedt bij het sorteren. Bijvoorbeeld: – Sort wordt voor dit veld niet ondersteund – Er worden teveel sorteersleutels / paramaters meegegeven om op te sorteren dan worden ondersteund.

Zie ook http://www.loc.gov/standards/sru/resources/diagnostics-list.html.

7 Collectiespecifieke informatie

Zoals in hoofdstuk 1 staat beschreven, kunnen er verschillende collecties worden bevraagd via SRU. Deels hebben deze collecties dezelfde veldnamen, zoals dcterms:title en dcterms:identifier (zie de website van OWMS voor meer informatie over de standaardisering tussen de verschillende collecties). In dit hoofdstuk wordt beschreven welke veldnamen er nog meer per collectie beschikbaar zijn.

7.1 Op welke velden kan worden gezocht? De explain-functie

De SRU heeft een explain-functie waarmee algemene gegevens over een collectie kunnen worden uitgevraagd. Hierin staat (in XML) de serverinfo en indextermen die kunnen worden uitgevraagd. Ook is hierin aangegeven op welke indexvelden kan worden gezocht en gesorteerd. De Explain voor Bekendmakingen is bijvoorbeeld op te vragen via de volgende URL:

http://zoekdienst.overheid.nl/sru/Search?x-connection=bm&operation=explain&version=1.2

Voor het uitvragen van andere collecties moet de x-connection= parameter worden aangepast met de afkorting die is te vinden in hoofdstuk 1.

Voor een uitgebreide inhoudelijke omschrijving van wat alle metadatavelden betekenen, wordt verwezen naar de IPM's (Internet Publicatie Modellen) van ieder product, zoals deze vindbaar zijn via de website van KOOP en Logius. Een kort overzicht van alle mogelijke metadatavelden van de collecties is tevens te vinden op de OWMS-website.

8 Vragen of opmerkingen over dit document

Indien u vragen hebt over de bevraging van afzonderlijke collecties, neem dan contact op met de collectiehouder via het mailadres.

Collectie	E-mailadres
Lokale regelingen	regelgeving@koop.wmrijk.nl
Lokale bekendmakingen	bekendmakingen@koop.wmrijk.nl
Lokale vergunningen	vergunningen@koop.wmrijk.nl
Overheidsorganisaties	servicecentrum@logius.nl
Producten en diensten	servicecentrum@logius.nl
Overheid.nl	servicecentrum@logius.nl

9 Bijlage: Voorbeelden van SRU-queries

Collectie	Voorbeelden
Lokale regelingen	http://zoekdienst.overheid.nl/sru/Search?x-connection=cvdr&operation=searchRetrieve&version=1.2&query=dcterms.title ="Algemene plaatselijke verordening" AND creator="Middelburg"
Lokale bekendmakingen	http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrie ve&x-connection=bm&startRecord=1&maximumRecords=50&query=keyword="open bare"
Lokale vergunningen	http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrie ve&x-connection=vg&startRecord=1&maximumRecords=50&query=producttype="milieuvergunning"
Overheids- organisaties	http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrie ve&x-connection=oo&startRecord=1&maximumRecords=50&query=keyword=""
Producten en diensten	http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrie ve&x-connection=sc&startRecord=1&maximumRecords=50&query= organisatie=Amersfoort and organisatietype=Gemeente
Overheid.nl	http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrie ve&x-connection=lnk&startRecord=1&maximumRecords=50&query=plaatsnaam ="Amersfoort"

10 Bijlage: Naamgevingsconventies van veldnamen

De metadata tussen de verschillende collecties is zo veel mogelijk gestandaardiseerd volgens OWMS. Om ook het uitvragen van deze metadata zo consistent mogelijk te maken, worden conventies voor 2 categorieën veldnamen onderscheiden.

→ **Leeswijzer**: deze informatie dient slechts als achtergrondinformatie. Het mechanisme om de mogelijke veldnamen per collectie uit te vragen staat beschreven in hoofdstuk 7.

10.1 Categorie 1: Veldnamen met een 1-op-1 relatie met metagegevens

Deze veldnamen zijn direct af te leiden uit de metadata. Dit geldt zowel voor OWMS-metadata als collectiespecifieke metadata. Conventies:

- De namespace wordt gescheiden van de veldnaam door een punt. Ofwel: <dcterms:creator> levert de veldnaam dcterms.creator.
- Bij het uitvragen van de metadata kan de namespace worden weggelaten maar dit hoeft niet. Ofwel: dcterms.creator is hetzelfde als creator.
- Voor dieper geneste veldnamen wordt doorgaans ouder en kind verbonden met '_'.
 Bijvoorbeeld:

10.2 Categorie 2: Verrijkte veldnamen

Deze veldnamen ontstaan door verrijking middels aanwezige schema-attributen, combinatie van meerdere metadatavelden of relaties.

Bijvoorbeeld:

Voor meerdere IPM's geldt dat met het element <dcterms:creator scheme="overheid:Gemeente"> een nieuwe veldnaam organisatieType wordt aangemaakt met in dit geval als waarde gemeente.

Deze veldnaam zijn zo veel mogelijk hetzelfde voor alle collecties.

11 Bijlage: Ondersteuning van de SRU-standaard

De webservice implementeert SRU versie 1.2.

Facetzoeken maakt geen onderdeel uit van SRU 1.2, maar maakt wel onderdeel uit van de SRU 2.0 draft. De draft van SRU 2.0 is daarom gebruikt als voorbeeld voor een extensie voor facetzoeken (zie http://www.loc.gov/standards/sru/specs/extra-data.html voor een uitleg over het extensiemechanisme).

Verder wordt het volgende ondersteund:

Onderdeel	Туре	Ondersteund door zoekdienst?
SearchRetrieve	Operation	Ja
Explain	Operation	Ja
Scan	Operation	Nee
Diagnostics	Response	Ja
Stylesheet	Request parameter	Nee
ResultsetTTL	Request parameter	Nee. Er zijn geen persistent result sets.
Recordschema	Request parameter	Nee. Er wordt alleen gebruik gemaakt van een default.

		recordschema gzd.xsd. *
RecordPacking = string	Request parameter	Nee. Alleen XML wordt ondersteund.
CQL	Contextual Query Language	Level 1. Zie www.loc.gov/standards/sru/s pecs/cql.html.
SRU via SOAP		Niet expliciet ondersteund.

^{*} De gzd maakt gebruikt van een enkel response-schema voor alle collecties. Dit betekent dat het schema open van opzet is. Voor details over de respons- XML van afzonderlijke collecties biedt de XSD van elke collectie afzonderlijk het meeste houvast aangezien doorgaans de originele data wordt teruggegeven.

12 Bijlage: voorbeeld XML-zoekresultaat

```
<?xml version="1.0" encoding="UTF-8"?>
<searchRetrieveResponse xmlns="http://www.loc.gov/zing/srw/"</pre>
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://www.loc.gov/zing/srw/ srw-types.xsd">
 <version>1.2</version>
<numberOfRecords>1</numberOfRecords>
 <records>
<record>
  <recordSchema>http://standaarden.overheid.nl/sru/</recordSchema>
  <recordPacking>xml</recordPacking>
  <recordData>
  <gzd xmlns="http://standaarden.overheid.nl/sru"</pre>
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xmlns:dcterms="http://purl.org/dc/terms/"
 xmlns:overheid="http://standaarden.overheid.nl/owms/terms/"
xmlns:overheidproduct="http://standaarden.overheid.nl/product/terms/"
 xsi:schemaLocation="http://standaarden.overheid.nl/sru gzd.xsd">
   <original Data>
 <overheidproduct:scproduct owms-version="4.0">
       <overheidproduct:meta>
               <overheidproduct:owmskern>
 <dcterms:identifier>http://www.zuidhorn.nl/afvaldoorzoekingsvergunning</dcterm</pre>
 s:identifier>
 <dcterms:title>Vergunning voor het doorzoeken van afval/dcterms:title>
 <dcterms:language>nl</dcterms:language>
 <dcterms:type
 scheme="overheid:Informatietype">productbeschrijving</dcterms:type>
 <dcterms:modified>2010-05-17T11:01:00+02:00</dcterms:modified>
 <dcterms:spatial scheme="overheid:Gemeente" resourceIdentifier =</pre>
 "http://standaarden.overheid.nl/owms/terms/Zuidhorn (gemeente)">Zuidhorn</dcte
rms:spatial>
```

```
<overheid:authority scheme="overheid:Gemeente" resourceIdentifier =</pre>
 "http://standaarden.overheid.nl/owms/terms/Zuidhorn (gemeente)">Zuidhorn</over
heid:authority>
             </overheidproduct:owmskern>
             <overheidproduct:owmsmantel>
<dcterms:audience scheme="overheid:Doelgroep">ondernemer</dcterms:audience>
<dcterms:abstract>Als u afval wilt doorzoeken, moet u een vergunning
hebben.</dcterms:abstract>
             </overheidproduct:owmsmantel>
             <overheidproduct:scmeta>
 <overheidproduct:productID>F25C2487-E593-40DE-A2CD-
AE34ECCF79DF</overheidproduct:productID>
<overheidproduct:onlineAanvragen>nee</overheidproduct:onlineAanvragen>
<overheidproduct:eenmaliqAanmelden>nee</overheidproduct:eenmaliqAanmelden>
<overheidproduct:uniformeProductnaam scheme="overheid:UniformeProductnaam"</pre>
resourceIdentifier =
"http://standaarden.overheid.nl/owms/terms/afvaldoorzoekingsvergunning">afvald
oorzoekingsvergunning
</overheidproduct:uniformeProductnaam>
             </overheidproduct:scmeta>
      </overheidproduct:meta>
      <overheidproduct:body>
             <overheidproduct:productHTML xmlns="http://www.w3.org/1999/xhtml"</pre>
 xsi:schemaLocation="http://www.w3.org/1999/xhtml
http://www.w3.org/2002/08/xhtml/xhtml1-strict.xsd">
                    Hier de opgemaakte tekst
             </overheidproduct:productHTML>
      </overheidproduct:body>
</overheidproduct:scproduct>
  </originalData>
  <enrichedData>
  <authorityScheme>Gemeente</authorityScheme>
<authorityUri>http://standaarden.overheid.nl/owms/terms/Zuidhorn (gemeente)</a</pre>
uthorityUri>
  <spatialType>Gemeente</spatialType>
 <spatialUri>http://standaarden.overheid.nl/owms/terms/Zuidhorn (gemeente)</spa</pre>
tialUri>
<uniformeProductnaamUri>http://standaarden.overheid.nl/owms/terms/afvaldoorzoe
kingsvergunning</uniformeProductnaamUri>
  </enrichedData>
 </gzd>
 </recordData>
 <recordPosition>1</recordPosition>
 <extraRecordData xmlns:version="http://standaarden.overheid.nl/sru/version"</pre>
xmlns:dtf="http://standaarden.overheid.nl/sru/dtf"
```

13 Bijlage: Doorontwikkeling van de SRU-webservice

De specificaties van de SRU-webservice zijn per 7 juli 2010 vrijgegeven.

Er zijn op het moment van schrijven geen grote doorontwikkelingen van de SRU-webservice voorzien. Mogelijkerwijs worden in de toekomst extra collecties via de SRU-webservice ontsloten.