

SOUBORNÁ DATABÁZE
KOOPERAČNÍHO SYSTÉMU ČLÁNKOVÉ BIBLIOGRAFIE
- OPTIMALIZACE INTEGRACE A SPRÁVY HETEROGENNÍCH DAT
Souhrnná zpráva za rok 2000

Předkládá: PhDr. Vojtěch Balík, ředitel NK

Zpracovala: PhDr. Ivana Anděrová, hlavní řešitelka



Praha, listopad 2000

OBSAH

A

Konstatační část

- A.1 Rešerše
- A.2 Současný stav ve světě a v ČR
- A.3 Cíl, vstupní data

B

Analytická část

- B.1 Vlastní řešení
 - B.1.1 Linka automatické indexace
 - B.1.2 TOPIC
 - B.1.3 Standardizace - bibliografická metadata ve formátu UNIMARC a metadata DUBLIN CORE v plných textech
 - B.1.4 Propojení bibliografického záznamu s plným textem
 - B.1.5 Souborná databáze ANL Kooperačního systému článkové bibliografie (KOSABI)
 - B.1.6 Česká národní bibliografie - řada Články v českých novinách, časopisech a sbornících na CD-ROM
 - B.1.7 Management Kooperačního systému článkové bibliografie (aplikace MNG KOSABI)
 - B.1.8 HW a SW podpora Kooperačního systému článkové bibliografie, bibliografické báze ANL a plnotextové databáze ANL FULL
 - B.1.9 Smluvní ošetření projektu a Kooperačního systému článkové bibliografie
- B.2 Přínos řešitele
- B.3 Posun znalostí

C

Návrhová část

- C.1 Výsledky řešení
- C.2 Závěr
- C.3 Návrhy opatření

D

Použití finančních prostředků

- D.1 Komentář
- D.2 Tabulky

E

Resumé a klíčová slova

- E.1 Resumé a klíčová slova v češtině
- E.2 Abstract and key words in English

A Konstatační část

A.1 Rešerše

Tištěné dokumenty:

Národní bibliografie - analytický popis : příručka pro zpracovatele / Ivana Anděrová [et al.] . - Praha : Národní knihovna, 1993. - 412 s. - Revize 1, 1993; Revize 2, 1997.

Současný stav a perspektivy kooperačního systému článkové bibliografie / Ivana Anděrová. - Národní knihovna : knihovnická revue. - ISSN 0862-7487. - Roč. 6, č. 1 (1995), s. 39-42.

Záznam pro soubornou databázi : UNIMARC. Fyzicky nesamostatné části dokumentů. Tištěné monografie a seriály / Pracovní skupina pro analytické zpracování, Rada pro katalogizační politiku. - 1. vyd. - Praha : Národní knihovna České republiky, 1999. - 45 s. - (Standardizace ; č. 19) . - Určeno k připomínkám.

Záznam pro soubornou databázi : Výměnný formát. Fyzicky nesamostatné části dokumentů. Tištěné monografie a seriály / Pracovní skupina pro analytické zpracování, Rada pro katalogizační politiku - 1. vyd. - Praha : Národní knihovna České republiky, 1999). - 39 s . - (Standardizace ; č. 20) . - Určeno k připomínkám.

Metodika popisu článků ve formátu UNIMARC - podklad pro interpretace AACR2R : verze 1.1 (14.4.2000) / Ivana Anděrová . 103 s. - Pracovní materiál.

Vyhledávání v databázích plných textů / Vlastimil Červený. - Národní knihovna : knihovnická revue. - ISSN 0862-7487. - Roč. 10, č. 1 (1999), s. 6-12.

K otázkám pojmu, třídění a typologie internetových a webovských informačních zdrojů / Eva Bratková. - Národní knihovna : knihovnická revue. - ISSN 0862-7487. - Roč. 9, č. 5 (1998), s. 262-276.

Topic : systém pro inteligentní vyhledávání dokumentů. - Praha : Tovek, 19?. - 77 s.

What is the hybrid library? / Charles Oppenheim, Daniel Smithson. - Journal of Information Science. - Vol. 25, no. 2 (1999), s. 97-112.

Metadata jako nový nástroj pro komunikaci webovských informačních zdrojů / Eva Bratková. - Národní knihovna : knihovnická revue. - ISSN 0862-7487. - Roč. 10, č. 4 (1999), s. 178-195.

Elektronické dokumenty:

Helsinská univerzitní knihovna přívětivá vůči každému / Filip Vojtášek, Iva Celbová. - Ikaros [online]. - Č. 9 (2000). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/2000/c09/helsinky.htm](http://ikaros.ff.cuni.cz/ikaros/2000/c09/helsinky.htm).

Knihovny zaujmou pozornost médií neobvyklými událostmi / Filip Vojtášek. - Ikaros [online]. - Č. 9 (2000). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/2000/c09/tyden.htm](http://ikaros.ff.cuni.cz/ikaros/2000/c09/tyden.htm).

Intelligence systémů zpracování textů / Zdeněk Jonák. - Ikaros [online]. - Č. 1 (2000). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/2000/c01/isko/z_jonak.htm](http://ikaros.ff.cuni.cz/ikaros/2000/c01/isko/z_jonak.htm).

Interconnectivity and the Hybrid Library / Inge HEIJTING. - Ikaros [online]. - Č. 10 (1999). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c10/ebsco.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c10/ebsco.htm).

Elektronické publikování / Martin Svoboda. - Ikaros [online]. - Č. 3 (1999). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c03/elpubl98/index.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c03/elpubl98/index.htm)

Trendy v rozvoji informačních služeb / Richard Papík. - Ikaros [online]. - Č. 8 (1999). Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_papik.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_papik.htm).

Když se řekne digitální knihovna ... / Daniela Tkačková. - Ikaros [online]. - Č. 8 (1999). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_tkacikova.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_tkacikova.htm).

Veřejné informační služby knihoven - nový program pro občany a knihovny / Michal Hora, Vít Richter. - Ikaros [online]. - Č. 8 (2000). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/2000/c08/visk.htm](http://ikaros.ff.cuni.cz/ikaros/2000/c08/visk.htm).

Sdílená katalogizace a CASLIN / Gabriela Krčmářová. - Ikaros [online]. 2000, č. 8. - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/2000/c08/caslin.htm](http://ikaros.ff.cuni.cz/ikaros/2000/c08/caslin.htm).

TEXTQUEST: software pro obsahovou analýzu / Zdeněk Jonák. - Ikaros [online]. 2000, č. 5. - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/2000/c05/text.htm](http://ikaros.ff.cuni.cz/ikaros/2000/c05/text.htm).

Elektronické zdroje publikované v síti Internet jako součást České národní bibliografie / Ludmila Celbová. - Ikaros [online]. 2000, č. 6. - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/2000/c06/elzdroje.htm](http://ikaros.ff.cuni.cz/ikaros/2000/c06/elzdroje.htm).

Pojem "informace" ve světě sdíleného pojetí skutečnosti / Zdeněk Jonák - Ikaros [online]. Č. 2 (2000). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/2000/c02/veda.htm](http://ikaros.ff.cuni.cz/ikaros/2000/c02/veda.htm).

Intelligence systémů zpracování textů / Zdeněk Jonák. - Ikaros [online]. - Č. 1 (2000). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/2000/c01/isko/z_jonak.htm](http://ikaros.ff.cuni.cz/ikaros/2000/c01/isko/z_jonak.htm).

Pokles důvěry ve vědu jako důsledek změny paradigmatu vědy : důsledky změny paradigmatu v informační vědě. Část 1. / Zdeněk Jonák - Ikaros [online]. - Č. 2 (1999). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c02/veda.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c02/veda.htm)

Reflektuje teorie informace a komunikace dostatečně na zvýšený zájem společenských věd o semiotické a komunikační aspekty života? / Zdeněk Jonák. - Ikaros [online]. - Č. 3 (1999). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c03/veda2.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c03/veda2.htm).

Krise mezilidské komunikace v období komunikační a informační exploze / Zdeněk Jonák. - Ikaros [online]. 1999, č. 5. - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c05/veda4.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c05/veda4.htm).

Vztah komunikační a obsahové struktury literárního díla / Zdeněk Jonák. - Ikaros [online]. - Č. 6 (1999). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c06/kom.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c06/kom.htm).

Právní aspekty poskytování knihovních elektronických a reprografických služeb / Jarmila Burgetová. - Ikaros [online]. - Č. 6 (1999). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c06/repro.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c06/repro.htm).

Elektronické časopisy a jejich vliv na infrastrukturu vědeckých znalostí / Jaroslav Pokorný. - Ikaros [online]. - Č. 8 (1999). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_pokorny.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_pokorny.htm).

"Computing in Humanities", čili: Táhneme, anebo jsme vlečeni? / Zdeněk Uhlíř. - Ikaros [online]. - Č. 11 (1999). - Dostupný z: [URL: http://ikaros.ff.cuni.cz/ikaros/1999/c11/computing.htm](http://ikaros.ff.cuni.cz/ikaros/1999/c11/computing.htm).

Projects at the Royal Library in Stockholm, Sweden [online] . Stockholm : Royal Library, updated July 1, 1999 . - Dostupný z: [URL: http://www.kb.se/ENG/projekt.htm](http://www.kb.se/ENG/projekt.htm).

Description of the Nordic Metadata project : Cataloguing, Indexing and Retrieval of Digital Documents [online] / Juha Hakala. - Helsinki (Finsko) : Helsinki University Library, [1996]. - Dostupný z: [URL: WWW: http://linnea.helsinki.fi/meta/projplan.html](http://linnea.helsinki.fi/meta/projplan.html).

Metadata [online]. - Bath (Anglie) : UKOLN, last updated 16-Feb-2000. - Dostupný z: [URL: http://www.ukoln.ac.uk/metadata/](http://www.ukoln.ac.uk/metadata/).

Cobra+ : Computerised Bibliographic Record Actions [online]. - Boston Spa (Velká Británie) : COBRA+, 1997. - Dostupný z: [URL: http://portico.bl.uk/gabriel/en/projects/cobra.html](http://portico.bl.uk/gabriel/en/projects/cobra.html).

Dublin Core Metadata Initiative [online]. - Dublin (Ohio, USA) : OCLC, 2000. - Dostupný z: [URL: http://purl.org/dc/](http://purl.org/dc/).

The Nordic Metadata projects [online]. Helsinki (Finsko) : Helsinki University, 1996, last updated 21 February 2000. - Dostupný z: [URL: http://linnea.helsinki.fi/meta/](http://linnea.helsinki.fi/meta/).

Dublin Core Metadata Template [online] / Traugott Koch, Maattias Borell. - Mattias.Lund (Švédsko) : Lund universitetsbibliotek, 1997, last update 1997-08-20. - Dostupný z: [URL: http://www.lub.lu.se/metadata/DC_creator.html](http://www.lub.lu.se/metadata/DC_creator.html).

Nordic Countries URN-generator : provided by the Nordic Libraries [online]. - Lund (Švédsko) : Lund universitetsbibliotek, 1997]. - Dostupný z: [URL: http://www.lub.lu.se/dc/urntest.pl](http://www.lub.lu.se/dc/urntest.pl).

DOI, the Digital Object Identifier System [online]. - Kidlington (Oxford, Velká Británie) : International DOI Foundation, 1998, updated 4 April 2000. - Dostupný z: [URL: http://www.doi.org/](http://www.doi.org/)

Uniform Resource Names (urn) Charter [online]. - Reston (VA, USA) : IETF, last modified 03-Jun-99. - Dostupný z: [URL: http://www.ietf.org/html.charters/urn-charter.html](http://www.ietf.org/html.charters/urn-charter.html).

Cataloguing Internet Resources [online] / Nancy B. Olson. - Dublin (Ohio, USA) : OCLC, c1997. - Dostupný z: [URL: http://www.purl.org/oclc/cataloging-internet](http://www.purl.org/oclc/cataloging-internet).

SICI Generator. - [Dostupný z: URL: http://www.ep.cs.nott.ac.uk/~sgp/sicisend.html](http://www.ep.cs.nott.ac.uk/~sgp/sicisend.html).

Serial Item and Contribution Identifier. - Dostupný z: [URL: http://sunsite.berkeley.edu/SICI/version2.html](http://sunsite.berkeley.edu/SICI/version2.html).

Sborník příspěvků ze semináře CASLIN '99 - Souborné katalogy: organizace a služby. - Dostupný z: [URL: http://www.caslin.cz:7777/caslin99/prispevky.html](http://www.caslin.cz:7777/caslin99/prispevky.html)

Výsledky práce **Kooperačního systému článkové bibliografie - báze ANL** lze nalézt WWW adrese: page.php3?page=oazp_www.nkp.cz

Výsledky práce společnosti **Anopress, s.r.o.** lze nalézt na adrese: <http://www.anopress.cz>.

A.2 Současný stav ve světě a v ČR

A.2.1 Obecně

Databáze citací článků a další sekundární zdroje informací patří v současné době mezi standardní služby, poskytované uživatelům knihoven spolu s přístupem do katalogů. Dostupné jsou z mnoha zdrojů rozptýlených po síti, z lokálních připojení, ze systémů CD-ROM. Technologie jako Z39.50 umožňují zavádění konzistentních uživatelských rozhraní pro širokou škálu databází přístupných po síti. Většina uživatelů používá databáze sekvenčně (vždy jen jednu), roste potřeba rozhraní, které by slučovalo záznamy, získané z několika databází do logické "souborné" databáze. Elektronické dokumenty jsou zpřístupňovány prostřednictvím nakladatelství, distributorských firem, informačních institucí či služeb a jejich produktů, dále pak prostřednictvím digitálních knihoven a služeb vznikajících na základě projektů. Přístup k plným textům je zajišťován přes různé formy bibliografií a soupisů, obsahů časopisů a plnotextových databází. Vyhledávání v plných textech zvyšuje komfort přístupu uživatelů k informacím. Elektronické dokumenty jsou zpřístupňovány v dohodnutých formátech, např. JPEG, GIF, PDF, TIFF, HTML. Služby knihoven jsou založeny na typu služby "document delivery". Poměrně dobře jsou zpřístupňovány plné texty novin, týdeníků aj. časopisů. Problém vytváření vazeb na primární obsah se v současnosti soustřeďuje na článkové databáze proto, že technologie přístupu k datům v síti celkem dobře umožňuje přístup k článkům v elektronické podobě, zatímco přístup k jiným typům dokumentů je problematičtější. Kromě vyhledávačů typu "search engines" (Alta Vista aj.) nebo předmětových katalogů Internetu (Yahoo! aj.) registrující informační zdroje v nestrukturované podobě a u nichž relevance jejich zpětného vyhledávání je značně problematická, se přímo v prostoru Internetu a webu objevují registrační systémy, které přistupují ke zpracování těchto zdrojů přes strukturované záznamy. Tyto údaje mohou být obsažené ve zdrojích samotných (metadata). Pro popis webovských informačních zdrojů

navržen formát Dublinské jádro (DC) jako základní soubor údajů pro popis zdrojů. [Dublin Core](#) může být vytvářen autorem, vydavatelem, distributorem těchto zdrojů.

A.2.2 Současný stav v zahraničí

Některé digitální knihovny, služby a projekty zabývající se zpřístupněním sekundárních informací o člancích a zpřístupněním plných textů

- [UNCoverWeb - A Current Awareness and Un Cover](#) je databáze registrující články z časopisů na základě obsahů přebíraných asi z 17000 titulů periodik. Databáze obsahuje stručné záznamy více než 7000000 článků, které vyšly v odborných a vědeckých časopisech od r. 1988. Hlavním cílem služby je poskytování plných textů článků - placená služba.
- Nejkompletnější přístup k elektronickým časopisům nabízí [OCLC FirstSearch Electronic Collection Online](#). Interface umožňuje prohledávání periodik i čísel podle různých kritérií, přístup k citacím časopisů je zdarma, přístup k abstraktům a plným textům je možný jen u předplacených titulů, OCLC podporuje konzorciální přístup včetně přístupu kombinovaného s individuálním a "document delivery".
- Ve **Švédsku** jsou články prezentované v systému [LIBRIS](#). Záznam článku je možno zobrazit ve zkrácené podobě i ve struktuře MARC. Formulář obsahuje hypertextové odkazy na knihovny, v jejichž fondu se titul nachází.
- V [DBC \(Dánském knihovnickém centru\)](#) se zpracovávají články a recenze v rámci báze **BASIS** (ročně 30000 článků a 20000 recenzí z 9000 dánských periodik). Toto centrum buduje souborný katalog [DANBIB](#), přes který lze zaslat objednávku elektronické kopie článku. Je propojený se švédským souborným katalogem [LIBRIS](#) a norským souborným katalogem [BIBSYS](#).
- Universitní knihovna v **Helsinkách** (plní funkci Finské národní knihovny) provozuje centrální knihovnický systém VTLS síť **Linnea**, v rámci které je zpřístupňována databáze článků [ARTO](#) z tisíce finských periodik.
- [NORDINFO](#) - projekt skandinávského souborného virtuálního katalogu zohledňuje aspekty specifické pro severské země - předmětová hesla, klasifikační systémy, národní systémy identifikátorů.
- [The Nordic Metadata Project](#) - kooperační projekt severských zemí jako jeden z prvních řeší problematiku metadat [Dublin Core](#) v rámci spolupráce Norska, Dánska, Švédska, Finska a Islandu.
- Systém knihoven [Oxfordské university](#) zpřístupňuje rozsáhlou sbírku elektronických dokumentů v rámci Electronic Reference Library. Záznamy článků obsahují krátké citace, abstrakty a možnost získání elektronické kopie ve formátu PDF.
- Program [PICA](#) v Holandsku zajišťuje přístup k centralizovaným bibliografickým databázím a zejména vytváří tzv. Otevřenou síť knihoven (OBN - Open Bibliothek Network), v rámci které je umožněno elektronické propojení knihoven s databází článků.
- V **Německu** je vyvinut projekt **JADE** (Journal Articles Database) ve spolupráci s British Library. Obsahuje pouze krátké citace článků (tato báze obsahuje i záznamy z NKČR). JADE je doplněna projektem [JASON](#) (Journal Article Send On Demand). JASON umožňuje dodávání článků v elektronické podobě z německé databáze časopisů.

A.2.3 Současný stav v ČR

Některé digitální knihovny, služby a projekty zabývající se zpřístupněním sekundárních informací o člancích a zpřístupněním plných textů

České **nakladatelské elektronické zdroje** na Internetu jsou ve stadiu vývoje a hledání podoby. Vztahy mezi uživateli, knihovnami a vydavateli/nakladateli nejsou dosud jasné z hlediska právního i obchodního, v budoucnu lze předpokládat v tomto směru vznik nových iniciativ.

- Nakladatelství [Economia, a.s.](#), nakladatelství ekonomické literatury vystavuje na Internetu plné texty Hospodářských novin a Ekonomů. Je možno předpokládat vystavení dalších titulů z produkce tohoto nakladatelství. Objevují se specializovaná elektronická nakladatelství (např. Sagit zpřístupňuje plná znění zákonů aj.).
- [Albertina icome](#) je česká soukromá společnost zaměřená na přístupnost profesionálních informačních zdrojů v elektronické formě a jejich využití v praxi. AiP nabízí přes 1000 elektronických titulů předních světových vydavatelství. Elektronické vydavatelství spolupracuje na vydávání ČNB na CD-ROM.
- V České republice existují některé oborové báze plnotextových informací, např. [ASPI \(Automatizovaný systém právních informací\)](#) zachycuje vývoj právní kultury, vztahující se k území současné České republiky i Slovenské republiky v rozsahu dvou století s výhledem na legislativu Evropských společenství.
- [Akademie věd ČR](#) zpřístupňuje na Internetu current contents a plné texty článků publikovaných pracovníky AV prostřednictvím jednotlivých redakcí časopisů (plné texty zatím ojedinele).
- V rámci [Parlamentní knihovny](#) se buduje systém, ve kterém jsou zpřístupněna v plné formě parlamentária.

Relativně velký rozvoj na Internetu nastal v nabídce **českých novinových a časopiseckých elektronických zdrojů** - jsou vystaveny deníky, týdeníky a časopisy s různou hloubkou retrospektivy a úplnosti od volně přístupných přes registraci a služby placené. V některých elektronických zdrojích lze vyhledávat plnotextově.

Např. [Seznam](#) katalogový a vyhledávací server se službou Kompas, která umožňuje plnotextové vyhledávání www stránek českého Internetu. V lednu 1999 se změnila základní podoba Trafiky - virtuální [Trafika](#) se mění v pravý český portál. Je možno zde najít informace o člancích z novin a časopisů, dále pak elektronické časopisy a magazíny vydávané M.I.A (Svět Namodro aj.). Trafika nabízí z vlastní produkce následující tituly aktuálně v portálové podobě: Mladá fronta Dnes, Lidové noviny, Právo, Slovo, Týden, Mladý svět, Respekt aj. Politika, metody a strategie vystavování těchto zdrojů na českém Internetu se často mění, u některých je však možnost vysledovat určitou stálost a uvažovat o propojení s analytickými záznamy. Propojováním s volně přístupnými zdroji na Internetu však musí být velmi obezřetné.

Na českém informačním trhu působí **dvě společnosti**, které se zabývají zpřístupňováním plných textů programově. Společnost [ANOPRESS, s.r.o.](#) a společnost [Newton I.T., s.r.o.](#) Obě společnosti získávají na základě smluv s jednotlivými vydavateli plná znění deníků a dalších periodik. Převod článků do tvaru vhodného k dalšímu zpracování se děje pomocí

vlastních patentových postupů a zajišťuje věrnost původní předlohy. Obě společnosti vlastní archiv titulů celostátních, regionálních a dalších včetně jejich mutací, dále pak přepisy televizních a rozhlasových pořadů. Poskytované služby obou společností se však liší.

- [Newton I.T., s.r.o.](#) poskytuje plné texty v rámci služby Media Monitoring na základě individuálních požadavků. Neumožňuje přímý přístup do celé databanky.
- [ČTK](#) je národní informační agentura a zabývá se sběrem, zpracováním a distribucí zpravodajství a informací ze všech oblastí lidské činnosti.
- [ANOPRESS, s.r.o.](#) umožňuje on-line přístup do databanky plných textů [TAMTAM](#), na jejíž bázi poskytuje následné služby. Společnost zpřístupňuje informace zákazníkovi na dané téma. [ANOPRESS, s.r.o.](#) umožňuje přístup do databanky novin on-line na základě licenčních smluv a umožňuje nákup celých titulů periodik. Společnost Anopress je výhradním zpracovatelem elektronické podoby [většiny českých regionálních titulů](#) (51 titulů nakladatelství Bohemia). Pro zpřístupnění plných textů ve veřejných knihovnách bylo založeno [Konzorcium Anopress](#). Společnost je výhradním zástupcem slovenské firmy SLOVAKIA ONLINE v ČR, která zpracovává elektronickou podobu slovenských tištěných médií. Kromě mediální části obsahuje databanka [TAMTAM](#), i část vědomostní, v níž jsou k dispozici pro fulltextové vyhledávání různé encyklopedie, příručky a další knihy referenčního charakteru.

Společnost vyvinula vlastní software ISA, který umožňuje všechna data dále analyticky zpracovávat, exportovat je v několika formátech, včetně HTML, pro Internet či Intranet. Vyhledávací systém [TOPIC](#), který [ANOPRESS, s.r.o.](#) používá k monitoringu a analýze informačních zdrojů, je v současnosti jediným interaktivním systémem na českém trhu.

Automaticky vyhodnocuje relevanci dokumentů a umožňuje jejich řazení podle důležitosti. Na rozdíl od zdlouhavého fulltextového vyhledávání jde v tomto případě o pojmové, tzv. inteligentní vyhledávání, šité přímo na míru požadavkům uživatele. Systém využívá fulltextovou technologii americké firmy Verity.

[ANOPRESS, s.r.o.](#) ve spolupráci s Národní knihovnou vytváří v rámci tohoto projektu technologii, která umožní propojit bibliografické záznamy knihovny s plnými texty článků z databáze Anopress, dále pak vkládat bibliografická metadata do analytických záznamů a metadata typu [Dublin Core](#) do plných textů.

[ANOPRESS](#) zpracovává cca 35 titulů, které odpovídají excerpční základně [Kooperačního systému článkové bibliografie](#).

Bibliografické zpracování článků v ČR je poměrně rozsáhlé co do zdrojů, které se analyticky zpracovávají, tak co do typů institucí, které tuto činnost provozují.

Národní knihovna ČR zpracovává výběrově bibliografické záznamy článků ze všech druhů seriálů (noviny, časopisy, odborná periodika, sborníky) v rámci [Kooperačního systému článkové bibliografie \(KOSABI\)](#).

ve kterém spolupracují stávající SVK a MZK, specializované odborné knihovny (STK, ÚZPI, SPKK-ÚIV, ČSAV). Na základě této spolupráce vzniká souborná databáze [ANL](#). V systému [LANIUS](#) se zpracovávají bibliografické záznamy článků v knihovnách na úrovni okresů. V budoucnu je třeba sladit systém KOSABI a LANIUS tak, aby nedocházelo k duplicitnímu zpracování. V současné době se postupně v rámci KOSABI aplikuje nebo plánuje přechod na nové SW vyšší generace, zatím probíhá ve většině SVK popis článků v ISISu. V SVK Kladno se články popisují v systému RAPID, v MZK v Brně v ALEPH. V době přechodu

spolupracujících institucí na různé nové integrované systémy (KP-SYS, TINLIB, RAPID apod.) je kvalitní automatizovaná správa souborné databáze nutná. Souborná databáze [KOSABI ANL](#) obsahuje přes 620000 záznamů, v NKČR se excerpuje se cca 210 titulů, 469 titulů ve spolupracujících institucích (278 specializované knihovny, 191 titulů v SVK a MZK). Přejít na zpracování v systému v ALEPH 500 v dubnu 2000 posunulo zpracování na úroveň mezinárodního formátu UNIMARC a pravidel popisu AACR2 s respektováním mezinárodních standardů věcného popisu - MDT-MRF pro oblast systematické indexace. V oblasti verbální věcné indexace se kombinují klíčová slova, věcné obecné kategorie a předmětová hesla. Vyváženost vazby mezi jednotlivými vrstvami popisu je klíčovým momentem. V rámci kooperačního systému byla stanovena pravidla pro výběr titulů k popisu (na základě územní gesce - tituly regionální a celostátní provenience a dále pak na základě odborného zaměření). Dále byly stanoveny zásady výběru článků co do úplnosti i co typů (viz Příloha F 4, s. XIII).

V posledních letech vzniká několik projektů, zabývajících se zpřístupněním analytických záznamů v kooperaci s ostatními knihovnami, jejich prezentací na Internetu a propojením těchto záznamů s plnými texty.

Zpřístupnění výsledků analytického zpracování prostřednictvím Internetu (kooperační projekt 13 knihoven v rámci RISKu, řešen v r.1998, hlavní řešitel Ivana Anděrová) umožnil konverzi analytických záznamů z CDS/ISIS do UNIMARCu. V rámci projektu byla vypracována a odzkoušena konverze tehdejší verze Tinlibu do UNIMARCu. V rámci průzkumu Internetu se ukázalo, že postupné propojení článků s některými plnými texty již vystavovanými na Internetu na různých serverech je krajně nespolehlivé (různá retrospektiva a úplnost vystavovaných plných textů, různá strategie vystavovatelů). Výběr spolehlivých zdrojů plných textů je možné řešení.

Výzkumný záměr NK

[Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů](#) (VAV, hlavní řešitel Ivana Anděrová, r. 1999-2003) - je projekt analyticko-koncepční a připravuje půdu pro praktickou realizaci účelového projektu popisovaném v této zprávě a dalších projektech. Cílem výzkumného záměru je optimalizace přístupu uživatelů k plným textům dokumentů domácí provenience (nikoli zahraniční). Základem je propojení analytických záznamů o článcích s plnými texty, které jsou dostupné na Internetu a/nebo CD-ROM. V rámci projektu v r. 1999 proběhlo v NK **výběrové řízení** a na základě výše uvedených faktů byla vypracována výzva k podání nabídky pro společnost Anopress. V rámci projektu bylo vyvinuta iniciativa k vytvoření **[Konzorcium Anopress s.r.o.](#)**, Smlouva byla podepsána mezi SKIP a Anopressesem v r. 2000. V r. 1999 bylo experimentálně propojeno cca 4000 záznamů s plnými texty, získanými od Anopressu a některá odborná knihovnická periodika. Periodikum **[Národní knihovna](#)** bylo v Anopressu převedeno do digitální formy a zpřístupněno na Internetu (v r. 1999 pouze technikou OCR, v r. 2000 se přistoupilo i k prezentaci obrázků).

Projekt Západočeský ANAL - Kooperativní zpracování periodické produkce západních Čech (SVK v Plzni a 11 městských knihoven, řešitel Jaroslava Hanzlíčková, RISK, podaný v r. 1999) se zabývá odstraněním duplicit při zpracování, metodikou excerpcí titulů a zpracování záznamů v jednotlivých okresech západočeského regionu).

Projekt Zavedení automatizovaného zpracování článkové bibliografie v systému T-Series (SVK v Ostravě, hlavní řešitel Alena Hrazdilová, VaV, r. 2000-2001) řeší problematiku bibliografického zpracování článků v tomto systému.

Právě podávaný projekt SVK Kladno je velmi významný z hlediska tvorby a rozvoje regionálních faktografických databází a souborů autorit. souborů autorit.

Analytické záznamy zpracovávané v rámci [KOSABI](#) jsou zpřístupňované také na **CD-ROM** vydávaném **AIP icode** v rámci ČNB jako [řada Články v českých novinách, časopisech a sbornících](#) od června v 2000 v UNIMARCu. CD-ROM je vydáván ve čtvrtletních aktualizacích, každý měsíc je bibliografie [aktualizována na Internetu](#).

A. 3 Cíl, vstupní data

A. 3.1 Cíl

Náplní projektu je optimalizace integrace a správy heterogenních dat souborné databáze Kooperčního systému článkové bibliografie. Bibliografické záznamy článků, publikovaných v českém periodickém tisku a zpracovávané spolupracujícími knihovnami, budou postupně propojované s elektronickou podobou článku a takto prezentované na Internetu. Obě části souborné databáze - vznikající databáze plných textů a báze bibliografických záznamů ve formátu UNIMARC - vyžadují permanentní kvalitní SW a HW podporu.

Budování, doplňování, správu a údržbu plnotextové databáze s možností vyhledávání zajistí informační agentura [ANOPRESS](#). Zároveň půjde o vývoj manažerského systému pro příjem a správu dat kooperčního systému. Hlavním cílem projektu je zkvalitnění bibliograficko-informačních služeb.

Cílem projektu v roce 2000 je průběžné doplňování báze ANL v rámci [KOSABI](#) klasickým způsobem a publikovat ji v rámci [ČNB - řada Články v českých novinách, časopisech a sbornících](#), **dále návrh řešení linky automatického přebírání plných textů a automatické indexace bibliografických záznamů** pro bázi ANL a tvorby URL, budování databáze plných textů ANL FULL s možností pojmového vyhledávání, dále pak řešení automatické správy - **managementu KOSABI**. V roce 2000 je třeba vybavit systém **odpovídajícím SW a HW a ošetřit smluvně**.

Návrh řešení pro rok 2000 spočívá v zavedení nových metod v rámci získávání informací (možnost konzorciálního nákupu), zpracování bibliografických záznamů (přebírání metadat) a zpřístupňování informací (propojení záznamů s plnými texty a vyhledávání v plných textech).

Řešení otázek **standardizace jmenného a věcného popisu bibliografických záznamů** a elektronických dokumentů a **možnosti přebírání metadat** do bibliografických záznamů ve formátu UNIMARC a plných textů v podobě [Dublin Core](#) je nutným předpokladem funkčnosti celého systému.

Cílem v roce 2000 je dále průběžné **ukládání plných textů** ze současné produkce deníků a některých odborných časopisů odpovídající profilu NK na serveru NK a částečná příprava k **dynamickému propojení přes komponentu URL** uloženou v propojovacím poli bibliografických záznamů.

(Plné texty článků regionální provenience budou pravděpodobně v budoucnu uloženy na serveru Anopressu - mohou být uloženy i na serverech jednotlivých knihoven. Záznamy z let minulých budou postupně propojovány během řešení celého projektu v letech 2001-2004 s plnými texty metodou off line).

Dalším cílem v roce 2000 je zpracování bibliografických záznamů v NKČR a napojení na plné texty s **úplnými statickými URL adresami** - záznamy především titulů z oboru knihovnictví a oblasti práva .

A. 3.2 Vstupní data

Vstupními daty pro bázi [ANL](#) jsou jednak bibliografické záznamy zpracovávané v r. 2000 v rámci KOSABI, v rámci [KOSABI](#) bibliografické záznamy zpracované v NKČR a doplňované částečnými URL pracovníky oddělení a k nim [plné texty pro bázi ANL FULL](#) stažené v rámci [Konzorcia Anopress](#) taktéž pracovníky oddělení. V roce 2000 se jedná zejména tituly zpracovávané v NKČR, a to analyticky zpracovaný výběr z celostátních deníků a některých časopisů (Týden, Ekonom, Respekt, Reflex). Vstupními daty pro bázi [ANL FULL](#) jsou tedy plné texty, které odpovídají profilu zpracování bibliografických záznamů.

Záznamy takto zpracované a zároveň zaindexované plné texty s [Dublin Core](#) jsou takto připraveny dynamickému propojení.

Vstupními daty jsou dále bibliografické záznamy zpracované v NKČR a napojované na plné texty s úplnými URL adresami - záznamy především titulů z oboru knihovnictví (Národní knihovna, U nás, Ikaros, Daidalos aj.) a oblasti práva aj. (Veřejná správa, Obchodní právo, právo a podnikání, Moderní obec aj. - napojení na plné texty zákonů).

Vstupními daty pro léta minulá jsou bibliografické záznamy a adekvátní plné texty (v roce 2000 je to výběr produkce bibliografických záznamu a plných textů za léta 1998, 1999).

Summa summarum:

Vstupními daty pro plnotextovou bázi ANL FULL jsou plné texty zejména celostátních deníků a některých odborných časopisů za rok 1999, 2000, 1998 a jim odpovídající bibliografické záznamy v tomtéž období pro bázi [ANL v rámci KOSABI](#).

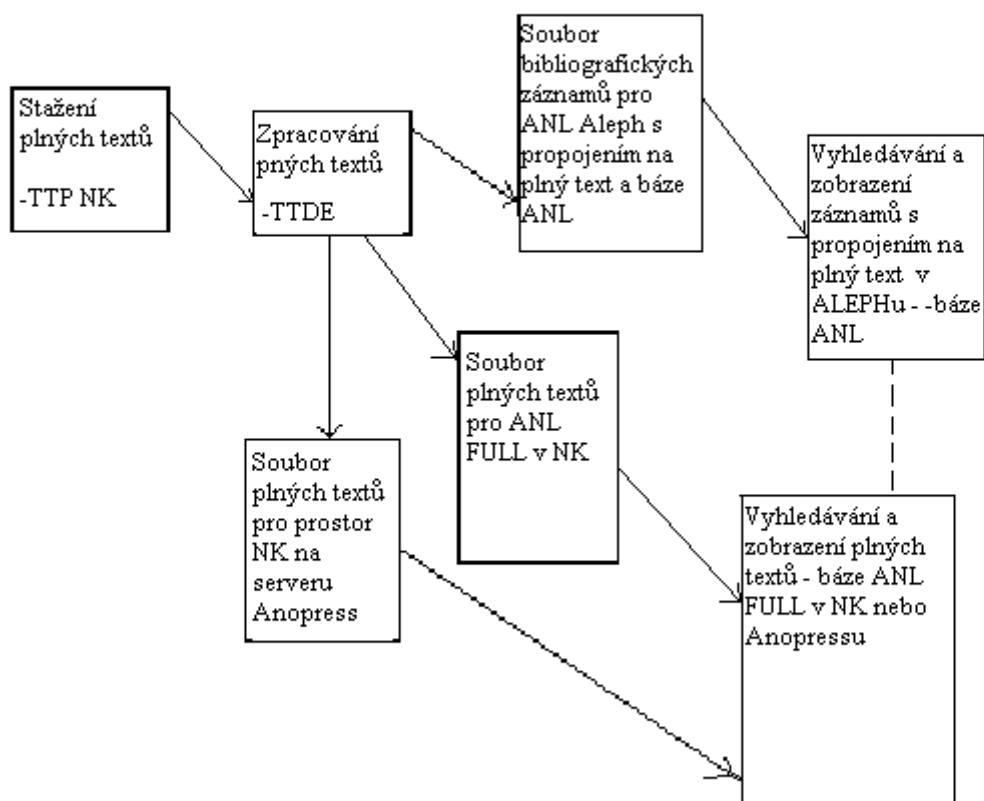
Vstupními daty pro návrh automatizované linky zpracování bibliografických záznamů jsou plné texty získané z databáze [TAMTAM](#). V rámci této linky vznikají vstupní data pro bibliografickou bázi [ANL](#) a fulltextovou databázi [ANL FULL](#).

B Analytická část

B.1 Vlastní řešení

B.1.1 Linka automatické indexace

Obr. 1 Linka automatické indexace



(Obr. 1)

Pro optimalizaci integrace a správy heterogenních dat souborné databáze kooperačního vyvinula česká firma [ANOPRESS](#) na podkladě analýzy a funkčního zadání návrh speciální technologie - linky automatického získávání plných textů, indexace bibliografických záznamů a plných textů, propojování záznamů na plné texty a jejich zpřístupnění. Řešení je progresivní a odpovídá nejnovějším trendům v této oblasti, je podpořeno kvalitním technickým a programovým vybavením. Jednotlivé moduly lze použít i samostatně. V rámci experimentu v r. 2001 je třeba ještě doladit technologii v rámci různých stádií aplikace.

Řešení spočívá ve speciální aplikaci v praxi již používané technologie firmy na získávání a zpřístupňování plných textů pro NK - **TAMTAM Profesional NK (TTPNK)**. Pomocí této technologie je možno **stahovat plné texty** článků z Internetu z báze **TAMTAM** založené na plnotextovém pojmovém vyhledávání systému **TOPIC**. Je možno stahovat více článků najednou na základě tématu, názvu článku, názvu zdrojového dokumentu aj. (Pro stahování je možné využít i verzi TAMTAM Standard - TTS).

Pro vlastní **automatickou indexaci článků a plných textů** - pro vytváření bibliografických záznamů v **UNIMARCU** na základě údajů uložených v plných textech a naopak pro vkládání metadat **Dublin Core** do plných textů je připravena technologie **TAMTAM Data Extractor (TTDE)**.

Bylo připraveno **6 hlaviček (headers)(Obr. 2)**, které se automaticky generují z plného textu:

1. **Formulář pro editaci (Obr. 2)**, do kterého se generují bibliografická data z plného textu. Data lze katalogizátorem následně upravovat a provádět tak korekce nejen ve Formuláři, ale automaticky také v **hlavičce UNIMARC-A (Obr. 3)**, **UNIMARC (Obr. 4)**, **Dublin hlavičce (Obr. 5)**. Obsahuje **údaje jmenného popisu**, které se přebírají z hlavičky plného textu - oproti původním údajům byly doplněny údaje roč., číslo, ISSN -, dále pak obsahuje **údaje věcného popisu** - předmětové kategorie, automaticky generovaná klíčová slova, automaticky generovaný abstrakt - extrakt, automaticky generovanou **URL** složenou z jednotlivých komponent odpovídající struktuře propojovacího pole 856 ALEPH a UNIMARC.
Volbou **Text** na horní liště je možno zobrazit plný text.
2. **UNIMARC-A hlavička (Obr. 3)** je hlavička s bibliografickými údaji **pro importní vstupní soubor záznamů pro ALEPH (řádkový UNIMARC)**, do které se automaticky generují tatáž data jako do Formuláře a úpravy zanesené do Formuláře, tato hlavička je také přístupná pro editaci samostatně.
3. **UNIMARC hlavička (Obr. 4)**, hlavička pro klasický UNIMARC s týmiž vlastnostmi jako hlavička UNIMARC-A - slouží k eventuelnímu importu pro systémy, které jsou založeny na UNIMARCU - řádkový UNIMARC. Do hlavičky jsou generované tytéž údaje jako do výše jmenovaných hlaviček.
4. **DUBLIN hlavička (Obr. 5)** s týmiž vlastnostmi jako předchozí dvě hlavičky sloužící ke generování metadat **Dublin Core** zpět do plného textu - slouží k zabudování těchto metadat do plných textů pro fulltextovou databázi - vychází z poslední verze Dublin Set Elements, obsahuje navíc automaticky generované **SICI (Seriál Item and Contribution Identifier a provizorní NBN (National Bibliography Number))**.
5. **Indexovací hlavička** obsahuje údaje jmenného popisu.
6. **Zobrazovací hlavička** slouží k zobrazení údajů v hlavičce plného textu.

Po odrážce různé je **možno nastavit tvar výstupní hlavičky pro UNIMARC-A (Obr. 3)** nebo **UNIMARC (Obr. 4)** a spustit **ruční vstup dat**.

Dále následuje **přesunutí UNIMARC-A hlavičky (Obr. 3)** do importu pro ALEPH (**báze ANL**) a umístění plných textů ve tvaru HTML na web server NK k indexaci do fulltextové databáze v NK nebo do Anopressu.

Pro **indexaci dat do fulltextové databáze (ANL FULL)** v NK byl vyvinut program **MkIndex (MkI)** . Tento program nalezená data automaticky **zaindexuje** , umožňuje jejich vyhledání ve fulltextové databázi a **zpřístupnění** . Plné texty jsou ve formátu HTML.

Pro **vyhledávání v datech ve fulltextové databázi** jsou vyvinuty [formuláře pro vyhledávání jednoduché, pokročilé, pokročilé s tématy \(Obr. 6\)](#). Vyhledávání probíhá v systému [TOPIC \(Search 97\)](#) a definice formulářů vychází z jeho filozofie.

V budoucnu bude Formulář pro editaci pravděpodobně rozšířen o některé údaje věcného popisu. Pro import do ALEPHu je třeba připravit převodní tabulky mezi kódem 1250 CP do Ansel, používaného v ALEPHu nyní a budoucím UNIMCODEM.

Pro **propojení s plných textů se systémem ALEPH** (doplnění URL adres do záznamů) byl vytvořen **skript mkdoc.htp** . Propojení probíhá na základě dynamicky generovaného odkazu na dokument. Program vyhledá požadovaný dokument dle identifikace (identifikační číslo),provede statistiku a v budoucnu bude provádět kontrolu autorizace a na jejím základě zobrazí plný text, abstrakt nebo nic.

B.1.2 TOPIC

TOPIC (pojmově orientovaný vyhledávací systém, concept based retrieval) je systém třetí generace založený na následujících principech: rozklad pojmu na podpojmy, vážení jednotlivých podpojmů (větví pojmového stromu), neostře vyhodnocování dotazů. Dotaz v systému třetí generace reprezentuje pojem, resp. ideu vyhledávaného tématu. Jádrem dotazu je stromová hierarchická struktura, která rozkládá hledané téma na podtémata a přiřazuje jednotlivým částem váhy, které vyjadřují do jaké míry příslušné téma přispívá k celkovému určení tématu. Systém dále vypočítá míru relevance vyhledaných dokumentů. Oproti běžně používaným operátorům **TOPIC** používá logický operátor ACCRUE se specifickými vlastnostmi. Tento operátor sbližuje operátory AND a OR. Každý topik obsahuje tedy tři charakteristiky - strukturu, váhy a operátory.

Nabízí se zde jistá formální analogie k hierarchickému selekčnímu jazyku systémové notace MDT. **Je však třeba zdůraznit, že topiky jsou tvořeny podle skutečnosti, MDT je víceméně taxativní systém jednotlivých oborů, nikoli témat. Proto je třeba k definici topiků přistupovat svébytně. Dotaz lze zadávat třemi způsoby: Prostý dotaz** je pouze seznam slov, které se mohou vyhledat. Všechna slova mají stejnou váhu - možno použít při hrubém hledání, kdy se přesně neví, co se má vyhledat. **Formulářový dotaz** slouží k přesnějšímu vymezení dané oblasti. Dotaz lze specifikovat dalšími atributy, jako např. autor, zdroj, datum atd. **Tematický dotaz** je nejpřesnější. Spočívá ve vytvoření topiku, kdy mohou být zadány všechny váhy. V roce 2000 byl vypracován experimentálně [topik pro obor Demografie \(Obr. 7\)](#) plánují se topiky další. V oddělení analytického zpracování při věcném popisu článků se používají k indexaci hrubých témat a podtémat předmětové kategorie, které připomínají svou podstatou topiky, resp. témata a skupiny témat v systému **TOPIC** v databázi Anopress. Je však třeba je sladit obsahově.

B.1.3 Standardizace - bibliografická metadata ve formátu UNIMARC a metadata [Dublin Core](#) v plných textech.

Struktura bibliografických dat respektuje formát [UNIMARC](#) a knihovnická pravidla [AACR2](#) v oblasti jmenného popisu. V oblasti věcného popisu se používá aktualizovaná verze [MDT-MRF](#). Verbální věcný popis obsahuje **předmětové kategorie**, které zasazují dokument do širších souvislostí v rámci databáze z hlediska obecných témat, jež by se měla sblížovat s tématy systému [TOPIC](#), dále pak **klíčová slova**, která jsou dále částečně řízená a **předmětová hesla**. V budoucnu se předpokládá intenzivnější využívání vznikajících souborů autorit jmenných i věcných. V záznamech určených k propojení s plným textem je zapracována **celá URL adresa (statická) nebo komponenta dynamické URL adresy**.

Struktura metadat v plných textech vychází z poslední verze [Dublin Core Metadata Set](#). K identifikaci plného textu je zabudováno [URL, SICI \(Serial Item and Contribution Identifier\)](#) - **SICI je automaticky generováno, NBN (National Identification Number)** zatím používající identifikační číslo plného textu Anopressu a **ISSN**.

[Dublin hlavička \(Obr. 5\)](#)

Všechny tyto údaje mohou sloužit k tvorbě dynamických adres jako komponenty a k tvorbě URN (Uniform Resource Name) a URI (Uniform Resource Identifier). Do HTML je také třeba zabudován **LINK tag** pro potřeby odkazu na webovský zdroj, v němž se nachází specifikace daného použitého souboru metadat.

B.1.4 Propojení bibliografického záznamu s plným textem

V databázi ANL se používají jednak **statické adresy** u propojení na plný text, kde se zdá strategie jeho vystavování poměrně stálá (knihovnické časopisy - Národní knihovna, U nás, Ikaros, Daidalos aj.), některé zdroje z oblasti práva, zejména plná znění zákonů. S těmito zdroji jsou propojovány např. záznamy z časopisů Veřejná správa, Obchodní právo, právo a podnikání, Moderní obec. Staticky byly propojovány také záznamy z deníků v r. 1999. V plánu je další propojování s plnými texty vystavovanými na Internetu.

V oddělení analytického zpracování byly připravené záznamy k propojení s plnými texty na základě komponentu URL adresy - identifikačního čísla plného textu. Tyto adresy mohou být doplněny o další komponenty "na stálo" nebo mohou být použity k dynamickému propojování v rámci otevřených URL adres. Zatím bylo takto připraveno cca 5500 záznamů a staženo odpovídající množství plných textů článků vydaných v 2. pololetí r. 2000. Dále bylo zakoupeno cca 17930 plných textů článků publikovaných v 1. pololetí r. 2000 a v r. 1998 (rok 1999 byl saturován z grantu Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů)

B.1.5 [Souborná databáze ANL Kooperálního systému článkové bibliografie \(KOSABI\)](#)

V roce 2000 přešlo oddělení analytického zpracování na ALEPH 500 a UNIMARC (do té doby probíhalo zpracování v CDS/ISIS a záznamy se konvertovaly do UNIMARCu a vystavovaly na WWW). Kooperující instituce přispívají pravidelně do souborné databáze (kromě SVK České Budějovice - čeká se na konverzi T-Series do UNIMARCu a MZK v Brně - souvisí s přechodem na novou verzi ALEPHu). Kromě toho instituce budují své lokální databáze.

Záznamy respektují metodický materiál [Záznam pro soubornou databázi : UNIMARC](#) a [Záznam pro soubornou databázi : Výměnný formát](#). Byla aktualizována pracovní verze metodické příručky pro zpracování článků v UNIMARCu.

V roce 2000 probíhaly rozsáhlé korektury báze. Chybovost v bázi je dána existencí dvou podob báze v minulosti - v ISIS a UNIMARC - a způsobem zpracování v CDS/ISIS. Opravy v databázi si vyžadují průběžnou pozornost, chybovost je stále velká na straně NK i spolupracujících institucí.

B.1.6 Česká národní bibliografie - řada Články v českých novinách, časopisech a sbornících na CD-ROM

V červnu 2000 vyšel první [CD-ROM](#) s články v UNIMARCU za spolupráce NKČR a AIP icome. V UNIMARCU vycházejí [aktualizace ČNB na Internetu](#).

B.1.7 Management Kooperačního systému článkové bibliografie (aplikace MNG KOSABI)

Na základě zadání pro řízení a správu kooperačního systému a za využití již vyvinutých řešení v rámci Souborného katalogu [CASLIN](#) probíhají práce na vývoji aplikace pro tento systém. V současné době je nainstalován na server ANL systém **LINUX a ORACLE**, ve stádiu řešení je aplikace pro příjem a automatizované zpracování dat (přijímání analytických záznamů, integrace stávajících programů pro konverzi analytických záznamů, globální úpravy analytických záznamů, vývoj programů na formálně logické kontroly kooperujících knihoven - test na UNIMARC pro analytické záznamy.

B.1.8 HW a SW podpora Kooperačního systému článkové bibliografie, bibliografické báze ANL a plnotextové databáze ANL FULL

Pro management kooperačního systému byl zakoupen PC Pentium III, 700 MhZ, ORACLE 8i server. Release 8.1.5 (5 licencí),

Pro správu a údržbu plnotextové databáze ANL FULL byl zakoupen server DELL - PowerEdge 6300 - Pentium III Xeon 500Mhz/512, Search Verity Information Server (TOPIC) v. 3.6 pro jednoprocessorový server Windows NT zatím pro ultranet (30 licencí), Windows NT v. 4.0. (server full.nkp.cz)

B.1.9 Smluvní ošetření projektu a Kooperačního systému článkové bibliografie (Viz příloha F5).

Pro automatickou indexaci, správu (údržbu) plnotextové databáze a dodávku plných textů byly uzavřeny dvě smlouvy s Anopressem a jeho pracovníky. Dále byla uzavřena Smlouva na vývoj aplikace pro management kooperačního systému.

V současné době je připravován **Dodatek č. 3 ke Smlouvě o sdružení pro Českou národní bibliografii**, který zabezpečuje fungování Kooperačního systému článkové bibliografie v situaci reformy státní správy.

B.2 Přínos řešitele

Přínos projektu spočívá v integraci elektronických zdrojů mezi tradičně zpřístupňované sekundární informace formou bibliografických záznamů. Jde o integraci heterogenních dat do [Kooperačního systému článkové bibliografie](#), v němž dochází k propojení tradičních knihovnických postupů a fondů s určitými prvky digitální knihovny.

Přínos projektu spočívá v postupném budování **plnotextové databáze** s možností kvalitního vyhledávání založeného na principu pojmovém vyhledávání (concept based retrieval) v kombinaci s metadaty, s možností dalšího doplňování, její správy a údržby.

Další přínos projektu spočívá v rychlém **zpřístupnění analytických záznamů** provázaných s plnými texty zdrojových dokumentů v rámci Kooperčního systému článkové bibliografie.

Nemenší přínos v spočívá v **revidování tradičních postupů** při zpracování české národní bibliografie v oblasti jmenného i věcného popisu v rámci návrhu linky automatické indexace bibliografických záznamů. Pro popis webovských informačních zdrojů je aplikován formát **Dublin Core**, který je v současné chvíli je využíván v mnoha systémech v zahraničí.

Od spolupráce mezi Národní knihovnou a ANOPRESSEM se očekává **ekonomický efekt** ve smyslu šetření **pracovních kapacit** Národní knihovny (generování některých bibliografických údajů a anotací). V oddělení je plánována do r. 2003 redukce o 6 pracovních úvazků ve prospěch jiných oddělení NK. V tomto roce byla realizována redukce o 2 pracovní úvazky.

Možnost vznik **duplicit či multiplicit** při zpracování je minimalizováno delimitací periodik mezi spolupracující instituce. Tím má projekt další i ekonomický význam.

Další přínos spočívá ve zvýšení **uživatelského komfortu** - v nalezení příslušného článku z novin či časopisu (v budoucnu snad i statě se sborníku) v elektronické formě Navigace k primárním dokumentům patří k základním trendům v oblasti knihovnictví a informatiky.

Možnost aplikace vypracovaných metod na některé spolupracující subjekty v Kooperčním systému článkové bibliografie v budoucnu.

Další přínos spočívá v nákupu plných textů v rámci Konzorcia Anopress s.r.o., což má nemalý ekonomický efekt.

Řešení navržená, programově realizovaná a částečně realizovaná v praxi v roce 2000 jsou příspěvkem do programu univerzální bibliografické kontroly (UBC) a všeobecné dostupnosti publikací (UAP) IFLA a jsou také součástí praktické realizace čl. 61 kulturní politiky o kooperaci knihoven a především čl. 63 o podpoře automatizace knihovnických systémů a propojení do globálních informačních struktur. Přispívají tak k optimalizaci veřejných informačních služeb (VIS).

B.3 Posun znalostí

- K významnému posunu znalostí došlo především v těchto oblastech:
- Poměrně dobrá orientace v nových trendech zpracování a zpřístupňování informací.
- Návrh praktické realizace těchto trendů na konkrétní fungující systém, která umožní jeho přetrvání v budoucnu (metadata, pojmové vyhledávání, propojování informací, vazba na klasický fond, fond elektronických dokumentů a registrace v ČNB,)
- Vytvoření předpokladů pro automatickou indexaci dokumentů.
- Propojení bibliografických záznamů s elektronickými zdroji na základě dynamických i statických URL adres, kombinace pojmového vyhledávání s metadaty jako předpoklad pro implementaci kooperčního systému do konceptu metaknihovny.

- Vytvoření předpokladů pro distribuované vyhledávání informací za současného využití tradičního slučování informací do souborné databáze založené na architektuře statických bází.
- Posílení vazby v rámci [KOSABI](#) v situaci reformy státní správy a samosprávy.
- Nová forma poskytování informačních služeb za spolupráce knihovnické a moderní informační instituce.

C Navrhovaná část

C.1 Výsledky řešení

- Vytvoření nástrojů pro optimalizaci integrace a správy heterogenních dat v rámci Kooperačního systému článkové bibliografie. Vznikl tak nástroj pro automatickou přípravu dat pro bibliografickou databázi založenou na UNIMARCu a plnotextovou databázi založenou na pojmovém vyhledávání systému [TOPIC](#).
- Návrh a realizace aplikace pro získávání dat TAMTAM PROFESIONAL NK (TTPNK).
- Návrh a realizace aplikace pro automatické generování bibliografických metadat do analytických záznamů a metadat [Dublin Core](#) a jejich umístění do plných textů - **TAMTAM DATA EXTRACTOR (TTDE)**
- Aplikace pro indexaci dat do plnotextové databáze [ANL FULL](#).
- Aplikace pro vyhledávání a zpřístupnění plných textů.
- Návrh aplikace pro management Kooperačního systému článkové bibliografie (MNG KOSABI).
- Standardizace týkající se popisu článků v UNIMARCu a implementaci standardu [Dublin Core SICI](#) do plných textů.
- Nastínění nových metod zejména ve věcném popisu (aplikace kategorií v popisu článků a témat při zpracování i vyhledávání informací o člancích).
- Průběžné doplňování báze [ANL](#) bibliografickými záznamy.
- Zakoupení plných textů pro plnotextovou databázi, průběžné doplňování databáze a vytvoření předpokladů pro automatizované propojování bibliografických záznamů s plnými texty na základě dynamických a statických URL adres. Průběžné propojování záznamů s plnými texty na základě statických URL adres.
- Zabezpečení HW a SW podpory kooperačního systému
- Právní zabezpečení projektu a kooperačního systému.

C.2 Závěr

Výsledkem řešení projektu v r. 2000 je návrh technologie linky automatizovaného získávání a zpracování informací o člancích a jejich následného zpřístupnění v rámci bibliografické souborné databáze [ANL](#) a fulltextové databáze [ANL FULL](#) založené na pojmovém vyhledávání systému TOPIC. Aplikace umožňuje přípravu importního souboru bibliografických záznamů (s automaticky generovanou URL adresou) pro ALEPH a jiné systémy, založené na UNIMARCu a doplnění plných textů o metadata [Dublin Core](#).

Návrh **aplikace pro indexaci** plných textů do fulltextové databáze **ANL FULL**, návrh **formulářů pro vyhledávání** a následné zobrazení výsledků vyhledávání umožní uživateli získat relevantní informace na základě **kombinace vyhledávání pomocí metadat a pojmového vyhledávání**.

Praktickým výsledkem v r. 2000 je příprava bibliografických záznamů **k propojení s plnými texty v bázi ANL FULL na základě komponenty URL adresy**. Takto připravená propojení se mohou realizovat klasickým propojením, tj. doplněním o dalších údaje URL adresy nebo v budoucnu v rámci **metaknihovny** na základě **dynamicky generovaných adres** na podkladě uživatelského dotazu. Výsledkem práce katalogizátorů v tomto roce je **dále vytvoření hypertextových odkazů statického typu** u dokumentů, které jsou propojovány klasickou cestou při zpracování záznamů na "stále" vystavené plné texty na Internetu.

Dalším cílem projektu v tomto roce je návrh řešení a realizace programové aplikace na provoz, správu a údržbu databáze článkové bibliografie, tj. **aplikaci pro management KOSABI**. Jde o vývoj aplikace v systému ORACLE na základě řešení aplikací v rámci Souborného katalogu CASLIN. Jde o automatizaci všech činností spojených se správou Kooperativního systému článkové bibliografie. Cílem je doplnit a posílit integrovaný knihovnický systém ALEPH.

V rámci projektu v tomto roce byla zajištěna **potřebná HW a SW** podpora, projekt je ošetřen po stránce smluvní. Po experimentálním odzkoušení systému automatické indexace v r. 2001 bude možno zahájit poloprovoz systému.

Je evidentní, že v budoucnu bude nutné nabídnout uživateli přímý přístup k elektronické formě článku i jeho tištěnou podobu. Nutným předpokladem je propojení bibliografické článkové databáze na fond časopisů v tištěné formě, ale také ve formě elektronické.

C.3 Návrhy opatření

- Zajištění financování projektu v roce 2001 a dále zajištění souvisejícího projektu Propojení analytických záznamů s plnými texty - optimalizace zpřístupnění plných textů, který tento projekt saturuje koncepčně a doplňuje finančně. Koordinace s ostatními stávajícími i budoucími projekty v NK i jinde.
- Posílení vazeb stávajícího KOSABI na existující kooperační systémy na nižší úrovni z hlediska správního, zejména systém LANIUS.
- Personální zajištění projektu z hlediska počtu pracovníků oddělení analytického zpracování v NK. Další redukce v oddělení jsou nežádoucí (v roce 2003 bude mít oddělení 11 úvazků z původních 17 v r. 1999). Vzhledem k redukcím oddělení nebude možno v dalších letech dodržet vyšší finančního vkladu oddělení do projektu.

- Existence souborů autorit v oblasti jmenného a věcného popisu jsou nutným předpokladem kvalitního zpracování a vyhledávání informací. Přesunutí kapacit do příslušných oddělení tuto situaci pomohou řešit.
- Přísnější výběr článků k indexaci deníků s cílem vyloučení subjektivního faktoru při excerpce deníků.
- Funkčnost propojovacích vazeb v systému ALEPH a možnosti expanze a spolehlivosti systému v tomto ohledu. Moderních informačních systémy jsou založeny na propojování sekundárních informací s primárními jak klasickými tak elektronickými, ale také na vzájemném propojování sekundárních informací o různých typech dokumentů. Nejde pouze o propojení záznamů s plnými texty, ale také o provázání seriálů a jednotlivých čísel na analytický rozpis článků obsažených v seriálu v rámci báze NKC či Souborného katalogu CASLIN, dále pak připojení článků - recenzí k recenzovaným dokumentům v rámci těchto bází. Navigační systémy na úplné obsahy čísel seriálů mimo záběr abstraktových databází jsou v zahraničí zcela běžné.
- Praktické odzkoušení aplikací v rámci experimentu v r. 2001.
- Nutnost řešení problému autorizace uživatelů z hlediska jednotného přístupu do informačního systému NK a ošetření přístupu uživatelů do plnotextové databáze v budoucnu.

D Použití finančních prostředků

D.1 Komentář

(Souhrnná zpráva podána 15.11.2000, od té doby další čerpání prostředků).

Využití investičních prostředků. Přiděleno 1 295 000,00 Kč. Čerpáno 1295 000,00 Kč.

Čerpáno do 15.11.2000: Dell Computer - 588 955,00 Kč, Search '97 - TOPIC - 523 979,00 Kč, PC-PIII-7000 Mhz - 77 958,00 Kč, ORACLE 8i - 42 941,60 Kč, ORACLE 8i - Upgrade - 2 520,00 Kč.

Čerpáno po 15.11.2000: PC - 58 646,40 Kč

Využití neinvestičních prostředků.

Přiděleno 1 025 000,00 Kč. Čerpáno 571941,00 Kč. Plánováno po 15.11.2000 453059 Kč.

V rámci **neinvestičních prostředků jsou odděleny placené služby, materiál, mzdy, licence.**

Čerpáno do 15.11.2000:

Služby: 442842 Kč - linka automatické indexace Anopress, plné texty Anopress, management kooperačního systému.

Materiál: 10050 Kč - CDR, diskety, pásky do tiskáren.

Mzdy (OON): 82200 Kč (bez pojištění) - příprava k propojení, propojení, stahování plných textů v rámci konzorcia.

Licence: 367849 Kč - Win NT 0.4

V tomto roce nebyla realizována plánovaná zahraniční stáž z důvodu pracovního vytížení řešitelského týmu a prospěšnosti investovat ušetřené prostředky do jiných typů služeb.

Prostředky z podnikových zdrojů a jiných zdrojů činí podle smlouvy mezi NKČR a MKČR 576 000 Kč. Z toho 120 000 Kč je plánováno jako **vklad Anopressu** v podobě jednoho přístupu k verzi TAMTAM Profesional během řešení projektu. Verze je zpřístupněna v NK od června 2000. Při měsíční ceně tohoto produktu (15 000 Kč) je to mnohonásobné překročení vkladu.

Pracovníci oddělení mají vložit do projektu v tomto roce 456 000 Kč.

Vzhledem k tomu, že projekt je náročný koncepčně i realizačně zároveň, je **vklad hlavní řešitelky a pracovníků oddělení** (rozsáhlé korektury databáze k opravě chyb, které byly částečně způsobeny dvojitou existencí báze - v CDS/ISIS a ALEPH) poměrně velký. Oddělení analytického zpracování provádí korektury záznamů nejen svých, ale i záznamů spolupracujících institucí. Konečné zúčtování vkladu bude provedeno po dokončení projektu v tomto roce.

E Resumé a klíčová slova

E.1 Resumé a klíčová slova v češtině

Resumé:

Náplní projektu je optimalizace integrace a správy heterogenních dat souborné databáze Kooperačního systému článkové bibliografie (KOSABI). Bibliografické záznamy článků, publikovaných v českém periodickém tisku a zpracovávané spolupracujícími knihovnami, budou postupně propojované s elektronickou podobou článku a takto prezentované na Internetu.

Výsledkem řešení projektu v r. 2000 je návrh aplikace pro získávání a automatickou indexaci bibliografických záznamů z plných textů a následné vytvoření importního souboru záznamů pro databázi bibliografických záznamů ANL a databázi plných textů ANL FULL v NK s implementací metadat v plných textech včetně automaticky generované URL adresy. Indexace plnotextové databáze v systému TOPIC umožňuje pojmové vyhledávání informací.

Plné texty článků byly v rámci Konzorcia Anopress průběžně stahovány a připraveny k dynamic-kému propojení s bibliografickými záznamy v 2. pololetí tohoto roku.

Plnotextová databáze byla dále průběžně doplňována články z deníků časopisů vydanými v r. 2000 a 1998. Průběžně byly staticky propojovány záznamy a plné texty z oblasti knihovnictví a práva.

V roce 2000 byla průběžně aktualizována souborná databáze kooperačního systému ANL a vydáván CD-ROM s Českou národní bibliografií.

Dalším výsledkem řešení v r. 2000 je návrh aplikace pro správu a údržbu KOSABI. Pro kooperační systém byla zakoupena kvalitní HW a SW platforma. Celý projekt byl ošetřen smluvně.

Klíčová slova:

Plné texty; TOPIC; analytická indexace; záznam; seriály; články; zpřístupnění; souborná databáze; propojování; Kooperační systém článkové bibliografie; Česká národní bibliografie; vyhledávání; automatická indexace; ANOPRESS; KOSABI; ANL; ANL FULL; plnotextová databáze; pojmové vyhledávání; CD-ROM; UNIMARC; Dublin Core; metadata

E.2 Abstract and key words in English

Abstract:

The contents of this project is optimization of integration and management of heterogenous data which are involved in union database of Article Bibliography Cooperative system. Bibliographical entries of articles published in Czech periodicals which are produced by single libraries linked with electronical form shall be presented via Internet.

The 2000 resulted in application resolution of acquisition and automated indexing of

bibliographical entries from fulltexts and creating of imported file for bibliographical database ANL. In the same time fulltext database ANL FULL with metadata in full text including automatically generated URL is originating in this process. Indexing of database on the TOPIC platform enables concept based retrieval.

In the frame of Cosortium Anopress fulltexts were bought and prepared for dynamic linking with bibliographical entries in the second half 2000. Futher fulltexts of articles published in 2000 and 1998 were bought. Simultaneously static links have been done (librarianship and law).

The union databaze of cooperative system has been regulary updated and published on CD-ROM as Czech National Bibliography.

Application for management of COSABI has been resolved, too. HW and SW support of a high quality was installed. The whole project has been treated with several agreements.

Key words:

Full texts; TOPIC; analytical indexing; entries; serials; articles; access; union database; linking; Cooperative system of Article Bibliography; Czech National Bibliography; searching; automatic indexing; ANOPRESS; COSABI; ANL; ANL FULL; fulltext database; concept based retrieval; CD-ROM; UNIMARC; Dublin Core; metadata

14. listopadu 2000

[PhDr. Vojtěch Balík](#), ředitel NK

[PhDr. Ivana Anděrová](#), hlavní řešitelka