

ZING:SRU/SRW - Web Search Technologies
<http://www.loc.gov/z3950/agency/zing/>

Charakteristika projektu

Protokol Zing SRU/SRW (Web Search Technologies) je součástí rodiny protokolů Z39.50 International - Next Generation. Tyto všechny vycházejí z protokolu Z39.50, jehož první verze byla vydána v roce 1988 a do dnešního dne má čtyři verze. Cílem původního Z39.50 bylo umožnění interoperability mezi dvěma počítači (systémy) nezávisle na jejich operačních systémech, databázích, dotazovacích systémech a aplikační oblasti. Snahou byla také obecnější abstrakce databázového vyhledávání než je vyhledávání pomocí SQL.

Z39.50 definuje velmi silný standard a nutí systémové komponenty striktně jej dodržovat. Specifikace původního protokolu Z39.50 byla (stále je) ale příliš složitá a bránila tak širšímu používání. Zejména z toho důvodu, a z potřeby realizovat vyhledávání v prostředí webu, vznikla iniciativa Next Generation.

Zing SRU/SRW je protokol, který slouží pro vyhledávání a získávání dat z databází prostřednictvím webu. Využívá k tomu protokoly HTTP, XML, URI, SOAP. Zjednodušuje původní Z39.50 a snaží se zachovat jeho výhody.

Doba řešení

První verze protokolu Z39.50 vznikla již v roce 1988 v současnosti je již ve verzi 4. Snaha zjednodušit tento protokol, zachovat jeho bohatou sémantiku, ale učinit jej více srozumitelným, jednodušším na implementaci a posunout ho do prostředí webu vyústila na konferenci v červnu 2001 v první "předverzi" Zing SRU/SRW. Následovaly konference v říjnu 2001, dubnu 2002, červnu 2002 a srpnu 2003 a v listopadu 2002 vznikla na CNI (Coalition for Networked Information) verze 1.0 protokolu SRW (Search/Retrieve Webservice).

Aktuální stav

Protokol SRU/SRW je aktuálně ve verzi 1.1 a definuje tři základní operace (služby) pro potřeby interoperability - **explain, scan, searchRetrieve** (původní Z39.50 jich definoval celkem jedenáct, ačkoliv implementace všech nebyly povinné). Na adrese [2] je možné nalézt seznam serverů aktuálně implementujících SRU. Jsou to například BIOME (Internet Resources in the Health and Life Sciences), Library of Congress Online Catalog (přístup k více jak 28 mil. záznamům), Oxford university a mnohé další.

V březnu letošního roku se chystá v Holandsku Integration WorkShop, kde se mají mimo jiné probírat témata integrace s Google, openSearch, bezpečnostní omezení protokolů, výběr služeb vhodných ke standardizaci a další.

Cíle projektu

Hlavním cílem projektu je dosáhnout integrace přístupu k různým zdrojům na síti a poskytnout interoperabilitu mezi distribuovanými databázemi za současné nízké ceny infrastruktury a snadnosti přidání nové komponenty interoperability. Dalším cílem je pokrytí oblasti komplexních dotazů na zdroji, jako to dělá například poměrně robustní dotazovací jazyk XQuery nad XML dokumenty, ale při současném zachování jednoduchosti a intuitivnosti použití (podobně jako služba Google).

Popis projektu a jeho výsledky

Na projektu pracují dobrovolní členové tzv. implementátorské skupiny SRU/SRW, na kterou dohlíží řídicí výbor [6], ale do diskuzí k projektu se může zapojit kdokoli.

Oba protokoly definují formát (strukturu) databázového dotazu zasíleného serverům (systémům) vybaveným komponentou SRW/SRU. Základním rozdílem mezi SRU a SRW je zakódování posílaného dotazu. SRW jej balí do xml obálky protokolu SOAP a SRU jej kóduje přímo do URL. Zpět oba dostávají odpověď kódovanou v XML. Existuje ještě varianta SRUP, která funguje skrze POST request protokolu HTTP, ale to je její jediný rozdíl oproti SRU.

SRU/SRW je možné použít spolu se širokým spektrem klientů – od Microsoft .NET platformy, Javy až po jednoduchý JavaScript a XSLT transformace. Oba protokoly používají pro dotazování CQL - Common Query Language. CQL je formální dotazovací jazyk sloužící k získávání informací nad webovými indexy, bibliografickými katalogy a různými datovými kolekcemi. Hlavní výhodou této technologie je “lidská čitelnost” a intuívnost i pro komplexnější dotazy. Je možné zapsat například jednoduchý dotaz typu “birds OR dinosaur” nebo složitější “author exact jones and title =sword”.

Protokol definuje, jak již bylo zmíněno dříve, tři základní operace. Operace **explain** umožňuje získat informace o databázi na dotazovaném serveru, jako je její umístění, popis obsahu, abstraktní přístupové body a serverem podporované vlastnosti komunikačního protokolu SRU/SRW. Je také většinou první operací, která se v rámci protokolu provádí.

Příkladem pro SRU může být:

```
http://srw.cheshire3.org:8080/15r?version=1.1&operation=explain
```

ekvivalentní operace pro SRW by byla

```
<SOAP:Envelope xmlns:SOAP="http://schemas.xmlsoap.org/soap/envelope/">
  <SOAP:Body>
    <SRW:explainRequest xmlns:SRW="http://www.loc.gov/zing/srw/">
      <SRW:version>1.1</SRW:version>
    </SRW:explainRequest>
  </SOAP:Body>
</SOAP:Envelope>
```

výsledky této operace je možné nalézt na [10] stejně jako u následujících příkladů.

Operace **searchRetrieve** umožňuje vlastní vyhledávání pomocí dotazů formulovaných v CQL a následně získávání výsledků v různých formátech specifikovaných v **explain** (DC, Marc, MODS atd.). Například následující searchRetrieve dotaz v SRU vrací metadata zdrojů ve standardu MODS jejichž autorem je pouze “jones” a titul je abecedně za “smith”. Jsou použity parametry *startRecord* = první záznam od kterého se má vracet výsledek, *maximumRecords* = maximální počet výsledků, které nás zajímají a *recordSchema* určující formát výsledku.

```
http://srw.cheshire3.org:8080/15r?operation=searchRetrieve&version=1.1&startRecord=1&query=dc.author exact jones and dc.title>=smith&maximumRecords=10&recordSchema=mods
```

Obdobná operace v SRW by vypadala:

```
<SOAP:Envelope xmlns:SOAP="http://schemas.xmlsoap.org/soap/envelope/">
  <SOAP:Body>
    <SRW:searchRetrieveRequest xmlns:SRW="http://www.loc.gov/zing/srw/">
      <SRW:version>1.1</SRW:version>
      <SRW:query>(dc.author exact "jones" and dc.title >= "smith")</SRW:query>
      <SRW:startRecord>1</SRW:startRecord>
      <SRW:maximumRecords>10</SRW:maximumRecords>
      <SRW:recordSchema>http://www.loc.gov/mods/</SRW:recordSchema>
    </SRW:searchRetreiveRequest>
  </SOAP:Body>
</SOAP:Envelope>
```

Operace **scan** slouží k získávání termínů z indexu databáze pro procházení rejstříku (browse), je možné ji použít také pro vizuální ověření negativní odpovědi searchRetrieve tak, že si necháme

vypsat jestli databáze vůbec podobné záznamy obsahuje. Vlastní operaci definuje parametr *scanClause* který může být CQL dotazem, relací, nebo startovním bodem v rejstříku. Funkce parametrů *maximumTerms* a *responsePosition* jsou podobné jako u *searchRetrieve*. Následující SRU scan operace vrací počty záznamů v nichž figuruje “jones” nebo autor s podobným jménem:

```
http://srw.cheshire3.org:8080/l5r?operation=scan&version=1.1&scanClause=dc.author any jones&maximumTerms=20&responsePosition=1
```

Zhodnocení a přínos

Projekt ZING:SRU/SRW hodnotím jako velmi přínosný.

Snaha o umožnění jednotné interoperability různorodých systémů na základě jednoduchého ale současně silného protokolu, je podle mého názoru velmi vhodná a do budoucna v duchu SOA (Servisně orientované architektury systémů) také nevyhnutelná.

Velmi mne zaujal příklad ze stránek projektu [9], kdy je současně s výsledkem dotazu nad nějakou databází posílán v XML formátu odpovědi také odkaz na formátovací XSL šablonu, která umožní velmi hezky zobrazit tento výsledek na klientu obsahujícím XSLT parser.

Počet reálných nasazení SRU/SRW a zejména některá jména institucí, kde se projekt již používá, svědčí o tom, že tento projekt bude mít zajímavou budoucnost.

Zdroje a odkazy

[1] Krátká historie SRW

http://www.cni.org/tfms/2004a.spring/presentations/CNI_Denenberg_SRW.ppt

[2] Některé ze serverů implementujících SRU.

<http://www.loc.gov/standards/sru/servers.html>

[3] SRW a OpenOffice.

<http://bibliographic.openoffice.org/srw.html>

[4] Index specifikací SRU/SRW

<http://www.loc.gov/standards/sru/specs-index.html>

[5] Developments resources

<http://www.loc.gov/standards/sru/resources-dev.html>

[6] Hlavní členové projektu

<http://www.loc.gov/standards/sru/board.html>

[7] Přednášky předmětu PV070 Digitální knihovny

<http://www.ics.muni.cz/mba/dl-fi05/>

[8] FAQ

<http://srw.cheshire3.org/docs/faq.html>

[9] Příklady použití

<http://www.loc.gov/standards/sru/stories.html>

[10] Příklady práce s protokolem

<http://srw.cheshire3.org/docs/sru-samples.html>

<http://srw.cheshire3.org/docs/srw-samples.html>

[11] Sémantika a syntax operací protokolu SRU

<http://www.loc.gov/standards/sru/index.html>

Metadata v Dublin Core

<i>Dublin Core atribut</i>	<i>Schéma</i>	<i>Hodnota</i>
DC.Title		ZING:SRU/SRW - Web Search Technologies
DC.Creator		Mahdal, Jakub
DC.Creator.Address		xmahdal@fi.muni.cz
DC.Subject		ZING:SRU/SRW
DC.Subject		Web Search Technologies
DC.Subject		Digital library
DC.Data.Created	W3CDTF	2006-01-21
DC.Type		Text
DC.Format	IMT	application/pdf
DC.Language	RFC3066	cze
DC.Source	URL	http://www.cni.org/tfms/2004a.spring/presentations/CNI_Denberg_SRW.ppt
DC.Source	URL	http://www.loc.gov/standards/sru/servers.html
DC.Source	URL	http://bibliographic.openoffice.org/srw.html
DC.Source	URL	http://www.loc.gov/standards/sru/specs-index.html
DC.Source	URL	http://www.loc.gov/standards/sru/resources-dev.html
DC.Source	URL	http://www.loc.gov/standards/sru/board.html
DC.Source	URL	http://www.ics.muni.cz/mba/dl-fi05/
DC.Source	URL	http://srw.cheshire3.org/docs/faq.html
DC.Source	URL	http://www.loc.gov/standards/sru/stories.html