

# System pro stahování bibliografických záznamů z českých a zahraničních zdrojů

**Martin LEDÍNSKÝ**

Univerzita Karlova v Praze  
Ústav výpočetní techniky  
martin.ledinsky@ruk.cuni.cz

INFORUM 2004: 10. konference o profesionálních informačních zdrojích  
Praha, 25. - 27.5. 2004

## *Abstrakt:*

V minulosti mohli zájemci z řad českých knihoven díky projektu Jednotná informační brána využívat funkci, která umožňovala stahování bibliografických záznamů z českých i zahraničních informačních zdrojů, zejména bibliografických databází, které jsou zařazeny jako prohledatelné zdroje v JIB. Tato funkce byla omezena tím, že formát záznamů knihovny, která záznamy stahovala a knihovny, od které se stahovalo, se musel shodovat. Od jara 2004 je spuštěna nová verze modulu, která dosavadní limity bortí. To je zvláště důležité při postupném přechodu českých knihoven z formátu UNIMARC na MARC 21. Díky robustnímu konvertoru vyvinutém v Ústavu výpočetní techniky Univerzity Karlovy v Praze pro potřeby Národní knihovny ČR je možná obousměrná konverze, záznamy je možné v reálném čase konvertovat jak z UNIMARC do MARC 21, tak opačně.

Stahování záznamů se odehrává přímo v prostředí knihovního systému, nezávisle na rozhraní JIB. Podmínkou je, aby knihovní systém disponoval Z39.50 klientem. Stahování záznamů ocení zvláště akvizitěři a katalogizátoři, kterým výrazně zefektivní práci. Tato funkce je k dispozici bezplatně pro všechny české knihovny. Zájemci mohou o informace o nastavení žádat na adrese [jib-support@cuni.cz](mailto:jib-support@cuni.cz).

## *O autorovi:*

Absolvent Vyšší odborné školy informačních služeb, obor Informační management, zaměření Služby knihoven. Nyní studuje 2. ročník bakalářského programu Ústavu informačních studií a knihovnictví Filozofické fakulty Univerzity Karlovy v Praze. Od roku 2002 zaměstnán v Ústavu výpočetní techniky UK, oddělení knihovnických aplikací. Zabývá se knihovními systémy T-Series a Aleph a portálem MetaLib.

## *Text příspěvku:*

Projekt Jednotná informační brána dnes již není nutné představovat. Vedle webového rozhraní JIB, které umožňuje paralelní prohledávání různorodých informačních zdrojů v jednotném prostředí, je možno JIB využívat i pro přebírání bibliografických záznamů. Tato funkce je určena pro české knihovny, konkrétně pro pracovníky knihoven z řad katalogizátorů a akvizitérů. Jaká je myšlenka této služby? To si ukážeme nejprve na příkladu s běžným pracovním postupem při pořízení nového dokumentu v knihovně. Nejprve se vytvoří akviziční záznam, poté, kdy se dokument dostane do knihovny, vytvoří se katalogizační záznam. Jmenný a věcný popis jednoho dokumentu obvykle trvá i několik desítek minut. Přitom mnohé knihovny mají podobný fond. A tak v momentě, kdy se akvizitér či katalogizátor pouští do práce, s největší pravděpodobností již hotový záznam existuje v jiné knihovně, velmi často dokonce v několika knihovnách. Tedy, dělá se množství zbytečné práce, pokud je možné stejně kvalitní záznam, který by jinak katalogizátoři sami vytvořili, již stáhnout hotový. Cílem funkce přebírání záznamů je šetřit limitované finanční i lidské zdroje v českých knihovnách.

Tato služba je pro české knihovny dostupná již od roku 2002. Během dvou let, kdy je služba v provozu, vyvstalo několik úskalí, které bránily širšímu využití. Hlavní úskalí, aktuální zvláště dnes, je v

tom, že záznamy se v bibliografických databázích vyskytují v různých formátech (UNIMARC, MARC 21, SUTRS, MAB ad.). Bez fungující konverze mezi formáty není možné záznamy z některých zdrojů přebírat. Není možné například využít záznam z Library of Congress ve formátu MARC 21, když v naší bibliografické databázi jsou záznamy ve formátu UNIMARC. Do této doby tato nejednotnost znamenala to, že nebylo možné pro přebírání záznamů prostřednictvím JIB využívat řadu kvalitních zahraničních zdrojů.

Částečně už nyní a hlavně v nejbližší budoucnosti by ale bez fungující konverze mezi formáty UNIMARC a MARC 21 nebyla plnohodnotná směnitelnost záznamů ani v českém prostředí. UNIMARC a MARC 21 jsou pro Českou republiku stěžejní formáty. Až na několik výjimek, jak bude patrné dále, používají české knihovny formát UNIMARC. Již několik let je ale jasné, že přechod na formát MARC 21 je nezbytný a nevyhnutelný, není již dlouho aktuální otázka "proč" ale "kdy". Národní knihovna České republiky se v posledních letech na přechod na MARC 21 důkladně připravila, tím zajistila podmínky pro přechod na nový formát pro knihovny v České republice. Samotná NK ČR plánuje migraci na MARC 21 na prázdninové měsíce letošního roku. Již v roce 2003 přešly na MARC 21 zároveň s výměnou knihovního systému knihovny Vysoké školy ekonomické v Praze (CIKS), Masarykovy univerzity v Brně a Vysokého učení technického v Brně. Ostatní české knihovny přechod na MARC 21 plánují a připravují, vedle NK ČR se letos uskuteční přechod u zhruba desítky českých knihoven. Tolik k úskalí s různými formáty bibliografických záznamů.

Méně významný, ačkoli také bránící plnohodnotnému využití služby přebírání záznamů, je problém znakových sad. Bibliografické záznamy mohou být v databázích ukládány v různých znakových sadách – například v UTF-8, CP 1250, CP 852, ISO 8859-2 atp. Při převzetí záznamu uloženého pomocí jedné znakové sady do databáze s jinou znakovou sadou dojde ke ztrátám – speciální znaky (např. znaky s diakritickými znaménkami) budou zobrazeny chybně, protože se v dané znakové sadě nevyskytují.

Aby byla služba přebírání záznamů plnohodnotně využitelná, bylo nezbytné na tato úskalí zareagovat a připravit takové úpravy, které by problémy různých znakových sad a formátů do nejvyšší míry eliminovaly. Modul pro přebírání záznamů využívaný v JIB proto doznal rozsáhlých změn, byl dovybaven funkcionalitou pro volitelnou konverzi mezi formáty s možností konvertování znakových sad. K tomu už scházel jen samotný konvertor s funkčními konverzními pravidly mezi formátem UNIMARC a MARC 21. Konvertor byl vyvinut pro potřeby Národní knihovny České republiky a ve spolupráci s NK ČR na půdě Univerzity Karlovy v Praze, konkrétně na pracovišti Ústavu výpočetní techniky. Robustní konvertor umožňuje konverzi marcovských formátů a je naplněn konverzními pravidly pro oboustrannou konverzi mezi formáty UNIMARC a MARC 21. Konvertovat lze záznamy všech typů dokumentů, funkční je i konverze autoritních záznamů. Stejně tak je možné využít konverze znakových sad. Po zajištění této funkcionality byla nová verze modulu pro přebírání záznamů spuštěna 1. března 2004. V těchto dnech využívá tuto službu téměř dvacet českých knihoven.

### **Jaké zdroje je možno využít pro přebírání záznamů**

Využít je možné jakýkoli zdroj, který je prohledatelný prostřednictvím Jednotné informační brány. Podmínkou je, aby tento zdroj byl připojený prostřednictvím protokolu Z39.50. Pokud jde o zdroj licencovaný, musíte k němu mít zakoupenou licenci. Výjimkou jsou zdroje databázového centra OCLC, které bohužel není možné pro přebírání záznamů využít. Záznamy v prohledatelných zdrojích musí být buď ve formátu UNIMARC nebo MARC 21, v ostatních případech bohužel nemůžeme zajistit jejich konverzi a záznamy byste tak mohli přebírat pouze v původním formátu. Konverze mezi formáty UNIMARC a MARC 21 probíhá v reálném čase, záznamy tedy dostanete již zkonvertované. Je nutné si uvědomit, že ne všechny bibliografické zdroje nabízejí záznamy stejné kvality. Mnoho záznamů se odchyluje od standardu MARC, což se samozřejmě projeví i při práci konvertoru.

### **Jak funguje služba přebírání záznamů**

Pro každou knihovnu, která projeví o službu přebírání záznamů zájem, vytvoříme profil, který sestává ze seznamu zdrojů, řazených podle důležitosti. Váš profil poté figuruje jako databáze externího Z39.50 cíle (s využitím Z39.50 serveru JIB). Vyhledávání prostřednictvím profilu pak probíhá sekvenčně – dotaz se zašle prvnímu zdroji v seznamu, pokud se v něm najdou záznamy, vyhledávání už dále nepokračuje. V opačném případě je dotaz zadán dalšímu zdroji v seznamu, dokud se záznamy nenajdou. Zdroje je tedy třeba řadit podle důležitosti – na první místo seznamu umístit zdroj, který se

strukturou fondu a bohatostí záznamu nejvíce blíží vaší knihovně. Pro každou knihovnu můžeme vytvořit více profilů, aby např. odrážely specifické potřeby jednotlivých oddělení. Katalogizátor nebo akvizitér pak záznamy přebírá například přímo z katalogizačního nebo akvizičního modulu (podle knihovního systému a organizace práce v konkrétní knihovně), bez použití webového rozhraní JIB.

### **Co je potřeba k tomu, aby službu přebírání záznamů mohla využívat i vaše knihovna?**

Služba přebírání záznamů je založena na protokolu Z39.50. Proto je nezbytné, aby váš knihovní systém disponoval nejlépe zabudovaným Z39.50 klientem, je možné ale využívat i externího Z39.50 klienta. Poté již stačí jen vyplnit a odeslat elektronický formulář se žádostí o zřízení profilu na adrese <http://jib-info.cuni.cz/prebirani>. Na základě žádosti vám zašleme informace s nastavením služby pro váš knihovní systém.

### **Tipy pro vyhledávání**

Zdroje do profilu vybírejte pečlivě. Volte ty, se kterými máte dobré zkušenosti a jejichž kvalita záznamů se nejvíce blíží vaší.

Čas, po který se záznamy vyhledávají, je dán hlavně rychlostí vzdálené databáze a také počtem záznamů, které se na dotaz vrátí. Zvláště při vyhledávání v zahraničních zdrojích, které jsou ve světě značně využívány (Library of Congress), je nutno brát v úvahu, že připojení k těmto zdrojům a následně vyhledávání v čase maximálního zatížení zdroje (špička), může trvat několikanásobně déle než vyhledávání mimo špičku nebo v tuzemských zdrojích. Pokud přebíráte záznamy s použitím konverze mezi formáty UNIMARC a MARC 21, zabere určitý čas i konvertování. Proto je třeba zadávat přesné dotazy, na které bude vráceno menší množství relevantních záznamů. Při zadávání příliš obecných dotazů, kterým budou vyhovovat stovky či tisíce záznamů, bude čas nutný ke stažení a zkonvertování trvat neúměrně dlouho.

Pamatujte, prosím, také na to, že komunikace se zdroji probíhá prostřednictvím protokolu Z39.50. Chování zdroje při vyhledávání přes Z39.50 se nemusí stoprocentně shodovat s chováním při vyhledávání v nativním rozhraní zdroje. To je dáno do značné míry tím, že zdroje využívají různé Z39.50 servery od různých producentů. Tyto servery mají různé možnosti, navíc konkrétní instalace Z39.50 serveru a nastavení pro konkrétní zdroje mohou být výrazně odlišné. Ze zkušeností při připojování zdrojů v JIB vyplynulo, že je třeba již sjednocovat chování samotných Z39.50 serverů, aby bylo dosaženo maximální shody při práci se zdrojem v nativním prostředí (např. OPAC) a prostřednictvím protokolu Z39.50. O tomto pojednává další příspěvek ve sborníku - Profil Z39.50 JIB : profil pro jednoduché vyhledávání a stahování záznamů.

Z našich zkušeností při vyhledávání vyplývá, že nejúčinnější způsob vyhledávání je buď podle samotného ISBN anebo, pokud není ISBN k dispozici, podle kombinace údajů autor, název a rok.

Katalogizujete-li dokument, kterému bylo přiděleno některé standardní číslo (ISBN, ISSN, ISMN), vyhledávejte zásadně podle něj. Pokud chcete knihu s číslem ISBN hledat v katalogu Library of Congress nebo Národní knihovny Kanady, zadávejte ISBN bez pomlček. Standardní číslo ISBN (nebo ISSN, ISMN) je jednoznačný identifikátor, a proto vyhledávání podle něj nekombinujte s žádným dalším polem.

Tímto způsobem dostanete přímo konkrétní záznam dokumentu, který držíte v rukou, a nemusíte se probírat množstvím vrácených záznamů.

Doporučujeme vyhledávání kombinací maximálně dvou polí – autor a rok nebo autor a název. U obou těchto případů však může docházet k této komplikaci:

Při vyhledávání pomocí kombinace polí autor a rok můžete narazit na problém, že některé zdroje vůbec nepodporují vyhledávání podle roku. Takový zdroj potom při vyhledávání z více polí spojených logickým operátorem AND (logický součin) v kombinaci s polem rok vrátí 0 záznamů.

## **Závěrem**

Věřím, že vám zlepšení služby přebírání záznamů udělá radost a že ji ve vaší knihovně využijete a bude vám při práci přínosem. Na základě vašich podnětů se dále budeme snažit nabídku zdrojů pro přebírání rozšiřovat. S vašimi případnými dotazy či náměty se obračejte, prosím, na adresu [jib-support@cuni.cz](mailto:jib-support@cuni.cz).

## **Bibliografie:**

1.

STOKLASOVÁ, Bohdana. Přežije formát UNIMARC rok 2003?. Ikaros [online]. 2001, č. 09 [cit. 2004-01-01].

Dostupný na World Wide Web: <<http://www.ikaros.cz/Clanek.asp?ID=200208358>>. ISSN 1212-5075.

2.

LICHTENBERGOVÁ, Edita. MARC21 v české katalogizační politice. Národní knihovna [online]. 2002, č. 2, s. 134-135.

Dostupný na World Wide Web: <<http://full.nkp.cz/nkkr/Nkkr0202/0202134.html>>. ISSN 1214-0678.

3.

LICHTENBERGOVÁ, Edita. MARC 21 a česká katalogizační praxe. Knihovnický zpravodaj Vysočina [online]. 2004, č. 1.

Dostupný na World Wide Web: <[http://www.knihzdar.cz/kzv/2004-01/lichtenbergova\\_marc21.htm](http://www.knihzdar.cz/kzv/2004-01/lichtenbergova_marc21.htm)>. ISSN 1213-8231.