

Претраживање и преузимање библиографских записа

Данијела Боберић-Крстићев,
Универзитет у Новом Саду
Природно-математички факултет
dboberic@uns.ac.rs

Сажетак

Развој библиотечких информационих система у последње време првенствено је окренут крајњем кориснику коме је потребно омогућити да на једноставан начин приступи фондовима различитих библиотека. Постојање интероперабилних система који омогућавају проналажење и размену квалитетних информација електронским путем постало је основа пословања сваке библиотеке. У циљу постизања ове интероперабилности осмишљени су различити библиотечки стандарди као што су, на пример, Z39.50 (*Information Retrieval: Application Service Definition and Protocol Specification*), SRU (*Search and Retrieve URL Service*), OAI-PMH (*The Open Archives Initiative Protocol for Metadata Harvesting*). У овом раду дат је преглед протокола SRU као и упитног језика који се користи у комбинацији са њим.

Кључне речи:

интероперабилност, библиотечки информациони системи, преузимање записа, протоколи.

Увод

Модерни библиотечки системи између осталог имају и следећа два циља: да што више својих услуга путем интернета учине доступним крајњим корисницима и да омогуће сарадњу са другим системима. Да би се ово остварило, потребно је да архитектура библиотечких информационих система буде флексибилна, како би скуп услуга које једна библиотека пружа временом могао да се проширује, али и да сервиси које она нуди другим системима буду стандардизовани. Стога се сервисно-оријентисана архитектура (COA) намеће као природни избор¹. Међутим, да би се ово остварило, потребно је да библиотечки систем не буде монолитан, већ да се састоји од више сервиса који ће моћи независно да се развијају и које ће моћи да користе и неки други системи. Сматра се да ће се све већи број пословних процеса које библиотека извршава аутоматизовати и да ће се развој библиотечких система заснивати на комбинацији већ готових сервиса који енкапсулирају одређену функционалност потребну за рад библиотеке².

Један од сервиса који је неопходан за електронско пословање библиотека је тзв. узајамна каталогизација. Ово је процес обраде библиографске грађе који се

састоји у следећем: библиотекар при обради нове библиографске јединице прво провери да ли је она већ обрађена у некој другој библиотеци у свету која има сервис за узајамну каталогизацију. Ако је пронађе, тада библиотекар електронски запис преузима у своју локалну базу библиографских записа. После тога, остаје му само да у тај запис унесе локацијске податке. На овај начин библиотекари утврде мање времена на сам процес каталогизације, али такође могу да провере квалитет својих библиографских записа са разним светским институцијама које се баве каталогизацијом, као што је, на пример, Конгресна библиотека у Вашингтону.

Да би библиотеке међусобно могле да размењују записе, неопходно је постојање интероперабилних система. Да би се постигла интероперабилност и омогућило претраживање различитих база података доступних преко интернета и преузимање библиографских записа, неопходно је постојање одговарајућих стандарда, односно комуникационих протокола који су независни од конкретног библиотечног информационог система који се користи у одређеној библиотеци.

Стандард Z39.50 (*Information Retrieval: Application Service Definition and Protocol Specification*)³ је један од првих стандарда који је дефинисао скуп сервиса за претраживање и преузимање података. Овај стандард дефинише само комуникацију између клијента и сервера и не

¹ Robert Fox, „The advent of twenty-first century library services”, *OCLC Systems & Services* br. 25, 1 (2009): 8–15.

² Xinya Yang и др., „System architecture of Library 2.0”, *The Electronic Library* br. 27, 2 (2009): 283–291.

³ ANSI/NISO Z39.50-2003, <http://www.loc.gov/z3950/agency/Z39-50-2003.pdf> (преузето 10. 8. 2011).

улази у детаље имплементације самог клијента, односно сервера. Значајан допринос овог стандарда је пре свега у апстрактном моделу који се користи за претраживање и преузимање. Односно, овај модел дефинише апстрактне префиксе за претраживање који не зависе од имплементације конкретне система који пружа услугу претраживања и такође дефинише формат у којем се подаци могу размењивати.

Још један од протокола који се често користи у библиотекарству је протокол OAI-PMH (*The Open Archives Initiative Protocol for Metadata Harvesting*)⁴. Он је намењен за преузимање библиографских записа и складиштење тих података у заједничку архиву. За разлику од протокола SRU (*Search and Retrieve URL Service*)⁵, протоколом OAI-PMH није дефинисан упитни језик и, самим тим, овај стандард не подржава претраживање, већ обезбеђује преузимање одређеног скупа записа које један систем поседује.

У овом раду описан је протокол SRU који се користи у библиотечким информационим системима за међусобну комуникацију два различита система у циљу претраживања и преузимања библиографских записа. Такође, у раду је дат и опис упитног језика који се користи у комбинацији са протоколом SRU.

1. SRU протокол

Један од најчешће коришћених протокола у размени библиотечких записа у протеклих двадесет година био је Z39.50, међутим уочено је да овај протокол поред свих својих предности има и неких недостатака. Због тога су корисници протокола Z39.50 покренули различите иницијативе у циљу да се он учини прихватљивијим већем броју корисника. Један од пројеката који има за циљ да се смањи сложеност протокола Z39.50, али да се очувају његове основне функционалности је протокол SRU (*Search and Retrieve URL Service*).

Протокол SRU покушава да задржи функционалности које су дефинисане протоколом Z39.50, али да омогући његову имплементацију помоћу тренутно актуелних технологија. Као и код протокола Z39.50, да би могао да се користи протокол SRU, потребно је да постоје клијентска и серверска апликација које комуницирају преко HTTP протокола⁶. Клијентска апликација (клијент) своје захтеве серверској апликацији (серверу) шаље или у оквиру GET и POST метода или у виду XML документа који је упакован у SOAP пору-

ку⁷. Размена порука у виду XML документа је једна од основних предности протокола SRU у односу на Z39.50.

Захтев који клијент може да пошаље серверу дефинисани су сервисима. За разлику од стандарда Z39.50 који је дефинисао 15 различитих сервиса, по стандарду SRU су дефинисана само три, пре свега јер је уочено да се само одређен број сервиса дефинисаних Z39.50 стандардом стварно користи у пракси. Сервиси дефинисани SRU стандардом су следећи:

- SearchRetrieve,
- Scan,
- Explain.

1.1. Сервис SearchRetrieve

Сервис *SearchRetrieve* задужен је за претраживање и преузимање података. Клијент шаље поруку серверу која садржи упит и захтева одређене записе. Сервер по пријему поруке процесира упит који је добио од клијента и враћа му поруку која садржи број пронађених записа. Уколико нема записа који одговарају задатом упиту, сервер ће вратити вредност 0 за број погодака.

Подаци које клијент добија од сервера могу бити у форми XML документа, што клијенту омогућава да употребом XSLT трансформација веома једноставно и уз минималну обраду прикаже резултате претраге. Уколико клијент жели да подаци буду у форми XML документа, може да дефинише у захтеву XML шему по којој очекује да му сервер врати записе. Постоји скуп регистрованих XML шема које су прихваћене од Конгресне библиотеке⁸. Клијент приликом слања поруке може да дефинише и максималан број записа које жели да добије од сервера, као и позицију првог записа који ће бити враћен из скупа пронађених записа. На овај начин, клијент може да преузима део по део пронађених записа, а не цео скуп.

Сервер може да имплементира опцију чувања пронађених резултата и тада сваком скупу резултата мора да се додели јединствено име. На овај начин је омогућено да клијент и после одређеног времена може поново да приступи скупу погодака које је раније пронашао.

1.2. Сервис Scan

Овај сервис омогућава клијенту да за одређени индекс по ком се претражује добије његов скуп вредности и евентуално, уколико сервер то подржава, за сваку вредност из индекса може се дати број погодака који би

⁴ Open Archives Initiative Protocol for Metadata Harvesting, <http://www.openarchives.org/OAI/openarchivesprotocol.html> (нпезето 10. 8. 2011).

⁵ Search/Retrieval via URL, <http://www.loc.gov/standards/sru/> (нпезето 10. 8. 2011).

⁶ Hypertext Transfer Protocol -- HTTP/1.1, <http://www.w3.org/Protocols/rfc2616/rfc2616.html> (нпезето 10. 8. 2011).

⁷ Simple Object Access Protocol (SOAP), <http://www.w3.org/TR/soap12-part1/> (нпезето 10. 8. 2011).

⁸ SRU Record Schemas, <http://www.loc.gov/standards/sru/resources/schemas.html> (нпезето 10. 8. 2011).

се добио уколико би се у претрази користио тај индекс са том вредношћу. На пример, овај сервис омогућава да погледамо које све наслове књига имамо у нашој бази података, а да при томе не морамо да знамо конкретан наслов књиге. Сервис *Scan* се обично користи у комбинацији са сервисом *SearchRetrieve* како би се избегло непотребно претраживање и добијање превеликог опсега података.

Порука коју клијент шаље серверу садржи параметар у коме је дефинисан индекс који се претражује. Поред овог параметра, клијент може да дефинише максималан број вредности које клијент може да прихвати.

Сервер враћа поруку која садржи листу вредности које су пронађене у датом индексу и за сваку даје и информацију о броју погодака у којима се посматрана вредност налази за задати индекс.

1.3. Сервис *Explain*

Овај сервис омогућава клијенту да добије информацију о елементима стандарда које сервер подржава, пошто он не мора да имплементира све детаље стандарда. Порука коју клијент шаље је веома једноставна и, поред обавезног параметра који означава верзију протокола, не мора да има ниједан други параметар.

Одговор сервера садржи обавезан параметар рекорд који представља XML документ у складу са спецификацијом *ZeeRex*⁹. Она служи за опис функционалности сервера који подржава протокол *SRU* и клијент може да користи добијену спецификацију сервера за самоконфигурисање корисничког интерфејса. Односно, на тај начин ће спречити корисника да користи неке од елемената стандарда које посматрани сервер не подржава. Увидом у XML документ, клијент који користи *SRU* сервис може да види на пример по којим индексима се може све претраживати, у којим форматима су доступни записи, и томе слично.

2. Упитни језик *CQL*

Упитни језик који користи *SRU* протокол за формирање упита назива се *CQL* (*Contextual Query Language*, Контекстни упитни језик)¹⁰. Он омогућава писање људских читљивих, интуитивних упита. Циљ *CQL*-а је да се постигне комбинација једноставности и интуитивности, али и спектар могућности које нуди упитни језик дефинисан стандардом *Z39.50*.

Упитни језик *CQL* може да се састоји или од једног операнда или од више њих повезаних логичким операторима. У оба случаја може да се наведе и критеријум

сортирања, навођењем кључне речи *sortBy*. Операнд може да се састоји од индекса, релације и терма који се тражи или само од терма.

Упитни језик *CQL* уводи као концепт појам контексни скуп (енг. *Context Set*)¹¹. Сви индекси по којима се може претраживати су груписани у контекстне скупове и ови скупови су регистровани у оквиру Конгресне библиотеке. Увођењем концепта контекстног скупа омогућено је да одређене заједнице, односно корисници, дефинишу своје индексе, релације, модификаторе, без бојазни да ће називи њихових елемената бити идентични елементима из другог скупа. Прецизније, могуће је дефинисати два елемента са истим називом, који припадају различитим контекстним скуповима и самим тим имају различиту семантику. Контекстни скупови имају аналогну улогу као простори имена (енг. *namespaces*) када је у питању XML.

Упитни језик *CQL* дефинише стандардне релације поретка које представљају везу између индекса и терма. Поред стандардних релација могу да се појаве и релације које су дефинисане контекстним скуповима. Оне могу да се модификују употребом једног или више релационих модификатора. Релација и модификатори су одвојени знаком *'/'*.

У наставку, дати су неки од типичних примера упита дефинисаних помоћу упитног језика *CQL*. Овај упитни језик дефинише различите концепте и велики спектар могућности за формирање упита, али стандардом *SRU* није прописано да сервер мора да имплементира све те концепте.

Пример 1:

CQL упит: *dc.title = Hamlet sortBy date/ascending*

Овим упитом се тражи терм *Hamlet* у индексу *title* и дефинисан је услов за сортирање по индексу *date* у растућем редоследу, при чему *ascending* представља модификатор индекса. Индекс *title* је квалификован (има префикс *dc*) и припада контекстном скупу Даблинско језгро (енг. *Dublin Core*)¹².

Пример 2:

CQL упит: *dc.title any "Julija Romeo"*

Ово је пример упита где индекс припада контекстном скупу Даблинско језгро, а релација *any*, пошто је неквалификована, припада контекстном скупу *CQL* и траже се они резултати који у наслову имају или реч Јулија или реч Ромео.

Пример 3:

CQL упит: (*dc.title=Hamlet and dc.creator=Šekspir*) or keyword=Danska

⁹ *Z39.50 Explain, Explained and Re-Engineered in XML*, <http://explain.z3950.org/> (преузето 10. 8. 2011).

¹⁰ *Contextual Query Language*, <http://www.loc.gov/standards/sru/specs/cql.html> (преузето 10. 8. 2011).

¹¹ *Context Sets*, <http://www.loc.gov/standards/sru/resources/context-sets.html> (преузето 10. 8. 2011).

¹² *Dublin Core Context Set*, <http://www.loc.gov/standards/sru/resources/dc-context-set.html> (преузето 10. 8. 2011).

Ово је пример упита у коме је више операнада повезано логичким операторима *and* и *or*. Такође, ово је пример да у једном упиту могу да се комбинују индекси из различитих контекстних скупова (индекс *title* и *creator* припадају контекстном скупу Даблинско језгро док *keyword* припада контекстном скупу CQL).

Закључак

У овом раду описане су основне карактеристике протокола SRU за преузимање библиографских записа који се све више користи у библиотечким информационим системима. Идеја је да протокол SRU замени протокол Z39.50. Међутим, још увек постоји паралелно подржавање оба протокола, јер појединачне библиотеке не

могу тако лако да се одлуче на прелазак на нови протокол док цела библиотечка заједница не почне да подржава комуникацију путем њега, а са друге стране, библиотечку заједницу чине појединачне библиотеке.

Основна предност протокола SRU је пре свега његова једноставна имплементација и постојање готових отворених софтверских решења који имплементирају клијентске, односно серверске стране овог протокола и могу да се интегришу у већ постојећи библиотечки систем. Чињеница да је задржао све добре особине протокола Z39.50, а да је при томе базиран на модерним интернет технологијама, чини га веома погодним да постане стандардни начин за остваривање интероперабилности између различитих библиотечких информационих система.

Литература

1. ANSI/NISO Z39.50-2003. <http://www.loc.gov/z3950/agency/Z39-50-2003.pdf> (преузето 10. 8.2011).
2. Dublin Core Context Set. <http://www.loc.gov/standards/sru/resources/dc-context-set.html>, (преузето 10. 8. 2011).
3. ZeeRex: The Explainable "Explain" Service . <http://explain.z3950.org/> (преузето 10. 8. 2011).
4. Search/Retrieval via URL. <http://www.loc.gov/standards/sru/> (преузето 10. 8. 2011).
5. Simple Object Access Protocol (SOAP). <http://www.w3.org/TR/soap12-part1/> (преузето 10. 8. 2011).
6. SRU Record Schemas. <http://www.loc.gov/standards/sru/resources/schemas.html> (преузето 10. 8. 2011).
7. Fox, Robert. »The advent of twenty-first century library services«. OCLC Systems & Services br.25,1 (2009): 8-15.
8. 6. Hypertext Transfer Protocol -- HTTP/1.1. <http://www.w3.org/Protocols/rfc2616/rfc2616.html> (преузето 10. 8. 2011).
9. Contextual Query Language. <http://www.loc.gov/standards/sru/specs/cql.html> (преузето 10. 8. 2011).
10. Context Sets. <http://www.loc.gov/standards/sru/resources/context-sets.html> (преузето 10. 8. 2011).
11. Yang, Xinya, Qunyi Wei and Xiaodong Peng. »System architecture of Library 2.0«. The Electronic Library br. 27, 2 (2009): 283–291.

Search and Retrieval of Bibliographic Records

Summary

The main subject of this paper belongs to the field of bibliographic records retrieval. One of the services necessary for the functioning of a modern library information system is called the shared cataloguing. It consists of the following: when processing a new bibliographic unit a librarian first has to check whether it has already been explained in another library in the world. If so, the librarian stores its electronic record to the local bibliographic records database of her/his library, and enters the specific location data. The basic idea is to make bibliographic records created by one library available to all the other libraries to download. In this way, librarians spend less time in the process of cataloguing books, and in addition they can also check the quality of their own bibliographic records consulting databases of various international institutions dealing with cataloguing, such as e.g. the Library of Congress in Washington. The process of exchanging records between libraries requires existence of the interoperable systems. In order to achieve interoperability and facilitate the search of various databases available over the Internet and download of bibliographic records, it was necessary to create appropriate standards and communication protocols, independent of the specific library information system used in a particular library. One of those protocols is described in this paper.

Key words:

interoperability, library information systems, bibliographic records retrieval, protocols.