

EEN GEÏNTEGREERDE TOEGANG TOT EUROPESE ERFGOEDCOLLECTIES

Binnen het European Library (TEL) project wordt gewerkt aan een systeem voor geïntegreerde toegang tot de erfgoedcollecties van de Europese nationale bibliotheken. Doel is het verlagen van de drempel voor deelname aan de European Library en het toegankelijk maken van nieuwe collecties. Als oplossing werd gekozen voor een nieuw protocol voor search en retrieval via URLs en een op functionaliteit gebaseerd metadatamodel. Theo van Veen beschrijft hoe dit veelbelovende concept tot stand is gekomen.

Bij de start van het European Library-project was het de bedoeling om te streven naar een operationele dienst die zou bestaan uit gedistribueerde portals. Deze zouden niet alleen toegang moeten bieden tot de metadata en collecties van de diverse deelnemers via het Z39.50-protocol, maar ook tot een centrale XML-database

Het European Library Project (TEL) is deels gesubsidieerd door de Europese Commissie als een "accompanying measure", zijnde een onderdeel van het Information Society Technologies (IST) research programma onder het "Vijfde Kaderprogramma". Het is een samenwerkingsproject van een aantal Europese nationale bibliotheken, onder auspiciën van CENL (Conference of European National Libraries). Het project werd eerder in algemene lijnen beschreven in een artikel in de serie "Europese Projecten".¹ Daarin werd in detail ingegaan op de ontwikkeling van een testomgeving, en met name het daarbij toegepaste SRU-protocol en het metadatamodel.

GEBRUIKTE TERMINOLOGIE	
Term	Betekenis
Application Profile	Op Dublin Core gebaseerd metadatamodel bestaande uit een verzameling van velden, verfijningen van velden en coderingsschema's uit verschillende "element sets".
Dublin core	Set van 15 basiselementen voor beschrijving van objecten.
DTD	Document Type Definition van XML documenten.
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting. Een protocol voor het verzamelen van metadata gebaseerd op http en XML.
OpenURL	Een mechanisme om via een standaard syntax op basis van beschikbare metadata – dynamisch en context afhankelijk – links te genereren.
Schema	Een set regels voor het beschrijven van een metadata record. Meestal betreft het een XML-schema voor het beschrijven van XML-records.
SRU	Search and Retrieve via URLs.
SRW	Search and Retrieve via the Web.
Stylesheet	Script voor het specificeren van een presentatie.
XML	eXtensible Markup Language, een manier om gegevens te structureren.
XSLT	eXtensible Stylesheet Language Transformation. Een scripttaal voor het transformeren van XML.
Z39.50	Het standaard protocol voor Search and Retrieval van bibliografische records.

via het http-protocol. Er was dus sprake van een hybride combinatie, waarbij gelijktijdig zou worden gezocht in gedistribueerde bestanden via Z39.50 en in een centrale index van gedistribueerde bestanden die niet via Z39.50 toegankelijk zijn.

Voor beide vormen zou een testomgeving worden ontwikkeld en aan het einde van het project zou er een testomgeving gerealiseerd worden voor de combinatie van beide. Uiteindelijk zou een commerciële portal geselecteerd worden op basis van het uit de testomgeving verkregen inzicht en de resultaten van de metadata-eisen. Deze selectie zou onderdeel worden van het advies aan de TEL steering committee omtrent de mogelijkheden voor een operationele TEL-service.

Deze uitgangspunten kregen gedurende het project echter een paar interessante wendingen.

ONTWIKKELING VAN DE ZOEKOMGEVING

Kort na de start van het project werd besloten om niet voor gedistribueerde portals te kiezen, omdat anders geen sprake zou kunnen zijn van een geïntegreerde toegang. Gekozen werd voor één centrale portal, die eventueel in een

XML EN XSLT

XML is een manier om gegevens gestructureerd op te slaan of te verzenden. Door deze structuur zijn de gegevens machineleesbaar. Een eenvoudig voorbeeld is:

```
<boek>
<titel>Dit is een boek</titel>
<auteur>Jan Jansen</auteur>
</boek>
```

Omdat XML alleen gegevens bevat en geen opmaak is het voor een server eenvoudiger om gegevens in XML beschikbaar te stellen dan in bijvoorbeeld HTML. Omdat XML erg veel gebruikt wordt, zijn er veel standaard hulpmiddelen voor het gebruik van XML. Zo kan de opmaak verzorgd worden met een speciaal hiervoor bestemde scripttaal: XSLT.

XSLT is een voor XML ontworpen scripttaal. Hiermee kunnen gegevens met een XML-structuur getransformeerd worden naar bijvoorbeeld een HTML-presentatie. Internet Explorer en enkele andere browsers hebben de mogelijkheid om via XSLT-scripts een XML-document om te zetten naar HTML zodat zo'n document gepresenteerd kan worden in de browser. Zo kan men dezelfde XML-gegevens via verschillende XSLT-scripts op verschillende manieren presenteren, afhankelijk van de toepassing.

XSLT-scripts worden ook wel stylesheets genoemd, niet te verwarren met CSS-stylesheets, die alleen bedoeld zijn voor het vastleggen van een speciale HTML-opmaak. In XSLT-stylesheets kan per XML-element vastgelegd worden wat er met de inhoud van zo'n element gedaan moet worden. Dit gaat dus verder dan alleen opmaak.

later stadium naar andere locaties gekopieerd zou kunnen worden.

Voor de centrale index werd besloten om het Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) te gebruiken voor het verzamelen van metadata, omdat dit gezien werd als de meest aangewezen standaard voor dit doel. Voor de zoektoegang tot de centrale index via http/XML zou een eigen protocol gespecificeerd worden. Hierbij werd uitgegaan van zoeken via URL's met respon-

sies geheel in XML. Zoeken via URL's is de meest logische vorm binnen een web-omgeving. Responsies in XML bieden de mogelijkheid de gegevens via een scripttaal te bewerken.

Tevens werd vastgesteld dat we ons in TEL niet tot één centrale index hoefden te beperken, maar dat het ook mogelijk zou zijn dat meerdere indexen via http/XML toegankelijk zijn. Vandaar dat het zinvol is om zowel voor de URL, waarmee gezocht wordt, als voor de XML-respons standaarden te specificeren, die eventueel ook buiten TEL bruikbaar zouden zijn. De reden om hiervoor niet Z39.50 te gebruiken is vooral de complexiteit van Z39.50 en het feit dat het pure gebruik van http beter aansluit bij webapplicaties.

Aardig in dit verband is dat in 1999 dit principe al in Informatie Professional voorspeld werd in het artikel: 'Z39.50; een gepasseerd station'.² Daarin werd geopperd dat de huidige standaard voor search en retrieval, namelijk Z39.50, vervangen zou moeten worden door standaard URL's met standaard responsies in XML.

Toen de specificatie van het eigen http/XML-protocol een feit was en aan de implementatie was begonnen, kwam de Z39.50 Implementers Group – de organisatie die zich met de ontwikkeling en het onderhoud van het Z39.50-protocol bezighoudt – met een initiatief voor een laagdrempelig alternatief voor Z39.50, namelijk SRW: Search and Retrieve via the Web.

Hierin worden twee toegangsmechanismen gebruikt: SRU (Search and Retrieve via URLs) en een op SOAP (Simple Object Access Protocol) gebaseerde toegang die de naam SRW bleef houden. De reden om met dit alternatief voor Z39.50 te komen was dezelfde als geschetst werd in het eerder genoemde artikel: Z39.50 is te ingewikkeld en sluit niet aan bij de moderne op het web gebaseerde standaarden. SRU leek veel op het protocol dat in het kader van TEL gedefinieerd was en besloten werd om voor TEL de SRU-standaard te volgen. Hiermee was de European Library ook meteen een van de eerste die dit protocol, althans een eerste versie, implementeerde.

SRU in een NOTENDOP

Een eenvoudige SRU-opdracht heeft de vorm van een webadres en ziet er uit als:

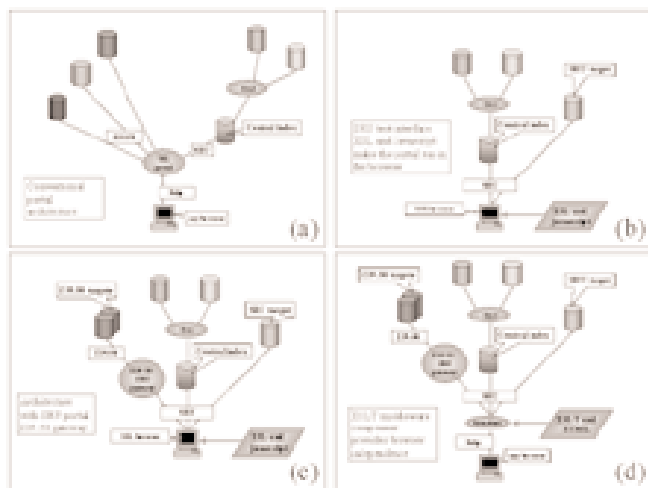
```
http://www.kb.nl/cgi-zoek/srw.pl?query=multatuli&maximum-
Records=10&recordSchema=DC
```

Voor het vraagteken staat de service waar de opdracht heen gestuurd wordt en achter het vraagteken staat de opdracht in de vorm van parameters met een waarde. Naast de parameters in het voorbeeld zijn er onder andere nog parameters als het startrecord, sorteersleutels en de naam van een XSL-stylesheet voor presentatie.

Een voor de leesbaarheid sterk vereenvoudigde XML-respons ziet eruit als:

```
<searchRetrieveResponse>
  <numberOfRecords>1</numberOfRecords>
  <records>
    <record>
      <recordPosition>1</recordPosition>
      <recordData>
        <title>Max Havelaar</title>
        <creator>Multatuli</creator>
      </recordData>
    </record>
  </records>
</searchRetrieveResponse>
```

Door het gebruik van een URL als zoekvraag en XML als respons is het zoeken en presenteren van records eenvoudig te realiseren als webapplicatie.



Figuur 1. Ontwikkeling van de architectuur van de European Library

In figuur 1a is de architectuur aangegeven zoals die op basis van de hier geschetste ideeën en in dit stadium van het project verwacht werd. Een centrale portal geeft toegang tot gedistribueerde bestanden via Z39.50 en tot een centrale index via SRU. Voor dit doel moest de centrale index voorzien worden van een SRU-toegang. Om deze SRU-toegang te testen werd een testinterface ontwikkeld, die gebruik maakte van de Internet Explorer-functionaliteit om XML om te zetten naar HTML met behulp van XSL-stylesheets. Tevens werd de mogelijkheid opgehouden dat er meer SRU-databases zouden zijn en werd deze test-interface zo ontwikkeld dat gelijktijdig gedistribueerd zoeken in verschillende SRU-databases mogelijk zou zijn. Het resultaat was dus een SRU-portal vergelijkbaar met commerciële Z39.50-portals, maar dan draaiend in de browser. Dit is geschetst in figuur 1b. De kracht van dit principe is dat er geen centrale verwerkingscomponent nodig is; de portal is niet meer dan een XSL-pagina met HTML en Javascript. In tegenstelling tot commerciële portals hebben we hierbij volledige controle over de gebruikersinterface en wordt integratie met andere web-services ongekend eenvoudig.

Zoals verderop beschreven was deze SRU-testinterface of testportal ook zeer goed bruikbaar voor de ontwikkeling van het te gebruiken metadatamodel. Ook de schaalbaarheid van dit concept was vele malen beter dan bij commerciële Z39.50-portals het geval is, omdat de verwerking gespreid plaatsvindt in de browsers van de gebruikers en dus niet één centraal systeem belast. Daarnaast biedt ook de personalisering van de interface ongekende flexibiliteit: met een stylesheet afkomstig van een willekeurige aanbieder of ontwerper naar keuze, gecombineerd met een lijst van collecties naar keuze, afkomstig van andere servers, kan iedereen zijn eigen portal creëren. De scheiding van data en presentatie maakt de ontwikkeling van nieuwe functionaliteit zeer goedkoop.

Ondanks de vele voordelen van dit concept ten opzichte van Z39.50, konden we niet verwachten dat veel collecties al binnen de termijn van het project via SRU toegankelijk

zouden worden. Echter, de inzet van een gateway die SRU-zoekopdrachten omzet naar Z39.50-zoekopdrachten en die de Z39.50-respons omzet naar een SRU-respons, bracht een omwenteling teweeg. In figuur 1c is deze nieuwe architectuur geschetst. Door de inzet van een gateway werden alle Z39.50-bronnen in de European Library in één keer via de test-SRU-portal toegankelijk en was er geen reden meer om naar een commerciële portal uit te zien. Het project kon zich verder richten op de SRU-portal in combinatie met één of meer SRU-Z39.50-gateways.

De software voor deze gateway wordt door de European Library vrijelijk ter beschikking gesteld. Technische details over de gateway zijn te vinden op <http://herbie.bl.uk:9080/>. Omdat Z39.50-servers niet altijd helemaal conform de regels zijn geïmplementeerd, is het de bedoeling dat data-aanbieders die hun data via SRU toegankelijk willen maken, hun eigen gateway inrichten. Voor Z39.50-servers die niet via SRU toegankelijk zijn – en waar TEL geen controle over heeft – biedt TEL voorlopig ook een generieke centrale Z39.50-gateway aan. De verwachting is echter dat de wereld van “search and retrieval” massaal het SRU-protocol zal adopteren vanwege de laagdrempelige implementatie en ongekende nieuwe mogelijkheden.

Het hele concept is vooral aantrekkelijk door het feit dat alle verwerking op het werkstation van de gebruiker plaatsvindt, waar toch voldoende verwerkingscapaciteit aanwezig is. De browser-omgeving van de gebruiker moet voor de toepassing van dit concept wel aan enige voorwaarden voldoen. Zo moet de browser XSLT-transformaties toestaan en in sommige gevallen is het nodig om de beveiliging van de browser iets minder streng in te stellen. Dit laatste heeft te maken met het feit dat Internet Explorer om toestemming vraagt om data, komend van één systeem, te presenteren met de XSLT-stylesheets van een ander systeem. In de European Library is dit echter ondervangen door het mogelijk te maken zowel de XML als de XSLT via hetzelfde systeem (een proxyserver) op te vragen.

Voor gebruikers die niet over een browser met volledige functionaliteit beschikken, is een extra component nodig die in figuur 1d geschetst is. Dit is een centrale service die XML transformeert naar HTML met XSLT-stylesheets. Hierbij verliezen we weliswaar het voordeel van de schaalbaarheid, maar dit beperkt zich tot de gebruikers van deze architectuur (een kleine groep). →

DUBLIN CORE

Dublin Core is een internationaal vastgelegde standaard set van vijftien elementen voor het beschrijven van objecten. Voorbeelden van die elementen zijn title, creator en identifier. Voor veel applicaties bieden deze vijftien elementen niet voldoende precisie. Daarom zijn ook nog eens meer dan veertig termen toegevoegd (qualifiers) als verfijning of coderingsschema bij de bestaande set van vijftien elementen. Bijvoorbeeld “alternativeTitle” is een verfijning van “title”, en “URI” is een coderingsschema van “identifier”. Het idee hiërarchisch is een set van veldnamen te hebben, waarvan de betekenis internationaal is vastgelegd, zodat deze velden altijd een herkenbaar label hebben.

Figuur 2. Schets van de relatie tussen functies en metadata

Het voordeel van het SRU-concept is dat we ons op een spectrum van verschillende doelgroepen kunnen richten: van gebruikers met een werkstation met volledige functionaliteit, tot gebruikers met een “mobile device”.

Deze ontwikkeling heeft het TEL-projectteam doen besluiten om niet meer naar een commerciële portal uit te zien, maar dat TEL zich helemaal zal richten op een SRU-portal in combinatie met een SRU-Z39.50-gateway. Dit advies is door de TEL steering committee overgenomen en verwacht wordt dat eind 2004 een op SRU gebaseerde operationele European Library-dienst van start zal gaan.

ONTWIKKELING VAN HET METADATAMODEL

Gelijktijdig met de ontwikkeling van de testportal werd ook gewerkt aan de ontwikkeling van een metadatamodel. Al in het begin van het project is besloten om XML als record syntax te gaan gebruiken voor zowel de Z39.50-omgeving als de http/XML-omgeving.

Als syntax voor bibliografische records is XML de meest logische keuze, vanwege de algemene toepasbaarheid en het feit dat er veel hulpmiddelen ter beschikking staan om XML te verwerken. De grote vlucht die het gebruik van XML in de wereld om ons heen heeft genomen rechtvaardigt deze keuze ten volle.

Dan blijft over het datamodel of recordschema. Het gaat hier om de keuze van velden, veldnamen en de structuur van de velden. Onder de deelnemers van het project waren heel verschillende formaten in gebruik. De werkgroep die zich bezighield met de ontwikkeling van het TEL-datamodel koos voor een op Dublin Core gebaseerd datamodel, omdat dit zeer algemeen bruikbaar is. Het was echter de verwachting dat dit model te mager zou zijn voor de functionaliteit die we in TEL mogelijk zouden willen realiseren. Het Library Application Profile – zie kader op p. 35 – is daarom gekozen als uitgangspunt voor het TEL-datamodel.

Besloten werd te onderzoeken of dit model voldoende was voor de functionaliteit die de TEL-service zou moeten gaan bieden. Om met de testomgevingen van start te kunnen gaan, werd Simple Dublin Core voorlopig als recordschema gebruikt, met het idee dat dit gedurende het project geleidelijk kon worden bijgesteld.

De toepasbaarheid van het Library Application Profile binnen de TEL-portal is onderzocht door de metadavelden

uit te zetten tegen gewenste functionaliteit en dan te bezien of er hiaten waren en er eventueel een TEL Application Profile gerealiseerd zou moeten worden. Er werd bekeken welke metadata nodig waren voor allerlei denkbare functies en omgekeerd werd gekeken welke functionaliteit mogelijk was op basis van aanwezigheid van specifieke metadata. Dit idee is voor slechts een paar velden geïllustreerd in figuur 2.

Het gaat erom dat bij ieder veld de portal aan de gebruiker specifieke functies moet kunnen aanbieden. Dit kan variëren van gewoon presenteren tot andere specifieke functies, zoals het aanbieden van een link of het vertalen van velden. Er werd hierbij naar de verschillende combinaties gekeken, maar ook naar welke velden nodig zijn voor een bepaalde functionaliteit en omgekeerd welke functionaliteit de portal kan bieden als een specifiek veld voorkomt.

De testportal, die in het kader van SRU ontwikkeld was, kon hierbij zeer goed gebruikt worden om de relatie tussen door de portal ontvangen metadata en aan de gebruiker geboden functionaliteit te testen of te demonstreren. Aan de hand van figuur 3 wordt dit verder geïllustreerd.

Op basis van deze exercitie bleek er inderdaad behoefte te zijn aan nieuwe velden, vooral op het gebied van linking en navigatie, wat resulteerde in een TEL Applicatie Profiel. Omdat ook op het gebied van collectiebeschrijvingen werd vastgelegd welke velden gebruikt zouden worden, werd tevens een applicatie profiel voor collectiebeschrijvingen gemaakt.

Tegelijkertijd ontstond het besef dat de huidige functionele eisen voor TEL snel verouderd zouden kunnen raken, zodat TEL als operationele dienst niet stil zou mogen blijven staan. Er was behoefte aan een mechanisme om het TEL Applicatie Profiel gecontroleerd te laten doorontwikkelen, zodat het eenvoudig zou zijn met nieuwe functionele wensen mee te groeien. Nieuwe collecties kunnen nieuwe functionele eisen met zich meebrengen, die soms alleen met behulp van nieuwe velden mogelijk zijn.

Het mechanisme hiervoor werd gevonden in de TEL metadata-registry. Dit is de combinatie van een bestand met de karakteristieken van metadata-elementen en bijbehorende hulpmiddelen. In deze registry worden karakteristieken opgenomen van velden die in gebruik zijn, maar ook van velden die voorgesteld en/of verworpen zijn. Dit geeft data-aanbieders de mogelijkheid te zien welke velden beschik-

Figuur 3. Schermafbeelding van een record met functies die bij een specifiek veld aangeboden worden

APPLICATIE PROFIELEN

Het concept van applicatie profielen komt voort uit de Dublin Core organisatie (DCMI). Een applicatie profiel bestaat uit velden uit verschillende sets van elementen uit specifieke applicatiegebieden, waarbij voor ieder veld een aantal karakteristieken gedefinieerd is, zoals de betekenis, de herhaalbaarheid, de best practice, etcetera. Een voorbeeld is het Library Application Profile dat specifiek gericht is op bibliotheekomgevingen en velden bevat uit onder andere de Dublin Core elementset en de Dublin Core Qualifier elementset. Het idee achter een applicatie profiel is dat velden, die al voor een bepaald applicatiegebied gedefinieerd en vastgelegd zijn, ook in andere gebieden gebruikt kunnen worden met dezelfde betekenis. In een applicatie profiel wordt dus ook per veld vastgelegd uit welke set van elementen een veld afkomstig is.

baar zijn, hoe ze gebruikt moeten worden, wat mogelijke alternatieven zijn en het biedt eventueel de mogelijkheid om zelf voorstellen tot nieuwe velden te doen. Uit deze metadata-registry kunnen dan applicatieprofielen voor verschillende doeleinden gegenereerd worden. De belangrijkste daarvan in TEL zijn het TEL Applicatie Profiel voor objecten en voor collectiebeschrijvingen.

De TEL metadata-registry bestaat uit een XML-file met allerlei karakteristieken van een term, zoals: is het veld herhaalbaar, is het verplicht, uit welk applicatiegebied komt het voort, wat zijn de labelvertalingen etcetera. Met behulp van diverse XSLT-stylesheets kunnen uit de metadata-registry allerlei producten, zoals bijvoorbeeld applicatieprofielen, gegenereerd worden.

Het principe van XSLT-stylesheets is dat in een scripttaal wordt aangegeven op welke plaats – bijvoorbeeld in een HTML-bestand – een specifiek XML-veld ingevuld moet worden. Elk bestand dat uit tekst bestaat en waarin op bepaalde plekken een XML-veld ingevuld moet worden, kan geschreven worden als stylesheet. Zo kan bijvoorbeeld een HTML-involformulier voor collectiebeschrijvingen geschreven worden als stylesheet, waarbij de velden en veldnamen afkomstig zijn uit een XML-bestand. Een ander voorbeeld is het gebruik van een stylesheet om een stuk code te genereren waarin voor elke veld een vertaling van het label wordt gegenereerd.

DE COMBINATIE VAN SRU EN TEL METADATA

In de TEL-portal zijn beide ontwikkelingen als volgt samengebracht. Een HTML-pagina met Javascript geschreven als XSL-stylesheet en een lijst met collectiebeschrijvingen in XML conform het TEL-applicatieprofiel voor collecties worden in de browser getransformeerd tot de TEL SRU-portal. De SRU-portal zet de zoekvragen van de gebruiker om naar SRU-opdrachten en verstuurt die naar de diverse zoekservices. De resultaten komen terug in XML conform het TEL-applicatieprofiel voor objecten en met een tweede stylesheet wordt een presentatie gegenereerd. In dit stylesheet is gespecificeerd wat er met elke metadata veld gedaan moet worden. En omdat de metadata van alle deelnemers aan dit applicatieprofiel voldoen, geldt dit voor alle collecties.

Een paar voorbeelden van functionaliteit die we kunnen bieden op basis van het voorkomen van bepaalde velden:

- Titel of samenvatting: indien deze velden voorkomen, kan een vertaalservice aangeboden worden om ze met één muisklik te vertalen.
- Identifier met URI of URN als coderingsschema: op basis hiervan kan een link naar het beschreven object aangeboden worden.
- Identifier met OpenURL als coderingsschema biedt de mogelijkheid tot Open-Linking.
- Identifier met SRUbaseURL als coderingsschema biedt de mogelijkheid om een collectiebeschrijving om te zetten naar een doelbestand voor nieuwe zoekacties! Dit is een unieke faciliteit die in de volgende paragraaf verder beschreven zal worden.
- IsPartOf biedt de mogelijkheid te navigeren naar een “parent” object, bijvoorbeeld van een artikelbeschrijving naar de beschrijving van het bijbehorende tijdschrift.
- Een onderwerpsveld biedt de mogelijkheid van nieuwe zoekacties in de database met de inhoud van dat veld. Ook een “multi-lingual subject search” op basis van specifieke onderwerpsontsluiting behoort tot de mogelijke functies.

Dit is slechts een greep uit de mogelijkheden. In figuur 3 is een voorbeeld te zien van een scherm waarin de gebruiker door zijn muis op een bepaald veld te zetten de functies te zien krijgt die met dit veld mogelijk zijn. Elk veld kan zo zijn specifieke functies krijgen.

Met de muis op het LCSH onderwerpsveld verschijnt een keuzemenu met de functies die voor dit veld mogelijk zijn, zoals verschillende soorten nieuwe zoekacties met de inhoud van dat veld als zoekvraag, of het opvragen van een vertaling van de inhoud. Zet de gebruiker de cursor op andere velden dan wordt een menu met andere functies getoond.

Het totaal aan functionaliteit is vastgelegd in het stylesheet, waarmee de browser de van de zoekservices afkomstige XML verwerkt. De stylesheets komen in dit geval van een centrale TEL-server, maar kunnen ook van andere servers en zelfs van de gebruiker afkomstig zijn. De verdere verwerking vindt echter geheel plaats in de browser: de portal draait in de browser. Omdat de TEL-portal niets meer is dan een stylesheet met Javascript (de programmeertaal voor webpagina's) vindt er dus ook niets achter de schermen plaats.

De kracht van het SRU-concept heeft te maken met de eenvoud waarmee bepaalde functionaliteit verwezenlijkt kan worden, zonder dat dit de centrale TEL-server belast en waarbij de gebruiker zelf zeggenschap krijgt over zijn zoek-interface. Het concept is toegankelijk en bruikbaar voor iedereen: benadert men de TEL-portal zoals elke andere webpagina, dan heeft men meteen een portal op zijn werkstation. Door het kiezen van het juiste stylesheet kan men als het ware zijn eigen portal creëren.

Commerciële portals wekken de indruk van personalisering door het gebruik van de prefix “my” voor sommige opties, bijvoorbeeld “myLibrary” en maken het mogelijk wat persoonlijke instellingen te bewaren. Het SRU-TEL-concept gaat heel veel verder dan dat. Persoonlijke presentaties en dynamische links naar specifieke diensten van eigen voorkeur (bijvoorbeeld OpenURL) zijn eenvoudig te

realiseren zonder dat de gebruiker daarvoor een programmeur hoeft te zijn, omdat het te verwachten is dat de gebruiker kan kiezen uit een groot aanbod.

COLLECTIES IN THE EUROPEAN LIBRARY

De collecties in TEL zijn beschreven in XML conform het applicatieprofiel voor objecten en deze beschrijvingen zijn via SRU toegankelijk zoals alle andere objectbeschrijvingen. Het bijzondere van de TEL-portal is dat deze het resultaat is van een transformatie van die XML-collectiebeschrijvingen door middel van een XSL-stylesheet. Deze beschrijvingen kunnen ook het resultaat zijn van een zoekactie. Zodra collectiebeschrijvingen via de portal gevonden worden, kunnen deze beschrijvingen via datzelfde XSL-stylesheet getransformeerd worden naar een nieuw portal-scherm met de bijbehorende collecties als doelbestanden voor nieuwe zoekacties.

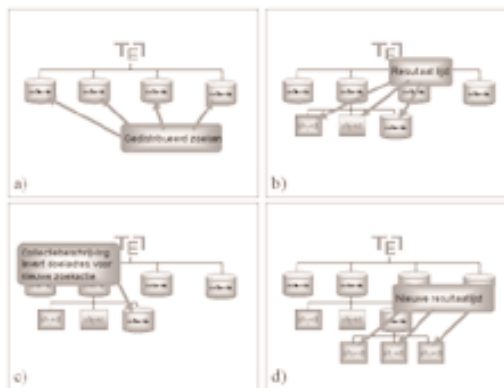
Dit concept is geschetst in figuur 4. Voordelen hiervan zijn:

- indien er veel collecties beschikbaar zijn, kan men niet meer verwachten dat de gebruiker een selectie maakt door in een lange lijst aan te vinken welke daarvan doorzocht moeten worden. In deze situatie is het handig als de selectie verkregen kan worden door een zoekactie met bepaalde criteria;
- de gebruiker is niet afhankelijk van de portal voor wat betreft de te doorzoeken collecties. Elke deelnemer in TEL kan via de TEL-portal een eigen lijst van collecties aanbieden. Het zoeken wordt via deze mogelijkheid vergelijkbaar met hyperlinks in webpagina's: een zoekactie kan nieuwe webadressen opleveren die weer een doorzoekbaar bestand vormen. Een vereiste hiervoor is dat de collectiebeschrijving als zodanig herkenbaar is. In TEL wordt dat gerealiseerd door de aanwezigheid van een "basis-URL" in de collectiebeschrijving. Door deze collectiebeschrijvingen te laten voldoen aan het TEL-applicatieprofiel voor collectiebeschrijvingen, kan de vindbaarheid van collecties geoptimaliseerd worden. Het maakt niet uit waar de records met collectiebeschrijvingen vandaan komen, om omgezet te kunnen worden naar een lijst nieuwe doelbestanden. Dit mag ook een persoonlijke lijst van een gebruiker zijn op zijn werkstation of beschikbaar via een server. Ook hier is dus vergaande personalisering mogelijk.

DE TOEKOMST: FLEXIBELE BENADERING

Een belangrijk aspect van het concept is dat de portal "metadata gedreven" is, dat wil zeggen dat voor elk veld in een stylesheet wordt vastgelegd wat daarmee moet gebeuren. Als een aanbieder van metadata een nieuw onbekend veld toevoegt, zal daar niet automatisch iets mee gebeuren. De aanbieder van de portal kan echter besluiten de stylesheets aan te passen om de bij het nieuwe veld behorende functionaliteit te bieden. Ook de gebruiker zelf kan besluiten een stylesheet van elders te gebruiken, waarin deze functionaliteit al geboden wordt.

Deze groeimogelijkheden kunnen optimaal benut worden, door bij het opvragen van records aan te geven, dat naast Dublin Core-velden ook uitbreidingen op de Dublin Core metadata elementset toegestaan zijn. Hiervoor zal gebruik gemaakt gaan worden van het Dublin Core eXtended schema (DCX), als verzamelnaam voor beschrijvingen die in principe uit Dublin Core-velden bestaan, maar ook nieuwe onbekende velden kunnen bevatten.



Figuur 2. Hoe het vinden van een collectiebeschrijving kan leiden tot zoeken in een nieuw doelbestand

Hiermee wordt het mogelijk om nieuwe collecties te kunnen raadplegen zonder a-priori op de hoogte te zijn van de exacte recordschema's of applicatie profielen behorende bij die nieuwe collecties. Aanbieders van metadata worden hierbij aangemoedigd nieuwe velden te registreren in de TEL metadata-registry om de karakteristieken voor anderen zichtbaar te maken en zodoende op nieuwe functionaliteit te kunnen inspelen.

TOEGANKELIJK CONCEPT

Het concept van een op SRU gebaseerde portal in combinatie met een metadatamodel op basis van Dublin Core applicatie profielen, biedt op de eerste plaats een portal die de aanschaf van een commerciële portal overbodig kan maken. Het concept is eenvoudig toe te passen en voor iedereen toegankelijk en beschikbaar.

Instellingen en gebruikers die gelijktijdige toegang willen tot gedistribueerde bestanden doen er goed aan de mogelijkheden van SRU goed te bekijken. Gaat het om toegang tot Z39.50-bestanden, dan kan bekeken worden of de SRU-Z39.50-gateway hiervoor bruikbaar is. Voor nieuwe bestanden waarvoor nog een toegangsmechanisme gecreëerd moet worden, is het raadzaam direct op het SRU-protocol over te stappen.

Voor zowel gebruikers als instellingen is het voordeel van SRU dat men – op termijn – niet meer afhankelijk is van centrale software. Heeft men een browser, dan heeft men een portal!

Aanbieders kunnen zich beperken tot het aanbieden van een lijst van collectiebeschrijvingen conform het TEL-applicatieprofiel voor collectiebeschrijvingen en de SRU-portal doet de rest. Gebruikers kunnen een eigen keuze maken uit aangeboden collecties en aangeboden stylesheets voor presentatie van de metadata en kunnen daarmee zelf de functionaliteit bepalen.

Noten

1. Informatie Professional 5/2001, nr. 5, p. 28-31.
2. 'Z39.50; een gepasseerd station', Informatie Professional 10/1999, p. 15.

Nadere informatie

Website van de European Library: www.europeanlibrary.org
 TEL Portal: <http://krait.kb.nl/coop/tel/portal>
 Z39.50-SRU-gateway: <http://herbie.bl.uk:9080/>

Theo van Veen is projectleider afdeling Research & Development, Innovatieve Projecten van de Koninklijke Bibliotheek. In het TEL project was hij leider van het metadata werkpakket en nam deel in het werkpakket rond de ontwikkeling van SRU portal.