

## DE STANDAARD

### SRU en OpenSearch

#### Marco Streefkerk

Wat doet een informatieprofessional als hij wil weten waar SRU voor staat? Hij zoekt het op via Google. De resultaten op google.nl (situatie 1 februari 2007) zijn niet eenduidig. Afgaande op de fameuze page ranking staat SRU allereerst voor Scottish Rugby Union. Slippery Rocky University is twee. De eerste Nederlandstalige verwijzingen vertellen dat SRU staat voor Synchronicity Research Unit. Maar daarna is er een resultaat uit een bron die de informatieprofessional zal aanspreken: het domein van de Library of Congress. Het acroniem blijkt ook te staan voor *Search and Retrieve via URL*. Even doorklikken en we lezen ook een definitie, vrij vertaald: 'SRU is een gestandaardiseerd zoekprotocol voor zoekacties over internet die gebruik maakt van CQL (*Common Query Language*) een standaard syntax om zoekacties te formuleren. De Library of Congress heeft de functie van beheerorganisatie voor deze standaard.

#### In den beginne was er .....Z39.50

SRU en broertje SRW (*Search and Retrieval via the Web*) bouwen voort op de lange historie van Z39.50. Z39.50 ontstond in de jaren zeventig van de vorige eeuw vanuit een behoefte om via internet aan *shared cataloguing* te doen. Het idee was om een protocol te ontwikkelen waarmee meerdere bibliografische databases konden worden doorzocht, zonder kennis van de gebruikte techniek en structuur van deze database.

In de jaren tachtig nam het gebruik van Z39.50 toe als gevolg van de groei van het Internet, de toegankelijkheid van computerfaciliteiten en een groot aanbod van ge-netwerk-te (commerciële) databases en bibliotheekcatalogi. Wetenschappers gebruikten het protocol voor literatuuronderzoek op afstand via citatiemanagement-programma's als EndNote. In de jaren negentig namen bibliotheken de zorg van het opzetten van individuele verbindingen met verschillende databases bij eindgebruikers over. Bibliotheekgebruikers kregen meer en meer te maken met een zeer heterogeen aanbod van (bibliografische) databases. In antwoord op deze ontwikkeling zetten bibliotheken nieuw beschikbaar gekomen portaalsoftware in om gebruikers de mogelijkheid te bieden alle databases via één webinterface te benaderen en, indien mogelijk, te doorzoeken. Deze portalen zijn nog steeds een middel voor bibliotheken om een centrale rol in de informatieverwerving binnen academische instellingen te behouden. Achter de schermen van het portaal wordt de door de gebruiker ingevoerde zoekactie vertaald naar Z39.50, en vervolgens via internet naar één of meerdere databases verstuurd. De heterogene resultaten van de verschillende databases worden door de portaalsoftware in sets opgehaald, genormaliseerd, samengevoegd en eventueel ontdubbeld aan de gebruiker op een uniforme wijze gepresenteerd.

Het Z39.50 protocol is, hoewel in principe universeel inzetbaar, niet echt buiten het domein van de bibliotheken, of voor iets anders dan het zoeken en ophalen van bibliografische data toegepast. Dat komt ook door de lange en diffuse ontstaansgeschiedenis van het protocol die geleid heeft tot een grote mate van keuzevrijheid bij het toepassen. Hierdoor is het protocol voor database-aanbieders lastig te implementeren. Bovendien leidt de keuzevrijheid tot een diversiteit waar bibliotheken weer mee worstelen als ze een database via Z39.50 willen benaderen.

Een andere beperking komt voort uit het feit dat Z39.50 nog dateert van ruim voor de opkomst van het World Wide Web. De standaard kon daarom niet profiteren van de enorme populariteit en brede toepassing van het HTTP protocol.

## Webservices

Een aantal jaren geleden werd daarom het plan opgevat om een (beperkt deel) van de Z39.50 semantiek op te zetten als webservices met als resultaat de SRW protocollen. Hiermee werd aangesloten bij een bredere ontwikkeling waarbij geprobeerd wordt computertoepassingen via het web met elkaar te laten communiceren op een gestandaardiseerde wijze. SOAP (*Simple Object Access Protocol*) is de bekendste uitwerking daarvan. Bij SOAP vindt de uitwisseling van data plaats in XML. SRW is opgezet als SOAP: de zoekactie wordt verpakt in XML en overgebracht via een HTTP post. Hiermee zijn complexe zoekacties mogelijk. Bij SRU wordt de zoekactie verpakt in een URL die ook met behulp van HTTP get kan worden overgebracht. De URL bestaat dan uit het adres van de SRU server en parameters. De parameters vallen uiteen in de zoekactie zelf (gebaseerd op CQL) en parameters die specifiek voor het protocol zijn zoals start record, gewenste formaat, aantal records, etc. SRU doet in feite hetzelfde als alle web zoekmachines met het verschil dat het antwoord niet is vervuild met HTML, maar alleen bestaat uit data gestructureerd als XML, overeenkomstig de SRU Namespace op:

<http://www.loc.gov/zing/srw/>

## Toegepast

Een praktijkvoorbeeld van toepassing van SRU als alternatief voor Z39.50 is het project Digitale Bibliotheek Amsterdam. Twee jaren geleden gaf de Stichting AdamNet de opdracht voor de ontwikkeling van een metazoekdienst.. In deze stichting werken 25 Amsterdamse bibliotheken, groot en klein, samen. De grote leden zoals de universiteitsbibliotheken van UvA en VU en de Openbare Bibliotheek Amsterdam boden al een metazoekdienst aan op basis van Z39.50. Deze wordt bijvoorbeeld gebruikt ten behoeve van de integratie van hun catalogus in het instellingsportaal of als directe zoektoegang voor citatiemanagement-programma's. Voor de andere bibliotheken was het opzetten van een dergelijke toegang op de eigen catalogus een voorwaarde voor het slagen van het project. Wel hadden vrijwel alle leden een online webcatalogus. Vanuit die uitgangssituatie werd de bibliotheken geadviseerd in samenwerking met hun softwareleverancier een SRU-toegang te ontwikkelen. In eerste instantie werd daarmee voorzien in de behoefte om te kunnen koppelen met de AdamNet metazoekdienst, maar een SRU-toegang zou in de toekomst ook voor andere server-to-server communicatie kunnen worden ingezet.

De resultaten van het project kunnen mooi dienen als illustratie van de SRU standaard. Wanneer een bezoeker van de Digitale Bibliotheek Amsterdam via de webinterface op

<http://digitaal.uba.uva.nl/V?func=change-portal-1&portal-name=DBA>

zoekt op publicaties met het zoekwoord 'Gogh' in de titel wordt de zoekactie door de zoekdienst omgezet in SRU: "?version=1.1&operation=searchRetrieve&query=title=gogh".

Een verplichte vermelding van de versie, het soort actie (in dit geval zoeken) en de zoekactie zelf. Stel het geval dat deze zoekactie door de bezoeker is gericht aan drie catalogi dan worden de volgende http verzoeken uitgezonden waarbij telkens alleen het adres van de SRU-server verschilt:

<http://www.ihlia.nl:9000/ihlia?version=1.1&operation=searchRetrieve&query=title=gogh>

<http://library.rijksmuseum.nl/srw/library.pl?operation=searchRetrieve&query=title=gogh>

<http://vangogh.adlibsoft.com/sru/wwwopac.exe?version=1.1&operation=searchRetrieve&query=title=gogh>

De drie catalogi sturen hun antwoord in XML terug, waarna het door de portaalsoftware wordt omgezet naar het interne formaat om aan de bezoeker te presenteren. Door één van de bovengestelde SRU-URLs direct in de browser aan te roepen kunnen we het antwoord in SRU-XML bekijken:

```
<?xml version="1.0" encoding="UTF-8" ?>
- <searchRetrieveResponse xmlns="http://www.loc.gov/zing/srw/" xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:mods="http://www.loc.gov/mods"
  xmlns:dcterms="http://purl.org/dc/terms/">
  <version>1.1</version>
  <numberOfRecords>38</numberOfRecords>
  - <records>
  - <record>
    <recordPacking>xml</recordPacking>
    - <recordData>
      - <rdf:RDF xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        - <rdf:Description rdf:ID="10142167">
          <dc:identifier>ID:10142167</dc:identifier>
          <dc:title>Vincent van Gogh : collection du Musée National Vincent van Gogh à Amsterdam</dc:title>
          <dc:publisher>Paris</dc:publisher>
          <dc:date>1971</dc:date>
          <dc:type>Book</dc:type>
          <dc:format>ill.</dc:format>
          <dc:format>4°</dc:format>
          <dc:description>Exposition à l'Orangerie des Tuileries Paris, 21 décembre 1971-10 avril 1972</dc:description>
          <dc:creator>Adhémar, Hélène</dc:creator>
          <dc:creator>Châtelain, Jean</dc:creator>
          <dc:creator>Gogh, Vincent van</dc:creator>
          <dc:creator>Meyer, Emile Rudolf</dc:creator>
          <dc:subject>Gogh, Vincent van</dc:subject>
          <dc:source>P Gogh 1971</dc:source>
        </rdf:Description>
      </rdf:RDF>
    </recordData>
    <recordPosition>1</recordPosition>
  </record>
</records>
  <echoedSearchRetrieveRequest />
</searchRetrieveResponse>
```

De oplettende lezer ziet dat voor dit plaatje nog wat extra parameters aan de URL zijn toegevoegd: een aanduiding van het aantal terug te bezorgen records, het nummer waarmee de set begint en het gevraagde formaat (in dit geval DublinCore).

### Operation=Explain

SRU heeft ook de 'explain' operatie van Z39.50 geërfd. 'Explain' dient om de beschrijving van de zoekmachine op te vragen. Voor de SRU-dienst die toegang geeft tot de bibliotheekcatalogus van het Van Gogh Museum kan dat met de volgende SRU-URL:

<http://vangogh.adlibsoft.com/sru/wwwopac.exe?version=1.1&operation=explain>

In de XML van het antwoord is te lezen over de zoekdienst: waar deze zich bevindt, wie hem aanbiedt, wat er in te vinden is, waarop je kunt zoeken en hoe de zoekresultaten kunnen worden geleverd.

De structuur van de SRU/SRW antwoorden in XML biedt de mogelijkheid om de presentatie te regelen via een stylesheet en zo zelfs een webpagina met functionaliteit te genereren die aansluit op de inhoud. Zo kan zelfs de beschrijving van de functionaliteiten van een SRU-server in het 'explain'-document met de juiste stylesheet-aanroep worden gepresenteerd in een formulier dat de zoekmogelijkheden direct aan de bezoeker beschikbaar stelt. Voor het Van Gogh Museum is het op deze manier gegenereerd web-zoekformulier hier te bekijken:

<http://vangogh.adlibsoft.com/sru/wwwopac.exe?version=1.1&operation=explain&stylesheet=explainResponse.xml>

**Library Van Gogh Museum**  
Adlib Internet Server Copyright (c) 1984-2006, ADLIB Information Systems

Library catalogue Van Gogh Museum Amsterdam, collection books, articles and periodicals on 19th century art. The Van Gogh Museum library serves as a major centre for research into art from the period 1800 to 1920. The emphasis of its collection is on Western European art, in particular on Dutch and French visual arts of the mid-19th to the early 20th century.

**Search**

Index	Relation	Term	Boolean
All Indexes	=		and
Anywhere	=		and
Creator	=		and
ISBN number	=		and
ISSN number	=		and
publication year	=		and
Subject	=		and
Title	=		and
Whole record	=		and

Record Schema:

Number of Records:

Record Position:

Result Set TTL:

Record Packing:

Record XPath:

Sort Keys:

**Browse**

Index	Relation	Term	Boolean
All Indexes	=		

Response Position:

Maximum Terms:

Een andere aantrekkelijke mogelijkheid is dat bijvoorbeeld bij gebruik van de SRU-koppeling voor een metazoekdienst de portaalsoftware op basis van de informatie in de Explain XML de juiste instellingen kiest (bijvoorbeeld welke velden doorzoekbaar zijn). Wijzigingen aan de kant van de database zouden dan automatisch door de portaalsoftware worden herkend en verwerkt. Dat zou gebruikers van bibliotheekportalen veel storingen besparen en de bibliotheken van een belangrijke beheertaak vrijstellen.

### Status van de standaard

De SRU/SRW standaarden zijn nog in ontwikkeling. De huidige versie van het SRU protocol is versie 1.1 uit februari 2004. Versie 1.2 die binnenkort zal worden vrijgegeven brengt slechts enkele kleine aanpassingen. Een planning of inhoud voor versie 2.0 is naar mijn weten nog niet bekend.

De adoptie van SRU/SRW door bibliotheken, uitgevers en andere spelers in het domein van de wetenschappelijke informatievoorziening verloopt vooralsnog traag. Wel duikt SRU de laatste jaren steeds vaker op in beleidsstukken van grote informatieorganisaties. Dergelijke instellingen hebben de afgelopen jaren veel geïnvesteerd in het opzetten en toepassen van nieuwe systemen voor archiveren, ontsluiten, publiceren en presenteren van informatie. Op dit moment staan die systemen naar het idee nog teveel als losse applicaties naast elkaar, met nog onvoldoende en bovendien moeizame interoperabiliteit. De uitdaging voor de komende periode ligt in het geleidelijk opdelen van deze gesloten kolommen van data en functionaliteit in open, flexibele en uitwisselbare componenten. SRU wordt in dergelijke plannen vaak in

één adem genoemd met OAI-MPH en OpenURL<sup>1</sup> als complementaire protocollen voor de uitwisseling van data tussen de verschillende componenten.

Concrete toepassingen van SRU zoals de Digitale Bibliotheek Amsterdam zijn nog schaars. Een reden daarvoor is wellicht dat het protocol veelal wordt gezien als opvolger van Z39.50 en dat Z39.50 voor ondersteuning van metazoecken vanuit citatiemanagers en bibliotheekportalen nog goed functioneert. Het veld heeft weinig prikkel om te investeren in een nieuwe techniek waarvoor de meerwaarde initieel gering en de toekomstige toepassing onzeker zijn. Wel bieden een aantal uitgevers en softwareleveranciers al toegangen op basis van HTTP en XML die kunnen worden gezien als eerste stap richting gestandaardiseerde webservices. OCLC-PICA ontwikkelde al in samenwerking met enkele bibliotheken een SRU-toegang, onder andere voor de Nederlandse Centrale Catalogus. Op dit moment is de SRU-dienst helaas nog niet voor productie beschikbaar. Op de achtergrond speelt ook de strijd om de gebruiker. In welke context verricht de onderzoeker zijn literatuuronderzoek: binnen de interface van de (commerciële) uitgever of in het portaal van de bibliotheek?

### **OpenSearch**

Inhoudelijk richt zich de kritiek op het SRU/SRW koppel zich vooral op de specifiek bibliotheekgerichte opzet ervan. Een eigenschap die acceptatie daarbuiten bemoeilijkt is het gebruik van de CQL als voorgeschreven zoektaal die als complex wordt gezien en opnieuw buiten de bibliotheekwereld weinig toepassing kent. Voor de toekomst wordt dan ook gekeken naar integratie van SRU met OpenSearch. OpenSearch werd ontwikkeld voor de A9-zoekmachine, een initiatief van Amazon. OpenSearch is ook een protocol om zoekresultaten te delen. De specificaties van OpenSearch zijn beschikbaar onder Creative Commons voorwaarden en er is een actieve gemeenschap voor het uitwisselen van ervaringen en programmaatjes. OpenSearch kent net als SRU een beschrijvend document in XML dat kan worden opgevraagd en waarin de beheerder van de zoekmachine beschrijft wat er op zijn site is te vinden en hoe er gezocht moet worden. In tegenstelling tot SRU schrijft OpenSearch (versie 1.1) geen zoektaal en structuur van de resultaten voor. Vaak verpakken OpenSearch bronnen hun resultaten in RSS<sup>2</sup> (*Really Simple Syndication*) aangevuld met specifieke OpenSearch elementen zoals het aantal treffers. Door de weinige eisen die OpenSearch stelt is er inmiddels een lange lijst van websites die met OpenSearch kan worden bevraagd, waaronder enkele bibliotheekcatalogi en ook PubMed, een belangrijke publieke database voor medici.

### **Toekomstverwachtingen**

Zoals de voorbeelden aantonen volstaat een standaard browser als applicatie om SRU (of OpenSearch) zoekacties te starten. Omdat de output in SRU voorgeschreven XML is en de moderne browsers ook de presentatie van XML met behulp van stylesheets ondersteunen is in potentie de huidige laag van de portaalsoftware zoals wordt gebruikt in de Digitale Bibliotheek Amsterdam overbodig.

Niet de beheerder van het portaal maar de individuele gebruiker bepaalt welke databases worden doorzocht. Niet de beheerder van het portaal maar de gebruiker bepaalt hoe de zoekresultaten worden gepresenteerd en verwerkt.

Een ander eveneens op het SRU-protocol gebaseerd project, de European Library, heeft een demonstratiemodel daarvoor opengesteld.

<http://krait.kb.nl/coop/tel/SRUportal/>

---

<sup>1</sup> Besproken in De Standaard, IP 2007(3)

<sup>2</sup> Besproken in De Standaard, IP 2007(1)

Hier is aardig inzicht te verkrijgen in de mogelijkheden om zelf als gebruiker de uitvoerdata te presenteren in de gewenste vorm. Ook is er iets zichtbaar van de potentie die webservices zoals SRU bieden bij het dynamisch koppelen van verschillende informatiebronnen en hun functionaliteit. Dit principe is één van de kenmerken van WEB 2.0 en de daarbinnen gebruikte techniek AJAX (Asynchronous Javascript And XML). Met behulp van AJAX is het mogelijk presentaties te veranderen op basis van handelingen door de gebruiker zonder dat de webpagina's zelf wordt herladen. Hiermee wordt de gebruikerservaring van standaard computertoepassingen benaderd.

Gegeven het feit dat bibliotheken als informatiedienstverleners in de huidige volledig genetwerk-te wereld voor de uitdaging staan om mee te gaan in de WEB 2.0 ontwikkelingen en daar in de vorm van Library 2.0 een eigen specifieke bijdrage aan te leveren, dan is de cirkel rond. SRU/SRW is daarmee een interessante ontwikkeling voor bibliotheken: zowel om bestaande activiteiten zoals metazoeken te verbeteren als om nieuwe WEB 2.0 activiteiten mogelijk te maken.

### **Geraadpleegde literatuur:**

LeVan, R; OpenSearch and SRU: A continuum of searching, in INFORMATION TECHNOLOGY AND LIBRARIES 25 (3): 151-153, 2006.

Lease Morgan, Eric; An Introduction to the Search/Retrieve URL Service (SRU), in Ariadne 40, 2004.

Lynch, Clifford A.; The Z39.50 information retrieval protocol: an overview and status report, in SIGCOMM Comput. Commun. Rev., V21, N1, P 58-70, 1991.

### **Op het web:**

Officiële website voor de SRW/SRU standaard met alle specificaties en verwijzingen naar implementaties en hulpmiddelen: <http://www.loc.gov/standards/sru/>

De startpunt voor alle informatie rond Z39.50: <http://www.loc.gov/z3950/agency/>

De webcommunity rond OpenSearch is te vinden op: <http://www.opensearch.org/>

### **Verder lezen:**

Er is nog weinig verdiepingsliteratuur over SRU/SRW. Onderstaande artikelen laten vanuit verschillend perspectief zien hoe SRU één van de bouwstenen kan zijn binnen de informatie-infrastructuur van de toekomst:

Sompel, Herbert Van der; An Interoperable Fabric for Scholarly Value Chains, in D-Lib Magazine, Vol 12, No 10, 2006.

Veen, Theo van; Serving Services in Web 2.0, in Ariadne, 47, 2006.

Sanderson, R, e.a.; SRW/U with OAI, Expected and Unexpected Synergies, in D-Lib Magazine 11, 2, 2005.

Marco Streefkerk is projectleider Ontwikkeling en Innovatie bij de sector Elektronische Diensten van de Universiteitsbibliotheek van de UvA