

ROBOTI V NÁS

Tomášek Jan, Ústav informačních studií a knihovnictví FF UK

Obor knihovnictví se stále potýká s výměnnými formáty MARC. Hlavními aktéry těchto „potyček“ jsou staří známí – formáty MARC 21 a UNIMARC. Oba formáty implementují shodnou normu ISO 2709, v čem je tedy potíž?

Zkratka MARC přítomná v obou názvech znamená MACHine-Readable Cataloging, v překladu „Strojově čitelná katalogizace“ – což je vcelku známý fakt. S podivem však musím konstatovat, že **ani taktó výmluvný** název neodrazuje knihovníky od snahy tento formát dokonale ovládnout – v ideálním případě se pak stát „strojovým guru“, schopným i o půlnoci s tužkou v ruce vyplodit přesný katalogizační záznam v tomto formátu včetně dopočítání kontrolních údajů v hlavičce záznamu. Ano, pochopení a orientace ve struktuře formátu rozhodně není k zahození a profesionálnímu katalogizátorovi může mnohdy dokonce vytáhnout „trn z paty“. Je však nutné se v tomto formátu tak dalekosáhle pitvat, přitom (většinou) bez schopností potřebných k jeho programové implementaci? Stále se ptám, oč lepší je umět formát zpaměti, pamatovat si každé čísílko a písmenko pro to i ono, oproti pouhému povědomí o tomto formátu a jeho aktuálně preferované podobě, a schopnosti najít si v případě potřeby jeho specifikaci. Je možné, že se krutě mýlím, ale mám pocit, že by nepomohlo ani přejmenování formátu na RORC, tedy Robot-Readable Cataloging.

Knihovníci, z pro mě nepochopitelných důvodů, stále stavějí formát, ve kterém své údaje sdílejí, nad obsah, který jim má být „spoután“. Knihovny, zdá se, trvají na tom, aby katalogizační software reflektoval strukturu polí a podpolí, tak jak velí preferovaný formát, a to přímo v zadávacím formuláři aplikace. Dokonce i zadávání některých údajů a jejich syntaxe se podřizuje konvencím takového formátu. Je otázkou, zda-li je výhodnější podřídít katalogizaci přenosovému formátu s využitím doporučení definovaných v AACR(2R), nebo pouze respektovat pravidla a formát při zadávání a ukládání dat ignorovat. V současnosti se uvažuje o přechodu z formátu UNIMARC k formátu MARC 21, tedy k formátu staršímu, jehož pole nejsou strukturována s takovou jemností, jakou známe z formátu UNIMARC. Vzhledem k faktu, že současný formát a tedy i zpracování údajů je rozčleněno na více polí než vyžaduje formát „nový“, jde spíše o krok zpět a konverze by měla probíhat celkem snadno.

Strojové zpracování umožňuje konverzi a sloučení polí UNIMARCU pro potřeby formátu MARC 21 vcelku bezproblémově.

Logickým řešením by tedy mělo být vytvoření konverzního softwarového modulu, který by obstaral konverzi do formátu MARC 21. Proč ale vůbec podřizovat katalogizační praxi exportním formátům? Je logické, že katalogizační softwary vyšly ze struktury přenosového formátu podpořené katalogizačními pravidly. Je však výhodné se těchto formátů držet „zuby nehty“? Pakliže pravidla popisu definují veškerá pole nutná pro popis dokumentu, není výhodnější v softwaru implementovat vlastní číselník s maximální jemností zpracování tak, aby pokrýval tyto položky bez ohledu na jemnost vyžadovanou právě preferovaným formátem pro sdílení? Vzhledem k rychlosti, jakou se nové technologie vyvíjejí, by bylo směšné domnívat se, že formáty MARC jsou konečnou stanicí. Již dnes se v katalogizaci vehementně prosazují formáty jako XML a nárůst nových medií, na kterých jsou dokumenty dostupné, již dnes poukazuje na nedostatky formátů typu MARC.

Absurdita tohoto stavu je zřejmá, pokud se oprostíme od emoce vzbuzujících pojmů „knihovnické vědy“ a celý problém posuneme do neutrálních vod. Což takhle demonstrace na použití formátu HTML. Formát je sice souborový, ale stejně jako MARC slouží k přenosu a standardizaci přenášených informací. Je bezpochyby výhodné znát strukturu tohoto formátu, kdo však píše HTML dokumenty v jejich souborové formě? Každý si (a troufám si tvrdit, že tak činí i profesionálové živící se psaním těchto stránek) otevře svůj oblíbený editor, napíše to, co má na srdci, zformátuje písmo a odstavce, „napere“ dokument tabulkami dle svého gusta a nechá software, aby jeho dílo uložil ve formátu HTML. Opravdový „fajnšmekr“ se poté sice často vrhne do útrob finálního souboru a provede drobné korekce, ale psaní celého dokumentu přímo v HTML se jistě rád vyhne. To, co chci vyjádřit, snad nejlépe vystihuje používání barev. Pokud chci v HTML dokumentu použít světlemodrou barvu s lehce zeleným nádechem, mohu sice zadat hexadecimální kód 17 A0 94, přesto však většina smrtelníků v takovém případě otvírá okno s paletou barev. Potud jsou výhody použití dobrého editoru zřejmé a role použitého formátu, myslím, také. Konec konců, kdo v knihovně katalogizuje přímo ve formátu MARC – editor je přece samozřejmostí.

Nyní však nastává „zrada“. Zatímco uživateli textového editoru nepříjde podivné uložit si ještě kopii dokumentu například ve formátu PDF, knihovník se možností uložit kopii záznamu v jiném formátu doslova děsí. „Co se stane s těmi informacemi, co se v jednom formátu ukládají do jednoho pole a v druhém do tří, jakpak se asi vytvoří hlavička?“ Osobně vidím problém právě ve vnímání smyslu formátu. Zatímco u textového editoru

nikoho ani nenapadne zkoumat, v jakém formátu nástroj data interně zpracovává, nýbrž ho zajímá pouze to, pomocí jakých formátů lze tato data exportovat, u knihovnických systémů se stále zdůrazňuje nutnost implementovat ten který formát a v něm data nejlépe rovnou i ukládat. Myslím, že o vyhocené situaci v požadavcích knihovníků na nástroje zpracovávající bibliografické záznamy svědčí i fakt, že, ačkoliv už ručně psaným katalogizačním lístkům „odzvonilo“, je stále nutné popisné údaje zadávat s nejobtížnějšími oddělovači a syntaxí danou historickou konvencí, která v některých implementacích MARCU stále přetrvává. V informační společnosti, jak dnešní dobu rádi nazýváme, je takový přístup zadávání údajů pro zpracování výpočetní technikou směšné, ne-li přímo šílené – o chybách, které takto vznikají, ani nemluvíme.

Věřím, že podstatou zavedení elektronické katalogizace nebyla pouze snaha o digitální zpracování a sdílení záznamů, ale také snaha o zjednodušení a zkválitnění takové práce. Ačkoliv výměnné formáty interně používají tyto „lístkové“ rudimenty a jiné atavismy, jaký je důvod k jejich implementaci do uživatelských rozhraní knihovnických systémů? Proč upřednostňovat číselný kód právě úspěšně a proto používané implementace MARCU před přirozeným názvem položky katalogizačního záznamu a studovat syntaxi historických oddělovačů, když právě toho by nás mohla výpočetní technika uchránit? Katalogizační pravidla jasně vymezují položky použitelné pro katalogizaci i jejich smysl, problém tedy není v tom, jak a čím věci popisovat, ale pouze v tom, jakou implementaci normy (tedy ISO 2709) použijeme pro export těchto údajů.

Software by měl umět definovat veškeré položky nutné pro katalogizaci v jejich nejjemnější formě (a to nikoliv ve smyslu obsahu polí výměnného formátu, ale ve smyslu jejich definice v pravidlech popisu) a následně být schopen tyto položky exportovat do obecného formátu (tj. vodítko, adresář a proměnná pole).

Za těchto předpokladů můžeme s klidem mávnout rukou nad tím, kolik na to software potřebuje vlastních číselníků, zásobníků, hejblat a jiných volovinek, a naopak se soustředit na to, jak zefektivnit a zpříjemnit práci s takovým softwarem. Otázka převodu do konkrétní implementace normy (v tomto případě tedy do formátů MARC 21 a UNIMARC) pak zůstává na úrovni konfigurace softwarového nástroje. Uživatelé budou v rámci katalogizace stále pracovat v kompaktním a homogenním prostředí nezávisle na aktuálně podporovaném a preferovaném formátu pro sdílení dat. Obstarání syntaktické správnosti v tom kterém formátu zajistí software. Detailní struktura nejenže eliminuje chyby způsobené špatným zadáním syntaxe, ale umožní také podrobnější vyhledávání informací v rámci takového systému. Změny stávajících a vývoj nových formátů pro přenos i tak zůstane kapito-

lou pro knihovníky. Implementace exportu dat do takových formátů pak bude „oříškem“ pro specialisty – především programátory. Při změně formátu bude nutné dokonfigurovat systém, nikoliv uživatele a jejich návyky.

Každý formát je populární, dokud má podporu výrobců nástrojů, které s daným formátem pracují. Naproti tomu popularita a životnost katalogizačních pravidel je dána jejich schopností dostát nárokům kladeným na jejich schopnosti. Formát i pravidla jsou konvence, které se dynamicky mění a vyvíjejí. Formát určuje formu, pravidla obsah sdělení. Formátem lze omezit či rozšířit možnosti obsahu sdělení, nikoliv však jeho prostředky tento obsah přímo definovat.

Je možné, že se pletu a všechno se má úplně jinak. Vše, nad čím se zde tak podivuji, ve skutečnosti neexistuje a to, co navrhuji, je již dávno zde. Myslím si, že knihovnický software určený ke zpracování katalogizačních záznamů by měl být vyvinut, nebo nakonfigurován tak, aby umožňoval dostát požadavkům na strukturu exportních formátů pro sdílení dat. Nikoliv však, se této potřebě podřizovat na úrovni vlastní (interní) struktury dat, především pak komunikace s uživatelem, katalogizátorem. Struktura výsledného záznamu by měla být, v ideálním případě zcela, v „rukou“ nástroje. Katalogizátor by měl zajistit věcnou správnost zadávaných položek, nikoliv správnost syntaxe jakou jsou tyto údaje dále propagovány, prostřednictvím formátů určených k jejich sdílení. Cesta, o které si myslím, že k tomu cíli vede, předpokládá oddělení (extrakci) jednotlivých stavebních prvků, definovaných v pravidlech popisu dokumentů, od jejich syntaktických konvencí. Nikoliv katalogizátor! Software by měl být tím robotem, zajišťujícím syntaktickou správnost propojení těchto prvků a jejich vsazení do výsledné formy.