

Web services til med udgangspunkt i katalogen

Adam Dickmeiss
Index Data

Overblik

- Typer af services
- Informationssøgning generelt
- Kort om A9 OpenSearch
- Gennemgang af SRW/U.

Service typer

- Informations søgning
- Bestilling (ILL)
- Administration (Opdatering m.m.)
- Informationsdeling (OAI, RSS)

Informationssøgning

- SRW/U
- A9 OpenSearch
- Google
- Amazon
- M.fl.

Egenskaber for søgeservices

- Udvælgelse af data
- Søgning med søgeforespørgsel (query)
- Søgeresultat / fejlhåndtering
- Poster
- Evt. skan
- Evt. sortering

A9 OpenSearch

- Udviklet af Amazon
- Beskrivelse af ressource via XML (OpenSearch Description)
- Udvælgelse og forespørgsel via URL (REST)
- Resultater for søgninger er poster af typen OpenSearch RSS (afledning af RSS 2.0)
- Integration i top, men .. poster og forespørgsler er mest egende mod almindelige web ressourcer.

SRW/U

Search/Retrieve Web Service

- Er to protokoller: SRW og SRU.
- Vedligeholdt af library of Congress.
- Deltagelser fra OCLC, Index Data, PICA
- Engagerede enkeltpersoner.

SRW/U Mål

- Erstatning for Z39.50
- Integration: webservice (og XML)
- Forenkling af forespørgsler
- Løsning af “explain”-problemet.

Z39.50 Erstatningen

- Søgninger med generelle søgeveje og angivelse af søgeattributter for fx trunkering m.m.
- Fleksibel postmodel, således at nationale/lokale postmodeller kan tilgodeses.

Forskelle fra Z39.50

- Tilstandsløs som udgangspunkt, men server kan evt. navngive et resultatsæt.
- Sortering altid i kombination med søgning.
- Ingen databaser.
- Ingen adgangskontrol i SRW/SRU.
- Ingen extended services – endnu!
- Kun et sprog til forespørgsler.
- Ingen ISO2709 poster!

SRW kontra SRU

- SRW benytter SOAP 1.1 og 1.2.
 - Beskrives ved brug af WSDL og XML Schema.
- SRU er en standard URL – altså REST.
 - Kan også beskrives med en WSDL, men dette benyttes kun sjældent i praksis

WS krav til SRW/U

- Postrepræsentation i XML (SRW+SRU)
- SRU kræver en notation for forespørgsler som kan URL-enkodes.

Forenkling af forespørgsler

- Type-1/RPN (fra Z39.50) har vist sig for svær at implementere.
- Sprog som fx SQL er for maskinnære og desuden ikke egnet til fritekst.
- CCL har samme struktur som Type-1/RPN i Z39.50 men er:
 - nemmere at forstå
 - strengbaseret.
- CQL (Common Query Language) blev født

CQL 1

- CQL består af
 - boolske operatorer and,or,not,prox
 - termer med evt. index+relation tilknyttet
 - gruppering med paranteser ()
- Eksempler:
 - hans
 - title = hans
 - hans and grethe
 - title=hans and grethe
 - (title=hans) and grethe

CQL 2

- Termer med blanktegn og specialtegn, herunder fraser, skal i plinger.
- Men der er ingen reservede ord!
- Eksempler:
 - “the art”
 - title = “the art”
 - and and **FEJL**
 - and and by

CQL 3

- Maskeringsværktøjer:
 - ? : Masker et vilkårligt tegn.
 - * : Masker et vilkårligt antal vilkårlige tegn.
 - ^ : Match start – eller slut på sætning.
- Eksempler:
 - vilk?rlige
 - vilk*e
 - vilkårlige tegn^
 - ^vilkårlige tegn

CQL context set 1

- CQL index angiver en bestemt søgevej.
- Et CQL index er knyttet til et CQL context set.
- Context set og index er således en “unik” størrelse (lokale attributter kan defineres).
- Context set identificeres ved en URI (ligesom namespaces m.m. i XML).

CQL context set 2

- Dublin Core context set:
 - identifier: info:srw/cql-context-set/1/dc-v1.1
 - med indexnavne: title, creator, subject, description, publisher,...
- CQL context set:
 - identifier: info:srw/cql-context-set/1/cql-v1.1
 - med indexnavne: resultSetId, serverChoice, anywhere, allRecords.

CQL context set 3

- I CQL benyttes konstruktionen >prefix=ID for at knytte et prefix til et context set (ID).
- Her knyttes dc til Dublin core set:
 - >dc=info:srw/cql-context-set/1/dc-v1.1
- Nu kan vi søge med
 - >dc=.... dc.title = x

“Explain problemet”

- Explain beskriver vores service
- WSDL beskriver grundstrukturen. Men vi vil have en stabil API!
- Derfor en ny supplerende struktur som beskriver:
 - søgeveje
 - postformater
 - supplerende services
 - beskrivelser
- SRW Explain/ZeeRex blev født

SRW/U Poster

- SRW/U returnerer XML poster.
- Klient angiver ønsket type, fx DublinCore, MARCXML eller MarcXchange.
- Ligeledes angiver klient postindpakningsmetode:
 - xml
 - string

Explain

- Explain er kendt som ZeeRex
- - udtales Ti-rex og står for Z39.50 Explain, Explained and Re-Engineered in XML.
- Ti-Rex er et XML dokument – er beskrevet ved en DTD.
- Eksempel på ZeeRex post:
 - [Explain record for LoC Z39.50 server](#)

SRW/U operationer

- Operationer:
 - explain
 - searchRetrieve
 - scan

SRU eksempler

- Explain:
 - <http://z3950.loc.gov:7090/voyager>
- searchRetrieve - ingen poster:
 - <http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=dinosaur>
- searchRetrieve – 1 post i svar:
 - <http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=dinosaur&maximumRecords=1>
- searchRetrieve – dublin core post:
 - <http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=dinosaur&maximumRecords=1&recordSchema=dc>

SRW/U erfaringer

- CQL er en udfordring for nogle.
- SOAP værktøjer kan ikke alle håndtere SRW.
- Ikke alle servere understøtter både SRW/SRU, hvilket giver (nye) problemer for klienterne.
- De fleste implementerer SRU.
- Bro mellem Z39.50 og SRW/U mulig.
- Modstand mod SRW grundet Z39.50's udbredelse (og nogen succes).

SRW/U i fremtiden

- Leverandører implementerer ikke SRW/U med mindre at det ønskes fra brugere.
- Nogle leverandører vil benytte YAZ Proxy (som giver SRW/U serverfunktionalitet).
- A9 OpenSearch får muligvis stor rolle på internettet generelt.
- SRW/U skal gøres mere tilgængelig! Specifikt bliver LoC's SRW/U hjemmeside lavet i en ny version snart.

SRW/U i drift

- Library of Congress' Z39.50/SRW/U server.
 - <http://www.loc.gov/z3950/lcserver.html>
- British Library's SRU server.
 - <http://herbie.bl.uk:9080/>
- Alle nye servere baseret på Index Data's YAZ toolkit er klar til SRW/U.
 - <http://indexdata.dk/yaz/>
- PICA
- OCLC Research