

Z39.50 versus (?) XML

Stanislav Psohlavec
AiP Beroun s.r.o.

Byl jsem požádán o napsání článku o Z39.50 a XML. Nečekejte technický rozbor – na ten se necítím dostatečně fundovaný a specializovaný – chci se s Vámi podělit spíše o své pocity. Je zajímavé všimnout si některých rysů geneze obou systémů. Srovnání je vlastně neúčelné, protože Z39.50 se zabývá komplexně komunikací mezi systémy – od vyhledávání po modifikaci dat, zatímco XML je v podstatě „jen“ o formě vytváření dokumentů a samotné komunikace se primárně netýká. XML se ale prvotně zabývá formou uchování informací i formou zachování jejich významu a tím dává nezbytný základ pro obecné řešení sdílení dat a následně i pro komunikaci.

Co si myslím o historii Z39.50.

Na počátku Z39.50 byla potřeba komunikace mezi konkrétními knihovnami, potřeba sdílení distribuovaných dat. Vznik protokolu iniciovali technici – přímo implementátoři, možná i programátoři. Neexistoval standard, jeho potřeba byla evidentní, možnost volné domluvy také. Proto v roce 1970 vznikl *Linked System Project*. Technici měli, co aktuálně potřebovali pro svou práci, věnovali se dál programování a implementaci. Péči o standard převzaly specializované organizace.

Standard byl (a je) průběžně vyvíjen jako knihovnický. Charakteristické je také to, že vývoj standardu a konkrétní implementace probíhaly v podstatě paralelně. Řekl bych, že technici na více oddělených pracovištích občas (spíše však často) předbíhali standard. To se projevilo v tom, že standard existuje, neexistuje však snad žádná kompletní implementace. Kompatibilita je částečná, protože jde o kompatibilitu již existujících systémů, které vznikaly pro splnění určitých úkolů, kterými bylo prvotně zajištění poskytovaných služeb (lokální knihovnické systémy).

Při vývoji knihovnických systémů na komerční bázi byla kompatibilita s jinými systémy celkem pochopitelně až druhotnou vlastností. Investice do zvýšení kompatibility jsou zvažovány s otázkou, co taková změna bude stát a co přinese, jak zlepší prodejnost. Odpověď na tyto otázky nebyla dosud příliš radostná, teprve v poslední době začíná být přikládán potřebný význam sdílení dat. Problém je v tom, že trh pro Z39.50 je relativně malý, pevně ohraničený knihovnami. Také účelnost využívání samotných technických možností závisí v tomto složitém prostředí na mnoha dalších těžko předvídatelných okolnostech.

Naopak, internet a jeho standardy knihovnické prostředí ovlivňují a pronikají do něj velmi snadno, protože nabízejí řešení vyvinutá pro mnohem větší trh a proto velmi levná. Vytvoření a udržení vlastního odděleného kni-

hovnického prostředí je drahé a je otázkou, na jak dlouho je vůbec reálné.

Vývoj i současný stav Z39.50 je charakteristický pro aktivitu vysoce odborné, specializované, ale fragmentované komunity, v níž praktická realizace občas předbíhá standardy.

Knihovnické prostředí bylo svou snahou o standardizaci komunikace a sdílení velkých databází dlouho na intelektuální špici ve veřejné sféře. Rozpoznání skutečné velikosti tohoto potenciálu však přišlo až s nástupem a rozšířením internetu. Internet si ale žije vlastním životem, málo ovlivněným knihovnickým prostředím. Přizpůsobovat se musí knihovny.

Potenciál Z39.50 je v knihovnickém prostředí zúročitelný. Je ale otázkou, jak dlouho bude efektivní jeho prosazování paralelně s globální standardizací komunikace, jejíž součástí je nepochybně i prosazování XML.

Jaká je situace Z39.50 v knihovnách?

Je dobře si uvědomit, že v knihovnickém prostředí působí dva subjekty:

1. Knihovny, které potřebují knihovnické systémy a mají stoupající potřebu komunikace a sdílení dat s okolím.
2. Poskytovatelé služeb, kteří nabízejí řešení v různé míře splňující požadavky.

Knihovnu samu nezajímá (nemusí zajímat) konkrétní prostředek, kterým je dosaženo cíle, ale kvalita, dlouhodobá perspektivnost a cena tohoto dosažení.

Pro poskytovatele je naopak chtě nechtě důležité prosazení svého řešení a zúročení investic do nabízeného řešení vložených. Protokol Z39.50 proto prosazují pochopitelně nositelé a implementátoři tohoto standardu. V současnosti jiné komplexní funkční řešení komunikace nad existujícími knihovnickými systémy neexistuje, takže přes všechny výhrady panuje shoda knihoven a poskytovatelů a vlastně je vše jasné – klid a pohoda. Nutno si ale uvědomit, že nositelé „know-how“ okolo Z39.50 mají minimální motivaci tento současný konsensus narušovat.

Některé knihovny se ale již nyní zabývají využitím XML, konverzí UNIMARCu do XML. Důvodem je dle mého názoru správný předpoklad, že je perspektivní soustředit se na využití standardu, který je prvoplánově koncipován pro globální komunikaci a sdílení dat. Lze očekávat, že v budoucnu tak bude možno využít prostředků, které vznikají nezávisle na knihovnách v novém informačním prostředí a které budou jeho nedílnou součástí. Speciální prostředky pro komunikaci mezi knihovnickými systémy (např. právě Z39.50) nemusí být v budoucnu nezbytné.

Ve vývoji, který předcházel vzniku XML, lze pozorovat působení zcela jiných mechanismů než při vzniku Z39.50. Tvůrci standardů, producenti dat a producenti aplikací nejsou svázáni tak, jako je tomu v knihovnickém prostředí. Proto autoři aplikací ve svém vlastním zájmu respektovali již existující standardy, takže kompatibilita a možnost sdílení dat mezi různými aplikacemi byla vždy velmi vysoká.

Dovolím si připojit krátkou informaci o XML všímající si i historického hlediska.

Standardizace komunikace byla nutným předpokladem vzniku internetu. Nejběžnějším komunikačním rozhraním na www (world wide web) je protokol HTTP (HyperText Transfer Protocol). Tento protokol slouží ke komunikaci mezi webovým serverem a webovým prohlížečem a k přenosu dokumentů, které jsou realizovány zpravidla ve formátu HTML (Hypertext Markup Language). Tento formát se vyvinul velice prudce s rozvojem internetu jako aplikace SGML (Standard Generalized Markup Language). Lze říci, že v současné době jsou HTML dokumenty základem pro tvorbu www stránek a jejich přenos. HTML je nástroj, který má poměrně silné prostředky pro grafickou prezentaci dokumentů (tabulky, obrázky, zadávací formuláře apod.), ale neobsahuje takřka žádné prostředky pro popis struktury dokumentu. Proto vyplynula potřeba standardu umožňujícího popisovat strukturu textového dokumentu a s touto strukturou dále pracovat. SGML vznikl už v 80. letech právě jako odezva na potřeby strukturalizace velkého objemu textových dat přidáním významových značek do dokumentů. Je obsažen v normě ISO 8879 z roku 1986. Tento jazyk je velmi obecný a z toho vyplývá i vysoká technická náročnost a tedy i cena prostředků využívajících tohoto standardu. Prakticky se proto uchytil ve velkých firmách a státních organizacích (americké ministerstvo obrany, Boeing, IBM atd.), ale masově se nerozšířil.

V 90. letech byl proto vyvinut skupinou XML Working Group pod konsorciem W3C jazyk XML (eXtended Markup Language), zobecněný značkovací jazyk založený na SGML. Tento jazyk, vycházející ze SGML a značně ovlivněný HTML oblíbeným díky své jednoduchosti, se jeví jako velice vhodný pro zpracování textových dokumentů (a to i strukturovaných) díky mnoha výhodným rysům:

- XML je formát textový, tzn., že dokument je možno vytvářet, přenášet a zpracovávat na libovolné platformě, jakýmkoli počítačovým systémem, který je schopen pracovat s textovými daty, jsou vyřešeny jazykové a kódové problémy.
- XML odděluje popis struktury dat od jejich vlastní prezentace (pomocí tzv. style sheetů). To umožňuje snadnou konverzi do jiných formátů, možnost prezentace dat různými způsoby, tj. pomocí jednoho style sheetu můžeme data převést do HTML pro zobrazení na monitoru počítače, jiným stylem můžeme vygenerovat poscriptový soubor pro tiskárnu atd., ale i jiné formáty tedy i UNIMARC nebo libovolný jiný textový formát.
- Každý dokument má definovanou svoji strukturu prostřednictvím tzv. DTD (Document Type Definition). Tato DTD může být unikátní pro každý dokument, ale může být také společná pro celou řadu dokumentů, které mohou vznikat kdekoli v rámci organizace nebo třeba celosvětově. Dnes již na internetu existují knihovny DTD, kde jsou publikovány obecné předpisy DTD pro konkrétní aplikace z různých oborů lidské činnosti, namátkou uvedme astronomii, účetnictví, bibliografii, tvorbu technické dokumentace, geografii, zdravotnictví... Toto umožňuje vytváření dokumentů

přenositelných mezi různými institucemi a aplikacemi. Při změně struktury dokumentu lze potom velmi snadno a tedy i levně přenést dokument do jiné struktury.

- XML je otevřený formát, jeho specifikace je k dispozici zdarma na stránkách konsorcia W3C a jeho použití a šíření není nijak omezeno licenčními poplatky.

Technik jistě doplní i další charakteristiky: XML je jednoduše použitelný na internetu, XML podporuje širokou paletu aplikací, XML je kompatibilní se SGML, XML dokumenty jsou snadno čitelné i vytvořitelné, počítačové programy pro zpracování XML dokumentů jsou velmi jednoduché.

Otevřená je i cesta k dalšímu pro mne těžko odhaditelnému využití. XML umožňuje více než jen popis prostých dokumentů. Dokumentem totiž může být i formulace SQL dotazu, není vyloučena ani možnost dotazu na Z39.50.

Aplikace využívající XML mohou být technicky snadno propojitelné, je však nutno zajistit i propojitelnost „významovou“. Tento fakt je v knihovnickém prostředí velmi dobře znám, neboť letité zkušenosti se standardy daly poznat (někdy tvrdě) význam jejich existence, dodržování i jednotnosti jejich interpretace.

Pro data v UNIMARCu je převod do XML bezztrátovou změnou formy (nikoli obsahu a členění). Přechod na XML není tedy zaváděním nového formátu, ale spíše interpretace těchč (již formátovaných) dat v jiné formě, která je přímo použitelná v novém informačním prostředí a která dovolí bez velkých investic do vlastního vývoje využívat vlastností a prostředků v tomto prostředí existujících a dále vznikajících. Tento trend je dle mého názoru jednoznačně perspektivní a je užitečné „být při tom“.

Před krátkou dobou se u nás ve firmě znovu krátce otevřela diskuse zda investovat do interface Z39.50 pro naše bibliografické produkty nebo toto přeskocit a orientovat se především na využití XML. Závěr byl jednoznačný.

Signifikantní může být i postoj firmy EX LIBRIS, která má implementován Z39.50 v rozsahu pokrývajícím poskytování dat, ale nepokládá zřejmě za perspektivní implementaci dalších vlastností nabízených Z39.50 (UpDate).

Při novém nasazování Z39.50 v knihovnách bych se při rozhodování rozhlížel po rozsahu, úspěšnosti a účinnosti jeho nasazení na západ od našich hranic. Hodně bych se zamýšlel i nad otázkami, v čem a proč může být tento vývoj předobrazem vývoje u nás.

K těmto úvahám mám však pouze zprostředkované informace, které si nedovolím předložit jako své názory.

Dovolím si ale na závěr citovat prezidenta firmy *Apache Software Foundation* pana Briana Behlendorfa, který řekl: „...kolem XML již byla udělána příliš velká mediální bublina, je to prostě další krok, který přišel po ASCII“ (pro jistotu – ASCII jedna z nejstarších norem souvisejících s digitálními dokumenty – binární kódování abecedy).

Pozn. autora: Příspěvek byl zpracován s využitím materiálů poskytnutých G. Krčmařovou a K. Kučerou.