

“Máte to?” - Summon jako základní vyhledávací nástroj NTK

Milan JANÍČEK

Národní technická knihovna, Praha

milan.janicek@techlib.cz

INFORUM 2013: 19. ročník konference o profesionálních informačních zdrojích
Praha, 21. - 22. 5. 2013

Abstrakt

Fondy zpřístupňované dnešními knihovnami už často přesahují množství postihnutelné klasickým katalogem. K fyzickým dokumentům se přidává obrovská masa licencovaných i nelicencovaných elektronických zdrojů, lokální dokumenty institucí v elektronické formě, výstupy digitalizace apod.

Pouhé “zpřístupnění” navíc už dnešním (a zítřejším) uživatelům nestačí, pokud poskytované dokumenty nejsou také snadno nalezitelné. Národní technická knihovna dbá o to, aby její zákazníci dokázali její fondy co nejlépe využít - a to ideálně kdykoli, odkudkoli a s vynaložením co nejmenšího úsilí. I proto patří mezi první české instituce, které do portfolia svých služeb zařadily vyhledávač fungující na principu agregovaného indexu - konkrétně Summon od firmy Serials Solutions.

V příspěvku budou popsána jak nejdůležitější východiska řešení, tak konkrétní zkušenosti knihovny s implementací a provozem vyhledávače.

Proč další systém?

Na otázku proč přidávat ke stávajícím službám další systém lze odpovědět poměrně snadno. Jedním ze základů služeb moderní knihovny je možnost snadno vyhledat (a získat) buď konkrétní dokument, nebo dokument k tématu o němž má uživatel knihovny zájem. Sbírkové zpřístupňované knihovny se mohou rozrůstat nebývale rychle - například titulů časopisů k nimž poskytuje přístup NTK jsou desetitisíce. K co možná nejefektivnějšímu využití zdrojů je ale nutné být schopen je i najít.

Nejlepším technickým prostředkem, umožňujícím přístup k velké šíři zdrojů, byl donedávna metavyhledávač. NTK dosud provozuje jednu z oborových bran - TECH - jejíž podstatnou součástí je metavyhledávač MetaLib. Jako významná bariéra jeho využití našimi běžnými uživateli se ale ukazuje nutnost být k vyhledávání v licencovaných zdrojích autorizován. To je ve srovnání se systémy používanými na webu velká komplikace. Další významnou nevýhodou je

způsob zpracování výsledku z různých zdrojů, kdy se využívá pouze několika málo (např. 30) záznamů z každého zdroje, a také celková nepřehlednost a pomalost systému.

Metavyhledávače mají ale i jednu nespornou výhodu - dotazují se vždy databáze v aktuálním stavu a tak dostávají "nejnovější možné" výsledky.

Předpřipravený index je oproti tomu optimalizován na vyhledávání a proto je schopen rychleji pracovat s větším množstvím dat - jeho chování se blíží tomu, na co jsou zvyklí běžní uživatelé Internetu. Větší péče se musí věnovat přípravě dat, tzn. tomu co do indexu zahrnout. Autentizaci uživatelů není obvykle nutné provádět před vyhledáváním samotným, ale stačí ji vyřešit až když se uživatel pokusí s vyhledaným dokumentem dále pracovat (např. získat plný text, nebo si rezervovat fyzickou knihu).

Do indexu se také často zahrnuje obsah katalogů knihoven a další specifické zdroje. Uživatelům pak můžeme nabídnout pochopitelný a snadno použitelný vstupní bod jak k fyzickým, tak k elektronickým zdrojům knihovny.

Summon v NTK

V březnu 2012 vypsala Národní technická knihovna (NTK) výběrové řízení na systém postavený na technologii agregovaného indexu. Mezi požadavky byly pokrytí elektronických zdrojů, indexace plného textu i metadat, možnost zapojení vlastních zdrojů, ale třeba i existence API pro propojení s dalšími systémy. Výběrového řízení se zúčastnili tři dodavatelé - EBSCO, Multidata a AiP se Summonem od firmy Serials Solutions, který byl nakonec vybrán jako vítězné řešení.

Nová služba přišla do zavedeného prostředí ostatních aplikací NTK. V případě Summonu je důležitá návaznost na link server (SFX od Ex Libris) a na proxy server pro vzdálený přístup k el. zdrojům (EZproxy od OCLC). Napojení na další systémy bude podrobněji popsáno níže.

Implementace

Summon je provozován jako služba - tzv. Software as a Service. V takovém případě nemá knihovna fyzický přístup k serveru. Zároveň to znamená, že odpadá starost o hardware a spravuje se jenom aplikace. Tento typ služeb lze poznat i podle adresy:

<http://techlib.summon.serialssolutions.com/>, nicméně není problém provést přesměrování, třeba jako v našem případě z adresy <http://summon.techlib.cz>.

První kroky

První kroky spočívají ve vyplnění několika implementačních formulářů (pro lokální zdroje a katalog) a nastavení prohledávaných elektronických zdrojů ve webové administraci. Nastavení

lokálních zdrojů je složitější činnost, při níž probíhá analýza poskytovaných dat v Serials Solutions, proto je větší i časová náročnost celého procesu. Nastavení elektronických zdrojů leží ve větší míře na knihovně - jen ona ví (v ideálním případě) jaké zdroje má k dispozici.

Elektronické zdroje

Základním úkolem je nastavení obsahu indexu. Elektronické zdroje se zapínají ve znalostní bázi udržované Serials Solutions buď na úrovni kolekcí / zdrojů, nebo na úrovni konkrétních titulů a jejich přístupnosti. Samotné nastavení lze provést buď ručně, nebo pomocí importu prostřednictvím CSV. Protože se indexuje jednou denně, je důležité zmínit, že se změny promítnou se zpožděním.

Katalog

Katalog lze indexovat dvěma způsoby - buď prostřednictvím pravidelných updatů nahrávaných na FTP server Serials Solutions, nebo automaticky s využitím protokolu OAI-PMH - v našem případě byl využit plný export dat (zhruba 600 000 záznamů) z Alephu a aktualizace.

Nahrání dat do systému jako je Summon často odhalí jejich nedostatky - došlo k tomu samozřejmě i v našem případě. Největším problémem je rozdílný obsah pole SYS v Alephu a 001 v bibliografickém záznamu. Rozdílnost těchto identifikátorů způsobovala problémy při propojování záznamů ze Summonu do Alephu.

Využití facet umožňuje snadno nahlédnout do katalogu novým způsobem - což je ještě umocněno využitím prázdného vyhledávacího řetězce, který vrací jako výsledek kompletní obsah katalogu.

Lokální zdroje

Do indexu jsme zapojili i dva zdroje o jejichž správu se stará NTK - Digitální knihovnu NTK (v systému Kramerius 4) a repozitář Národního úložiště šedé literatury (v systému Invenio). Z obou zdrojů se harvestují záznamy ve formátu Dublin Core prostřednictvím protokolu OAI-PMH, jehož podpora je v obou systémech implementována.

Invenio bylo zapojeno zcela standardně a bez modifikací připojení z naší strany. Pro rozlišení dostupnosti plného textu je rozhodující záznam z Invenia - některé záznamy v tomto repozitáři sice obsahují odkaz do dalších repozitářů (ve kterých se plný text může, nebo nemusí nacházet), to ale není podle metadat z Invenia rozpoznatelné.

U systému Kramerius 4 jsme pak definovali hlavně kritérium "přístupné online". Kramerius

rozlišuje přístupová práva “public” (veřejné) a “private” (neveřejné; obvykle dostupné ve studovnách knihoven). Z hlediska uživatele jsou online jen takové dokumenty, které může okamžitě prohlížet odkudkoli, tzn. s přístupovými právy “public”.

... a vcelku

Summon NTK v tuto chvíli nabízí cca 45 000 000 výsledků - z toho zhruba 600 000 záznamů z katalogu, 100 000 z repozitáře NUŠL a 1 000 z Digitální knihovny NTK.

V kontextu ostatních systémů

Summon je pouze jednou z řady služeb se kterými se náš uživatel na cestě za požadovaným dokumentem setká. Zároveň je ale výchozím bodem na nějž musí být co nejlépe navázány ostatní služby - link server, katalog, proxy server pro vzdálený přístup, elektronické zdroje - a to ideálně takovým způsobem, aby uživatel vůbec nevnímal přechody mezi nimi. Tomu v našem konkrétním případě velmi pomáhá využití Shibbolethu jako autentizačního mechanismu pro všechny naše systémy (single sign-on). Uživatel se tak musí přihlásit jenom jednou a z pohledu ostatních služeb už přihlášen je.

Shrnutí

Se Summonem získali naši uživatelé systém umožňující

- snadno zjistit co je (zprostředkováno knihovnou) k dispozici
 - a ideálně se k tomu i snadno dostat
- snadno plnotextově vyhledávat v EIZ
- zjednodušit přístup do koncových systémů (licencované EIZ, katalog, lokální zdroje)
- možnost nabídnout další služby (prostřednictvím SFX) i když není dostupný plný text
- jednoduše nastavovat vyhledávací dotazy
 - zúžit hledání pomocí facet
 - možnost hledání nad rámec sbírek knihovny
 - je možné navázat služby dodávání dokumentů

Budoucnost

V této chvíli samozřejmě naše práce nekončí. Měli bychom se zaměřit na co nejlepší integraci služeb Summonu do služeb knihovny - přičemž bude potřeba ověřit i starší postupy a procesy. Využití řešení vyhledávače bude jedním z jeho nejdůležitějších stavebních prvků.

