

Nové aspekty zpracování knižních fondů

PhDr. Bohdana STOKLASOVÁ
Národní knihovna ČR, Praha

Úvod

Retrospektivní konverze katalogů v Národní knihovně České republiky (dále jen NK ČR) postoupila ze stádia vývoje a testování technologie a shánění finančních prostředků do stádia realizace. Technologie byla dokončena a odzkoušena, grant Mellonovy nadace (150 000 US\$) stačí na nákup základního vybavení, dotace Ministerstva kultury (4 000 000 Kč) pokryje práce v roce 1995.

V souvislosti se zahájením prací na retrospektivní konverzi hlavních katalogů Národní knihovny (NK) ČR s využitím originální, moderní technologie se velmi často setkávám s otázkou: "Jak je možné, že špičková technologie pro retrospektivní konverzi katalogů, na kterou se dnes jezdí dívat a o jejímž využití uvažuje řada knihoven doma i v zahraničí, vznikla právě v NK ČR, o níž bylo ještě v minulém roce známo pouze to, že má v souvislosti s retrospektivní konverzí katalogů obrovské problémy?" Dalo by se velmi zjednodušeně odpovědět, že moderní, špičková technologie vznikla právě zde mimo jiné díky tomu, že jsme začali až nyní, v době, kdy máme k dispozici dokonalé technické vybavení a softwarový základ, které dříve prostě neexistovaly. Kdybychom začali dříve, začali bychom jinak a i do případného využití moderních nástrojů by se promítlo dědictví minulosti, od něhož bychom se nemohli odpoutat, takže by nás brzdilo. Je to tedy šťastná náhoda, že začínáme právě dnes, "na zelené louce" a s mladými programátory, jimž jsou nejnovější technologie blízké. Ale jak známo, náhoda přeje připraveným. Supermoderní technika by zůstala pouze hromadou železa a špičkové softwarové produkty jen mrtvým základem, pokud by neoslovily tým připravených, pro věc zapálených lidí.

NK ČR byla ještě v minulém roce v oblasti retrospektivní konverze svých katalogů v opravdu zoufalé situaci. V průběhu posledních let zde byly analyzovány ve světě nejrozšířenější metody retrospektivní konverze, kterými jsou metoda **OCLC** a **SAZTEC**. Procento záznamů obsažených v generálním katalogu univerzálního knihovního fondu (dále jen GK UKF), které byly nalezeny v bázi OCLC, bylo nízké. Metoda bude vhodná pro určité části katalogu, nikoli pro katalog jako celek. Metoda SAZTEC se jevila jako vhodnější. Při částce 60 Kč za záznam by však retrospektivní konverze GK UKF v rozsahu 3 000 000 lístků stála 180 000 000 Kč. NK ČR neměla k dispozici ani zlomek této částky a neexistovala reálná naděje na její získání. Retrospektivní konverze katalogu byla zahájena v roce 1992 metodou pro NK ČR finančně přijatelnou - lístky byly připravovány v NK ČR, přepisovány ve Vodních stavbách Tábor a kontrolovány opět v NK ČR. Na NK ČR zůstávaly stále dvě třetiny práce, poměr se ani po letech (navzdory očekávání) nezlepšil. Práce postupovala pomalu, části katalogu byly po dlouhou dobu vyřazeny z provozu. Spolupráce s Vodními stavbami Tábor byla ukončena v červnu 1995. **Všem jmenovaným metodám je společné jedno: katalog lze ve strojem čitelné podobě zpřístupnit až po dokončení celé akce, což zároveň znamená vyplacení celé částky.** S ohledem na silně limitované finanční zdroje nelze počítat se získáním tak vysoké částky dříve než v horizontu zhruba deseti let. "Úzké hrdlo" představuje u všech zmíněných technologií obrovský objem lidské práce, kterou je třeba v souvislosti s retrospektivní konverzí 3 000 000 lístků vykonat a pochopitelně zaplatit. Nebyla naděje na získání finančních prostředků na práci najednou, zřizovatel přislíbil příspěvek po menších částech (řádově několik miliónů - v žádném případě desítek či stovek miliónů - za rok). Existoval příslib grantu od Mellonovy nadace - ovšem nikoli na práci, ale na zakoupení technického vybavení. S blížícím se dokončením výstavby depozitáře NK ČR v Hostivaři nabyla na aktuálnosti nutnost odstěhovat GK UKF do Hostivaře. Podmínkou pro jeho odstěhování však byla existence provozně využitelné (nejen čistě archivní) kopie na náhradním nosiči. Přitom naděje na pořízení mikrofišové kopie zdarma se rozplynuly v důsledku nedostatku zájemců o její zakoupení. Bylo nezbytné nutně odstěhovat katalog a zpřístupnit ho na náhradním médiu, které zabere minimum prostoru. Jedinou nadějí na získání finančních prostředků byl příslib Mellonovy nadace - grant omezený na nákup základní techniky. **Bylo nutné najít metodu, která umožní ušetřit lidskou práci, a nahradit ji levnější prací strojovou a zároveň přesunout úzké hrdlo, které představuje kvalifikovaná lidská práce, již se nelze vyhnout, na konec, za fázi zpřístupnění katalogu ve strojem čitelné podobě.**

Z této složité situace a zmíněných limitů vzešla **originální technologie pro retrospektivní konverzi**

katalogů. Technologie je stavebnicová a univerzální, využitelná v kterékoli české nebo zahraniční knihovně jako celek nebo po jednotlivých modulech. Je výsledkem analyticko-projekčních zkušeností pracovníků NK ČR (funkční zadání, testování, připomínky), zkušeností se skenováním katalogu v Knihovně Akademie věd ČR, finanční podpory Mellonovy nadace a pomoci prof. A. Lasse při jejím získání a především zkušeností firmy COMDAT se zpracováním a využitím textových i obrazových dat a ochoty vyvinout na vlastní náklady komplexní technologii pro retrospektivní konverzi katalogů. Téměř roční vývoj technologie představoval pro firmu značné riziko ve smyslu návratnosti značného objemu vložených prostředků a energie. Technologie je natolik revoluční, že bylo těžké odhadnout, zda ji knihovny přijmou. V případě NK ČR nebylo dlouho jasné, zda si vzhledem ke svému omezenému rozpočtu bude moci realizaci projektu dovolit. S finančními zdroji jsou na tom ostatně podobně i další české knihovny. Většina z nich se svým příškrčeným rozpočtem ztěžuje zvládat nejnужnější provoz. Budou mít vůbec prostředky a síly vrátet se do minulosti? **Retrospektivní konverze ale není jen náročnou a obtížnou cestou do minulosti. Je nezbytnou podmínkou zpřístupnění fondů a efektivní spolupráce knihoven v přítomnosti a budoucnosti i jejich plnohodnotného připojení ke stále častěji skloňovaným informačním superdálčím.**

Technologie

Pro účely tohoto textu postačí velmi hrubé schéma a charakteristika jednotlivých kroků:

I. SKENOVÁNÍ

II. PŘEPIS

automatizovaný (OCR) / manuální

III. STRUKTUROVÁNÍ (TAGOVÁNÍ)

automatizované / manuální

I. Skenování

V první fázi jsou lístky (nebo tištěné bibliografie) naskenovány špičkovými skenery, které zaručují kvalitní výsledek nezbytný pro další zpracování. Díky speciálním programům je možné katalogy využívat již po naskenování. K záznamům je možné doplnit do pomocné báze dat záhlaví, popřípadě další soupisné a popisné údaje (rozsah pomocné báze závisí na velikosti katalogu, personálně-finančních možnostech knihovny a prioritách - časových horizontech).

II. Přepis

Ve druhé fázi jsou lístky převedeny z obrazové podoby (soubory TIFF) do nestrukturované textové (ASCII) podoby. Přepis probíhá u kvalitních záznamů automaticky (OCR), u méně kvalitních manuálně.

III. Strukturování (tagování)

Ve třetí (závěrečné) fázi jsou nestrukturované textové soubory převáděny do podoby strukturované (UNIMARC) pomocí speciálního programu. V závislosti na kvalitě zdrojového materiálu probíhá strukturování buď automatizovaně (je-li možné převod spolehlivě algoritmovat), nebo manuálně. Speciální programové vybavení poskytuje maximální komfort, šetří kvalifikovanou lidskou práci a umožňuje dosahovat vysoké rychlosti i kvality.

Priority

GK UKF obsahuje lístky různé kvality, jejichž retrospektivní konverze bude různě náročná. Podstatnou část katalogu tvoří **záznamy české knižní produkce**. Retrospektivní konverzi těchto záznamů je třeba provést v první fázi. Pro zapojení do mezinárodní spolupráce bychom měli mít v pořádku vlastní produkci. Dalším důvodem je skutečnost, že české knihy tvoří největší část fondů většiny českých knihoven, které rovněž plánují (nebo již zahájily) retrospektivní konverzi svých katalogů. Hrozí zde zbytečná multiplicitní práce a plýtvání personálními i finančními zdroji, které by měly být využity racionálněji. Po české produkci bude následovat produkce zahraniční, postup bude od jednodušších

záznamů ke složitějším.

Česká knižní produkce obsažená v konzervačním fondu NK ČR je od roku 1922 systematicky zachycena v Bibliografickém katalogu. Od roku 1983 existuje ve strojem čitelné podobě. Jedná se tedy o převod zhruba 400 tisíc záznamů z let 1922-1982. Přidáme-li počátek století, celkový objem bude zhruba půl miliónu záznamů. Při převodu se bude pracovat téměř výhradně se soubory vytvořenými automatickým přepisem pomocí OCR, možnost algoritmizace pro strukturování se v jednotlivých časových vrstvách výrazně liší.

GK UKF obsahuje 3 milióny lístků, z nichž pouze část bude možné přepsat pomocí OCR, možnost algoritmizace závisí, podobně jako u Bibliografického katalogu, na použitých pravidlech popisu. V porovnání s Bibliografickým katalogem je u GK UKF možnost algoritmizace snížena prolínáním záznamů vytvořených na základě různých pravidel.

Finanční nároky, časové horizonty

GK UKF bude zpřístupněn po oskenování (zhruba za 1 rok) v jednoduché podobě, v úplné (UNIMARCOvé) podobě zhruba do deseti let. Odhad vychází z objemu finančních prostředků (a počtu pracovníků, s nimiž lze pro akci v dnešní době reálně počítat). Bude-li k dispozici více finančních prostředků (a pracovníků ochotných se na akci podílet), celková doba se zkrátí.

V první fázi (rok 1995) je nutná investice do základního technického vybavení (grant Mellonovy nadace) a skenování (příspěvek Ministerstva kultury). Ostatní činnosti (přepis, tagování) budou v roce 1995 probíhat pouze v omezeném rozsahu. V následujících letech bude nezbytné mít k dispozici především dostatek prostředků na zaplacení velkého objemu lidské práce (přepis, tagování), pro zpřístupnění katalogu po naskenování a jeho zálohování bude nutná investice do diskových polí (v roce 1996).

Česká knižní produkce 20. století by měla být zpřístupněna v úplné podobě do roku 2000. Pro odhad finančních a personálních možností platí totéž co pro GK UKF. Zvýšení objemu dostupných finančních prostředků a počtu pracovníků bude znamenat i zde urychlení akce.

Personální zajištění, kooperace

Retrospektivní konverze generálního katalogu knihovny je především záležitostí té které knihovny. Nejinak je tomu s GK UKF a NK ČR zajistí převod jeho částí specifických pro fond NK ČR vlastními silami. Ostatní knihovny budou moci výsledný produkt samozřejmě využít. Některé části ve větší, některé v menší míře. Jak již bylo zmíněno, česká knižní produkce 20. století bude využitelná velmi dobře ve většině českých knihoven. Proto se na ně NK ČR obrací se žádostí o spolupráci.

Má-li být převod české knižní produkce 20. století ukončen do roku 2000, je třeba sehnat 10 pracovníků na plný úvazek, nebo (což je patrně reálnější) $n \times$ více pracovníků na $n \times$ kratší úvazek. NK ČR je schopna zajistit zhruba polovinu pracovní kapacity, pro druhou polovinu je potřebná spolupráce českých knihoven.

Retrospektivní konverze Bibliografického katalogu (má-li být výsledek kvalitní) vyžaduje profesionály se znalostí mezinárodních standardů, aby byla zajištěna mezinárodní využitelnost výsledného produktu. Náklady na převod jednoho záznamu odhadujeme v průměru na 40 Kč. Lze je rozložit takto:

skenování a OCR: 10 Kč

tagování: 15-20 Kč

revize a kontrola záhlaví: 10 Kč

Cena je o třetinu (s ohledem na rozsah záznamu patrně o polovinu) nižší než od firmy SAZTEC při dosažení vyšší kvality (oproti prostému přepisu navíc aplikace principů ISBD a AACR2, sjednocení záhlaví s využitím souborů autorit).

Česká knižní produkce - podmínky kooperace

Požadavky:

- znalost standardů (UNIMARC, ISBD, AACR2) - bude prověřeno testem,
- schopnost strukturovat záznam (rozeznat v záznamu jeho jednotlivé části a přidělit jim ekvivalentní pole/podpole v UNIMARCu, popřípadě je modifikovat podle doporučení ISBD/AACR2),
- schopnost práce se speciálním programem (pod MS Windows, nutnost práce s myší),
- pečlivost, přesnost.

Nabídka:

- zaškolení, příručky, konzultace,
- 15-20 Kč/záznam,
- pro knihovny, které otagují minimálně 30 tisíc záznamů do tří let celý výsledek retrospektivní konverze české knižní produkce 20. století zdarma,
- pro knihovny, které otagují minimálně 30 tisíc záznamů do tří let a jejichž zřizovatelem je Ministerstvo kultury, kompletně vybavenou pracovní stanicí, kterou si budou moci po ukončení akce ponechat. Rozšíření této podmínky i na knihovny spadající pod jiné zřizovatele závisí na přijetí/nepřijetí vládního programu, který se zabývá vytvářením podmínek pro informační superdálnice a zahrnuje retrospektivní konverze katalogů jako významný informační zdroj (iniciátorem a předkladatelem je Ministerstvo kultury, účinek by měl být nadresortní),
- knihovny, které se zúčastní kooperace, budou mít prioritu při přidělování finančních prostředků na další aktivity související s retrospektivní konverzí (skenování vlastních katalogů atp.).

Toto je prozatím strategie Ministerstva kultury - věříme, že osloví i zřizovatele ostatních knihoven.

Podmínky kooperace byly vyhlášeny na semináři o retrospektivních konverzích katalogů v Hradci Králové v červnu 1995. Již několik knihoven se oficiálně ke spolupráci přihlásilo. Z jejich dopisů vyplývá, že chápou výhodnost i nutnost spolupráce při retrospektivní konverzi katalogů českých knihoven a šanci využít optimálním způsobem omezené finanční i personální zdroje, které jsou u nás pro retrospektivní konverzi k dispozici. Většina knihoven chce však co nejdříve dát čtenářům k dispozici vlastní katalog, čekání na výsledky spolupráce jim připadá zdoluhavé. Je to pochopitelný argument. **Je nutné zvážit, v jaké podobě čtenářům katalog zpřístupnit, je třeba usilovat o omezení multiplicitních prací, které zbytečně odčerpávají již tak omezené finanční i personální zdroje. Nová technologie díky své modulárnosti může poskytnout řešení, pokud ji dokážeme plně využít. K největšímu plýtvání penězi i energií dochází tam, kde jsou multiplicitně vytvářeny záznamy na tentýž dokument. Protože jsou většinou vytvářeny podle různých pravidel, naděje na jejich mnohonásobné využití je malá.** Vytváření mnoha různých záznamů pro tentýž dokument obsažený ve fondech řady českých knihoven je evidentním plýtváním v nedostatku. Čím rozsáhlejší záznam, tím větší komfort pro uživatele jednotlivých knihoven, ale zároveň tím větší plýtvání z hlediska celku. Optimálním modelem je bezesporu spolupráce, společné vytvoření "jádra" (podle jednotných pravidel), jeho obohacení o vlastní lokační údaje a záznamy, které "jádro" přesahují. I tyto záznamy by měly samozřejmě respektovat jednotnou metodiku, aby byly použitelné pro ostatní knihovny. Již byla zmíněna pochopitelná potřeba jednotlivých knihoven dát vlastní katalog k dispozici čtenářům co nejdříve. Nestačila by však v první fázi jednoduchá podoba? Máme špičkovou modulární technologii, která nabízí různé varianty knihovnám, které chtějí zpřístupnit svůj katalog čtenářům co nejdříve, ale zároveň by rády omezily plýtvání finančními a personálními zdroji a podílely se na spolupráci. Následující text by měl sloužit jako inspirace pro knihovny, které se retrospektivní konverzí svých katalogů zabývají nebo hodlají zabývat. (Varianty A a B přicházejí v úvahu pro knihovny, které s retrospektivní konverzí svých katalogů ještě nezačaly, varianta C pro knihovny, kde již byla stručná retrospektivní konverze katalogu dokončena, nebo je již v pokročilém stádiu).

Varianta A1:

- Naskenování vlastního katalogu a využití v obrazové podobě (TIFF).

Poznámka: Pro velké knihovny se vyplatí nákup techniky, pro menší knihovny je výhodnější zadat skenování jako zakázku. Akci lze provést velmi rychle (cca 5 sec/záznam) a levně (cca 3-4 Kč/záznam - podle rozsahu a charakteru katalogu).

- Spolupráce na převodu českých knih 20. století (zvládnutí metodiky).
- Postupné přebírání záznamů vzniklých v kooperaci, doplnění informací o vlastním fondu a záznamů české produkce ve výsledném souboru neobsažených (podle jednotné metodiky). Totéž platí pro záznamy zahraniční produkce.

Varianta A2:

- Naskenování vlastního katalogu a využití v obrazové podobě (TIFF).
- Postupné přebírání záznamů vzniklých v kooperaci, doplnění informací o vlastním fondu a záznamů české produkce ve výsledném souboru neobsažených (podle jednotné metodiky). Totéž platí pro záznamy zahraniční produkce.

Varianta B1:

- Naskenování vlastního katalogu a využití v obrazové podobě (TIFF).
- Doplnění záhlaví (popř. dalších soupisných a popisných údajů) do pomocné báze dat.
- Spolupráce na převodu českých knih 20. století (zvládnutí metodiky).
- Postupné přebírání záznamů vzniklých v kooperaci, doplnění informací o vlastním fondu a záznamů české produkce ve výsledném souboru neobsažených (podle jednotné metodiky). Totéž platí pro záznamy zahraniční produkce.

Varianta B2:

- Naskenování vlastního katalogu a využití v obrazové podobě (TIFF).
- Doplnění záhlaví (popř. dalších soupisných a popisných údajů) do pomocné báze dat.
- Postupné přebírání záznamů vzniklých v kooperaci, doplnění informací o vlastním fondu a záznamů české produkce ve výsledném souboru neobsažených (podle jednotné metodiky). Totéž platí pro záznamy zahraniční produkce.

Varianta C1:

- Využití produktu vlastní retrospektivní konverze (stručných záznamů).
- Spolupráce na převodu českých knih 20. století (zvládnutí metodiky).
- Postupné přebírání úplných záznamů vzniklých v kooperaci, doplnění vlastních stručných záznamů z tohoto zdroje a dodání do společné báze dat záznamů české i zahraniční produkce ve výsledném souboru neobsažených (podle jednotné metodiky).

Varianta C2:

- Dokončení vlastní retrospektivní konverze (stručných záznamů).
- Využití produktu vlastní retrospektivní konverze (stručných záznamů).
- Spolupráce na převodu českých knih 20. století (zvládnutí metodiky).
- Postupné přebírání úplných záznamů vzniklých v kooperaci, doplnění vlastních stručných záznamů z tohoto zdroje a dodání do společné báze dat záznamů české i zahraniční produkce ve výsledném souboru neobsažených (podle jednotné metodiky).

Varianta C3:

- Využití produktu vlastní retrospektivní konverze (stručných záznamů).
- Postupné přebírání úplných záznamů vzniklých v kooperaci, doplnění vlastních stručných záznamů z tohoto zdroje a dodání do společné báze dat záznamů české i zahraniční produkce ve výsledném souboru neobsažených (podle jednotné metodiky).

Varianta C4:

- Dokončení vlastní retrospektivní konverze (stručných záznamů).
- Využití produktu vlastní retrospektivní konverze (stručných záznamů).

-Postupné přebírání úplných záznamů vzniklých v kooperaci, doplnění vlastních stručných záznamů z tohoto zdroje a dodání do společné báze dat záznamů české i zahraniční produkce ve výsledném souboru neobsažených (podle jednotné metodiky).

Všechny varianty C lze ještě obohatit o naskenování katalogu knihovny, přičemž produkt vlastní retrospektivní konverze může naplnit pomocnou bázi dat. Čtenář tak získá možnost vidět ihned všechny údaje obsažené na katalogizačních lístcích, které nebyly do retrospektivní konverze zahrnuty.

Závěr

Pro retrospektivní konverzi katalogů českých knihoven je dnes k dispozici špičková modulární technologie, která je nejen silným nástrojem, ale i výzvou ke kooperaci. Vývoj a aplikace originální technologie zařadil české knihovny na přední místo v mezinárodním kontextu. Dokážeme se domluvit a díky dobře zorganizované spolupráci technologii maximálně využít? Věřím, že ano. **Ukončení plýtvání v nedostatku, spojení prostředků a sil a respektování standardů je totiž jedinou možností, jak přestat v oblasti (nejen) retrospektivních konverzí katalogů přešlapovat na místě, stěžovat si na nedostatek všeho (kromě problémů) a začít i dokončit toto nesmírně významné dílo, které umožní využít výsledky práce našich předchůdců současným i budoucím čtenářům doma i v zahraničí a přispěje ke skutečnému otevření českých knihoven.**