



DIGITALIZACE MORAVSKÝCH KNIHOVNÍCH SBÍREK

Moravská zemská knihovna v Brně

**Zpráva za 2. rok řešení projektu
2009**

Zpracovali:

Doc. Dr. Jaromír Kubíček, CSc., řešitel projektu

Dr. Blanka Sapáková, spoluřešitelka projektu

Brno, listopad 2009

OBSAH

	A	Konstatační část
	3	
A.1	Rešerše	3
A.2	Současný stav ve světě a v ČR	3
A.3	Východiska – cíl	4
B	Analytická část	5
B.1	Vlastní řešení	5
B.1.1	Analytický rozpis deníku	7
B.1.2	Analytický rozpis časopisů	9
B.1.3	Informační technologie	10
B.2	Přínos řešitele	11
B.3	Posun znalostí	11
C	Návrhová část	11
C.1	Výsledky řešení	11
C.2	Závěr	11
C.3	Návrhy opatření	11
D	Použití účelové podpory	12
D.1	Tabulky a komentář	12
D.2	Opravy a udržování	12
D.3	Cestovné	12
D.4	Mzdové náklady	12
D.5	Zákonné sociální pojištění	12
D.6	Zákonné sociální náklady	12
E	Resumé a klíčová slova	12
E.1	Resumé a klíčová slova v češtině	12
E.2	Abstract and key words in English	13

A Konstantační část

A.1 Rešerše

Literatura zpracovaná k tématu v r. 2009

- Allgemeines europäisches Journal 1794–1798 / Miloš Papírník // Brno : Moravská zemská knihovna, 2009. – 82 s. – Rejstříky moravského tisku. Sv. 4.
- Týdeník, listy ponaučné a zábavné 1848–1849 / Jaromír Kubíček // Brno : Moravská zemská knihovna, 2009. – 36 s. – Rejstříky moravského tisku. Sv. 5.
- Nový život 1896 – 1907 / Miloš Papírník // Brno : Moravská zemská knihovna, 2009. – V tisku. – Rejstříky moravského tisku. Sv. 6.
- K problematice české bibliografie / Jaromír Kubíček // Sdružení knihoven České republiky : rok 2009. – Brno, Sdružení knihoven ČR, 2009. – s. 8-12
- Die deutsche Presse in tschechischen Ländern ihre Zustand und Bearbeitung / Jaromír Kubíček // Referát na mezinárodní konferenci Digitale Quellensammlungen, Bratislava, 6.–8. 7. 2009, in www://indico.ulib.sk/MaKaC/conference/Display
- Německý tisk v českých zemích, jeho stav a zpracování / Jaromír Kubíček // Vlastivědný věstník moravský. – Roč. 61, č. 4 (2009), s. 362-369.

A.2 Současný stav ve světě a v ČR

Ochrana a zpřístupnění dokumentů obsažených v primárních informačních zdrojích ve sbírkách paměťových institucí jsou spojeny s digitalizací. Ta je prioritou nejen z pohledu záchrany dokumentů tištěných na papírových nosičích, ale také z důvodu masového rozšíření internetu a jeho využívání ve všech formách vzdělávání a výzkumu. O současném stavu ve světě vypovídala mj. mezinárodní konference *Digitale Quellensammlungen: Erstellung-Archivierung-Präsentation-Nutzung*, které se za Moravskou zemskou knihovnu aktivně účastnil řešitel tohoto projektu J. Kubíček. Pořadatelem konference byla Univerzitní knižnice Bratislava a Digitales Forum Mittel- & Osteuropa München. Tato německá společnost organizuje digitalizaci významných titulů německých novin vydávaných ve státech mimo NSR a dokončuje titul *Pressburger Zeitung 1764–1929*. V rámci této společnosti je připravován projekt ve spolupráci s MZK na digitalizaci deníku *Tagesbote aus Mähren 1850–1945*. Příspěvky na této konferenci podaly přehled o výsledcích a záměrech digitalizace knihovních sbírek v Maďarsku, Německu, Rakousku a Slovensku. Příkladný postup zvolila Österreichische Nationalbibliothek v projektech ANNO (AustriaN Newspaper Online) a ALEX (sbírky rakouských zákonů), které jsou svými uživatelsky velmi příjemnými portály virtuálními čítárnami rakouských periodik. Pro náš projekt bylo pozoruhodné, jak si ÖNB za rok (o stavu v r. 2008 informovala vedoucí odd. digitalizace ÖNB Ch. Müller na konferenci k 200. výročí MZK v Brně a účastnila se i konference v Bratislavě v r. 2009) zdokonalila možnosti vyhledávání článků podle ročních obsahů časopisů.

Významný je projekt TELplus (The European Library), který nabízí přístup ke sbírkám více než 45 národních knihoven Evropy. Národní knihovna ČR je rovněž řešitelem tohoto projektu a zavázala se přispět 3,4 mil. stran v podobě textových souborů, z toho 1 mil. stran periodik.¹ Mezinárodní provázanost vede nejen ke snadnějšímu vyhledávání informací, ale i přebírání bohemikálních titulů (např. ÖNB digitalizovala a zpřístupnila *Prager Tagblatt 1877–1938*). Výchozím materiálem pro evropské knihovny je *Rozhodnutí Evropského parlamentu a rady č. 456/2005/ES ze dne 9. března 2005 o zavedení víceletého programu Společenství pro usnadnění přístupu k digitálně šířenému obsahu a jeho používání a využívání v Evropě* a z něho vyplývající *Iniciativa „Evropská digitální knihovna“* ze dne 18. 4. 2007. V ČR v souladu s mezinárodními iniciativami pracuje NK a MZK na společném

projektu *Národní digitální knihovna*. Národní knihovna ČR (NK) a Moravská zemská knihovna (MZK) pracují v současnosti na společném projektu *Národní digitální knihovna*. Byla odevzdána studie proveditelnosti. S tím souvisí analýza a testování softwarů spojených s masovou digitalizací, vývoj nových verzí stávajících systémů a jejich propojování. V NK byl testován program DocWorks od firmy CCS, který dokáže rozčlenit např. stránku novin na jednotlivé články, rozezná nadpisy, prostý text, obrázky, poznámky pod čarou apod. a vše uloží do struktury XML. Výstupem z tohoto programu je dokument ve formátu METS/ALTO. Pokračují projekty některých knihoven, které jsou financovány z tzv. Norských fondů (Národní knihovna, Městská knihovna v Praze, Vědecká knihovna v Olomouci). Digitalizační centrum Knihovny AV ČR je stále nejlépe vybaveným a nejproduktivnějším digitalizačním pracovištěm v rámci českých knihoven, za pět let svého působení zdigitalizovalo přes 2 mil. stran knihovních dokumentů.² Snahám krajských knihoven o digitalizaci regionálních dokumentů se daří v těch případech, kdy datové úložiště je řešeno ve spolupráci s krajským úřadem. Takové začátky se podařilo realizovat Krajské knihovně Františka Bartoše ve Zlíně, která ve své digitální knihovně vystavuje 12 titulů, z nichž neobsáhlejší je týdeník *Zlín* 1918–1945. Některé specializované knihovny digitalizují a zpřístupňují oborová periodika, jako knihovna Národního filmového archivu (97 periodik o filmu) nebo Středisko vědeckých informací Fakulty sociálních studií UK (15 titulů novin k dějinám médií).

¹ Srov. FOLTÝN, Tomáš. *Účast NK ČR v projektu TELplus – vytváření OCR souborů*. In Knihovny současnosti 2009. Brno, Sdružení knihoven ČR 2009, s. 159-166.

² RYŠÁNKOVÁ, Anna, PŠENÍČKA, Michal. *Digitalizační centrum Knihovny AV ČR*. – Informace Knihovny AV ČR, č. 1 (2009), s. 5-13.

A.3 Východiska – cíl

Periodika jsou zajímavým materiálem pro digitalizaci. Zvláště u novin je výskyt úplných řad jednotlivých ročníků poměrně vzácný, jak svědčí údaje zpracované v rámci České retrospektivní bibliografie, v nichž bibliografický popis každého titulu je doplněn údaji o jejich výskytu v paměťových institucích.³ Nejen degradace novinového papíru způsobená jeho kyselostí, ale také sporadický výskyt ospravedlňuje poměrně vysoké náklady spojené s digitalizací.

Procesy digitalizace dokumentů zasahuje dnes již do řady správních agend a oblastí a knihovny se do nich zapojily především z důvodu ochrany svých sbírek. Technologické postupy vlastní digitalizace, spočívající ve skenování mikrofilmů nebo v hybridním snímkování, jsou již rutinou záležitostí. Takové práce probíhají dodavatelským způsobem od specializovaných firem nebo i budováním pracovišť s vlastní technologií, jak z knihoven má nejlepší Knihovna AV ČR v Jenštejně. Na technologie digitalizace navazuje problematika uložení výsledků digitalizace v datových úložištích. Tyto problematiky ovšem nejsou předmětem našeho výzkumného záměru.

Pro proces digitalizace a zpřístupňování digitálních dokumentů byly Národní knihovnou organizovány a postupně vyvíjeny speciální programy (Systém Kramerius, Sírius). Jejich stav představuje základnu, která se postupně rozšiřuje změnami technologie i požadavky na zpřístupnění digitalizovaných dokumentů podle možností i z hlediska požadavků uživatele. Do vývoje procesů spojených s digitalizací se zapojily i další organizace. Knihovna AV ČR garantuje další rozvoj a rozšíření systému Kramerius, firma INCAD rozvíjí systém registrace RD.CZ a také využívání konverze obrazových souborů pomocí OCR, která umožňuje vyhledávání jak v metadatech, tak v plných textech a při digitalizaci je u nás využívána od roku 2007.

Oblastí, kterou se zabývá náš výzkumný záměr je problematika zpřístupnění obsahu digitalizovaných dokumentů využitím možností stávajících programů pro systém Kramerius.

Vedle konverzí obrazových souborů pomocí OCR, popisem technických a administrativních metadat, sledujeme i jiné možnosti spočívající v propojení bibliografického záznamu ve struktuře formátu MARC 21 do digitalizovaných plných textů. Nejbližší k řešení tohoto záměru je výzkum na Masarykově univerzitě v Brně, kde se úspěšně řešil projekt Česká digitální matematická knihovna a výsledky jsou aplikovány v rámci projektu Metadatový editor pro systém Kramerius.⁴ Náš projekt směřuje k tomu, aby datový obsah digitální knihovny byl vytvářen vložením skenovaných stránek, jejich hierarchickým začleněním pomocí metadat (bibliografického popisu) jednotlivých částí.

Metody připravované masové digitalizace směřují k postupné konverzi periodik i monografií do elektronické podoby. U periodik jsou k dispozici metody, které nástrojem OCR umožní po zadání tematických hesel vyhledávat relevantní plné texty z digitalizovaných dokumentů. Vedle toho je k dispozici souběžná nebo retrospektivní analytická bibliografie, jejichž záznamy jsou a budou vyhledávány, zejména v oblasti společenských věd. Takové vazby, mezi záznamy bibliografických soupisů a digitalizovanými periodiky, sleduje z pohledu vývoje systému Kramerius předkládaný výzkumný záměr.

³ V edici *Česká retrospektivní bibliografie*, kterou zpracoval kolektiv bibliografů organizovaný J. Kubičkem v rámci Sdružení knihoven ČR dosud vyšlo: *Noviny České republiky od počátku do roku 1918*. Brno, 2008. – 2 sv. *Noviny České republiky 1919–1945*. Brno, 2004. – 2 sv. *Časopisy České republiky 1919–1945*. Brno, 2006–2007. – 4 sv. Dokončený a v tisku je poslední soubor *Časopisy České republiky od počátku do roku 1918*. Brno, 2010. – 3 sv.

⁴ ŠÁRFY, Martin. *Metadatový editor pro digitální knihovny*. In *Knihovny současnosti 2009*. Brno, Sdružení knihoven ČR 2009, s. 140–154.

B Analytická část

B.1 Vlastní řešení

Výzkum periodik, ať již z hlediska jejich vývoje v českých zemích od 17. stol., nebo uložení a dochování v paměťových institucích, či bibliografického zpracování, patří k dlouholetým prioritním úkolům řešitelů tohoto projektu a Moravské zemské knihovny. Především došlo k završení úkolu, který spočíval ve zpracování retrospektivní bibliografie periodik vy excerpováním katalogů knihoven paměťových institucí a zpracováním databáze PER – česká a moravská periodika do roku 1945. Tento výzkumný úkol byl řešen postupně v rámci schválených projektů Grantové agentury ČR č. 409/01/1194 a č. 409/06/1172 od roku 2001. Databáze PER je spolehlivým východiskem informujícím o výskytu jednotlivých ústředních a regionálních titulů periodik a tím je také významným zdrojem při kompletaci dokumentů připravovaných k digitalizaci. Databáze PER svými s více než 50 tisíci bibliografickými záznamy je nejrozsáhlejším informačním zdrojem bohemikálních periodik. V národním programu Kramerius zpracovala MZK již 24 titulů novin a časopisů, převážně vlastní technikou na mikrofilm. Ve své digitální knihovně vystavuje 17 titulů, z nichž nejrozsáhlejším a také nejvíce vyhledávaným titulem je deník *Lidové noviny*, vydávaný v Brně od roku 1893, s více než 220 tis. stranami.

Odborná literatura člení výzkumné metody v knihovní a informační vědě na kvalitativní a kvantitativní, které potom samostatně nebo kombinovaně aplikuje do jednotlivých procesů získávání, zpracovávání a zpřístupňování dokumentů. Procesy digitalizace se rozšiřuje zdrojová základna každé knihovny systému. Problematika digitálních knihoven je široká, nabízí řadu témat pro výzkumy. Náš výzkumný úkol se věnuje digitálním knihovnám jako prostředku pro prezentaci kulturního dědictví a je spojen s problematikou komunikace člověk-počítač (HCI = human-computer interaction). Přitom jde o využití metod analytického bibliografického zpracování a sestavení optimálního uživatelského rozhraní jako prostředku mezi informačním systémem a uživatelem.

Obor HCI se zabývá především vyhledáváním informací v informačních systémech a různých bázích dat. Je příznivě ovlivňován rozvojem informačních technologií a v našem případě jde o možnosti propojení koncového uživatele s obsahovými daty uloženými v digitálních knihovnách.

Nastupující etapa digitálních knihoven, jak spočívá v masové digitalizaci s konečným lapidárním cílem „všechno bude v počítači“, si vyžaduje z obrovského množství dokumentů a dat věnovat zvláštní přístup v oblasti periodik takovým titulům, které svým obsahem jsou významné pro dějiny národa, a to ať již z pohledu celku nebo jednotlivého regionu.

Cílem našeho výzkumu je rozšířit možnosti vyhledávání informací z vybraného souboru periodik prostřednictvím digitální knihovny Kramerius, kterou vytváří Národní knihovna ČR a Knihovna Akademie věd ČR (<http://kramerius.nkp.cz>). Přitom postupy našeho výzkumu jsou koordinovány s postupy jiných výzkumných úkolů napojených na projekt systému Kramerius ať již osobními kontakty řešitelů nebo sledováním výstupních zpráv (např. projekt Česká digitální matematická knihovna, řešený na MU a KNAV, projekt Optimalizace nástrojů pro digitalizaci, řešený NK ČR, projekt Zpřístupnění a ochrana kulturních, uměleckých a vědeckých zdrojů, řešený KNAV, Program pro tvorbu metadat pro systém Kramerius a Národní systém pro obohacování bibliografických záznamů, řešený v MZK). Ze zahraničních digitálních knihoven jsou inspirativní výsledky digitální knihovny ANNO v Rakousku (<http://anno.onb.ac.at>) a v Polsku Wielkopolska Digital Library (<http://wbc.poznan.pl/dlibra>).

Díky rozšířeným digitálním aktivitám jsou v poslední době stále více zdokonalovány technické prostředky spolu s digitalizačními postupy. Významnou oblastí při digitalizaci dokumentů jsou metadata, jejichž hlavním smyslem je poskytovat přidanou informační hodnotu k primárním datům. Stávající systém Kramerius je centralizovaný a při vyhledávání rozlišujeme z nabídky: typ dokumentu (monografie / periodika), dále u periodik název dokumentu, jeho bibliografický popis s přehledem ročníků, následuje rozpis dat u ročníku a nakonec stránka v zobrazení (djvu). Metadata u periodik tvoří celkový bibliografický záznam o periodiku a dále rozpis pro každou jednotku (datum vydání a číslo, svazek).

Ve výzkumu univerzit se objevují nové trendy, které směřují prohlížení digitálních knihoven pomocí využití nových informačních a vizualizačních nástrojů, umožňující načítání informací, které mají být zpracovány nebo anotovány. Tak např. univerzity v Curychu a Basileji představily v r. 2006 tzv. DELOS (Digital Library Management System), jenž kombinuje textové s audiovizuálním vyhledáváním a to je založeno na ISIS (Interactive Similarity Search), jenž podporuje obsahově založené vyhledávání v multimediálních sbírkách a je vhodné např. pro muzea.

Z hlediska knihovních dokumentů shrnul možnosti využití open source nejznámějších systémů pro digitální knihovnu Martin Lhoták a podrobně popsal vlastnosti českého open source systému Kramerius, který je vyvíjen ve spolupráci Národní knihovny ČR, Knihovny AV ČR a firmy Qbism technologies.⁵ Probíhá vývoj systému Kramerius verze 4, který bude založen na digitálním repozitáři FEDORA a měl by později sloužit jako prezentační vrstva pro digitální dokumenty. Opouští se od dosavadní struktury metadat (DTD pro periodika a monografie) a přechází se na kombinaci kontejnerového formátu METS (strukturální metadata) a schématu MODS (deskriptivní metadata), které lépe vyhovuje popisu digitálních dokumentů a je v souladu s rozhodnutím o ukládání a zpřístupňování dokumentů v projektu Národní digitální knihovna. V digitální knihovně Kramerius MZK bude možné vyhledávat články na základě metadat extrahovaných z knihovních lístků. S tím souvisí rozvoj metadatového editoru, který umožní zveřejnit digitální dokumenty a jejich metadata včetně těch článkových v digitální knihovně. Obojí je provázáno s Registrem digitalizace, do kterého se MZK účastní jako koordinátor krajských knihoven, které by se do projektu rády zapojily. Velice účinné by mohlo pak být vyhledávání propojení článků v časopisech na bibliografické

soupisy, jak již úspěšně se provádí při řešení projektu České digitální matematické knihovny (DML-CZ).⁶ Jde o projekt, v němž článkové metadata jsou provázána s matematickými referenčními databázemi Mathematical Reviews a Zentralblatt MATH.

Řešený projekt Digitalizace moravských knihovních sbírek je zaměřen na článkové metadata, která při očekávané masové digitalizaci by měla usnadnit vyhledávání prioritních informací národní kultury. Ve druhém roce řešení projektu se pracovalo na analytickém popisu deníku *Lidové noviny* v rubrikách kultury a ze souboru časopisů byly zpracovány měsíčník *Allgemeines europäisches Journal* (Brno, 1794–1798), *Týdenník, listy ponaučné a zábavné* (Brno, 1848–1849) a měsíčně vydávaná revue *Nový život* (Nový Jičín-Prostějov, 1896–1907). Výzkumné práce jsou popsány v následujících kapitolách.

Řešení výzkumného úkolu na příkladu analytického rozpisu periodik sleduje také možnosti využití článkových národních registrujících nebo speciálních bibliografií a jejich propojení s plnými texty digitalizovaných novin a časopisů.

⁵ LHOTÁK, Martin. *Open source pro digitální knihovnu*. In Automatizace knihovnických procesů, 11. Praha, ČVUT 2007, s. 65-72.

⁶ BARTOŠEK, Miroslav, KREJČÍŘ, Vlastimil. *Jak se dělá digitální matematická knihovna*. In Automatizace knihovnických procesů, 11. Praha, ČVUT 2007, s. 55-63.

B.1.1 Analytický rozpis deníku

Přesto, že v českých zemích vycházelo do r. 1918 celkem 2000 titulů novin a v období 1919–1945 to bylo 2276 titulů, není jich zase tak mnoho, které lze pokládat za stěžejní. Nesporně však pro 20. století hrají prim *Lidové noviny*, vydávané od r. 1893 s krátkými přestávkami až dosud. Zejména v meziválečném období byly moderním nestranným listem, jehož význam byl dán schopností šéfredaktorů (Arnošt Heinrich, Eduard Bass) a redaktorů a dopisovatelů tvořených osobnostmi vědeckého a kulturního života.

Lidové noviny, resp. články z tohoto deníku věnované kultuře, byly předmětem bibliografického analytického zpracování již před řešením tohoto výzkumného úkolu. Do té doby byly excerpovány články z let 1893 až 1903, ale práce byly pozastaveny. Zpřístupnění tohoto titulu internetem v systému Kramerius nabízelo vrátit se k původnímu záměru analytického rozpisu událostí vztahujících se ke kultuře, postupně na excerpci dále pracovat a získané záznamy propojit a tím rozšířit vyhledávací možnosti v systému Kramerius.

Pro *Lidové noviny* byly zpracovány zásady výběru článků se zaměřením na kulturní události. Excerpují se původní (ne agenturní) články referátového charakteru z jednotlivých oblastí kulturního života (divadlo, hudba, výtvarné umění) a kulturních institucí, literární texty v plném rozsahu.

V prvním roce řešení projektu byla excerptce relevantních článků z deníku provedena z let 1893 až 1903 a z let 1912 až 1914. Výsledkem bylo zpracování 8 700 záznamů a tyto retrokonvertovány z podoby psacím strojem nebo i rukou psaných katalogizačních záznamů do podoby čitelné strojem.

Analytický rozpis článků z Lidových novin za roky 2008 a 2009

Přehled podle tematických celků a roků zpracování

Tematické celky: I kultura, II jazykověda, III literatura, IV umění, V školství

Obory: I II III IV V celkem za rok
záznamy na listcích psané strojem

2008	1893	3	0	7	15	0	25
2008	1894	448	0	48	205	0	701
2008	1895	194	1	46	182	0	423
2008	1896	221	0	37	128	0	386
2008	1897	303	3	36	150	0	492
2008	1898	358	0	88	145	0	591
2008	1899	0	1	127	156	0	284
2008	1900	499	0	264	213	0	976
2008	1901	752	5	221	267	0	1245
2008	1902	0	1	184	277	0	462
2008	1903	378	2	41	225	0	646
		3156	13	1099	1963	0	6231

nově excerpované záznamy

2008	1904	275	0	197	205	0	677
2008	1905	0	0	202	187	0	389
2008	1906	319	0	227	139	0	685
		594	0	626	531	0	1751

záznamy na listcích částečně psané rukou

2009	1907	446	0	0	53	0	499
2009	1908	473	8	0	0	0	481
2009	1909	0	0	324	0	0	324
2009	1910	886	0	0	0	0	886
2009	1911	341	0	0	0	0	341
		2146	8	324	53	0	2531

záznamy na listcích částečně psané rukou

2008	1912	0	0	241	242	0	483
2008	1913	484	0	484	100	0	1068
2008	1914	538	0	190	190	0	918
		1022	0	915	532	0	2469

nově excerpované záznamy

2008	1915	115	8	130	75	28	356
2008	1916	120	6	136	163	36	461
2008	1917	107	14	121	150	54	446
2008	1918	240	13	134	159	76	622
		582	41	521	547	194	1885

nově excerpované záznamy

2009	1919	380	7	176	271	195	1029
2009	1920	557	18	593	511	149	1828
2009	1921	907	37	1112	592	191	2839
		1844	62	1881	1374	535	5696

záznamy na listcích psané strojem

2009	1940	594	68	2361	1818	8	4849
2009	1941	440	155	2229	1492	6	4322
2009	1942	435	193	1882	1620	8	4138
		1469	416	6472	4930	22	13309

Na zpracování článků v Lidových novinách roce 2009 pokračovaly práce podobně jako v 1. roce řešení projektu. Odborný knihovník – bibliograf excerpoval ještě nevyexcerpované ročníky LN 1919 – 1921 a zpracovával záznamy do připravené tabulky. Třídění do tematických celků se dodržovalo stejné, jako bylo v už vyexcerpovaných záznamech: kultura, jazykověda, literatura, umění, školství. Záznamy kromě bibliografické citace obsahují věcné třídění, příp. osobu pro předmětový rejstřík.

Z dříve provedené excerpcce kulturních článků *Lidových novin* byly z lístkové kartotéky zpracovány ročníky 1907 – 1911 a 1940 – 1942, záznamy na lístcích mají různou kvalitu a některé byly psány na psacím stroji a některá období rukou. Ručně psané záznamy byly přepisovány do formuláře ručně, ostatní byly naskenovány na lístkovém skeneru a převedeny do textu nástrojem OCR, ručně upraveny a zformátovány. Ve všech případech vytváření záznamů, excerpce, ručním přepisem nebo OCR převodem je výstupem strukturovaný formát záznamu v připravené .xls tabulce, která obsahuje sloupce: věcné třídění, autor, název článku, podnázev, koho se článek týká, bibliografická citace.

Zpracování záznamů se provádí do excelových tabulek. Ukázka tabulky, jejíž data budou na základě analýzy převáděna pomocí jednoduchého programu do .xml ve formátu DTD pro periodika, který je schopný pojmut systém Kramerius.

A	B	C	D	E	F	G	H	I	J
1	Slaměník, František	Ludvík Kellr mrtev	Předseda berlínské Společnosti Komenského - nekrolog	23	19150322	74	ranní	9	
3	Digamma [=František Votruba]	Ze Slovenska	Fejeton	23	19150310	69	ranní	2	
3	Dr. Š.	Jak se uzdravují	Fejeton	23	19150326	78	ranní	1	
3	št. [=Bohuslav Štěchovský]	V novinách nad čarou...	Fejeton	23	19150403	86	odpol.	2	
3	-is [=Josef Hais?]	Polární zář	Fejeton	23	19150302	61	ranní	2-3	
5		Poslední Mohykáni	Reakce na fejeton Obec pensistů z 21. 4.	23	19150505	117	ranní	1-2	
3	Kompiš, Peter ml.	Jako ujček dedil synov	Povídka	25	19170114	12	ranní	1-2	
3	Rameau, Jean	Mařenčin román	Román na pokračování	23	19150409- 19150702	91- 175	odpol.	5-6, 3	
3		Alois Kulíšek, domobranec		23	19150328, 19150404, 19150418, 19150425, 19150504, 19150516	80, 87, 93, 100, 107, 116, 128	ranní, ranní, ranní, ranní, ranní, ranní, ranní	1-2, 1-2, 1-2, 1-2, 1-2, 1-2, 1-2	

Popis tabulky: sloupec A: téma (1 kultura, 2 jazykověda, 3 literatura, 4 umění, 5 školství), sloupec B: autor, sloupec C: název, sloupec D: podnázev, sloupec E: charakteristika článku, sloupec F: ročník, sloupec G: datum ve formátu RRRRMMDD, sloupec H: číslo, sloupec I: vydání – ranní, odpolední, sloupec J: rozsah stran

B.1.2 Analytický rozpis časopisů

Do projektu bylo zahrnuto 10 titulů periodik a v 1. roce řešení projektu byly zpracovány a v edici *Rejstříky moravského tisku* vydány tiskem bibliografické soupisy *Moravsko-slezský Časopis pro lid*. Brno, 1842–1846 – 627 záznamů, *Magazin nützlicher und angenehmer Lektüre aus verschiedenen Fächern für Leser aus allen Ständen a Mährisches*

Magazin. Brno, 1782 a 1789 – 159 záznamů, *Večery : beletristická příloha Lidových novin*. Brno, 1911–1914 – 828 záznamů. Ve 2. roce řešení projektu byly zpracovány a tiskem vydány další tři soupisy:

Allgemeines europäisches Journal. – Roč. 1794, Bd. 1 (Juli 1794) – 1798, Bd. 12 (Dezember 1798) – 926 záznamů [Měsíčník, kdy každé číslo mělo 220 stran, přetiskoval převážně články ze zahraničních časopisů, byl šířen do všech zemí bývalé monarchie.]

Týdenník, listy ponaučné a zábavné. – Roč. 1, č. 1 (6. 1. 1848) – 2, č. 3 (25. 1. 1949) – 484 záznamů [První český list vydávaný v Brně pravidelně týdně se dvěma přílohami, redigovaný Janem Ohéralem a Františkem Matoušem Klácellem.]

Nový život. – Roč. 1, č. 1 (leden 1896) – 12, č. 12 (prosinec 1907) – 2793 záznamů [Měsíčně vydávaný orgán Katolické moderny, vydávaný a redigovaný Karlem Dostálem-Lutinovem.]

Analytický rozpis obsahu byl proveden pro všechny články, zatím co v prvním roce bylo ve třech soupisech zpracováno 1614 záznamů, ve druhém roce zdalejších tří soupisů bylo získáno 4203 záznamů. Pro každý z těchto časopisů bylo shodné, že jejich úplnost bylo obtížné dosáhnout. Zcela se to nepodařilo u nejstaršího časopisu, který v brněnských knihovnách je dochován torzovitě a bylo jej třeba dokonpletovat ze zahraničních sbírek. Rovněž pro nejmladší časopis *Nový život* bylo třeba z jiných knihoven vypůjčit ročníky k získání ilustrovaných obálek, které se u exemplářů vydávající knihovny nedochovaly. Každý soupis byl doplněn referátem o vývoji časopisů a po věcně seřazených záznamech následoval jejich autorský a předmětový rejstřík. Bibliografické soupisy byly zpracovány ve wordu a záznamy formátovány do tabulky .xls, aby byly využity pro doplnění metadat v systému Kramerius.

B.1.3 Informační technologie

Vzhledem k tomu, že v první fázi projektu bylo možno vystačit s existujícími nástroji, zkoumání se zaměřilo na možnosti prezentace vytvářených dat. Ve spolupráci s firmou INCAD řešíme zapojení Moravské zemské knihovny do projektu registru digitalizace. Tato spolupráce je zatím v přípravné fázi, předpokládáme ale výrazný pokrok ještě v tomto nebo začátkem příštího roku, ve kterém se také budeme intenzivněji věnovat možnostem rozvoje systému Kramerius v souvislosti s jeho přechodem na platformu Fedora Commons. V průběhu podzimu proběhly i první testy možnosti využití nedávno vydaného image serveru djatoka ve spolupráci s Krameriem, v této oblasti očekáváme do konce roku další pokrok s cílem eliminovat nutnost používat pro zpřístupnění digitalizovaných dokumentů formát djvu. Došlo i k testům možnosti automatické extrakce údajů z dokumentu do formátu METS/ALTO, které byly poměrně úspěšné. Většímu využití softwaru firmy CCS však zatím brání jeho cena. Zajímavé přitom bylo zjištění, že dvouvrstvé pdf, generované systémem CCS má při stejném stránkovém rozsahu a pro čtení srovnatelné kvalitě poloviční velikost než jednovrstvé pdf, obsahující jen obrazy stránek. Tento test také jednoznačně ukazuje, že dosavadní způsob tvorby OCR při digitalizaci do podoby samostatně stojícího textového souboru není dostačující, protože neumožňuje vytváření dvouvrstevných pdf, které jsou pro uživatele výrazně komfortnější. Této problematice se budeme dále věnovat.

B.2 Přínos řešitele

V souladu s ročním programem byly excerpovány záznamy z deníku Lidové noviny a zpracovány rozpisy vybraných periodik s cílem posoudit technické možnosti převodu bibliografických záznamů do strojově čitelné podoby.

V programátorské práci se sledovaly projekty směřující k prohlubování vyhledání dat v systému Kramerius a přínosem našeho projektu byla analýza článkových metadat a

zjišťování, jak se která metadatová schémata chovají při převodu bibliografického popisu článků.

B.3 Posun znalostí

Propojení bibliografických záznamů do skenovaných textů je v zásadě možné a k tomu směřují i práce na nové verzi 4 systému Kramerius. I když řešení problému si může vyžadovat větší granularitu schématu bibliografického záznamu, jak vyplynulo z analýzy excelové tabulky, jde především o provázání záznamů v národních nebo oborových bibliografiích na plné texty dokumentu.

C Návrhová část

C.1 Výsledky řešení

Výsledky řešení jsou v souladu s cíli projektu stanovenými pro rok 2009.

V analytickém zpracování bylo z celkového plánovaného počtu 60 000 bibliografických záznamů dosud zpracováno 39 689 záznamů, z toho 25 739 v r. 2009. Současně byly vydány tři soupisy v edici *Rejstříky moravského tisku*.

Úprava systému Kramerius pro zpracování metadat článků představuje náročnou programátorskou práci a je zásadně řešena projekty garantovanými Národní knihovnou ČR a Knihovnou AV ČR. Přínos našeho projektu spočíval v provedení analýzy převodu dat a ve spolupráci s firmou Incad, která verzi Kramerius 4 vyvíjí.

C.2 Závěr

Ve 2. roce trvání projektu byly splněny proponované práce v excerpci souboru periodik, jak je představují deník *Lidové noviny* a vytipované tituly časopisů z 18. a 19. stol. Je nadějně, že systém Kramerius v nové verzi umožní vyhledávat digitalizované texty podle citací článků. Zpřístupnění verze 4 systému Kramerius je plánována na r. 2010 a programátorské práce našeho projektu přispívají k takovému řešení.

C.3 Návrhy opatření

K dokončení plánovaného projektu jsou dány předpoklady dosavadními výsledky jak v úseku bibliografické heuristiky, tak spoluúčastí na programování nové verze Kramerius 4. V programování proto chceme přispět analýzou a převodem metadat do formátu METS pro strukturální a MODS pro deskriptivní metadata a prověření možností schématu METS/ALTO.

Pro naplnění cílů projektu proto navrhujeme zajistit kontinuitu v dalším financování projektu ve shodě s projektem.

D Použití finančních prostředků

E Resumé a klíčová slova

E.1 Resumé a klíčová slova v češtině

Digitalizace jako proces reformátování klasických informačních zdrojů má vedle ochrany papírových nosičů usnadnit přístup k dokumentům. Postupně se objevují projekty realizované velkými knihovnami u nás i v zahraničí, jejichž záměrem je digitalizace knihovních sbírek a jejich přenos do el. databází přístupných internetem. Cílem projektu je propojení metod bibliografie do digitalizace tak, aby došlo k novým možnostem ve zpřístupnění vybraných titulů periodické literatury. Bibliografické záznamy se stávají součástí digitalizovaných plných textů. Cílem projektu je provázání bibliografického popisu/metadat ve vyhledávacích možnostech systému Kramerius a to postupně na vzorku vybraných titulů časopisů a deníku *Lidové noviny*. S respektováním vyhledávacího nástroje OCR půjde o přístup na jednotlivé bibliografické záznamy analytického rozpisu souboru dokumentů. Spolupráce při vytvoření editačního a konverzního nástroje na tvorbu metadat a rozšíření funkcionality externího modulu systému Kramerius o možnost přejímat a využívat data (zejména bibliografické záznamy) vytvořená v jiných systémech (knihovních).

Klíčová slova:

Digitalizace, Digitální knihovny, Systém Kramerius, Informační technologie, Knihovní sbírky, Bibliografické texty, Metadata

E.2 Abstract and key words in English

Digitalization as a process of reformatting of classical information sources next to the preservation of printed materials has to facilitate an access to the documents. There are gradually appearing projects realized by large libraries both in our country and abroad that are focused on digitization of librarian collection and their subsequent conversion into the electronic databases accessible via Internet. The aim the project is to join the method of bibliography with the digitization and in this way to create new possibilities for availability of select titles of the periodical literature. Bibliographical records become a part of digitized full text. The goal of the project is to link the bibliographical description/metadata in the search fields of the systém Kramerius on the sample of select titles of journals and daily *Lidové noviny*. Respecting the searching tool OCR the access to the individual bibliographical records of analytical description of the documents file will be possible. Cooperation for editing and converting through expanded functionality of the external program unit of the „System Kramerius“. The functionality to receive and make use of data (especially bibliographic records) generated in other systems (librarian systems) should be achieved.

Key words:

Digitization, Digital Library, System Kramerius, Information Technologies, Librarian Collections, Bibliographical Records, Metadata