

Memoria Mundi Series Bohemica – z trezoru na Internet

Ing. Stanislav Psohlavec
AiP Beroun s.r.o.

Pilíře projektu MMSB	1
Digitalizace, digitální dokumenty, digitální knihovna	1
MASTER	1
Využívání výsledků MMSB	2
Rukopisy MMSB na internetu	2
Příprava	2
MANUSCRIPTORIUM	3
Nástroje pro vyhledávání	3
Nástroje pro prohlížení rukopisu	3
Programové vybavení	4
Základ XML, DTD MASTER	4
Zdroje dat	4
Volný přístup k informacím	4
Licencovaný přístup k informacím	4
Definice práv a povinností uživatele dat	4
Cena licence	5
Další vývoj	5

Pilíře projektu MMSB

Digitalizace, digitální dokumenty, digitální knihovna

Základním pilířem projektu MMSB je již v roce 1993 započatá digitalizace vzácných dokumentů. Za dobu deseti let bylo získáno množství zkušeností a lze říci, že v současnosti je již digitalizace kompletních rukopisů a starých tisků rutinní záležitostí. Tvorba digitálních obrazů technologicky dosáhla úrovně, která je pro cíle projektu MMSB a obecně pro zastoupení originálních dokumentů optimální a existují i dostatečné rezervy. Technické vlastnosti digitálních obrazů umožňují pořídit dokonalé reprodukce originálů. Také z hlediska bezpečnosti samotného procesu digitalizace je dosaženo vyhovujících podmínek. Je samozřejmostí, že obrazy se skládají do podoby digitálních dokumentů, které umožňují práci s obrazem zastupujícím vzácný originál. V současnosti existuje fond více než osmi set kompletně digitalizovaných rukopisů a starých tisků, který se pravidelně zvětšuje

MASTER

S obrazovými daty vzniká množství odborných textových informací, které je nutno napojit na obrazová data. Také tyto informace jsou hodnotné a je žádoucí zachovat je bezzbytku do budoucnosti. Zatím jsou víceméně statické a mají pevnou, předem dohodnutou strukturu. Dosavadní nástroje jsou tomu přizpůsobeny. Takto vznikající data mohou mít vlastnosti zaručující neomezenou morální životnost, nezbytné je ale akceptování pevně dané struktury záznamu. Pořizování dat odpovídá obvykle vyplňování předpřipraveného dotazníku nebo tabulky. To platí v podstatě pro všechny dosud používané nástroje. Tento stav je zcela

vyhovující pro činnost inventarizační a katalogizační. Je nezbytným základem na kterém lze dál stavět, není však konečným cílem. Neposkytuje potřebnou volnost nezbytnou pro badatelskou činnost skutečně využívající digitální prostředí.

Tyto potřeby daly vzniknout projektu, na kterém se od devadesátých let podílejí odborníci z mnoha významných světových knihoven. Tím je na technologii XML založený projekt MASTER. Definuje datovou strukturu a metodiku, která umožňuje zapsání libovolně složitých informací vztahujících se k rukopisům a zároveň umožňuje zpracování těchto informací počítačovými systémy. Tento projekt dospěl do stádia, kdy lze jeho výsledky prakticky využívat.

Možnosti záznamu informací pomocí XML s DTD podle projektu MASTER slibují vznik skutečně badatelského prostředí, protože tato volná forma přináší možnost záznamu jakýchkoli informací souvisejících s rukopisy. Tato vlastnost je využitelná a může přinést užitečné výsledky již nyní, protože se potvrzuje, že prakticky jakékoli informace o rukopisech mohou být beze ztráty (obsahové i strukturální) převedeny do této formy.

Co přináší XML kromě této volnosti ?

Do vývoje standardu byly investovány značné prostředky, všichni významní výrobci software investují další těžko představitelné prostředky pro zajištění schopnosti svých produktů zpracovávat a produkovat informace ve formě XML. XML vznikl s akademicky definovanými cíli a jde nepochybně o velmi dobrou myšlenku. Sílu a jistotu **technologické kontinuity** v tomto směru vývoje mu však dávají teprve **investice** do vývoje obecně použitelných softwarových nástrojů a **všeobecné akceptování** tohoto směru vývoje silnými komerčně úspěšnými výrobci.

Projekt MMSB využívá od počátku formu ukládání dat sice v pevné struktuře, ale v podstatě na stejných principech (DOBM je aplikací SGML). Díky faktické připravenosti bylo pro projekt MMSB logické výsledky projektu MASTER ihned využít a dostupná související data převést do XML.

Využívání výsledků MMSB

Prvním krokem projektu MMSB je vytvoření digitální podoby originálního dokumentu, který má být dále používán především (a raději výhradně) ve své digitální podobě. Výsledkem digitalizace a připojení popisných dat je nyní především sada CD-R disků, které pak lokálně nahrazují originál. Je zajištěna centrální péče o zachování těchto CD-R disků v Národní knihovně ČR. Pouze zde je také (opět lokálně) možno studovat v digitální podobě dosud digitalizované rukopisy.

Rukopisy MMSB na internetu

Příprava

Zpřístupnění rukopisů na Internetu předcházelo několik kroků. Významným krokem bylo zpřístupnění databáze MEMORIA, která na Internetu informuje o existenci již digitalizovaných rukopisů a starých tisků a přináší jejich podrobnější popis. Nepřináší však přístup k těmto dokumentům, zpřístupňuje jen ilustrativní ukázky. Paralelně byly provedeny možnosti převodu nejrůznějších typů dat na společnou platformu MASTER a možnosti interpretace těchto XML záznamů na Internetu. Vyzkoušely se také převody různých dostupných dat počínaje inventarizačními soupisy, přes klasické bibliografické záznamy, data MMSB, až po import záznamů primárně vznikající ve formě XML v projektu MASTER. Potvrdilo se, že všechny tyto cesty jsou možné a že mohou přinést zajímavé výsledky.

MANUSCRIPTORIUM

Výsledky těchto předchozích prací nyní představujeme ve formě databáze nazvané Manuscriptorium. Je to nástroj pro práci s daty v XML s využitím DTD dle projektu MASTER. Protože jakákoli data související s rukopisy lze převést do této formy aniž by došlo k jakékoli ztrátě informace, je možné tímto nástrojem zpřístupnit ve společném prostředí data z různých zdrojů, která jsou původně k dispozici v různých formách a formátech. Nové vlastnosti dat převedených do formy XML se stanou důležitými v budoucnu. Dovolí nasazení dalších obecně dostupných nástrojů nastavených dle potřeb badatelů, které vyplynou z nových zkušeností a potřeb.

Zřetelněji než dříve je vyjádřena rozdělovací linie mezi samotnými informacemi a nástroji na jejich využívání. Součástí řešení databáze Manuscriptorium je na jedné straně úkol vytvořit nástroje pro převod v podstatě libovolných dat do formy XML a na straně druhé úkol nabídnout nástroje pro kvalitní zpřístupnění těchto dat a výsledků digitalizace na Internetu.

Nástroje pro vyhledávání

Jako rešeršní nástroj je použit systém Tornádo. Jde o mimořádně výkonný rešeršní nástroj, jenž je využíván mimo jiné pro zpřístupnění České národní bibliografie na Internetu..

U jazykově velmi variabilních dat vynikají možnosti práce s heslovými a slovními rejstříky, možnost vyhledávání nepřesným dotazem a automatické variace dotazu s ohledem na různé možnosti zápisu slov ve středověkých textech.

Výsledky vyhledání jsou k dispozici formou přehledů základních informací o dokumentech vyhovujících dotazu a dále samozřejmě ve formě podrobných informací o konkrétním nalezeném dokumentu. Formu tohoto podrobného zobrazení si badatel může volit ze široké nabídky podle svých potřeb.

Nástroje pro prohlížení rukopisu

Chceme-li si prohlédnout konkrétní rukopis, už nepotřebujeme rešeršní prostředí. Plynule přejdete do použití jiného, pro daný účel optimalizovaného nástroje. Tento nástroj vznikl především na základě rozboru práce s digitalizovanými rukopisy, část inspirace pochází od různých nástrojů pro lokální práci s obrazem. Potřebné vlastnosti jsou však poprvé přivedeny na Internet.

Co tento nástroj umožňuje? Je samozřejmě možné zobrazit všechny existující doprovodné textové informace vztahující se k dokumentu i informace mající návaznost na jednotlivé stránky. Těžištěm jsou ale pochopitelně možnosti práce s obrazem. Nástroj je optimalizován tak, aby přenosy dat byly minimální, aby rychlost práce na Internetu byla co největší.

Z nabízeného spektra možností lze volit takové nastavení, které odpovídá reálným možnostem připojení.

Orientace v dokumentu je usnadněna přehledem stránek, který je k dispozici v textové formě nebo ve formě galerie miniaturních náhledů na stránky. Příjemná je možnost vložení záložek u jednotlivých stránek a listování jen po takto označených listech.

Navigace na konkrétní stránce je snadná díky provázanosti malého náhledu s hlavním, velkým zobrazením stránky dokumentu. V náhledu je rámečkem naznačen výřez zobrazovaný v hlavním okně. Výřez lze měnit přesunutím rámečku nebo přesunem přímo v hlavním obraze. Snadno lze měnit i měřítko zobrazení, které zůstává při listování zachováno.

Úpravy obrazu jsou vzhledem k velmi různorodému materiálu řešeny zcela originálně. Pro každý digitalizovaný dokument je vytvořena sada obrazů určených pro zpřístupnění na Internetu. **Individuálně pro každý rukopis** je zvolena optimální komprese obrazu a rozlišení tak, aby přenášený objem dat byl minimální při maximálním informačním obsahu. Nyní jsou k dispozici tři typy obrazů :

- normální – poskytující možnost spolehlivého čtení,
- vyšší – dovolující detailnější prohlížení,
- černobílý se zvýrazněnými detaily. Tento obraz je vytvořen speciálním postupem, má vyšší rozlišení a zachovává detaily lépe, než by bylo možno dosáhnout prostou manipulací s jasem a kontrastem obrazu. Přitom je soubor s obrazem dosti malý, takže jeho zobrazení je rychlé. Pro čtení textů je tato volba optimální.

Programové vybavení

V předkládané verzi nepředpokládáme využití jiných, než pro tento účel volně použitelných formátů. Proto není nutno instalovat jakékoli programy nebo doplňky, neboť všechny funkce využívají standardních vlastností Internetovských prohlížečů.

Pokud praxe ukáže, že by využití dalších formátů bylo významným přínosem (např. DjVU), bude formát zařazen do výše uvedené nabídky.

Základ XML, DTD MASTER

Jak již bylo řečeno základem je použití XML, s využitím DTD projektu MASTER. Používané DTD je složeno ze sekce přejímající přesně původní DTD projektu MASTER a ze sekcí umožňujících záznam technických informací a připojících digitální obrazy celého dokumentu.

Zdroje dat

V databázi jsou nyní zpřístupněny tyto informace:

- Data vzniklá v projektu VISK 6. Pouze tato data mají také připojené **kompletní obrazy digitalizovaných dokumentů**. Protože obrazová data se pro toto zpřístupnění individuálně generují z originálních dat, budou se v databázi objevovat postupně.
- Záznamy primárně vzniklé v projektu MASTER. Část těchto dat se překrývá s daty projektu VISK 6.
- Třetím zdrojem dat jsou data inventárního soupisu zámecké knihovny v Kynžvartu. Je to zajímavá ukázka možnosti zpřístupnit data zcela jiného původu, která vznikla v souvislosti s jinými úkoly a cíly.

Připojení dalších zdrojů dat se připravuje.

Volný přístup k informacím

Data o rukopisech mající charakter bibliografických informací a základních popisů budou v databázi MANUSCRIPTORIUM zpřístupňována volně a bez omezení, podobně jako doposud v databázi MEMORIA.

Stejně volně budou mít zajištěn přístup ke kompletním obrazům všech digitalizovaných dokumentů všichni účastníci projektu VISK 6.

Licencovaný přístup k informacím

Sbírka více než 800 významných rukopisů v digitální podobě je již hodnotný zdroj informací. Jeho vybudování si vyžádalo nemalé prostředky. Podobné i podstatně méně rozsáhlé projekty ve světě poskytují data k využití na základě poskytování licence.

Proto i zde bude přístup k plným obrazovým datům umožněn jen na základě poskytnutí licence na jejich využívání. Licence má v podstatě dvojí význam:

Definice práv a povinností uživatele dat

Licence definuje práva a povinnosti uživatele dat, která jsou poskytnuta na základě licence. Uživatel získává přístup k datům a zároveň se přijetím licence zavazuje s daty nakládat podle

podmínek této licence. Víme, že spolehlivá fyzická ochrana proti nežádoucímu kopírování a zneužívání obrazových dat neexistuje. Myslím, že je správné důvěřovat a spolehnout se na právní povědomí solidních uživatelů, kteří akceptují přijaté dohody. Vůči těm, kteří přijaté dohody nedodržují, je možno využít toho, závažné nedodržení licenční smlouvy je trestný čin a může být postižitelné. Jiný přístup není prakticky možný. Alternativou je jen zcela volné zpřístupnění nebo naopak žádné zpřístupnění.

Licence tedy prostě řečeno závazně předepisuje, jak se s daty smí nakládat. To samo o sobě ještě nesouvisí s tím, zda je licence placená či neplacená.

Cena licence

Obvyklým záměrem poskytnutí licence bývá (kromě definice práv k nakládání s licencovanými daty) získání úplaty za tuto službu.

Záměrem projektu VISK 6 je zpřístupnění digitalizovaných vzácných dokumentů. Lze právem očekávat, že v dosahu institucí, které se digitalizace svými dokumenty účastnily a podílely se na financování by mělo být poskytnutí licence na přístup co nejvýhodnější, nejlépe volně. Na druhou stranu je trochu luxus, aby zájemci o tato data, kteří nemají podíl na vzniku tohoto fondu, jej využívali bezplatně. Předpokládáme, že placené licence přispějí na zajištění a rozšiřování nabízených služeb a při úspěšnosti této strategie napomůžou i financování další digitalizace.

O rozsahu poskytnutí bezplatných licencí rozhodnou hlavní účastníci projektu VISK 6 a zástupci Ministerstva kultury.

Další vývoj

Dosavadní pořizování popisných dat k digitalizovaným dokumentům splývá postupně s pořizováním dat dle projektu MASTER. Již nyní se popisy digitalizovaných dokumentů provádějí v této formě. V krátké době bude tento postup jediným používaným. Školení jak připravovat popisy v XML proběhla v lednu tohoto roku a budou pokračovat.

Byly vytvořeny prostředky pro formální kontrolu takto vytvořených dat a byly dány volně k použití (validátor - viz www.rukopisy.cz).

Po dokončení generování obrazových dat pro zpřístupnění na Internetu a po zařazení všech digitalizovaných rukopisů do databáze MANUSCRIPTORIUM ztratí samostatná informační databáze MEMORIA význam a její provoz bude ukončen.

Volně budou dány k dispozici prostředky pro konverze různých dat do XML a správci souvisejících dat budou mít možnost svá data zpřístupnit pomocí této databáze.

Vývoj příštích speciálních nástrojů nebo způsobů použití jiných nástrojů vyvinutých pro obecnou práci s XML záznamy budou určovat uživatelé existujících dat na základě svých praktických zkušeností.

Doufáme, že databáze MANUSCRIPTORIUM dá badatelům dobrou příležitost tyto zkušenosti získat a pomůže k formulování nových cílů.

Stanislav Psohlavec, AiP Beroun, 2003-05-1