

**SOUBORNÁ DATABÁZE KOOPERAČNÍHO SYSTÉMU ČLÁNKOVÉ
BIBLIOGRAFIE – OPTIMALIZACE INTEGRACE A SPRÁVY
HETEROGENNÍCH DAT
Závěrečná zpráva za rok 2000-2004**

Předkládá: Mgr. Vlastimil Ježek, ředitel NK

Zpracovala: PhDr. Ivana Anděrová, hlavní řešitelka



Praha, leden 2005

OBSAH

A	Konstatační část	
A.1	Rešerše	4
A.2	Současný stav ve světě a v ČR	13
A.3	Cíl, vstupní data	30
B	Analytická část	
B.1	Vlastní řešení	31
B.2	Přínos řešitele	78
B.3	Posun znalostí	79
C	Návrhová část	
C.1	Výsledky řešení	81
C.2	Závěr	81
C.3	Návrhy opatření	81
D	Použití finančních prostředků - institucionální podpory (tabulky, komentář)	83
E	Resumé a klíčová slova	
E.1	Resumé a klíčová slova v češtině	104
E.2	Abstract and key words in English	105
F	Přílohy	108
F.1	Báze ANL v NKČR (ALEPHU), JIB CASLIN a ANL/ANL FULL, ČNB- vývoj báze ANL od UNIMARCu k MARCu 21 a aplikaci konspektu pro zapojení do oborových informačních bran	
F.2	Linka zpracování bibliografických záznamů z plných textů pro lokální pracovní stanice – Windows TamTam Data Extractor (Windows TTDE) - bibliografická metadata pro bázi ANL (ALEPH ve formátu UNIMARC) a plné texty s Dublin Core pro bázi ANL FULL (TOPIC)	
F.3	Linka zpracování bibliografických záznamů z plných textů – WWW TamTam Data Extractor	
F.4	Zpřístupnění plných textů v systému TOPIC. Dokumenty. Architektura systému získávání, zpracování a zpřístupnění plných textů v systému TOPIC (báze ANL FULL a portál Periodika na WWW) – detailní ukázky formulářů, formátů, výstupů. Administrace (opravy a statistiky)	

- F.5 Aplikace pro správu a údržbu Kooperačního systému článkové bibliografie (architektura systému, jednotlivé aplikace pro příjem a zpracování dat, báze titulů) – Management Kooperačního systému článkové bibliografie**
- F.6 Excerptní základna Kooperačního systému článkové bibliografie**
- F.7 Statistiky**
- F.8 Windows TTDE a WWW TTDE , TOPIC – dokumentace – hlavní programy a funkce s plánem do budoucna**
- F.9 Konverze UNIMARC, DC, M21; Dublin Core a ANL Core - tabulka návěstí pro vyhledávání v poli dotaz; tabulka formulářových polí, topiky, citace článku**

A Konstatační část

A.1 Rešerše

Poznámka:

V r. 2005 má dojít k rekonstrukci www stránek NKČR. Uvádíme současné adresy na serveru www.nkp.cz

Literatura (chronologicky):

CÍGLER, I., Königová, M., Lukavec, P., Vacek, V. Hodnocení efektivnosti informačních systémů. *Systémová analýza v informatice*. ČVTS, 1974. S. 98-115.

SARACEVIC, T. The concept of relevance in information science : a historical review. *Introduction to Information Science*. New York : Academic Press, 1976. S. 79-137.

ZEMANOVÁ, I. Problematika relevance a pertinence. Vývoj a současný stav. Diplomová práce. Praha. FFUK , 1977. 164 s.

Analytický popis. Praha : Národní knihovna v Praze, 1991. 2 sv. + disketa. (MAKS)

KOUDELKOVÁ, L. NÁDVORNÍKOVÁ, M. BAJÁK, M. Návod pro tvorbu a využívání báze záznamů dokumentů. Verze 1. Praha : Národní knihovna v Praze, 1991. 71 s. (MAKS)

STOKLASOVÁ, B., ANDĚROVÁ, I., KREMEROVÁ, J. Specifikace údajů pro bázi záznamů dokumentů. Verze 1. Praha : Národní knihovna v Praze, 1991. nestr. (MAKS)

ANDĚROVÁ, I. Pravidla zápisu údajů pro analytický popis. Praha : Národní knihovna v Praze, 1992. 217 s. + příl.

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

BÍNOVÁ, J. Regionální bibliografická činnost v SVK - možnosti spolupráce s okresními knihovnami. *Čtenář*, roč. 46, č. 2, 1994, s. 45-48.

NÁDVORNÍKOVÁ, M. Spolupráce na úplnosti národní bibliografie z pohledu regionálních vědeckých knihoven. *Knihovny současnosti '96*. 1. vyd. Brno : Sdružení knihoven, 1996, s. 134-139.

1996 Glenda Browne. To be published in [Online Currents](#), the AusSI Newsletter 20(6):4-9, July 1996 and LASIE 27(3):58-65

Doporučení pro popis částí dokumentu na základě mezinárodního standardního bibliografického popisu (ISBD). 1. vyd. Praha: Národní knihovna ČR, 1997. 32 s.

Cobra+ : Computerised Bibliographic Record Actions [online]. Boston Spa (Velká Británie) : COBRA+, 1997. Dostupný z: <<http://www.ddb.de/gabriel/cobra/>>.

KOCH, T. and BORELL, M. Dublin Core Metadata Template [online]. Lund (Švédsko) : Lund universitetsbibliotek, 1997, last update 1997-08-20. Dostupný z: <http://www.lub.lu.se/metadata/DC_creator.htm >.

Nordic Countries URN-generator : provided by the Nordic Libraries [online]. Lund (Švédsko) : Lund universitetsbibliotek, 1997. Dostupný z: <<http://www.lub.lu.se/cgi-bin/nmurn.pl> >.

OLSON, N. B. Cataloging Internet Resources [online]. Dublin (Ohio, USA) : OCLC, 1997. Dostupný z: <<http://www.oclc.org/support/documentation/worldcat/cataloging/internetguide/> >.

BRATKOVÁ, E. Bibliografické a plnotextové báze dat americké firmy H.W.Wilson pro společenské a humanitní obory: vyhledávání informací v systému WilsonWeb. Infomedia [online], 1998. Dostupný z: <<http://www.inforum.cz/infomedia98/pdf/wilson.htm> >.

BRATKOVÁ, E. K otázkám pojmu, třídění a typologie internetových a webovských informačních zdrojů. *Národní knihovna : knihovnická revue*, 1998, roč. 9, č. 5, s. 262-276. Dostupný též z: <<http://full.nkp.cz> >.

JONÁK, Z. Inteligentní nástroje pro práci s texty na Internetu. *Ikaros* [online]. 1998, č. 09 [cit. 1998-09-01]. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200208003> >. ISSN 1212-5075.

BRATKOVÁ, E. Metadata jako nový nástroj pro komunikaci webovských informačních zdrojů. *Národní knihovna : knihovnická revue*, 1999, roč. 10, č. 4, s. 178-195. Dostupný též z: <<http://full.nkp.cz/nkkr/Nkkr9904/9904178.html> >.

ČERVENÝ, V. Vyhledávání v databázích plných textů. *Národní knihovna : knihovnická revue*, 1999, roč. 10, č. 1, s. 6-12. Dostupný též z: <<http://full.nkp.cz/nkkr/Nkkr9901/9901006.html> >.

BARTOŠEK, M. Vyhledávání v Internetu a DUBLIN CORE. *Zpravodaj ÚVT MU*. ISSN 1212-0901, 1999, roč. 9, č. 4, s. 1-4.

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. Dostupný též z: <http://www.nkp.cz/pages/page.php3?page=fond_anal_unim_opr.htm>.

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. Dostupný též z: <http://www.nkp.cz/pages/page.php3?page=fond_ann_vf_opr.htm >

JONÁK, Z. Krize mezilidské komunikace v období komunikační a informační exploze. *Ikaros* [online]. 1999, č. 05 [cit. 1999-05-01]. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200205066> >. ISSN 1212-5075.

PAPÍK, R. Trendy v rozvoji informačních služeb. *Ikaros* [online]. 1999, č. 08 [cit. 1999-08-01]. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200208571> >. ISSN 1212-5075.

OPPENHEIM, Ch.; SMITHSON, D. What is the hybrid library? *Journal of Information Science*, 1999, vol. 25, no. 2, s. 97-112.

BURGETOVÁ, J. Právní aspekty poskytování knihovních elektronických a reprografických služeb. *Ikaros* [online], 1999, č. 06. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200205087>>.

HEIJTING, I. Interconnectivity and the Hybrid Library. *Ikaros* [online], 1999, č. 10. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200205142>>.

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

Sborník příspěvků ze semináře CASLIN '99 - Souborné katalogy: organizace a služby. Dostupný z: <<http://www.caslin.cz:7777/caslin99/prispevky.html>>.

TKAČÍKOVÁ, D. Když se řekne digitální knihovna ... *Ikaros* [online], 1999, č. 8. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200208578>>.

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

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

Uniform Resource Names (urn) Charter [online]. Reston (VA, USA) : IETF, last modified 2001-07-31. Dostupný z: <<http://www.ietf.org/html.charters/OLD/urn-charter.html>>.

MOENS, M.F. Automatic indexing and abstracting of document texts. Boston : Kluwer Academic Publishers, 2000. 265 s.

PAPÍK, R. Competitive Intelligence, informační služby, Internet a informační profese. *Ikaros* [online]. 2001, č. 04 [cit. 2001-04-01]. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200208281>>. ISSN 1212-5075.

JONÁK, Z. Intelligence systémů zpracování textů. *Ikaros* [online]. 2000, č. 1 [cit. 2000-01-05]. Dostupný z: <<http://ikaros.ff.cuni.cz/ikaros/2000/c01/isko/zjonak.htm>>. Pozn.: nepřístupný

ANDĚROVÁ, I. Programový projekt MK ČR : Souhrnná zpráva za rok 2000 (upraveno). *Ikaros* [online]. 2000, č. 08 [cit. 2000-08-01]. Dostupný na World Wide Web: <<http://www.ikaros.cz/Clanek.asp?ID=200301003>>. ISSN 1212-5075.

Biblink [online]. Bath (Anglie) : UKOLN, last updated 12-Jul-2000 [cit. 14. 3. 2001]. Dostupné na World Wide Web: <<http://hosted.ukoln.ac.uk/biblink/>>.

CELBOVÁ, L. Elektronické zdroje publikované v síti Internet jako součást České národní bibliografie. *Ikaros* [online]. 2000, č. 06 [cit. 2000-06-01].

Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200208144> >. ISSN 1212-5075

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

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

VOJTÁŠEK, F.; CELBOVÁ, I. Helsinská univerzitní knihovna přívětivá vůči každému. *Ikaros* [online], 2000, č. 9. Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200208199> >.

ŽABIČKA, P. Dublin Core - metadata pro popis elektronických dokumentů. Předneseno na konferenci DATASEM 2000, konané 21. až 24. října 2000 v Brně. Dostupný z: < <http://webarchiv.nkp.cz/datasem2000.pdf> >.

STOKLASOVÁ, B.: Budování a zpřístupnění fondů. *Daidalos* 2000. Dostupný z: http://daidalos.ff.cuni.cz/2000/prosinec/bs_ifla02.php. Pozn.: nepřístupný

NEDLIB : Networked European Deposit Library [online]. Hague (Nizozemí) : Koninklijke Bibliotheek, c1998, last updated 11-March-2001 [cit. 28. 4. 2001]. Dostupný z: < <http://www.kb.nl/nedlib> >.

Networked European Deposit Library [online]. Hague (Nizozemí) : Koninklijke Bibliotheek, last upd. 11-March-2001 [cit. 14. 3. 2001]. Dostupný z: < <http://www.kb.nl/nedlib> >.

VEJLUPEK, T. SPEIS - koncept jednotného využívání a jednotné nabídky informačních zdrojů a informačních služeb od různých poskytovatelů. Praha, 2001. 18 s.

ANDĚROVÁ, I. Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů. Souhrnná zpráva za rok ... [online]. Dostupný z: < http://www.nkp.cz/pages/page.php3?page=oazp_propoj04.htm >.

ANDĚROVÁ, I. Souborná databáze Kooperačního systému článkové bibliografie - optimalizace integrace a správy heterogenních dat. Souhrnná zpráva za rok ... [online]. Dostupný z: < <http://full.nkp.cz/nkdb/docs/dok.htm> >.

ANDĚROVÁ, I. Kooperační systém článkové bibliografie a propojení analytických záznamů s plnými texty - východiska a současný stav. *Národní knihovna : knihovnická revue*. 2001, roč. 12, č. 1, s. 26-37. Dostupný též z: < <http://full.nkp.cz/nkkr/NKCR0101/0101026.html> >.

PAPÍK, R. Competitive Intelligence, informační služby, Internet a informační profese. *Ikaros* [online]. 2001, č. 04 [cit. 2001-04-01]. Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200208281> >. ISSN 1212-5075.

ANDĚROVÁ, I. Metodika popisu článků ve formátu UNIMARC [online]. 2001. Dostupný z: < <http://www.nkp.cz/pages/page.php3?page=metodika.htm> >.

CASLIN 2001. Popis a zpřístupnění dokumentů : nová výzva. Beroun, 27.-31. května 2001 [online]. Dostupný z: <<http://www.caslin.cz:7777/caslin01/index.htm>>.

CELBOVÁ, L. Informace o projektu registrace domácích internetových zdrojů nově na serveru WebArchiv. *Ikaros* [online]. 2001, č. 5 [cit. 2001-05-01]. Dostupný též z: <<http://www.ikaros.cz/Clanek.asp?ID=200208297>>. ISSN 1212-5075.

CVRČKOVÁ, R. Služba GILS jako nástroj pro řízení informačních zdrojů z oblasti řízení státní správy USA. *Národní knihovna : knihovnická revue*. 2001, roč. 12, č.2, s. 99-113. Dostupný též z: <<http://full.nkp.cz/nkkr/NKKR0102/0102099.html>>.

SCHWARZ, J. Praktické aspekty hodnocení kvality a konzistence indexace. *Ikaros* [online]. 2001, č. 2 [cit. 2001-02-01]. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200303002>>.

H.W. Wilson Company Selects Verity to Power the New WilsonWeb Site - the Premier Reference Resource for Librarians and Researchers [online]. Dostupný z: <<http://www.verity.com/company/press/releases/release.jsp?pressID=723>>.

Integration Heterogeneous Resources : 25 Library Seminar, Prague 6-8 June 2001 [online]. Dostupný z: <URL : <http://www.stk.cz/elag2001/ELAG2001.html>>.

ANDĚROVÁ, I. Báze ANL FULL v systému TOPIC. Textová verze. *Inforum 2002*. Dostupný z: <<http://full.nkp.cz/nkdb/docs/dok.htm>>.

ANDĚROVÁ, I. Báze ANL FULL v systému TOPIC. Prezentace PPT. *Inforum 2002*. Dostupný z: <<http://full.nkp.cz/nkdb/docs/dok.htm>>.

Knihovny současnosti 2002. Sborník z 10. konference, konané ve dnech 24.-26.září 2002 v Seči u Chrudimi. Brno : Sdružení knihoven ČR, 2002. 401 s.

ANDĚROVÁ, I. Kooperační systém článkové bibliografie – KOSABI. (Vývoj a současný stav metodiky zpracování, zpřístupnění, organizace kooperace, perspektivy). Prezentace PPT na konferenci Knihovny současnosti 2002, Seč 24.-26.9.2002. Dostupný z: <<http://full.nkp.cz/nkdb/docs/dok.htm>>.

BÍNOVÁ, J.: Bibliografická sekce sdružení knihoven České republiky v letech 1995-2002. *Knihovny současnosti 2002*. Sborník z 10. konference, konané ve dnech 24.-26.září 2002 v Seči u Chrudimi. 2002, s. 182-183.

NÁDVORNÍKOVÁ, M. Nové formy a metody práce při poskytování regionálních bibliografických informací. *Knihovny současnosti 2002*. Sborník z 10. konference, konané ve dnech 24.-26.září 2002 v Seči u Chrudimi. 2002, s. 186-189.

MIKA, J. Regionální bibliografie a faktografie – příklad spojení tradičního a moderního přístupu ke knihovnické práci. *Knihovny současnosti 2002*. Sborník z 10. konference, konané ve dnech 24.-26.září 2002 v Seči u Chrudimi. 2002, s. 195-200.

KAŇKA, J. Koncepce krajského bibliografického systému. *Knihovny současnosti 2002*. Sborník z 10. konference, konané ve dnech 24.-26.září 2002 v Seči u Chrudimi. 2002, s. 195-201-205.

HRAZDILOVÁ, A. Analytické zpracování v systému T-Series v Krajské moravskoslezské knihovně v Ostravě : Výsledky řešení programového projektu Ministerstva kultury ČR. *Čtenář*, roč. 54, č. 4, 2002, s. 116-117

68th IFLA Council and General Conference, August 18-24, 2002. Glasgow. Dostupný z: <<http://www.ifla.org> >.

HADDAD, P., GATENBY, P. Providing bibliographic access to archived online resources: the National Library of Australia's approach. 68th IFLA Council and General Conference, August 18-24, 2002. Glasgow. Dostupný z: <<http://www.nla.gov.au/nla/staffpaper/2002/gatenby.html> >.

DAGERSTEDT, S.: Cataloguing and organizing library workflow – New wals. 68th IFLA Council and General Conference, August 18-24, 2002. Glasgow. Dostupný z <<http://www.ifla.org> >. Pozn.: nepřístupný

SMITH, R. The European Library Project: managing bibliographic standards at the European level. 68th IFLA Council and General Conference, August 18-24, 2002. Glasgow.

EDVARDESEN, JONNY. Newspapers at the National Library of Norway. News from the IFLA Round tabel of Newspapers. 2002, No. 10. Dostupný z: <<http://www.ifla.org/VII/s39/broch/no10.pdf> >.

SCHWARZ, P: Současný stav a trendy automatické indexace dokumentů. Přehledová studie. 2002. Dostupný z: <<http://full.nkp.cz/doc/dok.htm> .

RICHTER, V. Návrh nové „Strategie rozvoje knihoven 2003-2005“ *Knihovny současnosti* 2002. Sborník z 10. konference, konané ve dnech 24.-26. září 2002 v Seči u Chrudimi. 2002, s. 259-265.

SVOBODA, M. Knihovny v elektronické záplavě. Automatizace knihovnických procesů . 9. ročník. Liberec. 2003. Dostupný z: <<http://knihovny.cvut.cz/akp2003/> >.

SKLENÁK, V. Sémantický web. Knihovny v elektronické záplavě. Automatizace knihovnických procesů . 9. ročník. Liberec. 2003. Dostupný z: <<http://knihovny.cvut.cz/akp2003/> >.

ŽABIČKA, P. OAI-PMH: Protokol pro metadatovou interoperabilitu. Automatizace knihovnických procesů . 9. ročník. Liberec. 2003. Dostupný z: <<http://knihovny.cvut.cz/akp2003/> >.

ANDĚROVÁ, I. Kooperační systém článkové bibliografie a báze ANL, ANL FULL. Infos 2003: zborník z 32. medzinárodného informatického sympózia, ktoré se konalo v dňoch 7.-10. apríla 2003 v Starej Lesnej. Sest. Alojz Androvič, Judita Kopáčiková. Bratislava : Centrum VTISR, 2003, S. 149-161.

KOVAČKA, M. Prvé výsledky a najbližšie perspektivy národného programu retrokonverze a konverze bibliografických databáz a knižničných katalogov Slovenskej republiky. Infos 2003: zborník z 32. medzinárodného informatického sympózia, ktoré se konalo v dňoch 7.-10. apríla 2003 v Starej Lesnej. Sest. Alojz Androvič, Judita Kopáčiková. Bratislava, Centrum VTISR 2003. S. 135-140.

BARTOŠEK, M. Aktuální oblasti výzkumu digitálních knihoven Infos 2003: zborník z 32. medzinárodného informatického sympózia, ktoré se konalo v dňoch 7.-10. apríla 2003

v Starej Lesnej. Sest. Alojz Androvič, Judita Kopáčiková. Bratislava, Centrum VTISR 2003. S. 84-90.

Congress: 69th IFLA General Conference and Council. Access Point Library: Media - Information – Culture.1 - 9 August 2003, Berlin, Germany. Dostupný z: < <http://www.ifla.org/IV/ifla69/index.htm> >.

BARTOŠEK, M. Digitální knihovny. Dostupný z: < <http://www.ics.muni.cz/mba/dl-fi03/dlfi03-1.pdf> >. Nepřístupný.

ANDĚROVÁ, I. Problematika novin a World Library and Information Congress : 69th IFLA General Conference and Council. Media - Information - Culture . Ikaros [online]. 2003, č. 11 [cit. 2003-11-01]. Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200311005> >. ISSN 1212-5075.

Ikaros, redakce. Program LI znamenal průlom ve financování informačních zdrojů. *Ikaros* [online]. 2003, č. 08 [cit. 2003-08-01]. Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200307009> >. ISSN 1212-5075.

SCHWARZ, J. Současný stav a trendy automatické indexace dokumentů : přehledová studie. *Ikaros* [online]. 2003, č. 03 [cit. 2003-03-01]. Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200303002> >. ISSN 1212-5075. Pozn.: Originální verze studie je zveřejněna na adrese <http://full.nkp.cz/nkdb/docs/studie/MAJobsah.html>, zkrácená a upravená verze studie byla prezentována na konferenci *Znalosti 2003* (viz Schwarz, J. Současný stav a trendy automatické indexace dokumentů. In Svátek, V. (ed.). *Znalosti 2003 : 2. ročník konference*, Ostrava, 19.-21. únor 2003. Sborník příspěvků. Ostrava : VŠB-Technická univerzita Ostrava, 2003, s. 212-221.

HARTMANOVÁ, D. Knihovna on-line a autorské právo. *Národní knihovna : knihovnická revue*, 2003, roč. 14, č. 2, s. 100-105. Dostupný též z: < <http://full.nkp.cz/nkkr/NKKR0302/0302100.html> >.

BROŽEK, I.; PĚNKAVOVÁ, P.; ŠTĚRBOVÁ, J.; HOREJSKOVÁ, M. Knihovny současnosti 2003. *Ikaros* [online]. 2003, č. 10 [cit. 2003-10-01]. Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200309009> >. ISSN 1212-5075.

SVOBODA, M. PQNext: Nová verze vyhledávacího prostředí služby ProQuest 5000. *Ikaros* [online]. 2003, č. 08 [cit. 2003-08-01]. Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200308004> >. ISSN 1212-5075.

CELBOVÁ, L. Automatizace knihovnických procesů podeváté. *Ikaros* [online]. 2003, č. 06 [cit. 2003-06-01]. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200305005>>. ISSN 1212-5075.

SVOBODA, M. Zpráva z cesty na seminář ELAG 2003. *Ikaros* [online]. 2003, č. 08 [cit. 2003-08-01]. Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200308001> >. ISSN 1212-5075.

MACHONSKÁ, J. Cenová politika databázových center. Historie a současnost. *Národní knihovna : knihovnická revue*, 2002, roč. 13, č. 3, s. 177-195. Dostupný z: < <http://full.nkp.cz/nkkr/NKKR0302/0302100.html> >.

- BRATKOVÁ, E. Citace odborné literatury jako nástroj rozvoje služeb a integrace digitálních knihoven. Dostupný z: < <http://platan.vc.cvut.cz/akp/clanky/12.pdf> >.
- Van de SOMPEL, H.- HOCHSTENBACH, P. Reference Linking in a Hybrid Library Environment. Part 1-3. *D-Lib Magazine* [online]. 2000, vol 5, no. 4, no. 10. Dostupný z: < <http://www.dlib.org> >.
- PAVLÍK, J. Linkování na relevantní přidané služby v rámci Jednotné informační brány Knihovny současnosti 2003. 11. konference konaná ve dnech 16.-18.září 2003 v Seči u Chrudimi. Dostupný z: < http://jib-info.cuni.cz/dokumenty/sec2003/sec2003_sfx.doc >.
- SKLENÁK, V. Vyhledávací stroje v prostředí Internetu – a co bude dál? Dostupný z: < http://platan.vc.cvut.cz/akp2003/sbornik/03_sklenak.pdf >.
- ANDĚROVÁ, I. Aktuální informace o Kooperačním systému článkové bibliografie SDRUK , 13. zasedání Sekce pro bibliografii. Hradec Králové, 5.-6.5. 2003. Dostupný z: < <http://full.nkp.cz/nkdb/docs/dok.htm> >, Rubrika Co je nového
- ANDĚROVÁ, I. Problematika novin a World Library and Information Congress : 69th IFLA General Conference and Council. Media - Information - Culture . *Ikaros* [online]. 2003, č. 11 [cit. 2003-11-01]. Dostupný z: < <http://www.ikaros.cz/Clanek.asp?ID=200311005> >. ISSN 1212-5075.
- BULÍNOVÁ, E. Souborné katalogy Ruska, Běloruska, Ukrajiny a pobaltských zemí. *Ikaros* [online]. 2004, č. 01 [cit. 2004-01-01]. Dostupný z <<http://www.ikaros.cz/Clanek.asp?ID=200401008>>. ISSN 1212-5075.
- JEDLIČKOVÁ, P. Tvořivé propojení technické inteligence a zájmu o humanitní disciplíny : rozhovor s Prof. PhDr. Marií Königovou, CSc. *Ikaros* [on line]. 2004, č. 01 [cit. 2004-01-01]. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200311018>>. ISSN 1212-5075.
- SKOLKOVÁ, L. OCLC láká prostřednictvím Google nové uživatele do knihoven. *Ikaros* [online]. 2003, č. 12 [cit. 2003-12-01]. Dostupný z: <<http://www.ikaros.cz/Clanek.asp?ID=200312002>>. ISSN 1212-5075.
- BALÍKOVÁ, Marie: Harmonizace věcní katalogizace v Česku, sen či realita. *Knihovny současnosti 2004* : sborník z 12. konference (Seč u Chrudimi : 14.-16. září 2004).- Brno : Sdružení knihoven ČR, 2004, s. 307-322.
- ANDĚROVÁ, Ivana. Aplikace konspektu v článkové bibliografii. [Online]. Praha, Národní knihovna ČR, posl. aktual. 10. 11.2004 [cit. 2004-12-10]. Dostupný z: <http://www.nkp.cz/pages/page.php3?page=oazp_metodika.htm>.
- ANDĚROVÁ, Ivana. Kooperační systém článkové bibliografie a excerpční základna. [Online]. Praha, Národní knihovna ČR, posl. aktual. 10. 11.2004 [cit. 2004-12-10]. Přístup z: http://www.nkp.cz/pages/page.php3?page=oazp_kooperace.htm
- Záznam pro soubornou databázi: MARC 21. Fyzicky nesamostatné části dokumentu. Tištěné monografie a seriály. [Online]. Praha : Národní knihovna ČR, duben 2004 [cit. 2004-12-10]. Přístup z: http://full.nkp.cz/nkdb/docs/mini2_ANL_M21.htm

Definition of Data Elements for Article Level Description. Proposal No: 2003-03. California Digital Library. [Online]. Berkeley : University of California Berkeley Library, April 1994 [cit. 2004-12-10]. Přístup z: <http://www.loc.gov/marc/marbi/2003/2003-03.html>

SICI : Serial Item and Contribution Identifier Standard. ANSI/NISO Z39.56-1996 Version 2. [Online]. Berkeley : University of California Berkeley Library, April 1994 [cit. 2004-12-10]. Přístup z: <http://sunsite.berkeley.edu/SICI>

ČERMÁK, F., SCHMIEDTOVÁ, V. Ústav českého národního korpusu FF UK. *Národní knihovna : knihovnická revue*, 2004, roč. 15, č. 3, s. 152-158. Dostupný z: < <http://full.nkp.cz/nkkcr/nkkcr0403/0403152.html> >.

RICHTER, V. Koncepce rozvoje knihoven v ČR 2004-2010. *Veřejná správa*, 2004, , roč. 15, č. 40, s. 12,21. Dostupný z: <http://www.mvcr.cz/2003/casopisy/vs/0440/konz_info.html>.

Některé globální odkazy

IFLA. Dostupný z:< <http://www.ifla.org/>>.

TOPIC

[Verity](#)

Standardizace

<http://www.nkp.cz> (katalogizační politika)

JIB

[Dokumenty - konspekt](#)

[Webarchiv](#)

[Statement of International Cataloguing Principles](#)

Server FULL.NKP.CZ . Dostupný z: < <http://full.nkp.cz> >

Báze ANL FULL .Dostupný z: < <http://full.nkp.cz> >

Management Kooperálního systému článkové bibliografie - MNG KOSABI.

Dostupný z: < <http://full.nkp.cz> >

Dokumenty. Dostupný z: < <http://full.nkp.cz/nkdb/docs/dok.htm>>.

Výsledky práce společnosti ANOPRESS IT. Dostupný z: < <http://www.anopress.cz>>.

Báze ANL [online]. Dostupný z: < http://sigma.nkp.cz/F/?func=file&file_name=base-list >.

Národní knihovna. *Knihovnická revue* [online]. Dostupný z:

<<http://full.nkp.cz/nkkcr/NKKR.html>>.

Excerptní základna: [Báze Titulů](#); [Krajské knihovny a MZK](#); [Národní knihovna](#); [Specializované – odborné knihovny](#) ; [Instituce přistupující do kooperace](#)

Zahraniční reference

General information about ANL - Articles in Czech newspapers, magazines and collections of works.Login procedures to ANL [FULL] - Articles published in Czech journals and newspapers (full texts). *Gabriel*. Dostupný z:< <http://portico.bl.uk/gabriel/index.html>>.

ANL FULL. Dublin Core Metadata Initiative. Dostupný z:< <http://dublincore.org/projects/europe.shtml#denmark>>.

Tel Digital deposits state of the art review. Marco de Niet, Koninklijke Bibliothek. With contribution from Liesbeth Pskamp, Koninklijke Bibliothek. 18 December 2001. 0.2 (Second draft version).D1.1/R/Report. DEL/007. Hague, Koninklijke Bibliothek 2001. 84 p. Dostupný z: <http://www.europeanlibrary.org/doc/tel_results_d11_v02.doc>.

IFLA-Directory of Serials Content Databases and Current-Awareness Services for Serials Content. [V přípravě]. *IFLANET*. Dostupný z: < <http://www.ifla.org/I/whatsnew/new2002.htm> >.

A.2 Současný stav ve světě a v ČR

A.2.1 Obecně

V současné době vznikají nové modely získávání, zpracování a zpřístupňování bibliografických informací v návaznosti na elektronické publikování na základě přehodnocení klasických knihovnických postupů s ohledem na budování digitální knihovny (interoperabilita jednotlivých komponent z hlediska technického, strukturálního, syntaktického a sémantického). **Propojují se různé informace z hlediska formy, druhu a obsahu**, strukturované a nestrukturované fulltextové báze, elektronické archívy. Propojují se katalogy knihoven, záznamy s plnými texty dokumentů, "síťové dokumenty", je podporována spolupráce s archívy apod. Hovoří se o popisu dokumentu v hierarchii jako manifestace díla (čtyřúrovňový model manifestace díla **FRBR**) - vztahy mezi dílem, jeho vyjádřením, projevem a exemplářem.

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, konzorcií a licencí. **Při zpřístupňování elektronických informací se stále více prohlubuje spolupráce mezi státním a soukromým sektorem.**

Předpokladem plnohodnotného zpřístupnění plných textů je implementace metadat do plných textů. Tato metadata (Dublin Core, MARC) mohou být vytvářena autorem, vydavatelem, distributorem, knihovníkem a zpřístupňována na webu pomocí XML/RDF s definovanou standardní strukturou DTD. Existují iniciativy, které se zabývají konverzí mezi DC, MARC a XML. DC je určen primárně pro otevřený web, lze ho užít i pro databáze tzv. hlubokého webu. Významný je OAI-PHP - protokol pro metadatovou interoperabilitu, umožňující automatizované získávání metadat a vzájemnou komunikaci archivů, digitálních knihoven, je založený na DC a XML. V současné době se velká pozornost věnuje protokolům pro komunikaci a sdílení dat - **Z39.50 a Bath Profile a tzv. otevřenému nebo dynamickému propojování pomocí tzv. OpenURL.**

Jednoznačná indentifikace plných textů je jednou z podmínek zpřístupnění plných textů. K identifikaci služeb, zdrojů a objektů na internetu slouží nestabilní **URL (Uniform Resource Locator), PURL (Persistent URL), Uniform Resource Name URN (Uniform Resource Name), DOI (Document Object Identifier), SICI (Seriaal Item and Contribution Identifier)** aj.

Informační brány a portály usnadňují přístup k heterogenním informačním zdrojům. Předpokladem plnohodnotného zpřístupnění dokumentů je standardizace a cílem je sémantický web.

Současným trendem, je extrahování, automatizované zpracování metadat, otevřené propojování informací, pojmové vyhledávání v plnotextových databázích a interoperabilita systémů založených na různých platformách.

Předkládaná zpráva je poměrně podrobná vzhledem k rozsáhlé výzkumné problematice. Jsou v ní použity některé pasáže z projektu Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů, se kterým bezprostředně souvisí. V některých charakteristikách odborných pojmů vychází z termiologie uvedené v databázi knihovnictví a informační vědy – TKDIV).

A.2.2 Současný stav v zahraničí

A.2.2.1 Některé trendy a nástroje v zahraničí

Dochází tak ke kombinaci tradičně katalogizovaných dat s katalogizací vzdálených zdrojů, s dodáváním metadat od autora, vydavatele/nakladatele/distributora i dat získaných na základě automatizovaného sběru.

Mění způsoby informačního chování uživatele při vyhledávání, ve středu zájmu je komunikace člověk - počítač (human-computer interaction). Na základě zpětné vazby relevance (relevance feedback) může uživatel zpřesňovat svůj dotaz a spolupracovat se systémem.

Kombinace bibliografických a plnotextových databází představuje efektivní přístup k plnému textu. Vyhledávání s přidanou hodnotou a intelektuální indexace věcná zvyšují možnost získání relevantních informací.

Zdroje se integrují do informačních bran, portálů, virtuálních, digitálních či elektronických knihoven.

Informační brány a portály usnadňují přístup k heterogenním informačním zdrojům. Předpokladem plnohodnotného zpřístupnění dokumentů je standardizace a cílem je sémantický web.

Základní trend: kooperace v rámci interoperability.

Interoperabilita je schopnost dynamické spolupráce mezi technicky různorodými a nezávislými komponentami z hlediska syntaktického, strukturálního a sémantického. Interoperabilitu kromě jiných umožňují i zde jmenované nástroje. **Různé úrovně interoperability z hlediska použitých prostředků: webové vyhledávače, silné standardy – MARC, Z39.50; metada, jejich sklizení a otevřené standardy; zdroje s metadaty volně zapojené do kooperace – DC, XML, RDF; interoperabilita v oblasti propojování zdrojů např. OpenURL, ERL, dále pak propojení citlivé na kontext uživatele (open context-sensitive linking) - OpenURL a SFX. Významné jsou aktivity v oblasti ontologií (systém konceptů a vztahů mezi nimi).**

Ontologie – metoda získávání znalostí. Dílčí ontologie souvisí s konceptualizací jednotlivých oblastí. Ontologie, resp. tzv. topikové mapy umožňují členit textové univerzum z hlediska sémantiky. Kategorie je třeba propojit s koncepty.

Na základě interoperability mohou vznikat snadněji elektronické archívy, souborné katalogy virtuální i reálné, brány a portály.

Cílem je propojování informací, distribuované vyhledávání, relevantní (pertinentní) informace pro uživatele a sémantický web. Tzv. sémantický web předpokládá postupnou transformaci současného WWW srozumitelného pro lidi na WWW srozumitelného pro počítače (znalosti označené značkovacími jazyky nebo extrahované z textu).

Reálný souborný katalog

Souborný katalog ve formě fyzicky existující databáze, do které jsou dodávány záznamy dokumentů jednotlivých účastnických knihoven.

Virtuální souborný katalog

„Technologie propojení nezávislých knihovnických katalogů pomocí jednotného uživatelského rozhraní, které umožňuje paralelní prohledávání jednotlivých katalogů a vytváří virtuální (reálně neexistující) **souborný katalog**. Základním předpokladem funkce virtuálního souborného katalogu je standardní vyhledávací protokol, jenž podporuje formulaci rešeršního dotazu a zpřístupnění záznamů (např. komunikační protokol Z39.50).“

Informační brána

„Služba v síťovém prostředí určená pro zprostředkování přístupu k vybraným online informačním zdrojům určitého oborového nebo tematického zaměření. Zpřístupňované **informační** zdroje procházejí procesem intelektuálního nebo automatického výběru a zpracování na základě definovaných formálních a kvalitativních kritérií. Součástí předmětové brány je obvykle klasifikační systém členící **informační** zdroje podle oborů“.

Informační brány řeší přístup k různým zdrojům z jednotného prostředí. Nejpoužívanějším standardem pro tvorbu metadat je **Dublin Core (DC)**. Jejich fungování závisí na existenci pokud možno homogenního nástroje pro věcnou indexaci harmonizací řízených slovníků a tezurů a hledají se cesty k řešení vícejazyčnosti. Jednou z řešených metod pro zpřístupnění elektronických informačních zdrojů je **metoda konspektu, předmětová kategorizace** pro popis informačních zdrojů. Spočívá hierarchickém uspořádání předmětových kategorií, na nejvyšší úrovni není propojena se systematickou klasifikací. Cílem je jednotný tematicky strukturovaný popis heterogenních informačních zdrojů pro potřeby koordinovaného budování knihovních sbírek a knihovních fondů a pro tvorbu nástroje určeného ke zpřístupnění kvalitních (zhodnocených) heterogenních informačních zdrojů v síťovém prostředí, tj. tematických bran. Slouží k tomu údaj o předmětové kategorii spolu s vybraným znakem MDT. Pro mezinárodní srozumitelnost je potřebná konkordance MDT a DDC.

Portál

Webové sídlo, které poskytuje širokou škálu služeb a informací, často s možností jejich přizpůsobení uživateli podle osobních potřeb a zájmů.

Specializovaný **portál** zpřístupňující informační zdroje zaměřené na určitou cílovou skupinu uživatelů, která může být vymezena např. geograficky nebo tematicky.

Standardy a některé nástroje interoperability

Jmenný popis

Na mezinárodní úrovni se mění tradiční pojetí seriálů, které bude mít vliv i na naši **katalogizační praxi** v této oblasti. Termín **seriál je revidován**. V současné době již existuje revidovaný standard **International Standard Bibliographic Description for Serial and other Continuing Resources ISBD (CR)**. Revize je ovlivněna novou kategorií **tzv. pokračujících zdrojů** (bibliografický zdroj, který je vydáván v čase s předem neurčenou dobou ukončení, zahrnuje integrující zdroje a seriály, povaha pokračujících zdrojů je dynamická, pokračující a měnitelná.). **Integrující zdroje** jsou zdroje, které jsou aktualizovány a tyto aktualizace nemohou existovat samostatně. **Dochází k harmonizaci Angloamerických pravidel AACR2R, mezinárodního bibliografického popisu ISBD a mezinárodního registračního systému mezinárodního čísla seriálových publikací ISSN**. Do kategorie pokračujících zdrojů patří též ukončené pokračující zdroje – vycházejí po částech, periodicky a jsou číslovány, ale jejich trvání je ohraničeno, dále reprinty seriálů. Pro praktické potřeby je navržena **definice seriálů**: seriál je pokračující zdroj, který je vydáván po oddělených částech, obvykle je číslován, nemá předem určenou dobu ukončení. Definice zahrnuje časopisy, magazíny, elektronické časopisy, pokračující adresáře, roční zprávy, noviny a monografické edice (těmito kategoriemi se zabývá kapitola 12 AACR2R). Pro popisu **elektronických zdrojů** je určen standard **ISBD (ER)** a kapitola 9 AACR2R. **Functional Requirements for Bibliographic records (FRBR)** e uvádí čtyřúrovňový popis, orientuje se na obsah dokumentu, nikoli na nosič, umožňuje integrovat elektronické dokumenty mezi dokumenty tradiční. Účelem studie je definovat funkce bibliografického záznamu určeného pro různé typy dokumentů, způsoby využití, pro různé uživatelské potřeby. V současné době probíhají aktivity, které zkoumají použitelnost pravidel AACR2 i formátu MARC vzhledem k FRBR a aktivity v oblasti přípravy mezinárodních katalogizačních pravidel.

V oblasti **věcného zpracování**: zjednodušení syntaxe LCSH za současného zachování lexiky, věcná kategorizace informací do určitého počtu skupin na několika úrovních (např. metoda konspektu), důraz na autority a jejich mezinárodní srozumitelnost (projekty projekt MACS a LEAF).

Dublin Core (DC) je standardizované metadatové schéma pro popis informačních zdrojů zejména na internetu. Dublin Core je tvořen souborem patnácti základních prvků (jednoduchý Dublin Core), které lze specifikovat kvalifikátorem prvku a hodnoty (kvalifikovaný Dublin Core). Dublin Core nepředepisuje závaznou syntaxi (jedná se o sémantický standard), je základem dalších metadatových standardů.

Každý prvek je volitelný a opakovatelný, nezáleží na jejich pořadí. Důležitou podmínkou interoperability je používání hodnot prvku z dohodnutých souborů autorit.

Vazba mezi metadatovým záznamem a zdrojem, resp. plným textem, který popisují, může být dvojitá: metadatový záznam je uložen samostatně a odděleně od zdroje nebo metadata jsou vnořena (embedded) přímo do samotného zdroje (pomocí značek u dokumentů v jazyce SGML, HTML aj.)

Z DC vychází např. metadatový standard OAI (Open Archives Initiative) primárně vyvinutý pro vědecké a akademické komunity. Projekty, které využívají DC, je možno najít na adrese <http://dublincore.org/projects>. V budoucnu by mělo dojít k možnosti konverze mezi národními metadatovými záznamy.

(DC se v České republice zabývají pracovníci ÚVT MU, materiály týkající se DC jsou prezentovány na http://www.ics.muni.cz/dublin_core/index.html, projekt [Webarchiv](#), [JIB](#), projekty týkající se článkových informací).

RDF (Resource Description Framework)

Poskytuje základ pro popis v různých aplikačních doménách. Jako modelovací jazyk používá entity, atributy, vztahy.

XML (eXtensible Markup Language)

Jazyk XML je, podobně jako jazyk HTML, prostředek sloužící k zapsání **strukturovaného** textu, zvláště pak textu určeného k šíření v prostoru www. XML odděluje popis struktury dat od jejich prezentace (pomocí tzv. style sheetů). To umožňuje snadnou konverzi do jiných formátů, možnost prezentace dat různými způsoby (HTML, postscript, UNIMARC apod. textový formát). Každý dokument má definovanou svoji strukturu prostřednictvím tzv. **DTD** (Document Type Definition). V současnosti probíhají [aktivita](#) v oblasti mapování formátu DC do MARC a opačně a převodu do XML.

OAI-PHP je protokol pro metadatovou interoperabilitu, umožňuje automatizované získávání metadat a vzájemnou komunikaci archivů, digitálních knihoven – otevřené technické řešení, dostupnost SW komponent. Založen na použití jednoznačných identifikátorů, metadatový standard je nekvalifikovaný DC, komunikace přes HTTP a využití formátu XML. Protokol aplikují souborné databáze, archivy volně dostupných vědeckých prací (arXiv.org) a výzkumných institucí (CERN), knihovny (Library of Congress). Služba DP9 – umožní indexaci metadat webovými roboty, funkce pro indexaci fulltextů. Další aplikací je projekt Open Citation – automatická tvorba citačních resjtríků. Protokol vyvinut primárně pro potřeby akademické komunity pro zveřejnění informací. Možno využít při importech d souborných katalogů. Lze použít na bibliografické databáze i souborné katalogy díky schématu MARCXML zveřejněném Library of Congress.

Připravují se **nástroje pro převod dat Dublin Core/MARC [Library of Congress: Dublin Core/MARC/GILS Crosswalk](#)**, které využívají prvky DC v katalogizaci.

Propojování

Automatické a dynamické propojování informačních zdrojů (Technologie „SFX“ Special Effects vyvinuté na Gentské univerzitě a Národní laboratoři v Los Alamos).

Propojování na základě citací - [ISI](#) buduje SCI (Science citation Index). Projekty založené na propojování archivů na základě citací – [LANL](#) (propojování na základě přidělovaného identifikátoru a formátu HyperTeX, Los Alamos National Laboratory) a projekt [OpCit](#) (The Open Citation Project). Významná je iniciativa v oblastí propojování [OAI](#) (The Open Archive Initiative).

Základ propojení by měl být uložen již v samém počátku vzniku bibliografických záznamů, jejich katalogizace jednodušší a efektivnější. - **FRBR Functional Requirements for Bibliographic Records** - relace mezi Dílem, Vyjádřením díla, Provedením díla, a Exemplářem díla (Work, Expression, Manifestation, Item). V současné době probíhají aktivity tímto směrem v oblasti katalogizace – [Ustanovení mezinárodních zásad katalogizace](#) (setkání IFLA, Německo 2003).

Propojení může být statické a dynamické, otevřené a uzavřené.

K identifikaci, propojení a zpřístupnění elektronických objektů, resp. plných textů na internetu tedy slouží: URL, PURL, URN a DOI, SICI a klasické identifikátory jako je ISSN, ISBN. Některé z nich mohou být součástí OpenURL.

Integrace strukturovaných a nestrukturovanýchází dat, inteligentní vyhledávací systémy, management znalostí, poloautomatické zpracování textu a extrakce dat, expertní systémy, znalostní systémy

Problematika získávání, zpracování, zpřístupňování a využívání **znalostí** je velmi aktuální v oblasti výzkumu informačních technologií. Významný **zdroj znalostí** jsou relační **databáze**. Dále jsou významné techniky tvorbyází formalizovaných znalostí na základě spolupráce znalostního inženýra s **lidským expertem**. Dalším zdrojem znalostí jsou (vedle databází a expertů) **textové dokumenty**. Jejich **indexace** a **vyhledávání** je založeno na statistických a lingvistických charakteristikách (**extrakci informací** z volného textu). Systémy využívající formalizované znalostní báze směřují k aplikacím složeným ze samostatných komponent - **agentů** - s vlastním mechanismem řízení a založené na možnosti **sdílet** a **znovu používat** znalosti založené na syntaktické a sémantické standardizaci (značkovací jazyky a znalostní ontologie tj. konceptualizace určité oblasti). Hlavním cílem těchto technologií je zdokonalení **znalostního managementu** v institucích, firmách V důsledku je efektivní vazba tzv. knowledge managementu na rozhodovací a plánovací procesy.

Mezi renomované znalostní systémy patří systém TOPIC (concept based system) a jeho nová verze Portal One, resp. K2 americké firmy Verity. Informace o systému a reference jsou na adrese společnosti [TOVEK](#). Dalším z těchto systémů je systém Convera Retrieval Ware společnosti Excalibur Technologies. Informace o systému jsou na adrese společnosti [INCAD](#). Předností TOPICu je především pojmové vyhledávání a vysoké interaktivní schopnosti, předností Convera vysoká tolerance chyb při vyhledávání bez ohledu na chyby. Oba systémy umožňují fuzzy vyhledávání.

Pro dnešní dobu je charakteristický trend propojování technologií zpřístupňujících strukturovaná data (relační databáze) s fulltextovými databázemi s nástroji podporujícími zpřístupnění semistrukturovaných dat.

Vyhledávání (cit. Červený, 1999)

Pro vyhledávání je charakteristický nepoměr mezi úplností a přesností vyhledávání. Úplností vyhledávání rozumíme kvantitativní údaj udávající poměr vyhledaných relevantních a všech vyhledaných textů. Přesností vyhledávání rozumíme poměr vyhledaných relevantních a všech relevantních textů. Ideálně je hodnota koeficientu úplnosti i přesnosti rovna 1.

Intelligence informačních systémů je schopnost nalézt shodné modely i v případě textů s vysokým počtem rozdílných prvků a naopak eliminovat texty zdánlivě shodné, obsahující vysoké procento pouze formálně, nikoliv obsahově shodných prvků. Inteligentní systémy využívající poznatky z lingvistiky či kognitivní vědy.

Interaktivní pojetí vyhledávání

Teoretický výzkum interaktivního vyhledávání se soustředí většinou na možnosti zjišťování informací o kognitivním stavu uživatele a jejich použití při konstrukci dotazu. Objevují se však pokusy přímo kognitivní stav uživatelů ovlivňovat. Oblast interaktivního vyhledávání informací je interdisciplinární povahy.

Z hlediska databáze plných textů je z moderní lingvistiky velice zajímavý směr, který se nazývá „textová lingvistika“. Jedná se o lingvistickou disciplínu, která považuje za základní jednotku jazyka text. Na vývoji moderní lingvistiky je zajímavé, jak se postupně přenáší zájem jazykovědců ke zkoumání stále větších celků, od hlásek, přes věty až k celým textům (další pravděpodobný krok bude

zřejmě od textu k hypertextu). Textová lingvistika již definuje některé pojmy sloužící k popisu textu jako celku. Některé z nich (Makrostruktura, Témata) do jisté míry korespondují s definicí topiku v systému TOPIC. Dalším směrem ve vývoji těchto systémů je aplikace umělé inteligence, zejména pak systému na porozumění přirozenému jazyku. Informační systém, který by byl založen na tomto principu, by nepotřeboval selekční jazyk a vyhledávání dokumentů by probíhalo dotazováním se systémem v přirozeném jazyce.

Poloautomatická indexace textu (cit. J. Schwarz, 2002)

Většina současných systémů (vč. komerčních) určených pro automatickou indexaci či **poloautomatickou indexaci (machine-aided indexing) (vč. komerčních) nepracuje plně automaticky, nýbrž funguje jako automatizovaná podpora intelektuální indexace.**

Zpřístupňování plných textů uživatelům a legislativně právní problematika, konsorcia, cenová politika

Zpřístupňované plné texty mohou mít v zásadě dvojí podobu: mohou existovat volně, nebo jsou licencované (přístup formou licencí). Výhodné je zpřístupňovat plné texty v rámci konzorcií.

Legislativa zatím ve většině zemí neumožňuje dostatečně bezproblémově zpřístupňovat elektronické zdroje, které nejsou volné - zákon o povinném výtisku ve většině zemí nezahrnuje elektronické publikace on line, většinou se poskytují tyto zdroje na principu dobrovolnosti a na základě individuálních smluv.

Někde jsou zahrnuty pouze off-line zdroje (USA, Rakousko, Německo, Francie, ČR, Švédsko. V Kanadě, JAR, Dánsko, Švédsko, Norsko, jsou do zákona o povinném výtisku zahrnuty i elektronické zdroje. Příprava pozměňovacích návrhů: Austrálie, Japonsko, Rakousko, Německo, Francie, Švédsko, VB (návrh nového zákona prošel v parlamentu), ČR. Existuje dokument CENL/FEP (Conference of European National Libraries/Federation of European Publishers) – Mezinárodní deklarace k odevzdávání elektronických dokumentů do konzervačních fondů). Se zpřístupňování plných textů souvisí i problematika autorských práv.

Problematikou autorských práv se zabývá EBLIDA (European Bureau of Library, Information and Documentation Associations) a WIPO.

A.2.2.2 Některé služby a systémy zpřístupňující a zpracovávající články v zahraničí

Články jsou zpřístupňovány zejména v severských zemích (Švédsko-Libris, Dánsko - Basis, Norsko-Bibsys, Finsko – Arto). Velké článkové databáze se budují v Holandsku, Španělsku, Německu, Gruzii, Litvě, Rusku, Slovinsku, Makedonii, zemích bývalé Jugoslávie, Lucembursku, aj. v zemích bývalého východního bloku. Jsou budovány katalogy virtuální i reálné založené na kooperaci, jenooborové i univerzální, v rámci jednoho státu i mezinárodní ([Eesti raamatukogude elektronkatalog](#), [Consensus Omnium: Корпоративная сеть библиотек Урала](#), [Ярославская корпоративная библиотечная сеть \(OPAC\)](#), [Открытая корпоративная библиотечная система Челябинского региона](#), [Catalogue collectif des bibliothèques luxembourgeoises](#), [LIBRIS](#), [Gegdir](#), [COBIB.SR](#), [COBIB.CG](#), [COBIB.BH](#), [COBIB.SI](#), [REBECA](#), [PORBASE](#) , [REBIUN](#), [Virtueller Katalog Kunstgeschichte](#) aj.).

Poměrně kompletní přístup k elektronickým časopisům nabízí **OCLC First Search Electronic Collection Online**. Kvalitně zpřístupňují plné texty Ingenta, **ProQuest**, **EBSCO**, **Wilson Web** aj. K dispozici jsou časopisecké zdroje na nakladatelských serverech, tituly vědeckých časopisů s volným přístupem k obsahům a abstraktům, někde i k plným textům, denní tisk a časopisy populární, popularizační i odborné včetně jednotlivých článků. Plné texty jsou zpřístupňovány jak soukromými společnostmi, tak knihovnami v rámci portálů, digitálních knihoven, archivů, multioborových i oborových databází. Jde o databáze konkrétních vydavatelů, agregátorů nebo servery konkrétních titulů.

Uživatelé v České republice mají dnes přístup k zahraničním informacím o člancích různého typu z hlediska formy i obsahu zejména díky programu MŠMT "Informační zdroje pro výzkum a vývoj" (LI) a programu 1N.. Infozdroje.cz obsahují informace o všech produktech a službách dostupných díky financování v rámci těchto programů.

Společnost [Ovid](http://Ovid.com) uvedla službu **Ovid OpenLinks** („universal link resolver“) - službu, která umožňuje generovat propojení mezi informačními zdroji a službami podporujícími standard .

Bibliografické záznamy přístupné v databázích od firmy [H.W.Wilson](http://H.W.Wilson.com) jsou propojeny pomocí funkce WilsonLink (OpenURL) s plnými texty článků v elektronické knihovně JSTOR, která obsahuje více než 320 časopisů zaměřených na humanitní a společenské obory. Firma aplikovala též vyhledávací technologii **Verity**.

JSTOR – Journal Storage

Mezinárodní nevýdělečné konsorcium zaměřené na digitalizaci a zpřístupnění klíčových amerických humanitních vědeckých časopisů (v současnosti je k dispozici databáze 117 časopisů z Arts & Science Collection, obsahující všechny články od prvního čísla časopisu, s retrospektivou do minulého století, až po současnost).

IBZ - Internationale Bibliographie der Zeitschriftenliteratur

Mezinárodní článková bibliografie ze všech vědeckých oborů je nově nabízena i na CD-ROM. Příprava její tištěné verze byla zahájena již v roce 1840 v Lipsku a k roku 1896 se datuje její první vydání. Od té doby je pravidelně vydávána až do současnosti. Pětiletá kumulace let 1989-1993 na CD-ROM obsahuje 60 svazků tištěné verze IBZ s více než 2,7 miliónu záznamů. Od roku 1994 vycházejí samostatné ročníky IBZ na CD-ROM nabízející přes 120 000 článků z více než 6 000 titulů periodik. Jednotlivé záznamy obsahují autorské a názvové údaje o článku, klíčová slova, popis periodika, oborové zařazení obsahu článku, ISSN a ISBN. Software umožňuje práci s databází v jazyce německém i anglickém. Retrospektiva od roku 1983 (on-line verze).

EIFL-Direct, EBSCO

Zpřístupnění plnotextových databází odborných zahraničních periodik na základě programu Open Society Institute EIFL-Direct - **plošná multilicence**. Plné texty celkem cca 3300 časopisů od r. 1990 a další inf. zdroje (abstrakty, zpravodajství, příručky) především z oblasti sociálních a humanitních věd od **EBSCO Publishing, jednoho z předních světových dodavatelů el. a tištěných časopisů**,

Další některé dílčí databáze

[ERIC](http://ERIC.com), [MEDLINE](http://MEDLINE.com), [Academic Search Premier](http://AcademicSearchPremier.com), [Regional Business News](http://RegionalBusinessNews.com), [Newspaper Source](http://NewspaperSource.com), [Business Source Premier](http://BusinessSourcePremier.com), [MasterFILE Premier](http://MasterFILEPremier.com).

OCLC FirstSearch Service

Služba OCLC FirstSearch s přístupem k plným textům OCLC Base Package with Full Text od organizace OCLC. Služba FirstSearch kombinuje funkce souborných katalogů, meziknihovnických služeb, dodávání dokumentů a přístupu k elektronickým plným textům dokumentů. Kromě souborného katalogu (WorldCat) je zahrnut přístup do cca 12 dalších databází: ArticleFirst (bibliografické citace článků z cca 13.000 periodik), ContentFirst (seznamy obsahů periodik), NetFirst (bibliografické údaje o odborných zdrojích na Internetu, včetně abstraktů a klasifikace), PapersFirst (referáty ze světových konferencí, kongresů, symposií, výstav a workshopů od 1983), ProceedingsFirst (seznamy obsahů sborníků z vědeckých konferencí), UnionLists (souborný katalog periodik s uvedenými lokacemi), WilsonSelect (plné texty článků z 800 periodik), WorldAlmanac MEDLINE (medicínská informace), ERIC (bibliografie literatury z oblasti vzdělávání) a další.

Služba OCLC **[CORC](http://CORC.com) Cooperative Resource Catalog** poskytuje nástroj pro automatickou katalogizaci elektronických zdrojů přímo na webu (vyhledávání, vytváření a editace záznamů) ve

formátech MARC a DC. Na jeho testování se podílelo více než 450 knihoven z celého světa. Databáze vznikla ze záznamů původně uložených v bázích OCLC InterCat a NetFirst.

GILS

V rámci služby kooperují kromě jiných institucí knihovny v jednotlivých státech USA. Tzv. lokátorové záznamy odkazují na zdroje, které uchovávají úřady státní a místní správy nebo státní instituce. Pomocí lokátoru se zjišťují, popisují a zpřístupňují informace o zdroji.

V rámci systému **GILS existuje trojí způsob tvorby záznamů: klasický způsob (ruční), konverze (převod záznamů z jinýchází dat), automatizovaný způsob (extrahování metadat). Posledně jmenovaný způsob spočívá v automatizovaném vytváření záznamů při prohledávání a indexaci zdrojů** (metadat a samotného textu). Z tohoto důvodu je nutné, aby se metadata stala součástí www dokumentů. **Dosavadní editory pro tvorbu webovských dokumentů neumožňují vytvářet a automaticky začlenit prvky metadat, je nutné formát pro metadata ručně vložit do zdrojového dokumentu. Předpokládá se, že formuláře nahradí efektivnější software pro generování metadat.**

V rámci projektu **Nordic Metadata** byl vytvořen konvertor dat, který je schopen generovat záznamy ve formátech MARC severovýchodních zemí a USMARC ze zdrojových údajů DC.

Projekt Renardus: Akademický tematický portál konsorcia 12-ti institucí. Řešen v rámci pětiletého rámcového programu EU "Technologie pro informační společnost". Renardus má umožňovat paralelní pohyb uživatele po tematických portálech (metadata DC, Z39.50, DDC).

DNER – Velká Británie (Distributed National Electronic Resource). Cíl: Národní digitální knihovna pro vyšší a další vzdělávání, distribuovaný zdroj informací pro vzdělávání a výzkum, řízený soubor zdrojů, heterogenní povahy, bibliografická data, obrázky, texty, video, dostupnost místní i dálková. Fondy jsou typicky ve formě sbírek: primárních dat, sekundárních dat (tematické portály, knihovní katalogy, databáze) (Z39.5, portály, Bath profil, XML).

Program Cobra a CoBRA+ v rámci EC se zaměřují na problematiku elektronických publikací a sdílení dokumentů v sítích - protokoly, standardy, uživatelské rozhraní, elektronické publikování, dostupnost a dlouhodobé uchování elektronických zdrojů, vícejazyčné indexování. Jedním z projektu je projekt **Biblink**, který se zabývá vybudování vazeb mezi národními bibliografickými agenturami a vydavateli elektronických zdrojů s cílem společně vytvořit informace o těchto dokumentech využitelné v obou oblastech.

Příklady systémů automatické indexace

Media On Line Project

Media On Line Projekt byl realizován v letech 1996-1999 v Belgii, jeho financování bylo zajištěno v rámci programu Vlaams Actieprogramma Informatietechnologie, řešitelem bylo ICRI-Interdisciplinary Centre for Law and Information Technology na Katolické univerzitě v Lovani (Katholieke universiteit Leuven) ve spolupráci s dalšími institucemi. **Projekt byl zaměřen na online publikaci článků z různých oborů (politika, ekonomika, finance, životní styl, umění, sport atd.), které bylo třeba z důvodu průběžného dodávání uživatelům rychle a efektivně indexovat. Vzhledem k tomu, že pro indexaci byly používány maximálně tři obecné deskriptory na každý článek, jednalo se spíše o klasifikaci. Kromě automatické indexace bylo řešeno i automatické abstrahování.** Souhrnem lze říci, že testování automatické indexace proběhlo v tomto projektu úspěšně, většina použitých metod byla ověřena jako účinná. Pro zvýšení efektivity systému by bylo ještě možné v první fázi aplikovat **lematizaci**.

NASA MAI Tool

Center for Aerospace Information (CASI) spadající pod americkou vládní organizaci NASA už od konce 70. let 20. století vyvíjí a rutinně používá MAI Tool (Machine-Aided Indexing Tool) pro automatickou indexaci technických zpráv a dalších dokumentů. Tento systém je typický příklad jednoduchého, ale robustního a účinného nástroje, který slouží jako automatická podpora intelektuální činnosti indexátora. Systém na základě termínů z přirozeného jazyka vybírá za pomoci znalostní báze

(knowledge base) deskriptory NASA tezauru (NASA thesaurus) a předkládá je k posouzení indexátorům, kteří provedou výsledný výběr a přiřazení deskriptorů k záznamům. V rámci statistické analýzy jsou jedno- a víceslovná spojení převzatá z textu dokumentu porovnávána s obsahem znalostní báze a na základě jednoduchých pravidel jsou navrhovány kandidáti na indexační termíny (deskriptory NASA tezauru).

Ve Francii aplikuje systém Verity [Institut de l'Information Scientifique et Technique](#).



INIST's catalog of Articles and Monographs

Article@INIST allows you to search for documents (Journal articles, Journal issues, Books, Reports or Conferences, doctoral dissertations) in our catalog at the **Institut de l'Information Scientifique et Technique**.

If it's your first visit or if you don't know the document type you are searching for, choose Easy search. Otherwise, choose Expert search.

Article@INIST works with [Verity K2[®] Enterprise](#).



[Easy search](#)



[Expert search](#)



Copyright INIST-CNRS 2005

Mezinárodní akce a související problematika s projektem

IFLA 2003

[IFLA a informace z některých příspěvků na 68th IFLA Council and General Conference, August 18-24, 2002. Glasgow](#)

Problematika bibliografie, zpřístupnění elektronických zdrojů, informačních technologií se řeší v příslušných sekcích a skupinách. Zpřístupněním seriálových publikací se zabývá **Serial Publications Section** (standards, kooperace, dostupnost a akvizice, copyright, archivace, rozvoj a management sbírek, vztah s nakladateli a dalšími organizacemi zabývajícími se vydáváním, zpřístupňováním, distribucí seriálů, reprezentace knihoven na „technological marketplace“).

Jednou z aktivit IFLA je zmapovat situaci týkající se zpracování a zpřístupnění seriálových publikací pomocí IFLA-Directory of Serial Content Databases and Awareness Services for **Seriálové Content**. V rámci tohoto průzkumu byly poskytnuty informace o zpracování článků v ČR.

Austrálie

Australské zkušenosti prezentované na konferenci: [National Bibliographic Database](#), de facto australská národní bibliografie, přistupuje k novému modelu bibliografické služby, která **kombinuje**

tradičně produkovaná bibliografická data s dodáváním metadat od autora či vydavatele. Koncept předpokládá získávat data z těchto zdrojů: National Bibliographic Database, National Discovery Service, národního repozitáře metadat. **National Bibliographic Database obsahuje: katalogizovaná data, katalogizační záznamy vzdálených zdrojů archivované v National Library, digitalizované dizertace, katalogizační záznamy od prodejců („vendors“, kteří mohou stát mezi producentem a kupujícím) elektronických zdrojů a služeb. The Rource Discovery Service obsahuje: metadata z elektronických archivů, metadata ze „subjekt gateways“, metadata elektronických služeb – výchova a vzdělání, kultura. Uživatel se tak bude moci vybrat tištěnou i elektronickou formu dokumentů.**

Švédsko

Inspirativní je model aplikovaný v [The Royal Library – National Library of Sweden](#). Národní bibliografie ve Švédsku je částí LIBRIS. Model je zatím aplikován u **elektronických forem knih, předpokládá se i pro periodika a noviny**. Metadata vyplňovaná do formuláře jsou posílána vydavateli do LIBRIS, konvertována do XML databáze a MRAC 21 (LIBRIS). Textový soubor je dále přes FTP poslán do **Royal Library's digital archive**. V Royal Library jsou záznamy z LIBRIS doplněny předmětovým popisem. V příspěvku se konstatuje, že **efektivnější by bylo obdržet metadata přímo od distributorů** bez dlouhého vyplňování www formuláře. Údaje od vydavatelů již existují v určité podobě a vyplňování formuláře je nadbytečné.

Zpracovaná data je možno opět poslat vydavateli. Formáty dat: MARC, ONIX, Dublin Core.

Slovensko

Na Slovensku vychází [Slovenská národná bibliografia](#). Slovenské články 1978 - 1997 (cit: Rozpisový rad článkov popisuje články a state z vybraných periodík a zborníkov slovacikálneho charakteru. Databáza obsahuje záznamy od roku 1978 a jej súčasťou od roku 1981 je aj rozpis článkov zo sérií C (mapy), H (hudobniny) a J (audiovizuálne dokumenty). **Slovenské články od roku 1998** (cit: Navazujú na predošlú databázu a obsahujú záznamy o článkoch zo slovenských novín, časopisov a zborníkov od roku 1998 do súčasnosti).

[The European Library Project \(TEL\)](#) je zaměřen na národní knihovny a CENL (Conference of European National Libraries), na přístup k sbírkám dokumentů na základě kooperace (spolupráce s nakladateli, povinný výtisk, business modeling, metadata, Z.39.50 a XML, standardy, služby). V r. 2001 řešitelka projektu dodala údaje pro **Questionnaire for the European national libraries to determine the current status of digital deposits**. Výsledkem dotazníkové akce je publikace **TEL Digital deposits state of the art review**. Dotazník mapuje situaci kolem povinného výtisku, spolupráce s vydavateli/nakladateli, zpracování el. dokumentů, zpřístupnění a archivace.

Z dotazníku: pouze 5 národních knihoven denně aktualizuje dokumenty pro digitální knihovnu on-line (Čeká republika, Dánsko, Německo, Nizozemí, Velká Británie). Pracovní linky jsou v ČR, Německu, Lotyšsku, Nizozemí, Švédsku. 3 knihovny jsou v kontaktu s IT společnostmi. 4 knihovny automaticky konvertují dodaná data (ČR, Německo, Makedonie, Nizozemí). Většina knihoven používá deskriptivní/bibliografická metadata.

V č. 10 /2002 **News form the IFLA Round Tabel of Newspapers** jsou publikované články o některých projektech týkajících se digitalizace a zpřístupnění novin. Národní knihovna v Norsku se zúčastní integrovaného projektu [LAURIN](#) pro digitalizaci a indexaci novinových výstřížků (Norsko, Austrálie, Itálie, Španělsko, Švédsko, Německo).

Projekt [TIDEN](#) (Norsko, Švédsko, Grónsko, Dánsko) - Newspaper Library on the Net.

[IFLA a informace z některých příspěvků na 69th IFLA Council and General Konference, August 1.-8. 2003. Berlin](#)

K těmto informacím připojuji informaci o zpřístupňování novin (cit **Anděrová, 2003**), která byla předmětem semináře **Newspapers for Libraries. Newspapers and the press in Central and Eastern Europe: access and preservation** (Berlin-Brandenburgische Akademie der Wissenschaften, 9.-

10. 8. 2003), kterého jsem se zúčastnila. V řadě zemí střední a východní Evropy existují specializovaná pracoviště zabývající se uchováváním a zpřístupněním novin (tj. deníků, týdeníků a čtrnáctideníků) v návaznosti na projekty týkající se digitalizace a zpřístupnění těchto materiálů v celé širší problematice (technika pro převod tištěných dokumentů na různá média včetně automatické strukturace textu a automatického zpracování metadat). Na mnoha serverech je možno najít portály, které zpřístupňují tyto dokumenty často podle regionálního hlediska ze všech konců světa. V současné době se věnuje velká pozornost samizdatové a exilové literatuře, jejímu shromažďování, uchovávání a zpřístupnění. Stále potřebnější je kooperace na národní i mezinárodní úrovni. Noviny jsou zpřístupňovány v samostatných sbírkách nebo spolu se seriály, v rámci portálů a virtuálních knihoven, jsou budovány jejich rozsáhlé archívy. V některých zemích existuje stanovená strategie budování sbírek těchto informačních zdrojů.

[Newspapers Section](#) - sekce se zabývá strategií, popisem, digitalizací, novými technologiemi pro zpracování a zpřístupnění novin. Na stránkách IFLY je vystavena doporučení pro popis novin [International Guidelines for the Cataloging of Newspapers](#) .

Příspěvky podrobně zmapovaly situaci z globálního pohledu s ohledem na specifika v angloamerické oblasti, ve východní Evropě a částečně střední Evropě. V **angloamerické oblasti** se věnuje velká pozornost **budování sbírek** těchto dokumentů. např. National Library of Australia vystavuje na svých stránkách [Australian Newspapers on line](#). Pozornost též zaslouží australská strategie budování sbírek [Collection Development Policy](#). Podobně buduje sbírku novin National [Library of Canada](#). Strategie budování sbírek v rámci veřejných, universitních, akademických knihoven, školních, národních a státních knihovnách v rámci USA je veřejně přístupná. Strategii sleduje i [British Library](#). V rámci projektu [COSEELIS](#) se zpracovává UNION list of Slavonic and East European Newspapers in British Libraries. Významné jsou i aktivity týkající se rozšíření práva povinného výtisku na elektronické publikace v rámci britského parlamentu. V Německu je budovaná [Zeitschriftendatenbank ZDB](#).

IFLA 2004

**World Library and Information Congress. 70th IFLA General Conference and Council .
Libraries: Tools for Education and Development. August 22th - 27th 2004, Buenos Aires,
Argentina.**

V r. 2004 se objevila na zasedání IFLY i problematika Knowledge Management. Zajímavý byl příspěvek [Knowledge Management in an Academic Library](#)

INFOS

V r. 2003 jsem se také zúčastnila sympózia [Infos 2003](#). Zaujal mne fakt, kolik úsilí se na Slovensku věnuje retrokonverzi článků.

ELAG

Aktuálním trendům v informační vědě se věnují tradičně konference ELAG.

[25th Library Systems Seminar](#) Prague (Czech Republic) 6-8 June 2001

Integrating heterogeneous Resources

[27th Library Systems Seminar](#) Bern (Switzerland) 2-4 April 2003

Cross language applications and the web

[26th Library Systems Seminar](#) Rome (Italy) 17-19 April 2002

Semantic Web and Libraries

[28th Library Systems Seminar](#) Trondheim (Norway) 9-11 June 2004

Interoperability: new challenges and solutions

V r. 2004 pořádal ÚISK FFUK minikonferenci Odborné knihovny a informační střediska ve službách vědy, průmyslu a obchodu. Byly zde nastíněny některé možnosti spolupráce v těchto oblastech a možnosti transformace dostupných informací ve využitelné znalosti.

A.2.3 Současný stav v ČR

Viz též B Analytická část

A.2.3.1 Trendy

Trendy v ČR jsou v souladu s těmi, zmíněnými v kapitole A.2.2.1.

A.2.3.2 Přehled organizace a kooperace při zpracování článků s perspektivou do budoucna

Následující materiál poskytuje přehled organizace a kooperace při zpracování a zpřístupňování informací o člancích, které jsou zpracovávány nebo perspektivně mohou být zpracovávány v kooperaci se zaměřením na českou provenienci a české plnotextové zdroje.

Zdroje z hlediska:

Organizace

A) Plnotextové zdroje zpřístupňované v rámci [Jednotné informační brány](#)

B) Zdroje zpřístupňované v rámci dalších informačních bran, portálů, dalších systémů

Provenience

A) České

B) Zahraniční

Přístupu

A) Licencované

B) Volné

A) Zdroje článků zpřístupňované v rámci [Jednotné informační brány](#)

Kooperační systém článkové bibliografie a jeho výstupy v JIB

Zdroje:

[Krajská knihovna Vysočiny*](#)

[Krajská knihovna v Pardubicích*](#)

[NKČR - Články v českých novinách, časopisech a sbornících \(NKČR-ANL\)](#)

[Uměleckoprůmyslové museum-katalog](#)

[Vědecká knihovna v Olomouci](#)

[Severočeská vědecká knihovna v Ústí nad Labem](#)

[Databáze knihovnické literatury](#)

Odkazy:

[ANL FULL](#)

[Bibliografia Medica Českoslovaca**](#)

[Česká geologická služba*](#)

[Krajská vědecká knihovna Liberec](#)

[Národní filmový archiv*](#)

[Středočeská vědecká knihovna v Kladně](#)
[UZPI - Zemědělská bibliografie](#)

* nově přístupující instituce, ladění importů do báze ANL

** není součástí báze ANL, je součástí KOSABI

Předpokládaná možná kooperace v budoucnu

[AVČR](#)

[AVČR - Ústav pro českou literaturu - Bibliografie české literární vědy \(od 1965\)](#)

[AVČR - Ústav pro českou literaturu - Bibliografie české literární vědy \(do 1945\)](#)

[Divadelní ústav - Bibliografie článků o divadle](#)

[AVČR - Ústav pro soudobé dějiny](#)

[AVČR - Knihovna historického ústavu - Český časopis historický](#)

[ASEP](#)

[Časopisy vydávané AVČR](#)

[Organizační schéma AV ČR](#)

[UK](#)

[Souborný katalog UK](#)

[Adresář vysokoškolských knihoven](#)

[Katalog vysokých škol v ČR](#)

[Uměleckoprůmyslové museum-katalog](#)

aj.

B) Zdroje zpřístupňované v rámci dalších informačních bran, portálů, dalších systémů (některé z nich též zahrnuté v bodě A)

Kooperační systémy

Souborné katalogy a databáze článkové (zdroje české provenience) - kooperace

[Souborná databáze Kooperačního systému článkové bibliografie ANL](#)

[SKAT článků – Kooperace Lanus](#)

Kooperační systém článkové bibliografie

Znakem * jsou označeny instituce spolupracující v kooperačním systému, ** nově přístupující instituce do kooperace, *** kooperující instituce nepřispívající do báze ANL. Výsledkem spolupráce je báze ANL a ČNB.

[Národní knihovna ČR*](#)

[Moravská zemská knihovna*](#)

Krajské knihovny

[Jihočeská vědecká knihovna v Českých Budějovicích*](#)

[Krajská knihovna F. Bartoše ve Zlíně**](#)

[Krajská knihovna Karlovy Vary**](#)

[Krajská vědecká knihovna v Liberci*](#)

[Krajská knihovna v Pardubicích**](#)

[Krajská knihovna Vysočiny**](#)

[Moravskoslezská knihovna v Ostravě*](#)

[Studijní a vědecká knihovna v Ústí nad Labem*](#)

[Studijní a vědecká knihovna Plzeňského kraje*](#)

[Studijní a vědecká knihovna v Hradci Králové*](#)

[Středočeská vědecká krajská knihovna v Kladně*](#)

[Vědecká knihovna v Olomouci*](#)

Specializované knihovny a instituce

[Knihovna ČGS**](#)

[Národní pedagogická knihovna Komenského - Ústav pro informace ve vzdělávání*](#)

[Národní lékařská knihovna***](#)

[Státní technická knihovna*](#)

[Ústav zemědělských a potravinářských informací*](#)

[Národní filmový archiv \(přístupující nově do kooperace, experiment\) **](#)

[Divadelní ústav \(přístupuje nově do kooperace\) **](#)

Plné texty v českých novinách a časopisech (báze ANL a ANL FULL - NKČR, ANOPRESS, WWW) zpřístupňované v rámci Kooperačního systému článkové bibliografie (výsledek řešeného projektu)

[ANL FULL, ANL](#)

[Statistika 1 – linky](#)

[ANOPRESS](#)

[WWW](#)

[ANL](#)

[Statistika 2 – linky](#)

[WWW](#)

[Regionální periodika \(volně na www\)](#)

[Oborová periodika \(volně na www\)](#)

Další zdroje (obsahují též zdroje zahraniční provenience a nejenom články)

Další souborné katalogy a databáze

[Centrální evidence zahraniční literatury \(CEZL\)](#)

[Souborný katalog ČR](#)

[Souborný katalog ČVUT](#)

<http://www.caslin.cz/Souborný katalog Univerzity Karlovy>

[Souborný katalog Univerzity Palackého](#)

[Souborný katalog odborné literatury veřejných knihoven \(SKAT\)](#)

[Vysokoškolské práce](#)

Informační agentury

[Newton IT](#) (i) (i)

[Anopress IT](#) (i) (i) (i)

[ČTK](#) (i) (i)

Parlament, Senát

[Dokumenty Senátu](#)

[Český parlament](#)

Zpravodajské servery

[České noviny](#)

[IDNES](#)

[Lidové noviny](#)

[iHNED](#)

[Právo](#)

[Mojenoviny](#)
[Volný](#)
[Seznam Dnes](#)

Obory, specializované knihovny a instituce
[Automatizovaný systém právních informací \(ASPI\)](#)
[Divadelní ústav](#)
[Knihovna uměleckoprůmyslového muzea](#)
[Literatura](#)
[Archiv literárních časopisů \(AVČR\)](#)
[E-archiv J. Peterky](#)
[Multikulturní centrum Praha](#)
[Národní filmový archiv](#)
[Organizační schéma AV ČR](#)

České vysoké školy
[Adresář vysokoškolských knihoven](#)
[Katalog vysokých škol v ČR](#)
[ČNK](#)

Akademie věd ČR
[Časopisy vydávané AVČR](#)
[Organizační schéma AV ČR](#)

Církevní knihovny
[Česká biskupská konference](#)
[Centrální katolická knihovna](#)

Podnikové knihovny ?

Muzea, archivy
[Muzea](#)
[Archivy](#)

Databáze autorit
[Databáze Národních autorit NKČR](#)
[Databáze REGO](#)
[Databáze REOS](#)
[Databáze autorit Centrální katolické knihovny](#)

Digitální knihovny
[Digitální knihovna \(VK v Olomouci\)](#)
[Digitální knihovna \(NKČR\)](#)
[Manuscriptorium \(NKČR\)](#)

Nakladatelství, vydavatelství, archiv webu
[Sagit](#)
[Tigis](#)
[Portál](#)
[Muzikus](#)
[Vydavatelství Economia](#)
[Akademie věd ČR](#)
[Nakladatelství Karolinum](#)
[Webarchiv](#)

Brány a portály aj.

[Agronavigator](#)

[Česká škola](#)

[Divoch](#)

[Elektronické informační zdroje na vysokých školách ČR \(AKVŠ\)](#)

[Infozdroje.cz](#)

[Econlib](#)

[Literární servery a jiné](#)

[Jednotná informační brána](#)

[MEDVIK](#)

[Moje škola](#)

[PEC](#)

[Portál ČGS](#)

[Portál STM](#)

[Portál veřejné správy](#)

[ŠkolaOnLine](#)

Elektronické dodávání dokumentů (není úplné)

[Virtuální polytechnická knihovna \(STK\)](#)

[Econlib \(CIKS -VŠE, CERGE-EI\)](#)

[Virtuální medicínská knihovna MEDVIK](#)

[Národní knihovna ČR \(DoDo\)](#)

[ČVUT](#)

[UMPRUM](#)

[Národní pedagogická knihovna J.A. Komenského](#)

[Vysoká škola báňská](#)

Internetové vyhledávače

<http://www.seznam.cz/>

<http://www.centrum.cz/>

<http://www.redbox.cz/>

<http://www.quick.cz/>

<http://www.atlas.cz>

<http://www.najdi.to>

<http://www.alenka.cz>

Zpravodajské servery

[České noviny](#)

[IDNES](#)

[Lidové noviny](#)

[iHNED](#)

[Právo](#)

[Mojenoviny](#)

[Volný](#)

[Seznam Dnes](#)

A.2.3.3 Akce

Nejnovější informace jsou tradičně prezentované na konferenci Inforum, Knihovny současnosti, Automatizace knihoven, Celostátních poradách vysokoškolských knihoven aj.

Některé důležité prezentace na těchto konferencích vzhledem k projektu:

[Realizace Koncepce rozvoje knihoven \(Vít Richter\)](#)

[Knihovny Akademie věd ČR](#)

[Vztah vysokoškolských knihoven ke knihovnímu zákonu a jejich role v systému českých knihoven](#)

[Zdravotnické knihovny v České republice](#)

[Dobré a špatné zkušenosti se stahováním záznamů v českých knihovnách](#)

[Bibliografické citace snadno a rychle prostřednictvím internetu](#)

[Harmonizace věcné katalogizace v Česku: sen či realita?](#)

[Bibliometrické indikátory SCI-ISI jako míra kvality vědecké práce a jejich alternativy pro evaluaci v českém prostředí – výhody a úskalí](#)

[Citační analýzy jako nástroj hodnocení národů v kontextu světové a evropské vědy.](#)

V rámci AKVŠ rozvíjejí činnost tři odborné komise - OK pro informační výchovu a informační gramotnost na VŠ, OK pro elektronické informační zdroje, OK pro spolupráci s Radou vysokých škol.

V r. 2004 pořádal ÚISK FFUK minikonferenci Odborné knihovny a informační střediska ve službách vědy, průmyslu a obchodu. Byly zde nastíněny některé možnosti spolupráce v těchto oblastech a možnosti transformace dostupných informací ve využitelné znalosti.

A.3 Cíl, vstupní data

Anotace (původní zadání - r. 1999)

Náplní projektu je optimalizace integrace a správy heterogenních dat souborné databáze Kooperačního 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 (přes 510000 záznamů) - vyžadují permanentní kvalitní SW a HW podporu. Budování, doplňování, správu a údržbu plnotextové databáze s možností vyhledávání zajistí informační agentura ANOPRESS (systém TOPIC). Zároveň půjde o vývoj manažerského systému pro příjem a správu dat kooperačního systému. Hlavním cílem projektu je zkvalitnění bibliograficko-informačních služeb.

Vstupní data

- Vstupními daty pro bázi ANL jsou bibliografické záznamy článků z titulů zpracované v letech 2000-2004 v rámci KOSABI spolupracujícími institucemi včetně titulů zpracovaných v NKČR
- Vstupními daty pro bázi ANL FULL jsou plné texty s metadaty vydané v letech 2001-2003, zpracované on-line v rámci linky zpracování bibliografických záznamů z plných textů (linka poloautomatické extrakce/indexace TTDE za současného generování formátů UNIMARC/M21, DC a jejich zpřístupnění v HTML XHTML a XML v bázi ANL FULL.
- Vstupními daty pro bázi ANL FULL jsou plné texty s metadaty vydané v letech 1997-2000 v regionálních titulech doplněné a propojené off-line.
- Vstupními daty pro Portál WWW periodik jsou periodika strukturovaná oborově a regionálně

B Analytická část

B.1.1 Vlastní řešení - základní linie

Cíle projektu byly realizovány ve třech základních liniích:

- a) Analýza základních trendů v oblasti integrace heterogenních dat, koncepčně teoretické práce
- b) Nástroje pro optimalizaci integrace a správy heterogenních dat Kooperačního systému článkové bibliografie s ohledem na integraci plných textů do systému – jejich vývoj a uvedení do provozu
- c) Kooperační systém článkové bibliografie – organizace, metodika, kooperace a jeho vymezení k ostatním systémům

Linie se navzájem prolínají.

Oblasti či rámce projektu byly koncepčně řešeny v souvisejícím projektu – výzkumnému záměru [Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů \(závěrečná zpráva za r. 1999-2003\)](#).

Základní teze výzkumného záměru: optimalizovat zpřístupnění plných textů znamená je propojovat s ostatními zdroji a službami. Úspěch propojení závisí na standardech a nástrojích používaných ve všech fázích informačního procesu. Týká se organizace informačního systému a legislativně právních otázek zpřístupnění elektronických zdrojů.

Základním předpokladem optimalizace zpřístupnění plných textů je nutnost nových modelů získávání, zpracování a zpřístupňování bibliografických informací a nutnost změn v chování uživatele (human computer interaction) s rozvojem elektronického publikování, které se týká autorů, nakladatelů, vydavatelů a knihovníků.

Trendy

Byly analyzovány a stanoveny následující trendy v získávání, zpracování a zpřístupňování plných textů:

- **Typy elektronických dokumentů:** podle **původu** - digitalizované, současně tištěné i elektronické, doplňky k tištěným, pouze elektronické; podle **obsahu** – různá hlediska dělení; **způsob a úroveň vydávání/popisu** - monografie, pokračující zdroje (seriály, integrující zdroje - např. webovská sídla), jednotlivé části.
- **Získávání - zdroje plných textů :** **plné texty přístupné na Internetu** (elektronické časopisy, elektronická vydavatelství /nakladatelství/informační agentury - archivy a databáze), **plné texty získané** od autora/ vydavatele/nakladatele/distributora, povinný výtisk, **automatický sběr dat** (harvesting).
- **Zpracování bibliografických záznamů a plných textů - způsoby :** **tradiční** v integrovaném knihovnickém systému (ALEPH apod.); v **lince poloautomatické extrakce/indexace bibliografických záznamů z plných textů, které obsahují metadata** za současného generování URL a metadat ve formě UNIMARC (perspektivně MARC 21), DC, XHTML, XML/RDF; formulář pro generování DC, XML/RDF
- **Metadata :** **bibliografický záznam (MARC), DC** - metadatový formát pro www zdroje s jednoduchou syntaxí a sémantikou, specifikováno 15 prvků, kvalifikované atributem či nekvalifikované.
- **Značkovací jazyky:** značky určují způsob interpretace textu, značky určují význam (deskripce) a prezentaci (zobrazení) - **HTML** (kombinace obou typů, nedostatky), **XML** (pouze význam) + **XSL** (zobrazení) + **Xlink, XPointer, XPath** (odkazy).
- **Implementace metadat:** **MARC, DC a HTML, XHTML, XML/RDF** v kvalifikované a nekvalifikované formě.
- **XML/RDF:** popis systému značek **DTD** (Document Type Description) tvoří tzv. **RDF** (Resource Description Framework) - Rámec pro popis zdrojů v XML, definována syntax a

sémantika - odkaz na definici RDF a metadatové schéma, tzv. jmenný prostor, prvek identifikován popisem (jmenným prostorem) a jeho názvem, možnost vkládat vnořené prvky.

- **Identifikace plných textů a seálů:** URL, PURL (databáze, PURL je přiřazeno danému URL), URN, SICI (pro tištěné i elektronické seriály, články), DOI (ochrana autorských práv), systém ISSN.
- **Propojení:** ruční/automatické, on-line/off-line, statické/dynamické, uzavřené/otevřené; **statická adresa, dynamicky generovaná adresa a OpenURL, SFX.**
- **Vyhledání a zpřístupnění plných textů:** internetové vyhledávače, protokol HTTP, XML, Z 39.50, SFX; OAI, brány a portály; "subjects gateways", konspekt, předmětové kategorie, témata; OPAC - bibliografické záznamy (UNIMARC/MARC 21) s URL adresou nebo dynamicky a SFX, pojmové vyhledávání a **fulltextové vyhledávání** (např. TOPIC, Excalibur) v kombinaci s rejstříky, **interaktivní využití topiků při vyhledávání a zpětná vazba relevance (relevance feedback); metatagy DC v HTML, XHTML, XML/RDF.**
- **Uložení plných textů:** archivy, databáze - servery (nakladatelé/vydavatelé, informačních agentury, knihovny).
- **Výstupy:** seznam článků (řazení dle relevance, do skupin - clustering aj.), **formáty** (citace/abstrakt/plný text, UNIMARC, DC, HTML, XHTML, XML/RDF). **Stahování, export a tisk metadat a plných textů** (formát text, RTF, HTML, XML, též UNICODE).
- **Uživatelé:** lokální/interní, externí, registrovaní, neregistrovaní, zkušební registrace.
- **Zpřístupnění plných textů z hlediska formy:** přes **bibliografie** a soupisy, **obsahy** časopisů, **plnotextové databáze**, digitální knihovny.
- **Zpřístupnění plných textů z hlediska institucionálního (zdroje):** **autoři**, prostřednictvím **nakladatelství/vydavatelství, distributorů, knihoven/informačních institucí.**
- **Zpřístupnění plných textů z hlediska organizačního, ekonomického, technického:** **konzorcia, licence, zdarma, placené** (předplatné – fixed-fee, platby za uskutečněné operace – pay-as-you-go), **přes IP, login, heslo, proxy server, volně, brány, portály, kooperační systémy, souborné katalogy a databáze, univerzální bibliografie i jednotlivé kolekce dle oborů**
- **Standardy:** Metodika popisu článků ve formátu UNIMARC - vychází z AACR2R, ISBD/CP; **harmonizace AACR2R a ISBD, ISBD (ER), ISBD(S) - ISBD (CR), FRBR – Funkční požadavky na bibliografické záznamy** (Functional Requirements for Bibliographic Records) - 4 úroňový model manifestace díla, nutnost existence mezinárodních katalogizačních pravidel a eventuelní participace na jejich vzniku), norma pro citování dokumentů ČSN **Bibliografické citace, věcný popis – autority, předmětová kategorizace, konspekt** .
- **Zákony:** nutná změna autorského zákona, zákona o povinném výtisku aj.

B. 1.2 Analýza zpřístupnění plných textů v ČR a Kooperační systém článkové bibliografie

Během pětiletého řešení projektu nastal poměrně rychlý vývoj ve zpřístupnění plných textů na českém internetu co do forem a organizace.

Plné texty jsou organizovány na Internetu do tematicky, regionálně nebo jinak utříděných služeb. Jsou přístupné přes vyhledávače, jsou utříděny do kategorií, zpřístupněny na různých portálech a zpravodajských serverech, v informačních branách, digitálních a virtuálních knihovnách v rámci kooperace příslušných institucí. Některé knihovny a instituce vytvářejí soupisy volných českých zdrojů plných textů na Internetu a zdrojů, které vydávají. Jsou zpřístupněny licencované zdroje zahraničních plnotextových databází a je založeno konsorcium Anopress pro multilicenční přístup k českým plným textům. V ČR jsou přístupné i zahraniční tzv. alerting services. V řadě knihoven je funkční elektronické dodávání dokumentů.

Nabídka **českých novinových a časopiseckých elektronických zdrojů** je relativně velká - jsou vystaveny deníky, týdeníky a časopisy . Vystavené texty jsou zachyceny s různou hloubkou retrospektivity (aktuální číslo, poslední čísla či roky), objevují se archivy volně dostupné či pouze registrovaným uživatelům. Některé zdroje obsahují citaci , abstrakt, objevují se current contents. Někdy lze vyhledávat podle základních formálních údajů, kombinovat dotazy pomocí logických

operátorů, vyhledávat plnotextově ([České noviny](#) , [IDNES](#) , [Lidové noviny](#), [iHNed](#), [Právo](#), [Mojenoviny](#)).

Na české Internetu se objevují **nakladatelské elektronické zdroje** Vztahy mezi uživateli, knihovnami a vydavateli/nakladateli a knihovnami či bibliografickými agenturami nejsou dosud jasné z hlediska právního i obchodního, v budoucnu lze předpokládat v tomto směru vznik nových iniciativ ([Vydavatelství Economia](#) , [Sagit](#), [Tigis](#), [Nakladatelství Muzikus](#), [Nakladatelství Portál](#)

Informační agentury

[Albertina icome](#) (projekt [Infozdroje.cz](#)), [ČTK](#) , [Newton I.T.](#), [Anopress IT, a.s.](#)

[Anopress](#) v současnosti poskytuje následující služby: **On-line monitoring vyhledávání na Internetu** (on-line fulltextové prohledávání obsáhlé mediální a vědomostní databáze), **Off-line monitoring elektronická výstřižková služba** (pravidelná dodávka monitoringu médií podle individuálních požadavků), **Informační servis Anopress (ISA)** (vyhodnocovací software pro další práci s dodanými nebo vyhledanými daty), **Quick monitoring** (jednoduchá on-line služba umožňující fulltextové vyhledávání v médiích pomocí více jak 200 předdefinovaných témat, **Mediální analýzy**, **Archívy zdrojů** , **Archívy zdrojů** (dodávka nebo přístup do kompletních zdrojů), **E-noviny informace na míru do vašeho mobilu** (vyhledávání v celostátním a regionálním denním tisku z Vašeho mobilu – realizovaná v r. 2003), **Sbírka zákonů ČR on-line** . V roce 2003 je také zpřístupněna služba významná hlavně pro střední školy Brána vědění.

Hlavním produkty jsou: on-line databanka novin a časopisů celostátních a regionálních, monitoring na zakázku, vědomostní databáze, archívy zdrojů. Anopress IT, a.s. umožňuje on-line přístup do databanky plných textů **TamTam**, na jejíž bázi poskytuje následné služby. Společnost Anopress je výhradním zpracovatelem elektronické podoby většiny českých regionálních titulů (nakladatelství Bohemia). Pro zpřístupnění plných textů ve veřejných knihovnách bylo založeno v r. 2000 **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. Databanka v současné době obsahuje texty článků a zpráv všech celostátních deníků, dále pak regionální deníky, celostátní a regionální časopisy, dále i textové záznamy zpravodajských, publicistických a diskusních pořadů rozhlasu a televize. V roce 2002 byly zpřístupněny archívy zdrojů od r 1996.

Společnost vyvinula vlastní software ISA, který umožňuje všechna data dále analyticky zpracovávat, exportovat je v několika formátech, včetně HTML a XML, pro Internet či Intranet.

Vyhledávací systém **TOPIC**, který Anopress používá k monitoringu a analýze informačních zdrojů, je v současnosti jediným interaktivním systémem na českém trhu. V únoru 2002 byla agentura vydražena v dobrovolné dražbě, v dubnu 2002 zaregistrována jako nový podnikatelský subjekt s názvem Anopress IT, a.s. Zpráva o Anopressu byla dle dohody podána řešitelkou předkládaného projektu na MKČR koncem května 2002. **Báze TamTam je zpřístupňována v rámci programu VISK v rámci konsorcia knihoven, probíhají práce na jejím zapojování do JIB.**

Automatická či poloautomatická indexace (J. Schwarz, 2002)

V bývalém Československu začal rozvíjet výzkum a vývoj v oblasti automatizovaného zpracování textu až od konce 60. a začátku 70. let 20. století. Vznikla řada nejen regionálně, ale i mezinárodně jedinečných systémů, které byly určeny k automatizovanému zpracování textu především v oblasti automatické indexace, automatické tvorby tezauru a automatického překladu. Na světové úrovni, zejména zásluhou P. Sgalla, se v Česku rozvinula také matematická lingvistika, označovaná také jako počítačová lingvistika, a zejména v druhé polovině 90. let 20. století rovněž i korpusová lingvistika. Většina uvedených systémů dnes slouží k výzkumným účelům, ale některé z nich jsou částečně dostupné i komerčně (týká se to např. systému LEGSYS).

Již od začátku 70. let byly vyvíjeny i systémy SEMAN (V. Smetáček) a MOZAIKA (Z. Kirschner). 70. a 80. léta představovala také období rozvoje matematické, počítačové a korpusové lingvistiky pod vedením P. Sgalla; vzniklo specializované pracoviště (dnes Ústav formální a aplikované

lingvistiky/Centrum počítační lingvistiky na MFF UK), znalosti a zkušenosti jeho pracovníků byly zúročeny i v současném projektu **Pražského závislostního korpusu, resp. Českého národního korpusu**.

ČNK

Texty pro Český národní korpus jsou získávány několika způsoby. Smlouvy zavazují ČNK k užití textů pouze pro budování ČNK a pro jejich nekomerční využívání. Je to unikátní banka plných textů zpracovaných a konvertovaných do jednotné podoby. Užívá se především k lingvistickým analýzám, jeho použití by jistě bylo zajímavé i v aplikacích informačních.

Automatická indexace sněmovních tisků v KPS PČR

Parlamentní knihovna jako odbor Kanceláře Poslanecké sněmovny Parlamentu ČR od r. 2000 indexuje sněmovní tisky.

Znalostní systémy – v ČR distribuované společností TOVEK (fulltextová technologie Verity – TOPIC, Verity Portal ONE, Verity K2 katalog) a INCAD (systém Convera RetrievalWare). Problematikou znalostních systémů a vytěžováním informací, vyhledáváním v plnotextových databázích aj. se zabývají projekty a renomované konference jako je Datakon a Znalosti.

Systém TOPIC je distribuován firmou TOVEK. Jedním z projektů je projekt tzv. konkurenčního zpravodajství. Konkurenční zpravodajství je důležitý faktor konkurenceschopnosti firem a ozdravení nejen podnikatelského prostředí. Vazba na knihovnicko-informační sféru a akademickou sféru je potřebná. Projekt by pomohl přiblížit sféru knihoven, vědeckovýzkumnou sféru jak na vysokých školách tak v Akademii věd se sférou knihoven a umožnit koloběh informací mezi těmito oblastmi.

Souborný katalog CASLIN se v počátcích řešení vyvíjel na základě vlastní aplikace s využitím databázového systému ORACLE. Bylo rozhodnuto o jeho přechodu pod systém ALEPH.

WebArchiv

Jeho cílem je připravit podmínky pro zpracování české národní bibliografie elektronických zdrojů, se zaměřením zejména na zdroje dálkově přístupné a zajistit dlouhodobé uchování domácích elektronických zdrojů. **Dublin Core Metadata Generator** je nástroj umožňující autorům webových stránek poloautomaticky nebo ručně vytvořit, editovat, konvertovat a ve zvolené syntaxi uložit metadata respektující pravidla kvalifikovaného Dublin Core. Protože platná česká legislativa neumožňuje v současné době zdroje zpřístupňovat, řeší se situace uzavíráním smluv s autory/vydavateli internetových zdrojů. Zpřístupnění archivu se má realizovat pomocí technologie fulltextového indexování a automatizované extrakce autorem vytvořených metadat. V rámci projektu byl lokalizován nástroj **Harvester** z projektu NEDLIB, v budoucnu budou probíhat sklizně pomocí systému Heritrix.

V rámci popisu elektronických zdrojů NK se zatím popisují vybrané servery „národní produkce“ (doména cz), výběr zdrojů je podle obsahu (odborné, umělecké), typu (seriály, konferenční příspěvky, výzkumné zprávy, dokumenty veřejné správy, formy (pouze elektronická forma), přístupu, formátů (všeobecně podporované) s ohledem na automatické získávání elektronických zdrojů (volně přístupné zdroje). V ALEPHu by měly být záznamy ukládány ve formátech MARC a DC (konverze generátor DC). Uložení v digitálním archivu. Počítá se s popisem seriálových zdrojů (periodika publikovaná v elektronické formě, webové stránky) a jejich zpřístupněním v JIB. Předpokládá se spolupráce s knihovnami v JIB. V roce 2003 byla zpřístupněna báze bibliografická WEB, obsahující některé on line elektronické zdroje. Cílem je, aby ze záznamů v databázi byl současně umožňován přístup do digitálního archivu (WebArchiv).

V budoucnu dojde patrně ke spolupráci mezi článkovou bibliografií a projektem Webarchiv.

Jednotná informační brána

Projekty JIB:

Součástí projektu JIB je:

Portál STM

Jednotná informační brána

Česko - slovenský virtuální katalog Caslin.

JIB umožňuje základní prohledávání informačních zdrojů, jejich tematický výběr, stahování záznamů.

Zdroje jsou volně přístupné a licencované, jsou zpřístupněné v JIB k vyhledávání nebo jako odkazy.

Základním nástrojem JIB je Metalib a SFX. Metalib je paralelní prohledávač heterogenních zdrojů (plné texty, databáze, archivy, katalogy aj.), komunikuje především prostřednictvím protokolu Z39.50 nebo HTTP. Služby jsou personifikované (autorizace a autentifikace uživatele je předpokladem pro přihlášení a nastávaní práv uživatele). Zpřístupňované **zdroje** jsou volné a licencované/placené.

Přístup: pro registrované a neregistrované uživatele. **Vyhledávání** je možné podle autor, názvu, předmětu, roku vydání, ISBN/ISSN. Metalib má **znalostní bázi** (informace o zdrojích a uživateli).

Pro zpřístupnění článků je třeba umožnit detailnější vyhledávání podle data vydání. V budoucnu se plánuje implementace OIP. SFX je standard pro propojení dokumentů a je využíván v Metalib k vytváření vazeb. SFX je nástroj k nabízení přidávaných služeb a je aplikací OpenURL. Propojuje informace.

Předpoklady pro fungování JIB: **jednotné věcné zpracování** (překlad změn MDT, tvorba záznamů věcných autorit, aktualizace notací MDT v záznamech věcných autorit, příprava harmonizace věcných autorit a **Polytematický strukturovaný heslář**, konkordance MDT s DDC. Soubor věcných autorit obsahuje tematické termíny, geografické a formální. Věcné termíny jsou propojeny s notacemi MDT a anglickými ekvivalenty a vazbou na **konspekt**. Skupiny konspektu obsahují název skupiny a přidělený znak MDT plus anglický ekvivalent. Soubory autorit jsou přístupné na adrese <http://www.nkp.cz> (nabídky **Katalogy a databáze**). Na adrese <http://aip.nkp.cz/mdt/> je publikován překlad UDC MRF. Přehled předmětových kategorií je aplikován v JIB jako základní tematické rozcestí (1. úroveň třídění), skupiny konspektu jsou obsaženy v bázi AUT.

Pro JIB byl vypracován Návrh požadavků na metadata používaná pro provoz oborových informačních bran v České republice. **Standardy používané v JIB** jsou užívány nebo se aplikují postupně v ČR. Aplikace metody pro popis zdrojů umožňuje paralelní vyhledávání záznamů o dokumentech i na úrovni sbírek, integrovaný přístup k heterogenním zdrojům ve vícejazyčném prostředí. Metalib a SFX je produktem fy **Exlibris**. V České republice bylo založeno konsorcium knihoven, které ALEPH používají. Dalšími produkty této firmy kromě ALEPHU je nástroj pro zpracování, archivaci a zpřístupnění elektronických sbírek **DigiTool**. Umožňuje i fulltextové vyhledávání.

Kooperační systém článkové bibliografie, plné texty a JIB

Článkové databáze přístupné v JIB viz kapitola A.2.3.

Zpřístupnění plných textů v JIB v rámci Kooperačního systému článkové bibliografie je v letech 2000-2004 jednak přes statické linky z báze ANL na volná www periodika, jednak přes statický link v bázi ANL do báze ANL FULL. Je možné přímo vstoupit do odkazovaného zdroje - báze ANL FULL. Plné texty v ANL a ANL FULL jsou v letech 2000-2004 zpracovávány v lince poloautomatické extrakce/indexace TTDE, koncem roku v NKČR rozhodnuto zpracovávat texty přímo v ALEPHu. Zpracování v rámci ALEPHu se patrně bude vyvíjet, podobně, jako se vyvíjelo zpracování v lince TTDE. Rozdíl: propojovat se bude přímo v JIB na bázi Anopress, která má být zde zpřístupněna (zpočátku přes ID, postupně přes SFX.)

Koncem roku 2004 bylo vypracováno doporučení pro aplikaci metody konspektu v článkové bibliografii a v NKČR se začal konspekt experimentálně používat. Analytické předmětové kategorie dosud používané v systému se používají dál, někde konspekt vhodně doplňují. Jsou však zatím většinou neřízené a o jejich další aplikaci či neaplikaci se rozhodne v budoucnu. **Aplikace konspektu umožní článkům se zapojit do tematických oborových bran.**

Bibliografické zpracování článků v ČR - Kooperační systém článkové bibliografie –
<http://full.nkp.cz/nkdb/docs/vstupmng.htm>;
http://www.nkp.cz/pages/page.php3?page=oazp_kooperace.htm

O kooperačním systému viz též dále v rámci bodu B, Příloha F 5, F6, F7.

V posledních letech vzniklo několik **projektů**, zabývajících se zpřístupněním analytických záznamů v kooperaci s ostatními knihovnami. Na výsledky některých z nich navazuje předkládaný projekt.

Zpřístupnění výsledků analytického zpracování prostřednictvím Internetu (r. 1998) - projekt řešil zejména konverzi článků do UNIMARCu. V rámci průzkumu Internetu se ukázalo že statické propojování na www zdroje je krajně nespolehlivé. Výběr relativně spolehlivých zdrojů plných textů je možné řešení.

Výzkumný záměr NK **Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů** (VaV, r. 1999-2003) - projekt analyticko-koncepční, připravil půdu pro praktickou realizaci programového projektu týkající se kooperačního systému a integrace dat. Cílem výzkumného záměru je optimalizace přístupu uživatelů k plným textům dokumentů domácí provenience (nikoli zahraniční). 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 **Konzorcia Anopress**, která byla podepsána mezi SKIP a Anopressem v r. 2000. V rámci projektu bylo zpřístupněné **periodikum Národní knihovna .** **Projekt Západočeský ANAL - Kooperativní zpracování periodické produkce západních Čech** se zabýval odstraněním duplicit při zpracování, metodikou excerptce titulů a zpracování záznamů v jednotlivých okresech západočeského regionu.

Projekt Zavedení automatizovaného zpracování článkové bibliografie v systému T-Series řešený v r. 2000-2001 se zabýval řešil problematikou bibliografického zpracování článků v tomto systému.

Velmi významný je z hlediska tvorby a rozvoje regionálních faktografických databází a souborů autorit je projekt **Rozvoj regionálních databází na základě retrospektivní konverze článkové bibliografie** řešený v letech 2000-2002 v SVK Kladno.

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 spolupracuje v současnosti 8 krajských knihoven a MZK, specializované odborné knihovny (STK, ÚZPI, NPKK-ÚIV).

V r. 2004 byl proveden úspěšně experiment s Národním filmovým archivem a čtyřmi novými krajskými knihovnami týkající se importu a následně uploadu dat do kooperačního systému jako předpoklad pro zahájení reálné spolupráce v r. 2005.

Nově přístupující knihovny stanovily předběžně tituly k excerptci pro budoucí spolupráci v rámci systému. Přístupující knihovny byly vedeny také po stránce metodické.

V systému **LANIUS**, resp. **Clavius** se zpracovávají bibliografické záznamy článků v knihovnách veřejných knihoven. V rámci aplikace pro management **KOSABI** byl vytvořen speciální formát pro příjem dat ze systému **Clavius** (krajská knihovna v Havlíčkově Brodě), **Clavius** používá techniku standardních polí UNIMARC, nikoli vazebních. Byla analyzována excerptční základna LANIA a **KOSABI** - v budoucnu by se měly, pokud k tomu budou vhodné podmínky a bude to mít pro oba systémy význam, omezit duplicity ve zpracování mezi těmito systémy.

Některé instituce spolupracující v rámci **KOSABI** přecházejí na nové SW. V **krajských knihovnách v Kladně a Liberci** se články popisují v systému **RAPID**, resp. **ARL**, v **MZK v Brně v ALEPH**, v **Olomouci** též. V **Ostravě**, v **Českých Budějovicích** a **Ústí nad Labem** v systému **T-Series**. V **Hradci Králové**, **ÚZPI**, **STK** a **NPKK** je v r. 2004 je užíván **ISIS**. **NPKK** přechází konce r. 2004 na **KP-Win**, v **ÚZPI** bude patrně v budoucnu implementován systém **ALEPH**. V **krajské knihovně v Plzni** se přechází na **ALEPH**. V nově konstituovaných krajských knihovnách

v Pardubicích a Zlíně se používá KP-Sys, resp. KP-Win, v Karlových Varech se přechází na ALEPH.

Spolupracující odborné instituce

V [ÚZPI](#) se zpracovává Zemědělská a potravinářská bibliografie. V budoucnu je třeba převést také zde zpracování článků do formátu UNIMARC/M21. Má být umožněno propojení záznamů na plný text v rámci technologické linky zpracování bibliografických záznamů z plných textů přes formát XML v návaznosti na existující redakční systém za účelem importu do bibliografické báze. [Agronavigator](#) je moderní technologie zpřístupnění zemědělských informací aplikovaná na webu UZPI.

V [ČGS](#) je zpracovávána Česká geologická bibliografie. V současné době knihovna přechází na Clavius.

V [ÚIV – NPKK](#) jsou články zpracovávány v systému CDS ISIS. V současné době knihovna přechází na KP-Sys, resp. KP-Win. Na internetu je vystavována Pedagogická bibliografická databáze.

Ve [STK](#) jsou články zatím zpracovávány v systému CDS ISIS. Provozuje [Portál STM](#). Elektronické dodávání dokumentů se realizuje prostřednictvím [VPK](#). STK Provozuje [Českou národní databázi ISSN](#).

Knihovny pracující v ALEPHu (Praha, Brno, Olomouc) přešly v polovině roku 2004 na ukládání v MARCu 21. Pro obousměrnou konverzi [UNIMARC MARC 21](#) byly v NKČR vypracovány tabulky a probíhali testy na malém vzorku záznamů. V r. 2005 je ještě třeba doladit experimentální pole 773q určené pro propojení přes Open URL.

Pro automatizovaný upload dat do systému byla vypracována nápověda. Upload odzkoušen všemi institucemi a zprovozněn a koncem r. 2004.

Zpracování článků v budoucnu je třeba koordinovat s AVČR, vysokými školami aj. institucemi, jejichž seznam je v kapitole A.2.3.

Metodika

Data jsou zpracovávána ve nebo konvertována do formátu UNIMARC/M21, respektují se pravidla popisu [AACR2](#), [ISBDs](#), [Guidelines for the application of the ISBDs to the description of Component Parts](#) s respektováním mezinárodních standardů věcného popisu - [MDT-MRF](#) pro oblast systematické indexace.

V r. 2004 byla vypracovány pravidla pro popis dat v MARCu 21, pracovní skupinou pro analytický popis byl schválen minimální záznam pro soubornou databázi v MARCu 21.

V oblasti verbální věcné indexace se kombinují klíčová slova, věcné obecné kategorie a předmětová hesla a kontrolují se proti souborům autorit (kromě knihoven pracujících v ISIS). Vytváření vazby mezi jednotlivými vrstvami popisu je klíčovým momentem. V r. 2004 bylo vypracováno doporučení pro aplikaci konspektu v kooperačním systému (v NK se začal užívat koncem roku).

Kromě knihoven pracujících v Tinlibu resp. T-Series s používají ještě analytické předmětové kategorie, které je třeba sjednotit.

Metodika na stránkách NKČR

Metodika je určena především pro pracovníky Oddělení analytického zpracování NK ČR a knihoven spolupracujících v Kooperačním systému článkové bibliografie.

❖ [Metodika popisu článků ve formátu UNIMARC](#) - podklad pro interpretace AACR2R/UNIMARC verze 1.2 (26.6.2001)

❖ [Přehled polí pro bibliografický záznam nesamostatných částí dokumentu](#) - tištěné monografie a seriály ve formátu UNIMARC (používaných jak v části dokumentu-analytické jednotce, tak ve zdrojovém dokumentu)

- ❖ [Věcný popis ANL](#) - struktura polí a podklady pro tvorbu řízených předmětových termínů
- ❖ [Doplnění Záznamu pro soubornou databázi : UNIMARC. Fyzicky nesamostatné části dokumentu. Tištěné monografie a seriály. Věcný popis ANL](#)
- ❖ [Příklady](#)
- ❖ [Metodika popisu článků ve formátu MARC21](#)
- ❖ [Příklady popisu ve formátu MARC21 \(i UNIMARC\)](#)
- ❖ **Ke stažení:** [Metodika popisu článků ve formátu MARC21 a Příklady popisu ve formátu MARC21 \(i UNIMARC\) ve formátu .doc](#)
- ❖ **Interpunkce ISBD**, kterou používá M21 a která je zabudovaná do pravidel pro popis článků, je na adresách:
 - http://www.ifla.org/VII/s13/guide/isbdg_wrd.pdf
 - http://216.239.59.104/search?q=cache:BxLC03J6-CkJ:www.ifla.org/VII/s13/guide/isbdg_wrd.pdf+ISBD+G&hl=en
- ❖ [Aplikace konspektu v článkové bibliografii](#)
- ❖ [AACR2R - příloha B Zkratky](#)
- ❖ [Záznam pro soubornou databázi : MARC 21. Fyzicky nesamostatné části dokumentu. Tištěné monografie a seriály](#)

Knihovny pracující v ALEPHu (Praha, Brno, Olomouc) přešly v polovině roku 2004 na ukládání v MARCu 21. Pro obousměrnou konverzi UNIMARC MARC 21 byly v NKČR vypracovány tabulky a probíhali testy na malém vzorku záznamů. V r. 2005 je ještě třeba doladit experimentální pole 773q určené pro propojení přes Open URL.

Pro automatizovaný upload dat do systému byla vypracována nápověda. Upload odzkoušen všemi institucemi a zprovozněn koncem r. 2004.

Excerptní základna a vývoj propojování

Excerptní základna

<http://full.nkp.cz/nkbd/docs/vstupmng.htm>

http://full.nkp.cz/nkdb/docs/tituly_napoveda.htm

<http://anl.nkp.cz:2100/r/SKK/p210/p10Init>

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). V rámci KOSABI jsou stanoveny zásady výběru článků co do úplnosti i co typů.

Zpracovávané typy titulů a výběr článků z hlediska úplnosti: v kooperačním systému jsou zpracovávány články a statě české novinové, časopisecké a periodické sborníkové produkce (sborníky vysokých škol, muzeí, archivů, materiály z konferencí, ročenky, odborné i polytematické časopisy, kulturně-politické časopisy a populárně-naučné časopisy). Jsou určeny kategorie A-D zdrojů z hlediska obsahu a formy vydávání a těmto kategoriím je stanovena úplnost excerptce .

Kategorie titulů

A: Seriály vydávané AVČR a vysokými školami v ČR (časopisy, sborníky, ročenky)

B: Seriály (časopisy, sborníky, ročenky) vydávané ostatními institucemi v ČR

C: Noviny (ústřední i regionální) a kulturně-polické časopisy

D: Populárně-naučné časopisy

- 1=excerpce je prováděna v úplnosti (100-80% počtu článků)
2=excerpce je prováděna částečně (80-25% počtu článků)
3=excerpce je prováděna výběrově (25%- počtu článků)

- A: 1
B: 1 (sborníky ročenky), 1-2 (odborné časopisy)
C: 2,3
D: 3

Báze ANL - obsahuje kategorie A-D, **báze ANL FULL** zatím kategorie C, z kategorie B periodikum Národní knihovna (též samostatná aplikace ve formátech HTML, PDF). **O kategoriích titulů viz Příloha F6.**

Typy článků: faktograficky přínosné články, články odborné, články s dokumentární a uměleckou hodnotou, oficiální dokumenty (projevy, vyhlášení, komuniké, usnesení), zprávy a informace o kulturních, sportovních, politických, vědeckých aj. akcích (konferencích, seminářích, festivalech, jednáních, zasedáních, výstavách, divadelních, rozhlasových a televizních inscenacích, besedách, sportovních soutěžích), souhrnné a hodnotící články, závažné polemiky, diskuze a komentáře. Dále jsou zachyceny reportáže, biografické články (životopisné, jubilejní, nekrology, rozhovory, profily osobností, vzpomínky), literární texty (básně, povídky aj. ještě nepublikované, nové překlady), fejetony, sloupky, úvodníky (pouze hodnotné), recenze publikací různého zaměření, gramofonových desek, kompaktních desek, filmů apod. Obě báze jsou výběrové.

V současné době se objevují ve větší míře **odborné tituly. Dá se předpokládat rozvoj a nárůst elektronického publikování na vysokých školách a v jiných odborných institucích, tituly vydávané AV jsou již poměrně dobře vystaveny v plnotextové formě. V budoucnu lze předpokládat poměrně velkou možnost propojení s těmito tituly pomocí SFX.**

Propojování na plné texty

Po analýze titulové základny a zdrojů na WWW bylo rozhodnuto:

- deníky a některé odborné časopisy zpracovávat v lince TTDE ve spolupráci s Anopressem
- odborná volně dostupná periodika propojovat zatím staticky na WWW z báze ANL
- vytvořit portál WWW periodik ve struktuře regionální a odborné – aplikace na WWW

Spolupráce s Anopressem

V letech 2000-2001, 1. pololetí byly bibliografické záznamy zpracované v systému ALEPH doplňovány číslem ID jako parametrem URL adresy příslušných plných textů v poli 856, které se zároveň ukládaly na lokální počítač, následně byla doplněna do záznamu do pole 856 globálně základní adresa plného textu na serveru full.nkp.cz a plné texty získané z Anopressu se naimportovaly do báze ANL FULL.

Od poloviny r. 2001 do r. 2004 listopad byly plné texty Anopress zpracovávány v on-lince poloautomatické extrakce/indexace umístěné na lokálních počítačích – Windows TTDE.

V prosinci r. 2004 bylo zahájeno zpracování v internetové verzi linky – WWW TTDE.

Důvod pro budování báze ANL FULL: identifikační číslo článků se mění po 14 dnech po přesunu do archivu. Přímé, aktuální a stabilní propojení článků zpracovaných lince na plné texty do báze TamTam nebylo tak možné.

Koncem roku 2004 rozhodlo vedení NKČR zpracování textů Anopress opět převést pod ALEPH, probíhají práce na zpřístupnění báze TamTam v rámci JIB. Pracovníci oddělení kopírují do pole 856 základní společnou adresu plných textů do báze TamTam, pracovníci v referenčním centru doplňují po 14 dnech ID plného textu jako parametr. Jako parametr je v adrese také identifikace knihovny, přes kterou se uživatel může připojit na plný text v rámci konsorcia. ID je jeden z údajů pro propojení přes SFX v rámci JIB.

Kromě zpracování on-line v lince, byly retrospektivně v r. 2002-2001 doplňovány plné texty do báze ANL FULL off-line dle zpracovaných bibliografických záznamů v báze ANL. Adresa byla do bibliografických záznamů doplněna ex post.

Tituly a statistiky propojení vzniklých v rámci linky TTDE a staticky – Příloha F7.

B.1.3. Dílčí cíle, nástroje a metody optimalizace integrace a správy heterogenních dat v rámci Kooperačního systému článkové bibliografie s ohledem na integraci dat a plných textů

Cíle v zadání z r. 1999

„Cílem projektu je integrace heterogenních dat - plných textů článků a bibliografických záznamů, zpracovaných Kooperačního systému článkové bibliografie - v rámci souborné databáze. Základním předpokladem je vývoj aplikace na budování, doplňování, správu a údržbu plnotextové databáze s možností vyhledávání. Dalším cílem je vývoj aplikace pro management Kooperačního systému článkové bibliografie. Dalším cílem je 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). Pro existenci a vývoj kooperačního systému a souborné databáze je nezbytná spolehlivá HW a SW podpora.“

1. Získávání a zpracování plných textů v rámci Kooperačního systému článkové bibliografie

Zajistit plné texty ke zpřístupnění v rámci vyvíjených aplikací (nákup plných textů, licence - přístup do databáze Anopress):

zajistit plné texty k propojování;

propojovat data o článcích s plnými texty - propojování off-line (program pro identifikaci plných textů v databázi TamTam určených k propojení bibliografických záznamů s plným textem a samotné programové propojení)

zajistit statické propojování bibliografických záznamů článků báze ANL na vytipovaná „stabilní“ www periodika/články (realizováno zejména v souvisejících projektech VaV Propojení analytických záznamů s plnými texty ... a Budování vzájemně kompatibilních Informačních systémů ...);

zajistit vývoj aplikace pro získávání, zpracování plných textů on-line a zpřístupnění plných textů v plnotextové databázi ANL FULL (linka poloautomatické extrakce/indexace - TTDE);

umístění metadat - metadata uvnitř plného textu, metadata vně plného textu;

charakter URL adresy k propojení: statická (pevná URL), **dynamicky** tvořená (v aplikaci TOPIC), **SFX** – vytvoření předpokladů pro propojování přes(reálně ale ne v tomto projektu).

Vývoj linky TTDE na zpracování bibliografických záznamů z plných textů, resp. linky poloautomatické extrakce z plných textů Anopress (případně i jiných textů) TTDE (TamTam Data Extractor), a to ve dvou podobách:

Windows linka lokální, tj. linka umístěná na lokálních počítačích bez napojení na autoritu (použitelná pro texty z Anopressu - vstup .txt forma);

WWW linka, tj. linka s možností stahování z báze národních autorit přes Z39.50

Toolkit Klienta (použitelná pro texty z Anopressu – vstup .txt forma a pro libovolné html WWW texty) ;

editace a doplňování dat v lince TTDE zejména dat věcného popisu včetně napojení na národní autoritu ve formuláři pro editaci (struktura formuláře navržena dle struktury UNIMARC), identifikace článků pomocí automaticky generovaného identifikátoru SICI, automaticky generovaná URL pro statické propojení bibliografických záznamů ANL na plné texty báze ANL FULL, automaticky generovaný Dublin Core, formát UNIMARC/M21

2. Vývoj dílčí aplikace pro import bibliografických záznamů (UNIMARC/MARC21) zpracovaných v lince TTDE do ALEPHu (báze ANL) a plných textů s metadaty DC a UNIMARC do báze ANK FULL (TOPIC)

3. Vyhledávání a zpřístupnění plných textů v systému TOPIC a na serveru full.nkp.cz s těmito možnostmi:

integrace vyhledávání pomocí řízených termínů, pojmového vyhledávání a neostrého vyhledávání;

3 typy **vyhledávacích formulářů** podle pokročilosti vyhledávání;

zobrazení metadat a plných textů v několika formátech včetně zpřístupnění DC v XHTML a XML;

aplikace pro stahování a export plných textů a metadat v několika formátech;

aplikace Periodika na www včetně portálu volných textů na Internetu ve struktuře regionální a oborové a přehledu Zdroje informací o článcích, plných textech a související zdroje na WWW v ČR s důrazem na české plnotextové zdroje;

aplikace pro administraci databáze ANL FULL a portálu volných textů na Internetu, vedení účtů, statistiky, opravy metadat v plných textech, opravy rejstříků aj.

4. Vývoj aplikace pro management Kooperačního systému článkové bibliografie, tj. aplikace pro příjem, kontrolu, správu a export dat do ANL a aplikace pro kontrolu URL.

V rámci aplikace MNG:

zajistit koordinovanou excerpci základnu systému s omezením duplicit ve zpracování ;

zkvalitnit výběru titulů s ohledem na seriály odborného charakteru;

budovat Soubornou bibliografickou databázi ANL s ohledem na různé SW vybavení spolupracujících knihoven;

zajistit průběžnou **integraci dat knihoven s různým SW** (KP-Sys, KP-Win, Tinlib resp. T-Series, Rapid resp. AREV, Lanius), formáty (iso, řádkový, UNIMARC, MARC 21, Výměnný Formát, Lanius aj.) a různých kódech (ISO 646 nebo ISO 5426, PC Latin 2, kód Kamenických, UNICODE UTF -8, ISO 8859-2, Ansel) automatizovaným způsobem;

zajistit **kontrolu aktuálních platných URL adres.**

Jednotlivé části aplikace pro MNG KOSABI:

aplikace pro administraci a zobrazení **Báze titulů** (zobrazení dle zpracovávající instituce, kategorie titulů, dle napojení na plný text , též detail záznamu titulu;

aplikace pro **upload záznamů** z kooperačního systému (v několika formátech a kódech, login a heslo pro každou instituci);

aplikace pro **zpracování a kontrolu záznamů** (formálně logické chyby, kontroly na UNIMARC/M21, statistiky chyb) a export dat pro bázi ANL (s ohledem na kritické chyby a chyby UNIMARC);

aplikace pro kontrolu aktuálních platných URL adres (adresy na full.nkp.cz a adresy na volné www články);

dokumentace k aplikaci pro management kooperačního systému (popis).

5. Podrobná dokumentace (charakteristika) vyhledávání v bázi ANL FULL

6. HW a SW podpora (též z projektu Propojení analytických záznamů s plnými texty ...)

Dosavadní vybavení fulltextového zpřístupněn článků v NKČR - báze ANL FULL (server FULL, WIN NT, TOPIC-Portal One-VIS)

Power Edge 6300 – Pentium III Xeon, 500 Mhz/512

RAM 1 GB

SCSI disky – 5x18 GB SCSI LVD 7200 1”

Řadič pole RAID 5 – řadič PERC2/SC, 1xLVD, 16MB cache

Síťová karta Intel Pro 100+RJ45
Zálohovací zařízení DAR 12/24 GB SCSI DDS
Search Verity Information Server (TOPIC) v. 3.7 – Portal One, neomezená licence
Windows 2000, Windows NT (30 licencí)
ScanJet+OCR
VIS – internetová a intranetová licence

Pro management Kooperačního systému článkové bibliografie MNG KOSABI – báze ANL KOSABI (server ANL, Oracle, Linux)
PC Pentium III – 700 Mhz
RAM 512 MB
HDD 27 GB
VGA S3 Trio 3D 4 MB
Síťová karta 10/100
APC Smart UPS 420i NET
ORACLE 9.2, Linux SuSE 7.3, Windows 98
(5 licencí)
Jdeveloper 1 licence

7. Vytvoření předpokladů pro integraci dat zpracovávaných v Kooperačním systému článkové bibliografie do JIB a zpřístupnění plných textů

B.1.4 Popis oblastí řešení a aplikací (linky TTDE, TOPIC, Aplikace pro MNG Kooperačního systému článkové bibliografie)

Názorná dokumentace řešení a praktických výsledků - Přílohy F1-F9

B.1.4.0 Analýza problematiky automatické či poloautomatické indexace a selekční úplnost topiků

Analýza problematiky automatické či poloautomatické indexace v bázi ANL FULL (cit. J. Schwarz, 2002) - teoretická analýza s praktickým doporučením je velmi významná pro vývoj projektu (dále jsou uvedeny citace z práce mírně modifikované).

Někteří autoři např. považují vyhledávání v plném textu dokumentu za nejjednodušší formu automatické indexace.

Trend do budoucna směřuje k vývoji takových vyhledávacích metod, které budou za pomoci technologií umělé inteligence pracovat přímo s plným textem a k automatické indexaci v původním slova smyslu nebude vůbec docházet; dojde k posunu od systémů založených na externí bázi pojmů či znalostí (knowledge-based systems) k systémům založených přímo na zpracování plného textu dokumentů (text-based systems).

Automatická indexace má těsnou vazbu na automatickou klasifikaci, automatické shlukování (*clustering*) a automatické abstrahování (*abstracting*) dokumentů. Automatická indexace také souvisí s automatickou tvorbou tezauru, příp. vytvářením sémantických sítí nebo znalostních bází, které jsou nezbytným předpokladem pro odpovídající funkčnost metod založených na automatickém přiřazování.

Typy a techniky automatické indexace

Automatická indexace se podle standardních koncepcí běžně rozděluje na dva základní typy:

- **automatická extrakce** je automatická indexace založená na extrakci indexačních termínů přímo z plného textu dokumentu (využívá se tak pouze přirozeného jazyka) (*automatic extraction indexing, selection of natural language index terms*)
- **automatické přiřazování** je automatická indexace založená na určování indexačních termínů z řízeného slovníku nebo znalostní báze na základě jejich srovnání s výrazy z plného textu dokumentu (*automatic assignment indexing, assignment of controlled language index terms*).

Automatická extrakce je jednodušší a snáze programovatelnou metodou, v současnosti na jejím základě pracuje většina funkčních systémů automatické indexace.

Automatická extrakce

1. lexikální analýza (*lexical analysis*) – identifikace jednotlivých slov a sousloví v textu dokumentu; někdy se identifikace sousloví uvádí vzhledem ke své složitosti jako samostatná procedura
2. odstranění nevýznamových a nespecifických slov - provádí se pomocí negativního slovníku (slovníku stopslov); někdy se tato procedura považuje za součást lexikální analýzy
3. lematizace (*stemming*) – redukce slov na jejich základní tvary, resp. kmen; opakem tohoto procesu (např. při vyhledávání) je derivace, kdy se k základnímu tvaru slova generují jeho inflexní tvary
4. srovnání slov, resp. jejich kmenů s termíny řízeného slovníku – jedná se o jednoduché srovnání termínu vybraného z textu s termínem řízeného slovníku bez ohledu na strukturu řízeného slovníku; ta bývá zohledňována pouze při automatickém přiřazování
5. vážení neboli stanovení vah termínů (*weighting*) – provádí se na základě různých, nejčastěji frekvenčních metod

Pokud je při extrakci termínů zohledněna jejich pozice ve specifické části textu (např. nadpis, abstrakt, první a poslední odstavec apod.), promítá se tato skutečnost i do výpočtu váhy, nejčastěji tím způsobem, že váha termínu, který se vyskytuje v obsahově nejdůležitější části textu (např. název) je násobena koeficientem o zvolené velikosti, a váhy termínů nacházejících se v obsahově méně signifikantních částech dokumentu jsou násobeny koeficienty o nižších hodnotách. Tak se dosáhne toho, že termín z obsahově důležité části dokumentu získá relativně nejvyšší váhu.

Ve srovnání s intelektuální indexací, je automatická extrakce absolutně konzistentní.

Jako podstatné omezení automatická extrakce lze chápat skutečnost, že použité metody neodpovídají postupu, jakým probíhá intelektuální indexace, která není většinou realizována jako extrakce, nýbrž jako přiřazování. **Z toho důvodu automatická extrakce obecně přináší horší výsledky než intelektuální indexace a přestože na základě automatické extrakce funguje většina současných systémů, řada z nich slouží z tohoto důvodu pouze jako automatizovaná podpora intelektuální indexace.**

Automatické přiřazování

Předpokladem funkčnosti automatického přiřazování je existence strukturovaného řízeného slovníku, který může být implementován ve formě tezauru, sémantické sítě nebo znalostní báze. Klasické tezaury se svými jednoduchými vztahy hierarchie, asociace a ekvivalence jsou sice pro automatické přiřazování použitelné, ale často bývají pro větší efektivitu celé procedury obohaceny o další dílčí typy vztahů a forem lexikálních jednotek nebo jsou užívány společně s dalšími externími slovníky či znalostními bázemi.

1. identifikace termínů v plném textu
2. srovnání termínů s relevantními profily pojmů z řízeného slovníku
3. určení indexačních termínů na základě profilu pojmů reprezentovaných termíny vybranými z textu

Hodnocení automatického přiřazování **prokázala, že účinnost automatického přiřazování je na 60-67% úrovni ve srovnání s intelektuálním zpracováním.**

Selekční úplnost topiků v bázi ANL FULL (cit. J. Schwarz, 2002)

Základním cílem analýzy je zjistit selekční účinnost topiků, tj. předdefinovaných komplexních dotazů, zejména z hlediska úplnosti výsledků vyhledávání, a předložit případná doporučení pro jejich úpravu, popř. pro úpravu jiných selekčních prvků databáze.

Pro účely analýzy byly stanoveny dvě množiny, které se porovnávaly:

množina záznamů vyhledaných pomocí předmětových hesel (RMD) a množina záznamů vyhledaných pomocí hesel a omezená pomocí topiku (TMD).

Výsledky komparací lze vztáhnout nejen na topiky, ale i na předmětová hesla, resp. popis dokumentů předmětovými hesly.

Dokumenty z RMD (D_R) byly srovnány s dokumenty z TMD (D_T) pomocí koeficientu úplnosti vyjádřeném v procentech:

R=100% v případech shodnosti obou množin (RMD a TMD).

R<100% v případech, že TMD neobsahuje všechny dokumenty obsažené v TMD.

Z porovnání výsledků vyhledávání vyplynulo, že 100% úplnosti dosahuje pouze polovina TMD (26 z 50, tedy 52%). Zbývá, menší polovina TMD (24 z 50) dosahuje menší než 100% úplnosti. Lze konstatovat, že úplnost vyhledávání na základě topiků je přes 80 % (84 %).

Hodnocení výsledků komparace

Možné nedostatky na straně topiků a na straně na straně předmětových hesel.

Vliv topiků na úplnost vyhledávání v bázi ANL FULL (J. Schwarz, 2003)

Analýza rozšiřuje studii z r. 2002 a navazuje na analýzu provedenou v r. 2002. Z výsledků vyplynulo, že 100 % úplnosti dosahují cca 2/3 TMD. Necelá třetina dosahuje menší než 100 % úplnosti (z toho 1/3 vyšší než 80 %, necelá třetina menší než 60 %). Ve srovnání s předchozím průzkumem (Schwarz, 2002) mají všechny hladiny úplnosti v podstatě stejný podíl, TMD v hladině od 60% se pouze přeskupily ve prospěch úplnějších TMD (od 80% úplnosti).

Průměrná úplnost všech neúplných TMD je cca 68%, medián (střední hodnota) je cca 73% (obě hodnoty jsou v podstatě totožné s výsledky z r. 2002).

Průměrná úplnost všech TMD je 90% (v r. 2002 84%). Oproti výsledkům průzkumu z r. 2002 se o 16 procentních bodů zlepšil i celkový podíl úplných TMD (68,4% úplných záznamů vůči 52% v r. 2002).

Cestou ke zlepšení tohoto stavu nemůže být pouze rozšiřování topiků o nová klíčová slova, ale i redefinice funkce a významu topiků pro koncového uživatele, která zajistí skutečně efektivní využití topiků při vyhledávání v databázi ANL FULL.

Určitým řešením, jak přiblížit topiky skutečným informačním potřebám uživatelů, je dlouhodobě sledovat (např. na základě logů) uživatelské dotazy do databáze a na jejich základě vytvářet topiky, které budou odpovídat nejčastějším dotazům kladených do databáze uživateli.

Zamýšlená implementace jazykového analyzátoru v projektu pro program 1N by vyhledávání v systému zlepšila.

Analýza automatické indexace a selekční úplnosti topiků v ANL FULL a její možný dopad na další praktický vývoj projektu na základě prací J. Schwarze.

V projektu jsou dosud užívány metody související spíše s automatickou extrakcí.

1. Extrakce či spíše generování údajů jmenného popisu a některých dalších polí UNIMARCu slouží k základní identifikaci dokumentu (minimální záznam) přímo z plných textů (získávání těchto údajů přímo z databáze Tamtam - TTSNK a jejich generování v lince TTDE do příslušných formátů). Rovněž je automaticky generován souhrn (první věty textu), v současné době je zprovozněna v nové verzi TOPICu i funkce shlukování (clustering) článků do skupin podle stejných klíčových slov.
2. V TTDE je funkční metoda automatického vkládání vybraných klíčových slov do formuláře TTDE a záznamu na základě požadavku při vyhledávání dokumentů v bázi TamTam. Tato slova by se musela indexátorem redigovat. Tyto metoda není při práci v lince využívána – není příliš kvalitní, dosud je věcný popis dokumentu zcela vytvářen katalogizátorem.
3. Systém TOPIC generuje také automatický souhrn k článku – zatím začátek textu dokumentu. Zjistit další možnosti v tomto směru.
4. Vzhledem k fuzzy vyhledávání v systému TOPIC a ještě nerealizovaným a neobjeveným dalším možnostem, které TOPIC má, se kloním spíše k využití funkcí dosud v projektu nezohledněných, ke zkvalitnění vyhledávání pomocí topiku na základě jejich ladění a zabudování prvků věcného popisu (kategorií a prvků předmětových hesel) přímo do topiků, spíše k automatizovanému než k zcela automatickému přiřazování věcných termínů. Užívají se metody např. automatizované kategorizace či klasifikace.
5. Hypoteticky: v systému je náběh na řízený slovník tematických kategorií, který by mohl být použit i k jejich automatizovanému přiřazování; použít v rámci topiků (a jejich generování ?) spolu se vstupním prvkem předmětového hesla.
6. Slovník bude pravděpodobně možno využít také jako pomůcku pro katalogizátora při přiřazování těchto kategorií.

7. Jak mohou být tyto kategorie spojeny s automatizovanou kategorizací nebo do jaké míry mohou být pomůckou při automatizované kategorizaci informací ukáže budoucnost.
8. V budoucnu zprovoznit hledání dokumentů s podobným obsahem - volný dotaz (Free Text Query), dotaz příkladem (Query By Example).
9. Topiky ladit na konkrétních dokumentech z příslušného oboru – určit jak obory, tak odpovídající dokumenty s kvalitním předmětovým popisem. V úvahu přichází periodikum Národní knihovna (zpřístupňované v projektu Propojení analytických záznamů s plnými texty a ke kterému vzniká řízený slovník) aj. odborná plnotextová periodika, která budeme mít pravděpodobně k dispozici, dále pak obory a dokumenty, o kterých lze tvrdit, že jsou kvalitně indexovány katalogizátorem (např. ekonomie/ekonomika, některé společensko politické časopisy).
10. K ladění topiků je nutný přístup ke struktuře jednotlivých topiků. Rovněž je třeba zkvalitnit samotnou věcnou indexaci pomocí předmětových hesel indexátory.

V r. 2003 a 2004 byly podány projekty v rámci 1N programu MŠMT Informační zdroje pro vědu a výzkum. V projektech šlo mj. o následující témata: automatizované kategorizace s využitím technologie LCR (Logistic Regression Classification) umožňuje vytvořit taxonomii založenou na analýze obsahu textu (Thematic Mapping). Optimalizace vyhledávání za podpory jazykového analyzátoru a zavedení konceptu tzv. sociálních sítí s využitím technologie TSA (Tensor Space Analysis). Technologie umožňuje analyzovat jak uživatelé vytvářejí, modifikují, vyhledávají a přijímají informace. Na základě toho pak provádí doporučení dokumentů a autorit, které mají vztah k dotazu, resp. profilu, uživatele (adaptivní řazení dokumentů – založené na obsahu i na tom, jak často byly dokumenty vyhledány a prohlédnuty v minulosti, doporučení dokumentů vzhledem k uživatelskému kontextu, objevování/nacházení expertů, tj. experti jsou rozpoznáni na základě informací, které vytvořili, modifikovali a/nebo ke kterým přistupovali; na základě společných znaků vznikají komunity uživatelů). V projektu je dále akcentována tvorba a ladění topiků a implementace taxonomií pro příslušné odborné zdroje v bázi ANL FULL s ohledem na oborové tezaury, využití systému pro sledování citačních vazeb.

B 1.4.1 [Windows a WWW TTDE linka](#) pro zpracování plných textů, zpřístupnění plných textů v systému TOPIC – Příloha F 2, F 3, F4

Společnost Anopress IT provozuje bázi TAMTAM na základě spolupráce s vydavateli novin a časopisů. Ze zdrojů této databáze jsou plné texty získávány a prostřednictvím výše popsané programové aplikace dále zpracovávány v Národní knihovně (věcně indexovány) v rámci linky a zpřístupňovány v systému TOPIC v bázi ANL FULL s ohledem na nejnovější trendy v této oblasti (plnotextové pojmové vyhledávání) pomocí tzv. topiků.

Údaje o článku jsou pomocí speciálních maker vkládány v Anopressu do hlavičky textu dokumentu a takto vybavené textové dokumenty jsou ukládány a indexovány a zpřístupňovány v databázi TamTam (TOPIC). Plné texty jsou **získávány na základě speciální aplikace TTSNK z databáze TamTam.**

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, zpracování bibliografických záznamů z plných textů (linka automatické či poloautomatické indexace), 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.

B.1.4.1.1 Architektura systému zpřístupňování plných textů, funkce systému a procesy. Linka TTDE, resp. WWW TTDE. Báze ANL, ANL FULL a systém TOPIC

1.On-line získávání dokumentů z databáze TamTam přes vstupní filtr pro linku zpracování - aplikace TT Special NK (TTSNK), textové dokumenty vstupují do linky se strukturovanou hlavičkou, kde jsou dohodnutými znaky a návěštými označeny příslušné údaje.

2. Extrahování údajů do formuláře a pracování údajů v lince Windows TTDE, popř. WWW TTDE

3. Zaslání souborů v UNIMARCu/M21 do bibliografické báze ANL, zaslání plných textů s metadaty na plnotextový server - báze ANL FULL (systém TOPIC) – možnost volby formátu (UNI/M21, kód Ansel, UTF-8, UNICODE)

4. Indexace a vyhledání informací , indexace, uložení a vyhledání informací (vyhledávání pomocí textu dotazu, formulářů, topiků, rejstříků – báze Access) - aplikační a datový server: Portal One SE - fulltextový vyhledávací systém,)

- **Konverze, generování dat v XHTML+-, XML+-, indexace, uložení a vyhledání informací (vyhledávání pomocí textu dotazu, formulářů, topiků, rejstříků) - aplikační a datový server: Portal One SE - fulltextový vyhledávací systém, server Dell Power Edge 6300 (1 GB RAM, 68 GB, Pentium III). MKINDEX - indexace, generování XML- a XML+, jednorázové aplikace/moduly (TTEXTPORT-TamTamExport, TTXMLGen -TamTam XML generator, TTDT-TamTam Data Tester, TTDF-TamTam Data filter, TTDC- TamTam Data Converter, MKDOC.ASP - propojení bibliografického záznamu s plným textem). Strukturované údaje jsou ukládány v Accessu, v budoucnu nutná SQL databáze pro metadata DC a XML, migrace TOPICu na platformu K2.**
- **Zobrazení a zpřístupnění informací - Internet Information Server Win 2000 - komunikace v síti, zobrazování dat a metadat v uživatelských formátech (citace bez/s odkazy, výsledky jednoduché, se souhrnem, seskupené) a pracovních formátech (klíčová slova, DC v HTML, XHTML+ -, XML+ -), tisk. Obecné základní informace, nápovědy, autorizace, statistika zobrazených informací přes Internet aj.**
- **Stažení a export záznamů - formáty a tvary: HTML, RTF, text, XML, UNICODE**
- **Administrace (nové v r. 2002) – úpravy a opravy textů (metadat), statistiky, údržba rejstříků, údržba portálu. V r. 2003 zabudována aplikace pro opravy a údržbu rejstříků/plných textů přímo do linky.**

B.1.4.1.2 TTDE – Příloha F2 a F3

Součástí Windows linky a WWW linky je vyvinutá aplikace pro přístup do báze TamTam – TamTam Special NK (TTSNK).

Linka zpracování bibliografických záznamů z plných textů, resp. Linka poloautomatické extrakce/indexace – Windows TTDE (TamTam Data Extactor) pro lokální počítače a WWW TTDE pro použití na Internetu

Vývoj TTDE

Windows TTDE

V rámci poloprovozu v r. 2002 byla laděna Windows TTDE linka a ověřována v poloprovozu, v r. 2003-2004 rutinní provoz linky. Do formuláře jsou extrahována data, která se kontrolují a doplňují věcným popisem. Automaticky je generována URL adresa, SICI, "provizorní" URN, které sleduje filozofii a strukturu stanovenou pro tvorbu URN.Z uložených a editovaných dat se generuje DC, UNIMARC pro ALEPH a klasický UNIMARC, hlavička zobrazovací a indexovací.

Automaticky lze generovat klíčová slova a extrahovat do formuláře. Na výstupu procesu zpracování zabudovaná funkce kódování výsledného souboru záznamů v Ansel, UNICODu a UTF-8. Po odeslání souboru do příslušného adresáře na server full.nkp.cz je tento v UNIMARCu (aplikací vyvinutou v NK) naimportován do báze ANL (ALEPH), dále zaindexován a zpřístupněn v systému TOPIC na serveru full.nkp.cz v podobě plného textu vybaveného metadaty s generovaným UNIMARCEm, DC v podobě HTML, XHTML, XML v kvalifikované a nekvalifikované formě.

Vyhledání je možno z plného textu, z metadat v něm uložených - tj. ze strukturovaných údajů (vyhledávání pomocí polí, formulářů, topiků a rejstříků). Zobrazení je možno pomocí tří uživatelských formátů. Je možno tisknout vyhledaný jak seznam článků, tak jednotlivé plné texty. Záznamy ANL jsou propojeny s plným textem automaticky generovanou adresou v lince zpracování, která je uložena jako statický link v ANL záznamech.

Autoritní tvary se ve Windows TTDE lince se kopírují z ALEPHu, nestahují se automatizovaně.

WWW TTDE

V roce 2003 byla navržena a částečně realizována WWW TTDE linka zpracování bibliografických záznamů z plných textů TTDE (TamTam Data Extractor) pro práci na Internetu – WWW TTDE. Jsou v ní zachovány základní funkce Windows linky a je rozšířena o funkce následující: možnost zpracovávat plné texty na Internetu s uploadem metadat i bez stažení plného textu, vstupní formáty zip, txt, html, propojení linky na rejstříky báze ANL FULL a na authority (zatím ve formě komponenty OLE), zabudování funkce Opravy přímo do linky. Formulář pro editování dat byl rozšířen o další pole/podpole UNIMARCU zejména s ohledem na propojení se soubory autorit.

V r. 2004 probíhaly dále práce na vývoji WWW TTDE - zejména práce na konverzi do MARCu 21, práce na úplném zabudování funkce stahování autoritních tvarů přes Z39.50 Toolkit Klienta uzpůsobené pro asp. rozhraní. Pro napojení linky na authority přes Z39.50 toolkit klient byly stanoveny atributy dle instrukce pro Z39.50 pro bázi AUT zveřejněné na <http://www.nkp.cz>. Během r. 2004 probíhaly rozsáhlé testy zejména konverzí dat, v 2. pololetí praktické ověřování ukládání dat v lince a přípravné práce pro ověřovací provoz linky v rámci oddělení analytického zpracování. V prosinci 2004 byla linka odzkoušena pracovníky oddělení analytického zpracování a uvedena na praxe (licence pro přístup k plným textům TamTam v r. 2004, říjen – prosinec, tj. přístup TTSNK TamTam Special NK byly financovány z projektu Budování vzájemně kompatibilních informačních systémů ...).

Dokumentace k lince TTDE (Windows i WWW) – viz Příloha F2, F3, F 8.

Koncem r. 2004 rozhodlo vedení NKČR dále nezpracovávat plné texty v lince TTDE a zpracovávat všechny dokumenty v systému ALEPH.

Katalogizace v lince TTDE

Windows TTDE

Údaje jsou extrahovány z plného textu a na jejich základě je generováno 5 hlaviček (headers). Data je možno upravovat také v hlavičce UNIMARC-A, UNIMARC. Automaticky se generují údaje v rozsahu minimálního záznamu stanoveného pro popis článků (kromě notace MDT), tj. údaje zejména jmenného popisu a kódované údaje. Automaticky se generuje souhrn článku (první věty textu). Pokud je zvolena příslušná funkce, je možno automaticky generovat do formuláře i klíčová slova z plného textu. Další údaje věcného popisu - předmětové kategorie doplněné MDT, hesla se doplňují. Automaticky se generuje URL, SICI.

Formulář má tři strany. První obsahuje jmenné a věcné údaje, druhá pouze věcné, třetí jmenné a věcné údaje. Je možno stahovat více článků najednou na základě tématu, názvu článku, názvu zdrojového dokumentu aj. údajů. Pracovník oddělení analytického zpracování NK nejprve vybere a vyhledá příslušné články ke zpracování v této databázi. Poté proběhne stažení vybraných článků na počítač příslušného pracovníka v podobě komprimovaného souboru. Po jeho dekompresi je text článku naimportován do příslušného adresáře. Po otevření formuláře ke zpracování se objeví v levé části formuláře seznam názvů článků. Po kliknutí na článku se automaticky objeví údaje o článku, které jsou obsaženy v jeho ve formuláři a příslušných hlavičkách. Po kontrole takto extrahovaných údajů z plného textu se

doplňují další bibliografická metadata zejména věcné povahy, popř. kopírují autoritní tvary z báze autorit. **Z extrahovaných a editovaných dat se generuje UNIMARC-A, Klasický UNIMARC, Dublin Core, indexovací a zobrazovací hlavička.** Poté je soubor zpracovaných článků odeslán.

Po doplnění formuláře je po odrážce různé **možno nastavit tvar výstupní hlavičky pro UNIMARC-A nebo UNIMARC a spustit ruční vstup dat, nastavit kód výstupních dat (Ansel, UNICODE, UTF-8). Standardně nastaven výstup UNIMARC-A v kódu UTF-8.**

Následuje odeslání záznamů pomocí volby odeslat na dolní liště, import do ALEPHu (program vyvinutý v NK), import na server full.nkp.cz.

WWW TTDE

Linka pro zpracování na WWW **zachovává základní funkce Windows linky, liší se designem, možností více typů vstupů co do formátů, většími možnostmi editace, obsahuje 13 hlaviček,** její formulář pro editaci obsahuje 4 hlavičky detailně strukturované zejména s ohledem na strukturu autoritních údajů (z formuláře vyřazeno SICI, SICI se generuje a automaticky vkládá do příslušných metadat, aniž by ho katalogizátor musel vidět ve formuláři). Lze stahovat a doplňovat metadata s paralelním uložením těchto metadat do plného textu a jeho stažením na server. Lze natahovat a editovat metadata určitého plného textu bez stažení tohoto textu do báze.

Z jednotlivých údajů je možno se napojit na rejstříky báze ANL FULL a na soubory autorit přes asp rozhraní varianty Z39.50 Tollkit klienta a údaje stahovat.

V rámci linky se navíc oproti Window TTDE generuje M21. Výstup je možný v formátu UNIMARC/M21 a v kódu Ansel, UTF-8, UNICODE.

Je implemenován modul pro opravy rejstříků v podobě další hlavičky.

Linka WWW TTDE – přehled funkcí:

Načíst soubory (stažené na lokální počítač):

Načíst soubory pro zpracování: pouze zip, txt, html a to: z databáze TamTam -až 10 souborů.

Poslat soubory na full.nkp.cz a spustit TTDE (otevřít formulář pro zpracování).

Načíst URL:

Stáhnout zadaná URL (přímo z WWW) – až 10 souborů.

Spustit TTDE (otevřít formulář pro zpracování).

Spustit Linku

Extrakce metadat, editace pomocí 4 hlaviček formuláře s napojením na soubory autority, hlavička pro zobrazení textu, hlavička UNIMARC - generovaný UNIMARC, hlavička UNIMARC A - generovaný UNIMARC pro ALEPH, MARC 21, zobrazovací a indexovací hlavička, hlavička Opravy, hlavička ... (tj. různé)

Hlavička Opravy: umožňuje opravy rejstříků ANL FULL a metadat v plných textech - komplet nová funkce (nahrazuje opravy původně připravené v Administraci) Po vyhledání údaje v příslušném rejstříku lze údaj zrušit, opravit nebo vypsát všechny články databáze, které tento údaj obsahují, tj. lze opravit záznamy a následně znovu přenést do databáze ANL FULL nebo nalezené články úplně zrušit. Tato funkce je ve stádiu experimentu.

Hlavička ... (tj. různé) - export dat v různých kódech a formátech, navíc možnost stahovat z URL.

Konečný stav a možné použití Windows TTDE a WWW TTDE

Windows TTDE je ve stádiu rutinního provozu (export dat v UNIMARCU pro bázi ANL a plné texty pro bázi ANL FULL) .

WWW TTDE je dotažena do stádia ověřovacího provozu. Ověřen je export dat v UNIMARCU (po té dochází k následné konverzi do M21 a k importu do báze ANL v M21; téměř dotažena konverze do M21 (export dat a import do ANL přímo v M21 zatím neověřen).

Linka (obě varianty) je použitelná po úpravě vstupním filtrem i na data existující v jiné databázi, event. v komunikaci mezi autorem, nakladatelstvím, bibliografickou agenturou a naopak. V případě vhodné situace je možno WWW TTDE po otestování použít v rámci kooperačního systému.

Pro data z www se předpokládá však do jisté úrovně strukturovaný vstupní text, ze kterého data mohou být extrahována.

V této struktuře mohou být zachyceny údaje nejen jmenné povahy (autor apod.), ale i povahy věcné (klíčová slova, abstrakt). Ty je potom možno převést pomocí vstupního filtru do linky automatické indexace k dalšímu zpracování. V neposlední řadě je možnost automatického zpracování závislá na způsobu organizace práce s těmito texty a jejich umístění v eventuelní databázi.

V případě neexistence těchto dat, data musí být komplet doplněna.

Linka zpracování TTDE zabudovaná do předpokládaného možného vývoje - viz Příloha F3.

B.1.4.1.3 Server full.nkp.cz a systém TOPIC – indexace, konverze do XML, XHTML, uložení, vyhledávání a zpřístupnění informací v systému TOPIC, topiky - Příloha F 4

Pro indexaci dat do fulltextové databáze (ANL FULL) v NK byl vyvinut program MKINDEX

TOPIC je produkt americké firmy Verity, v současné verzi Portal One. V ČR TOPIC a další produkty založené na stejné technologii dodává firma TOVEK, s.r.o.

Základní charakteristika systému

- Fulltextový pojmově orientovaný vyhledávací systém, pojmové vyhledávání (concept retrieval) pomocí strukturovaných dotazů (topiků)
- Hodnocení důležitosti, relevance vyhledaných dokumentů vzhledem k dotazu (relevance ranking), váhy
- Kvantifikace obsahu dokumentů
- Shlukování dokumentů podle společného kontextu (clustering) a vytváření automatické anotace - sumarizace (summarization)
- Interaktivní vyhledávací systém – hledání dokumentů s podobným obsahem - volný dotaz (Free Text Query), dotaz příkladem (Query By Example)
- Dotaz formulovaný uživatelem - výraz složený ze slov a frází, který hledáme v dané databázi
- Topik - je předem definovaný uložený strukturovaný dotaz, resp. téma, které je tvořeno slovy, frázemi, operátory a modifikátory; obsahuje informace o předmětu hledání
- Topik má podobu pojmového stromu, na jeho nižších úrovních (větvích) jsou množiny dalších pojmů, resp. témat, která jsou tvořena dále nedělitelnými klíčovými slovy (listy)
- Pojmový strom tvoří vyhledávací podmínku pro dokumenty týkající se určitého tématu
- Topik je konceptuální popis znalosti o dané problematice ve formě znalostního stromu
- Definice topiků tvoří tzv. znalostní bázi .

Konstrukce topiku

- jednotlivé větve topiku, pojmy resp. témata, podpojmy, resp. podtémata a klíčová slova jsou připojena k vyšší úrovni operátory
- důležitost pojmů resp. témat, podpojmů resp. podtémat vzhledem k ostatním je určena váhami
- topik se vytváří speciálním editorem

- **předpoklad: dobrá znalost operátorů a orientace v dané oblasti.**

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 svěbytně.**

Operátory

listové operátory: WORD, STEM, SOUNDEX, WILDCARD, THESAURUS, SUGGEST a TYPO

proximitní operátory: PHRASE, SENTENCE a PARAGRAPH

koncepční operátory: AND, OR a ACCRUE

logické operátory: ANY a ALL a relační operátory: rovnost '=', větší než '>', větší nebo rovno '>=', menší než '<', menší nebo rovno '<=', MATCHES, SUBSTRING, CONTAINS, STARTS, ENDS
ACCRUE sblíží operátor AND a OR: "čím více různých klíčových slov nalezeno, tím je dokument důležitější", řeší dilema mezi přesností a úplností.

Váhy

Váhy lze použít pouze u uzlů připojených pomocí koncepčních operátorů (AND, OR a ACCRUE); pokud váhu u připojovaného uzlu neuvedeme, použije se standardní váha - pro AND a OR 1.00, pro ACCRUE 0.50.

Přiřazením váhy k určitému uzlu určujeme, jak (od 0.01 do 1.00) se určitý uzel podílí na celkovém skóre daného dokumentu při výběru; změnou vah je možno změnit pořadí dokumentů v seznamu dokumentů dle skóre relevance.

Charakteristika serveru <http://full.nkp.cz> a báze ANL FULL, Periodika na WWW - portál - Příloha F 4

Server full.nkp.cz je určen primárně ke zpřístupňování plnotextových dokumentů, zejména těch, které jsou obsaženy v periodicky vydávaných dokumentech české provenience.

Základní členění stránek - základní a dílčí nabídky

Stránky jsou tvořeny horním pruhem a základní částí (hlavním prostorem). Horní pruh (frame) je přístupný stále, obsahuje v horní části dynamické rozbalovací menu s jednotlivými nabídkami základních a dílčích funkcí. V základní části se zobrazují požadované funkce.

Obecné (základní) informace

Charakteristika db (projekty, charakteristika serveru, charakteristika báze ANL FULL)

Další databáze (databáze NK - charakteristika báze ANL)

Periodika na WWW (volně dostupná periodika na www)

Vstup do db (ANL FULL)

Jak se zaregistrovat (komentář)

Registrace (formulář)

Přihlášení (formulář)

Vyhledání (ANL FULL)

Jak vyhledávat (metodika)

Formuláře (i stahování e export plných textů)

Rejstříky (i stahování a export plných textů)

Charakteristika databáze ANL FULL (obsah, zdroje)

Databáze [ANL FULL](#) - fulltextové vyhledávání ve vybraných článcích z novin a časopisů zatím s retrospektivou od r. 1997 (periodikum Národní knihovna od r. 1990). Databáze je provozována na serveru full.nkp.cz v systému TOPIC.

Plné texty jsou vybaveny citací (bibliografický popis, resp.metadata), automaticky tvořeným souhrnem (začátek článku).

Vyhledávání je možné z každého slova z textu, pomocí slov i frází, pomocí bibliografických údajů, tj. metadat, dále pak tzv. topiků (definovaných dotazů, resp. témat v rámci jednotlivých oborů). Topiky v r. 2003: tematické oblasti (8), skupiny témat (28), detailní témata (zatím cca 197).

Báze obsahuje následující tematické oblasti (8):

- **Ekonomika, obchod, finance**
- **Hospodářství, výroba**
- **Kultura a umění**
- **Země světa a geografie**
- **Humanitní a společenské vědy**
- **Matematika a přírodní vědy**
- **Lékařství, zdravotnictví, technika**
- **Sport volný čas**

Výběr plných textů a tematický záběr koresponduje s výběrem článků ke zpracování pro bázi ANL (kategorie titulů, úplnost excerpce a typy článků, tituly viz dříve). Výběr titulů - v současné době celostátní deníky (celkem 6) a některé časopisy (Bankovníctví, Ekonom, Euro , Kapitál, Magazín Práva, Pátek Magazín LD, Profit, Respekt, Reflex, Týden, Květy - celkem 17 - viz Příloha 7.

Další databáze - ANL - Příloha F1

[Periodika na WWW](#) - portál – Příloha F4

Nabídka volně přístupných periodik na Internetu se dále bude doplňovat, event. jejich uspořádání měnit. Nejsou zde zahrnuta periodika firemní, inzertní, bulvární, propagující hnutí potlačující lidská práva, sportovní, zpravodajská (zprávy z tiskových agentur bez dalšího kontextu) aj. periodika efemérní povahy.

U každého periodika je možno uvést tyto údaje: oficiální název, www adresa, popis obsahu, vydavatel, tel., fax, e-mail, adresa, ISSN. V budoucnu bude možno podle těchto údajů vyhledávat.

Nabídka není ještě zdaleka proporcionální, postupně se ladí. V roce 2004 je v portálu zahrnuto cca 90 regionálních periodik, 90 oborových periodik.

Díličí nabídky:

- **regionální periodika** (uspořádaná podle krajů)
- **oborová periodika** (uspořádaná podle oborů)

Oborová periodika jsou členěna do skupin a v rámci nich do jednotlivých oborů a oblastí:

Kultura, umění

- film, televize
- hudba
- divadlo
- literatura

Odborná periodika

- architektura
- astronomie
- přírodní vědy
- filozofie
- psychologie
- ekonomika
- historie
- knihovnictví
- lékařství
- stavebnictví
- technika, IT
- právo

Společnost

- náboženství (církve, religionistika)
- životní styl
- politika (politická scéna, zprávy, komentáře a glosy událostí)

Zahraničí, zajímavosti, cestování

Zájmy (hobby)

Ostatní

- **Zdroje v ČR**

Další zdroje periodik dostupných na WWW (některé informační agentury, zpravodajské servery, nakladatelství, vyhledávače, brány a portály).

[Jak se zaregistrovat](#) (komentář k registraci v databázi ANL FULL)

Bibliografické údaje o článcích, resp. metadata a **plné texty jsou k dispozici čtenářům Národní knihovny.**

Pro **externí uživatele** jsou běžně k dispozici pouze bibliografické údaje o článcích. **Zpřístupnění externím uživatelům je omezeno na jeden den. Rutinní zpřístupňování plných textů bude možné po vyřešení autorskoprávních,** popř. ekonomických otázek souvisejících se zpřístupňováním plných textů.

Tito uživatelé se musí pro získání plného textu **zaregistrovat pomocí formuláře v nabídce Registrace.** Po vyplnění povinných údajů je těmto uživatelům zasláno e-mailem potvrzení registrace s aktivačním odkazem a heslem - po odeslání tohoto odkazu je registrace aktivována a prostřednictvím přiděleného hesla zajištěn bezplatný zkušební přístup na 1 den. Nabídka **Přihlášení** slouží k autentifikaci registrovaného uživatele. Ve formuláři pro **přihlášení je třeba vyplnit jméno a přidělené heslo.**

Plně funkční je aplikace pro hromadné stahování a export dat v několika formátech. Běžně však není využívána – dosud nevyřešené autorsko právní vztahy.

[Jak vyhledávat](#) a způsoby zobrazení (stručný přehled)

Báze ANL FULL obsahuje jak bibliografické popis v různé míře podrobnosti podle vývoje systému, tak plný text, přičemž obě tyto části jsou indexovány a lze z nich paralelně vyhledávat a docílit tím větší míru relevance výsledku vyhledávání k položenému dotazu. Vyhledávat lze též pomocí rejstříků. Systém TOPIC navíc umožňuje pojmové vyhledávání podle témat

TOPIC umožňuje velmi sofistikované kladení dotazů vyžadující určitou zkušenost. Na druhé straně je možné položit dotaz velmi jednoduchým způsobem.

[Formuláře](#) - tři základní formuláře podle pokročilosti vyhledávání - **základní, rozšířený, rozšířený s tématy, resp. s topiky** (viz dále).

[Rejstříky](#) - nadefinováno **17 rejstříků**, podoba rejstříků se bude dále optimalizovat (viz dále).

[Dotazy](#) - druhy dotazů se liší náročností formulace a možnostmi ovlivnit výsledek vyhledávání (viz dále) .

Prostý dotaz (odpovídá možnostem základního formuláře pro vyhledávání)

Obsahuje slova nebo fráze oddělené čárkami nebo logickými spojkami, formulace mohou obsahovat různé konvence. Při vyhledávání se všechna slova skloňují a mají stejnou váhu.

Formulářový dotaz (odpovídá možnostem rozšířeného formuláře pro vyhledávání)

Obsahuje kromě možnosti pro zadání hledaného slova nebo fráze jako u prostého dotazu i pole pro zadání podmínek pro jednotlivé položky strukturované části textové databáze. Jde o rozšíření prostého dotazu.

Tematický dotaz, resp. topik - definuje jednotlivá témata (odpovídá možnostem rozšířeného formuláře pro vyhledávání s topiky)

Topik je typ dotazu, na jehož základě se vyhledají dokumenty, které se týkají určitého tématu, oboru nebo problému.

Topiky jsou rozděleny v bázi do tří úrovní. Lze vyhledávat podle některých nadefinovaných topiků (označeny šipkou vlevo). jednotlivé je třeba ladit. Tematický dotaz je vytvořen v podobě nadefinovaného pojmového stromu v bázi. Skládá se z pojmů - slov, charakteristických pro danou

oblast. Ty pak tvoří listy stromu. Témata lze samostatně zároveň používat v prostých i formulářových dotazech. Napojení listů/témat na další úroveň je zajištěno operátorem a přidělením váhy tomuto listu/tématu. Váhy ovlivňují skóre dotazu. TOPIC postupuje od listů až k vrcholu stromu a postupně vypočítává skóre. Topiky viz dále.

Pole

Pole dotaz, resp. text dotazu – obsaženo ve všech formulářích.

Implicitní pole - obsažena v rozšířeném formuláři a v rozšířeném formuláři s tématy .

Vazby mezi polem dotaz, dalšími poli a tématy

- vazba dotaz-pole - vztah mezi polem dotaz a dalšími poli v rozšířeném formuláři pro vyhledávání
- vazba mezi poli - vztah mezi jednotlivými dalšími poli v rozšířeném formuláři pro vyhledávání
- vazba mezi tématy - vztah mezi zaškrtnutými tématy v rozšířeném formuláři pro vyhledávání s tématy
- vazba dotaz - téma - vztah mezi polem dotaz a tématy jako celkem v rozšířeném formuláři pro vyhledávání s tématy

Způsoby zobrazení seznamu výsledků (názvů vyhledaných článků)

V záhlaví seznamu výsledků uveden počet vyhledaných článků, možnost listovat v seznamu výsledků.

Seznam výsledků (názvů vyhledaných článků):

- **jednoduchý** (standardní zobrazení)
- **se souhrnem** (standardní zobrazení plus souhrn)
- **seskupený** (standardní zobrazení plus články uspořádané do skupin podle shody v klíčových slovech **podle společného kontextu, clustering**) – ve verzi TOPIC 3.7 se tato funkce v současné době zprovožňuje

Třídění seznamu výsledků (názvů vyhledaných článků)

Názvy článků je možno zobrazit od určité hranice skóre relevance, v určitém počtu na stránku a třídít dle skóre, názvu, zdroje, data, strany, a to vzestupně či sestupně.

Zobrazení údajů o článku – viz dále.

- uživatelské formáty
- pracovní formáty

Vyhledávání – metodika – Příloha F8

Vyhledávat lze z pole dotaz (horní řádek přítomen ve všech typech formulářů, zde je také možno pomocí návěští pro jednotlivé strukturované údaje vyhledávat stejně jako podle rozšířených formulářů, dále možnosti contains, =), z rozšířeného formuláře pro vyhledávání, z rozšířeného formuláře s topiky.

K vyhledání strukturovaných údajů lze také využít rejstříky.

Možnosti vyhledání v poli dotaz a formulářích

Vyhledání probíhá fulltextovou formou, tj. z plného znění jednotlivých dokumentů a z jejich citace, resp. metadat, resp. bibliografického popisu. Systém rozeznává pádové koncovky podstatných a přídavných jmen. Lze vyhledávat podle jednoduchého slova nebo fráze. **U rozšířeného formuláře a u rozšířeného formuláře s tématy lze nastavit** (ve spodní části formuláře) pro implicitní pole formuláře vyhledávání podle výskytu **řetězce** kdekoli v poli pomocí **operátoru <contains>**. Pokud je vyžadováno **přesné zadání** i porovnání celého řetězce **použijeme =**.

Tyto operátory lze zapsat také přímo do pole dotaz, resp. hledat v textu. Operátor <contains> lze v tomto případě zapsat jednoduše jako #. V tomto případě nastavení operátorů nemá vliv při hledání z polí pomocí návěští a při použití různých konvencí (saved, *, atd.). Doporučujeme však v případě nejistoty nastavení <contains>.

Základní formulář

Nabídky:

Dotaz - hledat v textu (formulace dotazu)

Období pro vyhledání od do

Typ seznamu výsledků (článků):

jednoduchý (pouze název s nabídkou zobrazovacích formátů)

se souhrnem (plus souhrn, který tvořený počátečními větami článku)

seskupený (skupiny článků uspořádané podle shody v klíčových slovech) viz též výše.

Rozšířený formulář

Nabídky:

Obsahuje **stejně nabídky jako základní formulář, navíc pak vyhledávání podle implicitních polí ve struktuře formuláře**. Je možno zvolit, zda vyhledávat v těchto implicitních polích způsobem <contains> nebo způsobem =. Pro seznam výsledků je možno navíc navolit oproti základnímu formuláři skóre relevance, od které zobrazovat články, zobrazit určitý počet článků na stránku, dále třídít dle skóre relevance, názvu, zdroje, data vydání, stran, a to sestupně či vzestupně.

Hledat v textu - odpovídá poli Dotaz v základní formuláři (formulace dotazu)

Období pro vyhledání od do

Implicitní pole:

název, autor (s nabídkami pro jednotlivé údaje)

číselné údaje (s nabídkami pro jednotlivé údaje)

předmět (s nabídkami pro jednotlivé údaje)

další nabídky: typ článku (s nabídkami pro jednotlivé typy článků)

zdrojový dokument (s nabídkami zdrojových dokumentů)

Způsob vyhledání řetězců (v implicitních polích):

<contains> (postačí výskyt řetězce v poli)

= je požadováno přesné zadání i porovnání řetězce

Typ seznamu výsledků (článků):

jednoduchý (pouze název s nabídkou zobrazovacích formátů)

se souhrnem (plus souhrn resp., který tvořený počátečními větami článku)

seskupený (skupiny článků uspořádané podle shody v klíčových slovech) viz též výše

Skóre (relevance, od které zobrazovat články)

Výsl./str. (počet článků na stránku)

Třídění: dle skóre (relevance)

dle názvu

zdroje

data (vydání)

stran

a to sestupně nebo vzestupně.

Rozšířený formulář s tématy

Nabídky:

Tento formulář má **stejně nabídky pro vyhledávání jako rozšířený formulář, umožňuje vyhledávat navíc podle témat, resp. topiků, resp. dotazů. Pro bázi ANL FULL jsou nadefinovány některé topiky, rozdělené do tří úrovní.** Topik slouží k tomu, že hledaný výraz vyplněný do textu dotazu a různě zpřesňovaný je navíc upřesněn topikem, tj. tématem. Např. hledáme-li v bázi výraz hvězdy, výsledek obsahuje několik tisíc dokumentů. Specifikujeme-li tento výraz pro vyhledávání v rámci astronomie, najdeme pouze několik set dokumentů.

Uživatel zvolí typ vyhledávacího formuláře popř. změní jeho implicitní parametry. Vlastní dotaz potom je možné zadat několika způsoby.

V poli dotaz, resp. text dotazu pomocí běžných konvencí nebo návěští pro pole (pole dotaz v základním formuláři, text dotazu v rozšířeném formuláři či text dotazu v rozšířeném formuláři s tématy). Jsou nadefinována návěští pro vyhledávání v jednotlivých polích bibliografických údajů, resp. metadat, resp. citace

V implicitních formulářových polích pomocí údajů obsažených v polích bibliografického popisu, resp. metadat, resp. citace a **zadáním volby <contains> nebo = pro způsob vyhledání řetězců**

Dotaz lze zadat v kombinaci s omezením období volbou pro **období od do** ve všech formulářích.

Pomocí témat - topiků (tématických dotazů) připravených v roletových menu v levém pruhu rozšířeného formuláře v tématy

Shrnutí základních způsobů vyhledávání pomocí typů formulářů :

Základní formulář - kombinace pole dotaz, období pro vyhledávání

Rozšířený formulář - kombinace dotazového pole, období pro vyhledávání, implicitní pole ve formuláři, nastavení <contains> (hledání po slovech) či = (hledání podle přesného řetězce - podle přesného řetězce je lépe hledat pomocí rejstříků)

Rozšířený formulář s tématy - kombinace dotazového pole, období pro vyhledávání, implicitní pole ve formuláři, nastavení <contains> či =, témata

Vyhledávat lze také podle **rejstříků** (17 rejstříků, obsahují metadata resp. strukturované údaje, které lze hledat také pomocí způsobů uvedených výše)

Metodika k některým způsobům vyhledávání

Pole dotaz, resp. text dotazu (horní řádek - přítomen ve všech typech formulářů)

Nejjednodušší dotaz tvoří jediné slovo, fráze. Pro zadávání složitějších nebo víceslovných dotazů je možno použít logické operátory (and, or, not a další), případně další konvence, např. zástupné znaky (wildcards) - viz dále bod PŘ.1-9. Uvedením návěští se vyhledávání omezuje na příslušné pole - PŘ. 10-12.

Základní operátory:

and - v poli jsou obsaženy všechny hledané výrazy

or - pole obsahuje alespoň jeden hledaný výraz

not- hledaný výraz nesmí být v poli přítomen

Konvence pro všechny formuláře - pole dotaz, resp. text dotazu

Př. 1

hvězdy

Vyhledají se dokumenty, které obsahují různé gramatické tvary zadaného slova (hvězdy, hvězdám...).

Př. 2

hvězdy, asteroidy, planety

Vyhledá dokumenty, které obsahují různé gramatické tvary slov „hvězdy“ nebo „asteroidy“ nebo „komety“ (čárky lze nahradit operátorem or nebo <accrue>, který je přesnější).

Př.: 3

(hvězdy, asteroidy) and komety

Vyhledá dokumenty, které obsahují různé gramatické tvary slov „hvězdy“ nebo „asteroidy“ a zároveň s nimi nebo některými z nich i slovo „komety“.

Př. 4

(hvězdy and asteroidy) not komety

Vyhledá dokumenty, které obsahují různé gramatické tvary slov „hvězdy“ i „asteroidy“ a zároveň neobsahují slovo „komety“.

Př.: 5

komety <near> kolize

Vyhledá dokumenty, které zároveň obsahují různé gramatické tvary slov „komety“ i „kolize“, a seřadí je podle textové vzdálenosti mezi těmito slovy.

Př. : 6

„meteorický roj“ or „padající hvězdy“ .

Vyhledá dokumenty, které obsahují různé gramatické tvary frází „meteorický roj nebo „padající hvězdy“.

Př.: 7

meteo*

Hvězdičková konvence: Vyhledá dokumenty, které obsahují slova začínající na „meteo“ (meteor, meteorický, apod.).

Př.: 8

*stvo

Hvězdičková konvence: Vyhledá dokumenty, které obsahují slova končící na „stvo“ (družstvo, mužstvo, apod.) .

Př.: 9

??běr

Otazníková konvence: Vyhledá dokumenty, které obsahují slova končící na „běr“ a sestávající z pěti znaků (výběr, záběr, apod.).

Návěští pro všechny formuláře (vyhledávání v polích) - Návěští zapisovaná v poli dotaz, resp. text

Př.: 10

Data

dat <contains> 8.10.2001

dat# 8.10.2001

dat=8.10.2001

Vyhledá všechny dokumenty vydané v tomto dni.

Pozn.: Datum vydání lze jednodušeji navolit v nabídce období od do ve všech formulářích. Zde je možno označit den či interval.

Vyhledá všechny dokumenty zpracované v tomto dni

Př.: 11

zdr=Respekt and naz=Rafinovaný odraz skutečnosti

src=Respekt and ti=Rafinovaný odraz skutečnosti

dc.source=Respekt and dc.title=Rafinovaný odraz skutečnosti

Vyhledávání podle obsahu polí: vyhledá dokumenty názvu zdroje (Respekt) s přesným názvem článku (Rafinovaný odraz skutečnosti). Návěští jsou v češtině nebo v konvenci Dublin Core nebo v angličtině. Text dotazu lze rozšířit o vyhledávání podle dalších polí (návěští viz Tabulka návěští dále). Na tento typ dotazu nemá vliv nastavení <contains> či = ve spodní části formuláře. Pro začátečníky je tento typ dotazu komplikovaný, doporučujeme vyhledávat přímo z rozšířeného formuláře, kde jsou implicitně pole nastavena. V tomto případě je třeba označit volbu <contains> či = v nabídce formuláře v spodní části rozšířeného formuláře.

Př. 12

zdr <contains>Respekt and naz <contains>odraz

src<contains>Respekt and ti<contains>odraz

dc.source<contains>Respekt and dc.title <contains>odraz

Operátor <contains> lze nahradit #:

zdr#Respekt and naz#odraz

src#Respekt and ti#odraz

dc.source#Respekt and dc.title#odraz

Vyhledá dokumenty z názvu obsahující slovo "Respekt" a z názvu článku obsahující slovo "odraz" .

Tabulka návěští pro pole dotaz a Prvky Dublin Core a Anl Core aplikované v bázi ANL FULL - viz Příloha F9

Využito 14 prvků DC, navíc přidán DC.Subtitle (vzhledem k minimálnímu záznamu pro ANL), 14 prvků AC (ANL CORE) definovaných pro vyhledávání a užívaných v popisu článku . Vzhledem k WWW TTDE a rozšíření formuláře pro editaci o některé prvky bude zřejmě potřeba rozšířit v budoucnu ANL Core.

Rovněž nepoužívejte návěští phnk, ale jeho alternativu.

Pozn.: při kombinaci návěští nemá na vyhledávání vliv nastavení <contains> a =.

Př: zdr=Lidové noviny

Lze vyhledávat podle tečkové konvence DC za použití operátoru contains nebo jeho náhrady #.

Př.:

dc.creator#Klaus

dc.creator <contains>klaus

Podle tečkové konvence lze vyhledávat údaje: dc.source, dc.date.issued, dc.date.modified, dc.title, dc.subtitle, dc.creator, dc.contributor, dc.publisher, dc.coverage.spatial, dc.coverage.temporal, dc.subject

Zvýraznění vyhledaných údajů pomocí dotazového pole

Údaje, které nejsou vyhledávány pomocí strukturovaných polí jsou zvýrazněny červeně v citaci článku i v textu článku.

Údaje, které jsou vyhledávány pomocí strukturovaných polí nejsou zvýrazněny červeně, jsou obsaženy v citaci článku.

Spíše než návštěví doporučujeme používat formulář s implicitně nastavenými poli.

Formulářová pole s implicitními údaji v rozšířených formulářích. Přehledná tabulka metodiky pro vyhledávání - Příloha F4, F 9

Jednotlivé údaje lze kombinovat pomocí operátorů **and, or, not**. **Ve spodní části obrazovky je možno nastavit vyhledávání <contains> - pro vyhledání daných řetězců kdekoli v poli, tj. po slovech, nebo = vyžadujeme-li přesné znění řetězce. Pro přesná znění je lépe využívat rejstříky. Všeobecně je lépe nastavit operátor <contains>.**

Údaje, které jsou vyhledávány pomocí implicitních polí, nejsou zvýrazněny v plném textu červeně.

Vyhledávání pomocí topiků - Rozšířený formulář s tématy - Příloha F4,F9

Tématem se rozumí předem připravený složitý dotaz. Jedno nebo více témat je možno vybrat označením v seznamu v pravém pruhu formuláře (maximálně 3 témata pomocí klávesy Ctrl). Témata lze kombinovat vzájemně mezi sebou i s dotazem v poli dotazu, resp. textu dotazu. Relace mezi tématy a poli je možné zaškrtnutím voleb příslušných voleb v pravé dolní části formuláře (volba and, or)

Tematický dotaz může obsahovat stovky slov. Je vytvářen speciálním editorem, který umožňuje vložit logické a další operátory dotazovacího jazyka, provádí kontrolu syntaxe a umožňuje proto plně využít výhod pokročilého vyhledávacího systému TOPIC. **Pro databázi ANL FULL jsou témata strukturována do tří úrovní: první obsahuje základní tematické oblasti (8), druhá skupiny témat v těchto oblastech (28), třetí obsahuje detailní témata ve skupinách témat (zatím 197). Z první úrovně nelze zatím vyhledávat, z druhé a třetí ano - tam, kde je topik nadefinován (označeno šipkou vlevo). Tematická nabídka je postupně rozšiřována.**

Topik slouží k tomu, že hledaný výraz vyplněný do textu dotazu a různě zpřesňovaný je navíc upřesněn topikem - tématem.

Přehled dosud nadefinovaných/navržených topiků strukturovaných do tří úrovní – Příloha F9

Rejstříky – Příloha F4

V rejstřících je možno listovat a vybrat výraz, podle kterého chceme vyhledávat. **Tyto výrazy jsou hypertextově propojeny v citacích s metadaty, resp. údaji bibliografického popisu. Podle rejstříků doporučujeme vyhledávat údaje zejména týkající se předmětu dokumentu (předmětové kategorie, hesla - věcná a geografická, osoby, korporace, akce, klíčová slova, dokument/dílo).** Vyhledané výrazy jsou v bibliografickém popisu, resp. metadatech, resp. citaci vyznačeny červeně.

Obsah rejstříku se zobrazuje v pravém okně, v levém okně se zobrazují vyhledané články.

Bylo nadefinováno **17 rejstříků**:

autoři, název zdroje, zdroj ročník číslo, datum vydání, předmětové kategorie, předmět, místo, hesla (věcná a geografická), osoby, korporace, akce, klíčová slova, dokument/dílo, typ článku, MDT, geografické vymezení, časové vymezení.

Některá doporučení, jak nejlépe vyhledávat

- Nejjednodušší vyhledávání je z pole dotaz pomocí jakéhokoli slova v kombinaci datem vydání
- Údaje obsažené v polích bibliografického popisu doporučujeme vyhledávat spíše pomocí implicitních polí v rozšířených formulářů než vypisováním návěští v poli dotaz resp. text dotazu
- V implicitních polích rozšířených formulářů vyhledávat pomocí nastavení operátoru <contains> či # než pomocí =
- Přesná znění řetězců vyhledávat spíše pomocí rejstříků
- Při vyhledávání autorů doporučujeme vyhledávat z nabídky všichni autoři v rámci nabídky autor,název v rozšířeném formuláři
- Ukončovat jednotlivé akce (nenechávat více otevřených formulářů více najednou, též výsledků vyhledávání)

Výsledky vyhledávání, zobrazení, tisk (všechny formuláře) - Příloha F4

Seznam výsledků (seznam článků)

Seznam vyhledaných dokumentů uvádí v záhlaví nadpis a údaje o výsledku hledání. Počet vyhledaných dokumentů na stránku je dán volbou v poli Výsl./str. V závislosti na této volbě se potom zobrazuje počet stran s možností listování.

Druhy seznamu výsledků (seznam článků):

U jednotlivých článků zobrazeno vždy, resp. standardně (zleva): tři formáty pro zobrazení údajů o článku, skóre relevance, datum vydání, název článku, velikost plného textu.

Seznam výsledků (názvů vyhledaných článků):

- **jednoduchý** (standardní zobrazení)
- **se souhrnem** (standardní zobrazení plus souhrn)
- **seskupený** (standardní zobrazení plus články uspořádané do skupin podle shody v klíčových slovech **podle společného kontextu, clustering**) - v současné verzi TOPICu **zprovožňována** - viz též výše

Třídění seznamu výsledků (názvů vyhledaných článků)

Skóre (relevance, od které zobrazovat názvy článků)

Výsl./str. (počet článků na stranu)

Třídění: dle skóre (relevance)

dle názvu

zdroje

data (vydání)

stran


a to sestupně nebo vzestupně.

Zobrazení údajů o článku - Příloha F4

Formáty zobrazení

Uživatelské formáty

Každý dokument je možné zobrazit ve **třech uživatelských formátech/variantách:**

 **Citace článku, souhrn - bibliografické údaje o článku, resp. metadata s odkazy pro další hledání**

 **Citace článku, plný text - bibliografické údaje o článku, resp. metadata s odkazy pro další hledání a plný text článku**

 **DC/HTML – Dublin Core Metadata**

Název článku hypertextově aktivní – zobrazení citace a plného textu bez odkazů

Hledané výrazy se zvýrazní červeně pouze vyhledávám-li z dotazového pole, podle topiků a rejstříků, nikoli podle nadefinovaných, resp. implicitních polí.

Pracovní formáty

V záhlaví plného textu jsou navíc různé pracovní formáty - **klíčová slova**, **UNIMARC** (komunikativní formát), **Dublin Core** – formát pro zpřístupňování elektronických dokumentů a **jeho aplikace**, která obsahují metadata (**DC/HTML/META**, **DC/XHTML kvalifikovaný a nekvalifikovaný**, **DC/XML kvalifikovaný a nekvalifikovaný**, které nejsou určené pro běžného uživatele, ale pro budoucí vývoj systému a zpřístupnění plných textů.

Metadata - Příloha F9

Ve všech variantách zobrazení je obsažena citace, resp. bibliografické údaje, resp. metadata.

Příklady uložených metadat v plném textu

Př.1:

Název:	Cesta mezi hlavou a rukou
Podnázev:	Když některé věci nenapíšu, nikdy se je nedozvím, říká publicista a spisovatel Pavel Kosatík
Hlavní autor:	Pavel Kosatík
Další autor:	Karel Hvízdala
Zdroj:	Mladá fronta Dnes
Zdroj-přl.:	Ekonomika
ISSN:	1210-1168
Roč.	12, č. 204 (1.9.2001), s. C/5
Rubrika:	Kultura - Pohledy
Předmět. ktg.:	politika: politici
	literatura: česká literatura
	hromadné sdělovací prostředky: novináři
MDT:	323-051, 070-051, 821.162.3-051
Osoba jako předmět:	Masaryk, Jan, 1886-1948
	Peroutka, Ferdinand, 1895-1978
	Kohout, Pavel, 1928-
Téma jako předmět:	politici-Československo-stol. 20.
	novináři-Československo-stol. 20.
	spisovatelé-Československo-stol. 20.
Typ dokumentu:	rozhovory

Př. 2:

Název:	Rafinovaný odraz skutečnosti
---------------	------------------------------

Podnázev:	Na pultech se objevil další titul singerovské řady
Hlavní autor:	Hana Ulmanová
Zdroj:	Respekt
ISSN:	0862-6545
Roč.	12, č. 41 (8.10.2001), s. 23
Rubrika:	KULTURA
Předmět. ktg.:	literatura: americká literatura
MDT:	821.111(73)-31, (070.447)
Osoba jako předmět:	Singer, Isaac Bashevis, 1904-1991
Dílo jako předmět:	Stíny nad Hudsonem (kniha)
Téma jako předmět:	anglicky psaná literatura
	americká próza
	spisovatelé-Spojené státy-stol. 20.
Typ dokumentu:	recenze

Tisky

Tisk seznamu výsledků – z nabídky Tisk seznamu, tisk plného textu – z nabídky Tisk detailu.
Tisk, stahování a export plných textů a metadat - přístup z nabídky Vyhledávání - Příloha F9

Aplikace pro stahování a export metadat a plných textů, která vychází z již vyvinuté technologie v Anopressu. Aplikace umožňuje stáhnout a export příslušných plných textů a metadat na lokální počítač v zip souboru v několika formách: HTML, RTF formát, textový tvar, UNICODE - vše se zvýrazněním klíčových slov a bez, XML formát, HTML formát se souhrnem. Je možno zobrazit detailní průběh zpracování. Exportovat plné texty a metadatum mohou zatím jen registrovaní vybraní uživatelé.

Postup:

1. Vyhledané články nejprve označit v Seznamu výsledků (levé okno)
2. Odrážka „Stáhnout označené“
3. Volba formátu výstupního souboru“
4. Odrážka „Připravit texty“
5. Odrážka „Stáhnout soubor“
6. Soubor otevřít nebo uložit

Administrace - Příloha F 4

Aplikace pro administraci databáze ANL FULL a portálu.

Úpravy textů článků – aplikace umožňuje opravovat zdrojové kódy článků (tj. i metadatum) na základě vyhledání podle identifikačních čísel. Umožňuje opravy rejstřík bez přímé vazby na záznamy. Opravené záznamy jsou uloženy do souboru a dále naimportovány do báze a zaindexovány.

Aplikace umožňuje provádět opravy a sledovat statistiky – zahrnuje tyto úkony: **úpravy textu článku (metadat aj.), údržba rejstříků, údržba portálu Periodika na WWW, údržba informačních zdrojů.** Aplikace umožňuje **podobně udržovat seznamy uživatelů, hesel, mailů a sledovat statistiku přístupů.**

Zcela nová aplikace pro opravy hesel a metadat v plných textech, která je zabudovaná přímo do linky WWW TTDE (viz též výše). Aplikace umožňuje vyhledání plných textů z rejstříků, jejich

zobrazení a zatím částečné načtení do pracovního listu k opravě. Po korekci metadat je plný text odeslán do báze, plný text je přepsán a zrušen. Aplikace je ve stádiu experimentu.

B.1.4.2 Kooperační systém článkové bibliografie standardizace, aplikace pro MNG KOSABI

B.1.4.2.1 Kooperační systém článkové bibliografie

V roce 2002 se poměrně hodně času věnovalo projektu **krajské knihovny v Ostravě** týkající se popisu článků v systému T-Series a jejich zpřístupnění. Byly specifikovány požadavky na konverzní program z T-Series do UNIMARCu, zejména v intencích minimálního záznam a požadavky na věcný popis. Řešitelka vypracovala připomínky začátkem roku 2002 a zaslala Mgr. Kybalovi (UK), který pracuje na konverzi do UNIMARCu.

Byly řešena také problematika řízeného slovníku pro věcný popis. V kooperačním systému článkové bibliografie se používal ještě za dob ISISu tzv. částečně řízený slovník klíčových slov, tj. byly řízeny nadřazené pojmy a většinou první podřazené slovo k nim. Kromě toho se klíčová slova vyvinula z předmětových hesel a zachovávala částečně i jejich strukturu, tj. bylo obsaženo téma, chronologický a geografický údaj. Pro osoby a akce, forma dokumentu byla zvláštní pole.

Řízená klíčová slova, tj. nadřazený a jemu podřazený pojem byly konvertovány do UNIMARCu a tvoří v UNIMARCu tzv. předmětové kategorie v poli 615.

Předmětové kategorie používají všechny knihovny kromě těch, které pracují v Tinlibu/T-Series. Slovník se musí v samotné bázi ANL zredigovat. Pro UNIMARC jsou zapisovány v poli 615, k odlišení od Konspektu s 2. indikátorem 9. V budoucnu by mohly dobře konspekt doplňovat.

Poměrně problematická situace vznikla v krajské knihovně v Českých Budějovicích,

V roce 1998 jsme minulosti navrhovali v zásadě dva způsoby, jak diferencovat zápis údajů věcném popisu tak, aby byly lépe převoditelné do UNIMARCu. Zápis příslušných údajů duplicitně do polí pro konverze nebo zápis čísel polí UNIMARCu přímo k údajům v pracovním listu. Takto by byly údaje připravené pro konverze do příslušných polí v UNIMARCu. Žádný z těchto způsobů zápisu však nebyl u článků v praxi realizován.

Mezitím se věcný popis článků vyvíjel v Českých Budějovicích tak, aby byl vstřícný pro uživatele i katalogizátory. K věcnému popisu byla používána pole pracovního listu – předmětové skupiny a výrazy tezaurů, v posledním období pouze pole předmětové kategorie v terminologii Tinlibu, resp. T-Series (něco jiného jsou předmětové kategorie UNIMARCu, pole 615). V roce 2002 se přistupuje k popisu do jemněji strukturovaného formuláře, podobně jako v Ostravě.

V roce 2002 šlo o rozhodnutí, jak pokračovat v zápisu článků v Českých Budějovicích. Řešitelka projektu Souborná databáze navrhla několik variant věcného popisu v Českých Budějovicích. Pracovnice krajské knihovny se překlony k následujícímu řešení: používat výrazy tezauru, který je aplikován v českobudějovické bázi pro knihy plus příslušná pole pro osoby, korporace, geografickou entitu. Pro své účely dále používat dosavadní systém s vazbami.

V Ostravě je situace jednodušší – do léta 2001 byly články zpracovávány v ISISu, koncem roku 2001 přechod na jemněji strukturovaný formulář T-Series.

V roce 2002 na **T-Series přešla i krajská knihovna v Ústí nad Labem.**

V červnu 2002 přešlo oddělení analytického zpracování v NKČR na verzi ALEPH 14.2.4, která umožňuje větší komfort pro opravy uložených dat. Na toto verzi přešly též v r. 2002 i knihovny v Brně a Olomouci. Pracovníci MZK ukládají záznamy přímo do báze ANL.

V r. 2003-2004 se ladila **struktura záznamů vzhledem k formátu UNIMARC ve spolupráci s krajskými knihovnami v Pardubicích a Zlíně (KP-Sys), v Karlových Varech a Havlíčkově Brodě (Lanius, resp. Clavius).** Zároveň byla delimitována excerpční základna vzhledem k novému státoprávnímu uspořádání a vstupu těchto knihoven do reálné kooperace. Podobně tomu tak bylo i u

Národního filmového archivu. **Tyto knihovny jsou de facto připraveny na reálnou spolupráci v r. 2005.**

V r. 2003 záznamy ANL respektují metodický materiál **Záznam pro soubornou databázi : UNIMARC a Záznam pro soubornou databázi : Výměnný formát.** Dle možností je **aktualizována** pracovní verze příručky pro zpracování článků v UNIMARCu - **Metodika popisu článků ve formátu UNIMARC a vystavena na Internetu s názornými příklady.** Většina knihoven vybavuje záznamy předměťovými kategoriemi, které jsou jednotícím prvkem souborné databáze. Oddělení analytického zpracování přistupuje k **intenzivnějšímu využívání souborů autorit v NK , korekturám báze ANL a připravuje data pro soubory jmenných a věcných autorit. Také kooperující instituce postupně používají národní soubory autorit.**

V rámci ANL i ANL FULL jsou formulovány základní principy věcného popisu, aplikovaného při zpracování článků v NKČR. **Pořadí kategorií, předměťových hesel odpovídá pořadí MDT.** MDT vztahující se k osobám a formě článku se uvádí na konec věcného popisu (osoby, forma). **Kategorie:** obecné zařazení tematiky dokumentu pod obecné hlavní téma (615a) a zpřesnění (615x). **Předmětové heslo:** vystihuje hlavní téma dokumentu. **Klíčová slova:** variantní selekční termíny k předmětovým heslům uvádíme zatím z důvodů postupného propojování báze ANL na soubory autorit, v současné době se eliminují - s postupným provázáním věcného popisu na autority i s ohledem ke zpřístupnění plných textů v bázi ANL prostřednictvím báze ANL FULL.

Ve většině **kooperujících institucí** se používá kombinace těchto tří prvků věcného popisu v různé míře. Předmětová hesla se aplikují pro osoby, korporace, typ dokumentu, geografické téma. Je třeba důsledněji používat pole 660 Geografický kód.

V r. 2004 probíhaly v NKČR práce na **konverzi UNIMARC M21**, knihovnám pracujícím v ALEPHu byly poskytnuty **šablony pro katalogizaci článků.** Skupina pro analytické zpracování schválila minimální záznam pro soubornou bibliografickou databázi MARC 21. V listopadu a prosinci byl odzkoušen a **veden do provozu upload záznamů** do kooperačního systému (viz dříve). Byl **vypracována pravidla popisu ve formátu M21** s respektováním AACR2R. Veškeré materiály jsou vystaveny na WWW.

Předmětové kategorie používají všechny knihovny kromě těch, které pracují v Tinlibu/T-Series. Slovník se musí v samotné bázi ANL zredigovat. Pro UNIMARC jsou zapisovány v poli 615, k odlišení od Konspektu s 2. indikátorem 9. V budoucnu by mohly dobře konspekt doplňovat.

Byl vypracován návrh postupné **aplikace konspektu v knihovnách kooperačního systému článkové bibliografie.**

Koncem r. 2004 se začal aplikovat v krajské knihovně konspekt. Při uploadu do kooperačního systému bylo zjištěno, že je třeba export pole vyladit. Konspekt se začal používat také v NKČR. Bylo

Údaje jmenného a věcného popisu jsou podle možností v bázi ANL revidovány a opravovány pomocí globálních a hromadných oprav, dále pak s využitím jmenných a věcných autorit.

B.1.4.2.2 Standardizace – shrnutí a perspektivy

Výstup z linky TTDE a ALEPHu respektuje formát UNIMARC, M21 a knihovnická pravidla AACR2 v oblasti jmenného popisu. Jmenný popis bude muset do budoucna reflektovat pohyb v oblasti standardizace (metodika popisu článků ve formátu UNIMARC - vychází z AACR2R, ISBD /CP; v budoucnu větší harmonizace AACR2R a ISBD, ISBD (ER), ISBD(S) - ISBD (CR)). Jak se dotkne popisu článků koncept FRBR – Funkční požadavky na bibliografické záznamy (**Functional Requirements for Bibliographic Records**) - 4 úroňový model manifestace díla - dílo, exprese, manifestace, jednotka) ukáže budoucnost. V bázi ANL FULL je zohledněna interpunkce ISBD jen částečně.

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 eventuelně s automatizovanou tvorbou kategorií. Předmětové kategorie do jisté míry konvenují **metodě konspektu** aplikované pro popis a mapování fondů. Od konspektu jsou odlišeny identifikátorem v rámci stejného pole 615. Dále se používají **klíčová slova**, která jsou dále částečně řízená a **předmětová hesla**. V rámci linky TTDE a ALEPHu je zajištěno také **propojení na soubory autorit** (ne zatím zcela pro katalogizátora pohodlné, pracuje se na propojení on-line - viz dříve) Automaticky je generován **souhrn článku**, automaticky lze také generovat klíčová slova, která však mají v současném stádiu "počítačovou formu". Koncem r. 2004 se začal používat v bázi ANL **konspekt**. Ve formuláři TTDE je pro konspekt také připraveno pole, konverze jsou ošetřeny. Není však zatím možné napojení na autority (neznámé atributy pro nastavení).

V záznamech určených k propojení s plným textem se v lince automaticky **generuje dynamická URL adresa, SICI**. Do zpracovávaných článků v ALEPHu jsou doplňované **statické adresy** volně dostupných dokumentů a WWW. Dále se automaticky generuje **URN** (do kterého je zabudováno identifikační číslo plného textu přidělené v Anopressu) a z velké části pole **LKR** určené k propojení se zdrojovým dokumentem.

Dublin Core používaný v lince TTDE obsahuje 14 z 15 definovaných údajů viz **Příloha F 9**. Je generován pro formát HTML, XHTML a XML ve kvalifikované i nekvalifikované formě. Do HTML je zabudován **LINK tag** pro potřeby odkazu na webový zdroj, v němž se nachází specifikace daného použitého soboru metadat. Bylo nadefinováno 14 údajů **Anl Core** vzhledem k detailnosti popisu a vzhledem k možnostem vyhledávání v současné verzi systému TOPIC a prvky ANL Core užívané pro vyhledávání.

B.1.4.2.3 Management Kooperačního systému článkové bibliografie, aplikace MNG KOSABI - Příloha F5

Management KOSABI se dělí na správu a údržbu: **bibliografické báze ANL(ALEPH) a serveru <http://anl.nkp.cz> (správy pomocné báze SKA pro bázi ANL, vedení koordinované excerpční základny systému - popsána dříve a metodické činnosti - popsány dříve, zálohování serveru anl.nkp.cz; serveru full.nkp.cz; plnotextové báze ANL FULL a serveru <http://full.nkp.cz> (popsána dříve).**

Perspektivy serveru <http://full.nkp.cz>

Server <http://full.nkp.cz> je v letech 2003-2004 je ve stádiu rutinního provozu. Báze ANL FULL má být dle rozhodnutí vedení NK určena k archivním účelům. Server full.nkp.cz má být převeden pro společnou doménu s ostatními NT servery NKČR, o jeho správu se má dělit fy Comdat z hlediska systému, a ing. Mattern (Anopress) z hlediska správy aplikace.

Server byl během řešení projektu jištěn na magnetických páskách, DVD, na serveru anl.nkp.cz jako zrcadlo k serveru full.nkp.cz. **Koncem r. 2004 je pravidelně zálohována na serveru NKČR.**

Aplikace pro správu a údržbu báze ANL je ve stádiu provozu koncem r. 2004 vzhledem ke konverzi M21, vzhledem k náročnému ladění různých importů a uploadu do KOSABI, nutností vytvořit nové formáty pro příjem dat a kontroly dat exportovaných z Clavia a pro M21.

V rámci MNG KOSABI je prováděna i správa serveru anl.nkp.cz (Linux, Oracle), na kterém běží aplikace MNG KOSABI a na které je umístěna báze SKA (báze, do které se importují – uploadují data od spolupracujících institucí).

Server anl.nkp.cz (Linux, Oracle) je pravidelně zálohován a jsou zde zálohy báze ANL FULL jako zrcadlo. Data a programy umístěné na serveru pro MNG KOSABI jsou jištěny pravidelně na CD. MNG KOSABI je názorně popsán v příloze F5.

Na základě zadání pro řízení a správu kooperačního systému a za využití již vyvinutých řešení v rámci Souborného katalogu CASLIN proběhly práce na vývoji aplikace pro KOSABI na serveru ANL(systém **LINUX a ORACLE**).

Ve stádiu provozu je aplikace pro evidenci a aktualizaci (dialogové zpracování báze) excerptní základny systému.

Ve stádiu provozu je aplikace pro příjem, automatizované zpracování a export dat do ALEPHu (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 a M21 pro analytické záznamy, test na kritické chyby, test na duplicitu klíčů).

Ve stádiu provozu je aplikace na kontrolu platných URL adres.

A. Báze Titulů – aplikace pro evidenci a aktualizaci excerptní základny systému - Příloha F6

Excerptní základna je evidována: v aplikaci Báze Titulů (aktuálně zpracovávané seriály), dále pak v samostatném soupisu aktuálně excerptovaných seriálů ve formátu .doc a titulů sborníkového charakteru ve formátu doc.

Aplikace pro evidenci zpracovávaných titulů – Báze Titulů (doplnění báze základními daty excerptní základny do nadefinovaných polí, které odpovídají UNIMARCu)

V bázi excerptní základny jsou obsaženy tituly všech institucí, které kooperují v systému článkové bibliografie. Tituly lze třídit podle zpracovávající instituce, podle abecedy, podle toho, zda jsou vybrané články titulů zpřístupňovány plnotextově, podle kategorie titulu. Pro jednotlivé tituly je definován krátký záznam dle polí UNIMARCu.

Formát SHORT pro bázi titulů

Název (200a)

Označení části (200h)

Název části (200i)

Obsahuje plný text (300a)

Dostupnost plného textu (310a)

Zdroj katalogizace - zpracovatelská instituce (800b)

Excerpt pro ANL (980a)

Excerpt pro vlastní potřeby (980b)

Kategorie titulu (980k)

Aplikace pro aktualizaci Báze titulů

(aplikace v Jdeveolperu, Java)

Tituly lze vyhledávat podle instituce i abecedně, aktualizovat jednotlivé záznamy ve stromové struktuře.

B. Aplikace pro příjem a zpracování dat v bázi SKA, export dat do báze ANL

1. Příjem dat pro systém

- způsoby příjmu dat **upload**, mail

Byla vypracována detailní nápověda pro užívání uploadu, pro každou instituci zvoleno heslo a jméno, po vyplnění se automaticky navolí nejběžněji používaný kód a formát a přístup do statistik souborů každé instituce.

Konvence označování souborů dodavatelů dat pro import do báze ANL

Nejčastěji formát a kód používaný institucí je implicitně nastavený, lze jej však změnit podle potřeby.

Struktura: ssssssKK.FFF ssssss sigla KK kódování FFF formát vstup. souboru
aba009kg.vfi

Zkratky pro kódování

- */ UM /* ISO 646 nebo ISO 5426
- */ GI /* GIZMO
- */ LG /* PC Latin 2 + GIZMO
- */ KG /* kód Kamen. + GIZMO
- */ UC /* UNICODE UTF 8
- */ SG /* ISO 8859-2 + GIZMO
- */ KA /* kód Kamen.
- */ AN /* ANSEL
- */ WN/* Windows 1250

Formát vstupního souboru

- */ DAT /* export ALEPH 300
- */ RUM /* řádkový UNIMARC
- */ UIS /* UNIMARC ISO 2709
- */ VFO /* vým.formát ISO 2709
- */ VFI /* vým.formát export soubor CDS/ISIS
- */ DTT /* export ALEPH 500
- */ CLA/* formát Clavius
- */ M21/* MARC21

- při importu se u každého záznamu vytvoří klíč na duplicitu záznamů (dále uložení klíčů v bázi a následně zobrazení klíče s počtem záznamů) – tzv. test na duplicitu klíčů (mezi základní klíče patří název článku, autor, název zdrojového dokumentu, datum vydání.)

2. Zpracování přijímaných dat

- přidělení identifikace pro zpracovávanou dávku (batch_id)
- formát a kódování dle názvových konvencí
- konverze do UNIMARC (u formátu, které to vyžadují)
- převod do interní struktury SQLLCB
- globální úpravy záznamů - 1. etapa (chyby v datech, které lze programově ošetřit, např. tvrdé mezery, tagy ve vazebních polích)
- test UNIMARC
- test na MARC 21
- uložení do baze SKA
- statistika zpracování (test na UNIMARC a M21 doplněn o test na kritické chyby a test na UNIMARC, podrobné statistiky chyb)

3. Export dat z baze SKA do adresáře určeného k opravám nebo do adresáře určeného pro ALEPH

- záznamy pro export určeny dle batch_id
- globální úpravy záznamu Etapa II. (oprava specifických "chyb" v dávkách od jednotlivých knihoven)
- konverze do kódování UTF-8 a formátu Aleph500
- uložení výstupního souboru do příslušného adresáře pro import do ALEPHu nebo do adresáře pro opravy
- statistika zpracování

K výše uvedeným funkcím slouží mj. tyto moduly:

conva - překódování
testuni - test na UNIMARC
logpst - statistiky
locvoc - modul pro práci se slovníkem
un2aleph - modul pro konverzi UNIMARCu do ALEPHu
crrr - test na kritické chyby

C. Test na kontrolu platnosti URL adres

Test se provádí na zadání administrátora . Testuje se na základě aktuálního souboru záznamů z báze ANL obsahujících pole 856.

Možné kombinace testu: testovat adresy vyjma báze ANL FULL nebo adresy ANL FULL, testovat ok adresy nebo testovat ERR adresy.

Na základě testu vznikají podrobné statistiky s výpisem chyb.

D. Dokumentace

Na serveru anl.nkp.cz je k jednotlivým aplikacím k dispozici dokumentace.

B.1.4.3 Česká národní bibliografie - řada Články v českých novinách, časopisech a sbornících na CD-ROM - Příloha F1

V letech roce 2000-2004 byla pravidelně poskytována data KOSABI pro měsíční aktualizace a **CD-ROM ČNB - řada Články v českých novinách , časopisech a sbornících. V r. 2004 proběhly práce na konverzi dat do M21.**

B.1.4.4 Akce

Kromě průběžných a pravidelných konzultací proběhly v **r. 2002** tyto akce: prezentace báze ANL FULL a KOSABI na **Inforu 2002**, prezentace KOSABI na konferenci **Knihovny současnosti 2002**.

Kromě průběžných a pravidelných konzultací proběhly v **r. 2003** tyto akce: prezentace báze ANL, ANL FULL a KOSABI na **Infos 2003**, [prezentace KOSABI](#) na zasedání **SDRUK**, sekce pro bibliografii v Hradci Králové. Řešitelka se pravidelně účastní bibliografických porad každoročně konaných v **STK**. Řešitelka se zúčastnila zasedání **IFLY 2003 v Berlíně**. Sekci pro noviny byly poskytnuty informace o KOSABI, ANL, ANL FULL.

B.1.4.5 HW a SW podpora pro správu KOSABI a pro bázi ANL FULL - vývoj

V r. 2000 byl management kooperačního systému zakoupen PC Pentium III, 700 MhZ, ORACLE 8i server. Release 8.1.5 (5 licencí).

Pro správu a údržbu plnotextové databáze ANL FULL byl v r. 2000 zakoupen server DELL - PowerEdge 6300 - Pentium III Xeon 500Mhz/512, Search Verity Information Server (TOPIC) v. 3.6 pro jednoprosesorový server Windows NT zatím pro ultranet (30 licencí), Windows NT v. 4.0. (server full.nkp.cz). V tomto roce byly analyzovány možnosti nové verze TOPICu - Portal ONE SE.

V r. 2002 byl proveden upgrade a update serveru ANL FULL. Instalace Windows 2000, instalace nové verze TOPICu v. 3.7 - Portal One SE. Byla zvýšena operační paměť a kapacita disku (na 1 GB RAM, 68 GB disk rozdělený na 3).

V r. 2002 byl proveden upgrade serveru ANL: operační systém Linux SuSE 7.3, verze Oracle 9.2.

V r. 2004 byl vyměněn jeden 18 GB disk.

V návrzích projektů v rámci 1 N je uvedena migrace TOPICu na platformu K2, implementace Z39.50 a OpenUrl pro zpřístupnění v JIB aj.

B.1.4.6 Smluvní ošetření projektu

Pro automatickou indexaci, správu (údržbu) plnotextové databáze a dodávku plných textů byly uzavřeny dvě smlouvy: s ing. I. Maternem a Anopressem dne 15.6.2000. Smlouva na vývoj aplikace pro management kooperačního systému s ing. Koktanem byla uzavřena dne 26.9.2000.

Dále zůstává v platnosti v r. 2001 Dodatek č. 3 ke Smlouvě o sdružení pro Českou národní bibliografii, který zabezpečuje fungování Kooperačního systému článkové bibliografie v situaci reformy státní správy.

Anopress, v současné době Anopress IT, a.s., plní všechny své závazky vůči projektům.

Byl učiněny 3 dodatky k základní smlouvě o projektu mezi NKČR a MKČR týkající se přesunu financí.

Dále zůstává v platnosti v r. 2001 Dodatek č. 3 ke Smlouvě o sdružení pro Českou národní bibliografii, který zabezpečuje fungování Kooperačního systému článkové bibliografie v situaci reformy státní správy.

Kooperace s odbornými knihovnami probíhá na základě smluv za symbolickou 1 Kč.

Je třeba nové smluvní ošetření Kooperačního systému článkové bibliografie a jeho vymezení v JIB i vzhledem k novým podpisovým pravidlům smluv v NKČR.

B.1.4.7 Integrace heterogenních dat a perspektivy Kooperačního systému článkové bibliografie

Organizace spolupráce

- Postupně reálně začlenit nově konstituované krajské knihovny v Pardubicích, Karlových Varech, Havlíčkově Brodě a Zlíně do KOSABI.
- Postupně sladit profil zpracovávaných titulů v KOSABI se Souborným katalogem článků systému LANius, resp. Clavius v případě oboustranného zájmu.
- Spolupráce s institucemi pedagogického a vědeckého zaměření (UK, ČSAV aj.) – zjistit možnosti.
- Spolupráce se společností Anopress a event. navázání spolupráce s dalšími nakladateli/vydavateli
- V r. 2003 jednáno předběžně s nakladatelstvím **Economia**, v rámci excerptní základny je možno určit skupinu titulů z hlediska této spolupráce. Je možné uvažovat o spolupráci s vydavateli titulů, které jsou dotovány MKČR. Pro návrh projektu rámci 1N MŠMT byly poskytnuty některé plné texty z oblasti zemědělství, ekonomie, náboženství, lékařství.
- Začlenění článků do oborových informačních bran

Metody zpracování, standardizace

- Doladit v případě potřeby převod výstupů z některých systémů do standardního UNIMARCu a báze ANL, v r. 2004 bylo v tomto směru děláno maximum
- Další harmonizace popisu článků v rámci KOSABI se soubory autorit zejména v oblasti věcného zpracování (předmětová hesla, předmětové kategorie a metoda konspektu).
- Korektury báze ANL, zejména zdrojových dokumentů.
- MARC 21 (ladit), pravidla pro popis článků vzhledem k tomuto formátu.
- Případně otestovat linku TTDE v instituci (ích) spolupracující (ích) v KOSABI.
- V NKČR se linka TTDE pro texty Anopress dle rozhodnutí vedení dále nebude používat, linka je v r. 2005 převedena pod Aleph. Budou se pravděpodobně aplikovat metody poloautomatického zpracování v rámci ALEPH/JIB.
- V případě zájmu by bylo možné linku TTDE použít při spolupráci s nakladateli event. vydavateli v budoucnu.
- Testovat metody, popř. navázat na metody získávání a zpracování elektronických plných textů (projekt Webarchiv).
- Sledovat vývoj harmonizace AACR2R, ISBD,FRBR, aktualizované a nové standardy
- **Postupně přecházet z tradičního zpracování článků obsažených v určených titulech na zpracování pomocí nových technologických metod. Přejít však musí být uvážlivý vzhledem k zachování kontinuity zpracování titulů.**

Organizace zpracování článků vzhledem k excerptovaným titulům z hlediska systémového

Do budoucna je možné uvažovat o několika subsystémech zpracování z hlediska typů titulů:

I. Zpracování sborníků a nepravých periodik aj. odborných seriálových informačních zdrojů (v rámci integrovaných knihovnických systémů), které zatím nejsou dostupné na Internetu.

II. Zpracování odborných seriálových informačních zdrojů, které jsou dostupné na Internetu v lince automatické extrakce/indexace i v návaznosti na harvesting.

III. Zpracování článků ze seriálů na základě spolupráce z vydavateli, nakladateli.

IV. Zpracování plných textů volně přístupných na www na základě automatického sběru dat event. metod vyvíjených v rámci projektu Webarchiv?

Dodržování zásad výběru článků.

Organizace zpracování článků z hlediska typů institucí:

Vzájemná kooperace mezi následujícími sítěmi knihoven:

- NK, krajské knihovny, základní knihovny
- Specializované knihovny
- Vysokoškolské knihovny
- Vědecké instituce.

Legislativně právní otázky zpřístupnění plných textů

Vyřešení autorsko právních aspektů zpřístupnění plných textů s příslušnými nakladateli/vydavateli a distributory a dohody s těmito subjekty za situace, kdy neexistuje v ČR právo povinného výtisku elektronických dokumentů, je značně problematické. Koncem r. 2004 byla podepsána smlouva mezi Anopressem a NKČR týkající se zpřístupnění báze ANL FULL na tři měsíce (uživatelé NKČR mohou plné texty prohlížet, tisknout a kopírovat, nikoli hromadně stahovat, externí uživatelé mají tytéž možnosti pouze po registraci a přihlášení na 1 den). Otázky zpřístupnění ANL FULL v rámci konzorcia by mělo být předmětem jednání. Je potřeba jasně vymezit a ošetřit přístup k plným textům, které jsou zpřístupňovány volně, přes konzorcia, interním a externím uživatelům institucí a způsob plateb (paušál, kredity, apod.).

Vzhledem k tomu, že v bázi ANL FULL jsou obsaženy články z příslušných zdrojů výběrově, domnívám se, že dohoda s příslušnými nakladateli/vydavateli by byla možná.

Dosavadní zpřístupnění báze ANL FULL – smlouva mezi NK a Anopressem je uzavřena na první tři měsíce r. 2005 (čtenáři NK k dispozici plné texty, externí uživatelé plné texty jen na 1 den).

Personální a finanční zajištění KOSABI

- KOSABI se již tradičně potýká s problémem personálního zajištění činností. Dobrá personální situace je podmínkou převedení systému do nového prostředí a pro aplikaci novým metod. Tyto metody pomohou práci zrychlit a zkvalitnit za menšího počtu pracovníků. Metody je však dobré nejdříve vyvinout a dobře otestovat při současném zachování kontinuity zpracování a zpřístupnění. V posledních dvou letech probíhají silné redukce oddělení analytického zpracování. V době řešení projektu, v letech 2000 – 2002 je oddělení redukováno o 5 pracovníků. Redukce byly plánovány na cílový počet 11 pracovníků v r. 2003. Zatím poslední redukce místa neproběhla, současný stav pracovníků je 12.
- Vývoj systému je po finanční stránce zajištěn do roku 2004 díky podpoře MKČR. V r. 2003 a 2004 byly podány dva projekty v rámci programu je podán návrh projektu v rámci programu 1N (neúspěšně). Analytické zpracování z pohledu ALEPHu je částečně obsaženo ve výzkumném záměru
Budování vzájemně kompatibilních informačních systémů.... . Dále jsou finančně podporovány projekty týkající se oborových informačních bran.

B.1.5 Plnění úkolů v jednotlivých letech

Plnění úkolů bylo možné vzhledem k poskytnutým finančním prostředkům jak v rámci samotného projektu Souborná databáze Kooperačního systému článkové bibliografie ... , tak vzhledem k paralelně běžícímu koncepčnímu záměru Propojení analytických záznamů s plnými texty, ze kterého programový projekt vychází.

Ve všech letech byl zajištěn praktický chod Kooperačního systému článkové bibliografie a průběžné plnění báze ANL, ANL FULL bez skluzů při zavádění nových technologií. Byl průběžně aktualizována excerpční základna, portál volných www periodik a vydávána ČNB.

Ve všech letech plnění byl zajištěn přísun plných textů a jejich propojení.

V letech 2000-2001, 1. pololetí byly texty, stahovány a off-line propojovány s připravenými záznamy k propojení (do záznamů doplněno pracovník ID, společná adresa doplněna globálně - tzv. simulovaná linka).

V letech 2000-2001 dále byly plné texty nakupovány a doplňovány off-line, programově propojovány s bibliografickými záznamy.

V letech 2001, 2. pololetí – 2004 byl zajištěn přístup do databáze TamTam v rámci 6-ti licencí a propojení je tvořeno automaticky.

V r. 2004, říjen-prosinec byly tyto licence hrazeny z nově přijatého projektu VaV v rámci programu 1N **Budování vzájemně kompatibilních informačních systémů** pro přístup k heterogenním informačním zdrojům a jejich zastřešení prostřednictvím Jednotné informační brány (hlavní řešitel B. Stoklasová).

Dále propojovány bibliografické záznamy ANL s **www texty** (relativně stálými) na **www** – v tomto projektu malý počet - cca 3000 (zejména však v projektu Propojení analytických záznamů a Budování vzájemně kompatibilních informačních systémů ..).

Báze ANL FULL byly také doplněna off-line digitalizovanými texty periodika Národní knihovna. (po r. 2001 samostatná aplikace v projektu Propojení analytických záznamů s plnými texty).

Po celou dobu projektu probíhaly práce na vývoji linky automatické extrakce/indexace TTDE, která v určitých ohledech vycházela z již vyvinutých aplikací Anopress. V počátečních stádiích (r. 2000) participace Anopressu, později linka vyvíjena ing. Matternem.

Po celou dobu řešení probíhaly práce na vývoji a správě serveru **full.nkp.cz a anl.nkp.cz**.

Údržba, správa, linka ze strany Anopressu spočívala v počátečních letech v poskytnutí textů off-line, převedení periodika NK do digitální podoby, od. r. 2001. 2. pololetí v poskytnutí speciálního přístupu pro NK do báze TamTam (6 licencí). Tento přístup TTSNK je součástí linky zpracování.

Statistiky propojení jsou uvedeny v bodě B.1.6.

Činnosti v jednotlivých letech detailně jsou uvedeny v kapitole D Použití finančních prostředků

Rok 2000

Zadání

V roce 2000 se bude realizovat nákup serveru DELL s příslušenstvím a dále dvou počítačů. Bude vyvinuta aplikace pro zařazování a indexaci dat do plnotextové databáze a správu plnotextové databáze. Dále bude vyvinut aplikačního software na vyhledávání v plných textech umístěných v Národní knihovně a vyhledávací internetovské prostředí. K tomuto účelu bude instalován software pro internetovský server MS IIS 4.0 (Microsoft Internet Information Server - možnost stažení z Internetu) a zakoupen Verity Information Sever (Topic). Ve čtvrtém čtvrtletí půjde o testování vyvinuté aplikace pro plnotextovou databázi. Databáze bude doplňována o další plné texty. Podle výsledků testování nové verze ALEPHu (verze 500) bude navržena a vyvinuta aplikace pro řízení a správu Kooperačního systému článkové bibliografie. Databáze bibliografických záznamů bude průběžně doplňována.

Plnění

Výsledkem řešení projektu v r. 2000 je návrh realizace modulární programové aplikace pro získávání, zpracování, indexaci a zpřístupňování plných textů (url, metadata a definice tří typů formulářů pro plnotextové vyhledávání podle pokročilosti) za současné automatické indexace bibliografických záznamů z plných textů ve formátu UNIMARC a DUBLIN CORE. V rámci linky poloautomatické extrakce/indexace TTDE (TamTam Data Extactor) tak vznikne importní 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 a SICI. Pro vyhledávání v plných textech je určen systém založený na pojmovém vyhledávání - TOPIC. Do tohoto systému jsou zatím vloženy hrubé definice topiků.

Plné texty článků byly v rámci konzorcia Anopress průběžně stahovány, připravovány k dynamickému propojení a následně propojeny s bibliografickými záznamy v 2. pololetí tohoto roku.

Plnotextová databáze byla dále průběžně doplňována off-line články z deníků časopisů vydanými v r. 2000 a 1998 z databáze TamTam.

Průběžně byly staticky propojovány záznamy a plné texty z oblasti knihovnictví aj. oborů. Další výsledkem řešení v r. 2000 je návrh programové aplikace pro správu a údržbu Kooperačního systému článkové bibliografie (MNG KOSABI), který vychází z aplikací vyvinutých v rámci SK CASLIN (přijímání záznamů, konverze, úpravy a kontrola záznamů). Pro celý systém byla zakoupena kvalitní HW a SW platforma (server Dell, TOPIC). Bylo zakoupeno 1 PC pro použití v analytickém oddělení.

Rok 2001

Zadání

V roce 2001 půjde o testování vyvinutých aplikací. Proběhne experimentální provoz systému. Dále bude posílena síť Národní knihovny, zakoupen laptop. Obě báze - báze bibliografických záznamů a báze plných textů budou průběžně doplňovány. Je plánována služební cesta.

Plnění

Výsledkem řešení projektu v r. 2001 je experimentální provoz lokální Windows linky pro získávání a poloautomatickou extrakci/indexaci bibliografických záznamů z plných textů (Windows TTDE) a následné vytvoření importního souboru záznamů pro databázi bibliografických záznamů [ANL](#) a databázi plných textů [ANL FULL](#) v NK s implementací UNIMARCu, metadat DC v plných textech včetně automaticky generovaných formátů HTML, XHTML, XML v kvalifikované i nekvalifikované formě. V r. 2001 byla navržena nová architektura a layout serveru FULL.NKP.CZ, pozornost byla věnována definici báze ANL FULL, zejména vyhledávání (prostý dotaz, formulářový dotaz, topiky, rejstříky), byl instalován formulář pro registraci a přihlášení uživatelů. Plné texty článků byly v rámci Konzorcia Anopress průběžně stahovány a připraveny k propojení off-line s bibliografickými záznamy v 1. pololetí tohoto roku.

Plnotextová databáze byla dále průběžně doplňována články z deníků časopisů vydanými v r. 1997 a 1998. Od května 2001 jsou zpracovávány záznamy z deníků a některých časopisů v rámci linky automatické indexace TTDE.

Průběžně byly staticky propojovány záznamy a plné texty z volně dostupných stabilních titulů na Internetu.. Byl založen portál pro zpřístupnění volně dostupných textů na Internetu ve struktuře tematické a regionální s popisem zdrojů. Další výsledkem řešení v r. 2001 je experimentální aplikace pro správu a údržbu Kooperačního systému článkové bibliografie (MNG KOSABI), která je zatím přístupná pouze v NKČR pro účely ladění. Vzhledem k potřebnosti investovat finance jiným způsobem, nebyla raalizována služební cesta, nebyl zakoupen laptop, ale bylo zakoupeno jedno PC, nebyla z finančních prostředků projektu posílena počítačová síť NKČR.

Rok 2002

Zadání

V roce 2002 půjde o další testování a odzkoušení mechanismu vyvinutých aplikací, proběhne ověřovací poloprovoz systému.

Plnění

Výsledkem řešení projektu v r. 2002 je poloprovoz Windows linky pro získávání a poloautomatickou extrakci/indexaci bibliografických záznamů z plných textů (Windows TTDE).

V r. 2002 byla navržena nová architektura a layout serveru FULL.NKP.CZ, poloprovozně byla ověřena aplikace pro registraci a přihlášení uživatelů, byla vytvořena aplikace pro administraci báze ANL FULL a portálu Periodika na WWW (opravy, statistiky, evidence), byla implementována a ověřena funkce pro stahování a export záznamů v několika formátech: text, RTF, HTML, XML, též v UNICODE. V rámci linky poloautomatické

indexace bylo v roce 2002 **zpracováno kolem 16 000 záznamů** z 15 titulů. V roce 2002 byl aktualizován portál, zejména jeho oborová část. V r. 2002 byly průběžně laděny vstupní soubory z některých systémů kooperujících institucí do KOSABI a báze ANL (T-Series, KP-Sys, KP-Sys aj.) a byl navázán kontakt s nově konstituovanými krajskými knihovnami, mezi některými knihovnami proběhla jednání o spolupráci při zpracování titulů. Další výsledkem řešení v r. 2002 je částečně poloprovozní aplikace pro **pro správu a údržbu Kooperačního systému článkové bibliografie (MNG KOSABI)** a vytvoření jednotného interface pro tuto aplikaci. **Byla vytvořena Báze titulů**, kterou lze třídit podle názvů, zpracovávajících institucí a podle toho, zda tituly obsahují zpracované články s plným textem. Bylo zjištěno, že **budoucím trendem je vývoj vyhledávacích metod, které budou za pomoci technologií umělé inteligence pracovat přímo s plným textem a k automatické indexaci v původním slova smyslu nebude vůbec docházet. Analýza problematiky automatické indexace však potvrdila v současné době nutnost intelektuální indexace věcné, která je stále kvalitnější než automatické přiřazování termínů.** Byla provedena důkladná **analýza zpracování článků v zahraničí** (metody, organizace) a ověření, že podobné a srovnatelné metody se vyvíjejí v rámci tohoto projektu a výsledky projektu byly prezentovány na konferenci Inforum 2002 a Knihovny současnosti 2002.

V r. 2002 byla provedena analýza selekční úplnosti topiků v bázi ANL FULL, která je přibližně 80 % a byly vyvozeny praktické závěry pro zkvalitnění funkce topiků (ladění a interaktivní využití topiků). V r. 2002 **byla navržena a částečně vyvinuta internetová verze linky – WWW linka poloautomatické extrakce/indexace (TTDE).**

V r. 2002 byl proveden upgrade a update serveru FULL.NKP.CZ - instalace Win 2000 serveru, instalace MS IIS 5, byl implementován Portal One SE - Topic verze 3.7, byly zprovozněny www stránek v novém prostředí, byl aplikován XMLHTTP a ASP skript, byly provedeny úpravy propojení do ALEPHu). V r. 2002 byl proveden **upgrade serveru ANL** (operační systém Linux SuSE 7.3 Oracle 9.2). Byly zakoupeny 2 upgrady PC, 1 PC, laptop, tiskárna, UPS, scanner a další SW vybavení. **Báze ANL FULL byla zpřístupněna v JIB jako odkaz, báze ANL byla zpřístupněna pro vyhledávání včetně propojení na plný text v bázi ANL FULL** (dostupný pro interní uživatele NK, pro externí volně dostupná pouze metadata, plné texty na 7 dnů po registraci).

Rok 2003

Zadání

V roce 2003 bude realizován poloprovoz systému, průběžné doplňování obouází.

Plnění

Výsledkem řešení projektu v r. 2003 je **rutinní aplikace – lokální Windows linky pro získávání a poloautomatickou extrakci/indexaci bibliografických záznamů z plných textů (Windows TTDE)** a následné vytvoření importního souboru záznamů pro databázi bibliografických záznamů **ANL** a databázi plných textů **ANL FULL** v NK s implementací metadat DC v plných textech včetně automaticky generovaných formátů HTML, XHTML, XML v kvalifikované i nekvalifikované formě – aplikace pro lokální stanice v NKČR. Automaticky se generuje URL.

V r. 2003 byla **částečně realizována internetová aplikace – WWW linka pro získávání a poloautomatickou extrakci/indexaci bibliografických záznamů z plných textů (WWW TTDE)** rozšířená o vstupní formáty HTML z libovolné URL adresy s možností ukládat metadata do báze ANL FULL (ANL) i bez uložení plných textů, dále s možností propojení na rejstříky báze ANL FULL s možností propojení linky se soubory autorit pře Z39.50 Toolkit Klienta. V této lince je zakomponován i modul pro opravy rejstříků a metadat v plných textech báze ANL FULL (částečně hrazeno z též projektu Propojení analytických záznamů). V r. 2003 je v **rutinním provozu** aplikace pro **administraci databáze ANL FULL** (statistiky, evidence uživatelů) a **aplikace pro stahování a export metadat a plných textů** (formát text, rdf, html, UNICODE, XML) – pouze vybraní uživatelé (problematika autorských práv). V lince zpracování bibliografických záznamů z plných textů (Windows TTDE) **bylo zpracováno cca 15 000 metadat** (bibliografických záznamů) a plných textů. V roce 2003 je dále laděna aplikace pro management kooperačního systému. Byl navázán pracovní kontakt se 4 nově vzniklými krajskými knihovnami (export/import v UNIMARCu, excerptní základna).

Jednotlivé činnosti podrobněji (ve vazbě na příslušné pracovníky) – viz kapitola D

Použití finančních prostředků v r. 2003.

Rok 2004

Zadání

V roce 2004 bude realizován plný provoz systému.

Plnění

V r.2004 je rutinní provoz aplikace pro získávání a poloautomatickou extrakci/indexaci bibliografických záznamů z plných textů (Windows TTDE) a následné vytvoření importního souboru záznamů pro databázi bibliografických záznamů [ANL](#) a databázi plných textů [ANL FULL](#) v NK s implementací metadat DC v plných textech včetně automaticky generovaných formátů HTML, XHTML, XML v kvalifikované i nekvalifikované formě – aplikace pro lokální stanice v NKČR. Automaticky se generuje URL

V r, 2004, říjen-prosinec bylo 6 přístupů do databáze TamTam financováno z již zmíněného projektu VaV Budování vzájemně kompatibilních informačních systémů Za tyto tři měsíce bylo uloženo cca 3 000 záznamů.

Koncem roku byla uvedena otestována a uvedena do provozu **internetová aplikace – WWW linka pro získávání a poloautomatickou extrakci/indexaci bibliografických záznamů z plných textů (WWW TTDE)** rozšířená o vstupní formáty HTML z libovolné URL adresy s možností ukládat metadata do báze ANL FULL (ANL) i bez uložení plných textů, dále s možností propojení na rejstříky báze ANL FULL s možností stahovat údaje ze souborů národních autorit přes Z39.50 Toolkit Klienta. Je v ní zabudována též konverze dat do M21 (vyžaduje ještě drobné ladění).

Aplikace, jejíž vývoj probíhal již v r. 2003 a která byla dlouhodobě laděna a testována řešitelkou projektu, byla otestována a uvedena do provozu pracovníky oddělení v prosinci 2004. Bylo v ní uloženo cca 900 záznamů. Tato WWW TTDE linka plynule navázala na Windows TTDE linku lokální.

V r. 2004 je v **rutinním provozu** aplikace pro **administraci databáze ANL FULL** (statistiky, evidence uživatelů) a **aplikace pro stahování a export metadat a plných textů** (formát text, rdf, html, UNICODE, XML) – pouze vybraní uživatelé (problematika autorských práv). Ve Window TTDE lince (TTDE) **bylo zpracováno cca 13 000 metadat** (bibliografických záznamů) a plných textů, ve **WWW TTDE lince cca 900 metadat a plných textů s propojením na národní authority**. V roce 2004 je **dále laděna a v provozu aplikace pro management kooperačního systému (Báze Titulů, uveden do provozu upload záznamů, testy na UNIMARC/M21, statistiky, test URL adres)**. Byly laděny a otestovány importy dat u nově přístupujících institucí (nové krajské knihovny, Národní filmový archiv). Celkem bylo automatizovaným uploadem naimportováno přes 4000 záznamů. V budoucnu není vyloučeno další ladění ad hoc.

Automatizovaný upload je funkční u všech spolupracujících a nově přístupujících institucí do kooperace, koncem r. 2004 patrně problém s poštou u NPKK (musí se posílat soubory v .zip formátu a administrátor báze ANL uploaduje sám). U krajské knihovny v Pardubicích problém s .iso formátem, je třeba zatím uploadovat řádkový tvar. V roce 2004 **proběhly intenzivní práce oblastí metodiky analytického popisu**: byl navržen **minimální záznam** pro soubornou databázi MARC 21, **byla vypracována metodika popisu článků ve formátu MARC 21** s příklady. Koncem roku 2004 byl vypracován návrh pro **aplikaci metody konspektu** v kooperačním systému článkové bibliografie a konspekt se začal zkušebně uvádět nejprve v záznamech produkce NK s ohledem na jeho rozšíření do celého systému v r. 2005. Probíhaly práce na obousměrné **konverzi UNIMARC a MARC21**, testy konverze. **Báze ANL byla v polovině r. 2004 zkonvertována do M21**. Do konverze bylo experimentálně zabudováno podpole 773q určené pro číselné údaje zdrojového dokumentu, které by v budoucnu mohly sloužit k propojení přes SFX a URL. Konverzi do tohoto pole je třeba ještě v r. 2005 znovu otestovat a popř. programově dál vyladit. Jinak konverze proběhla díky dlouhodobému čištění báze celkem uspokojivě.

Koncem r 2004 bylo rozhodnuto dále texty Anopress v lince TTDE nezpracovávat a převést zpracování pod ALEPH a bázi ANL FULL dále o tyto texty nedoplňovat.

Plné texty jsou přístupné uživatelům v prostorách NKČR – lze je prohlížet, tisknout, kopírovat, nikoli však hromadně stahovat. Externím uživatelům jsou pro účely testování přístupné jen na jeden den se stejnými možnostmi. Koncem r. 2004 byla uzavřena smlouva mezi NKČR a Anopress o možnostech zpřístupnění plných textů získaných od této společnosti z databáze TamTam. Další zpřístupnění plných textů co do forem a možností bude řešeno v souladu s řešením legislativně právních otázek zpřístupňování elektornických plných textů v ČR a NKČR.

Jednotlivé činnosti podrobněji (ve vazbě na příslušné pracovníky) – viz kapitola D
Použití finančních prostředků v r. 2004.

B.1.6 Některé globální (statistické) ukazatele

Některé statistické údaje týkající se Báže ANL FULL, ANL a Kooperačního systému článkové bibliografie za r. 2000-2004

- [ANL FULL](#) - cca **130 944** plných textů s metadaty -viz tabulka Příloha F7

Báze ANL FULL	Počet plných textů
Texty zpracované v TTDE (2001-2004)	53493 (licence do TamTam)*
Texty připravené k propojení (2000-2001)	11441 (zakoupené extra) a naimportované off-line, příprava k propojení v ALEPHu **
Texty naimportované off-line	66010 (zakoupené extra) a naimportované off-line **
Celkem	130944***

* v r. 2004, měsíc říjen-prosinec – přístup do TamTam hrazen z projektu VaV Budování vzájemně kompatibilních informačních systémů ...

** v tomto projektu hrazeno cca 42 359 plných textů, hrazeno též z projektu Propojení analytických záznamů ...

*** určité velmi malé procento článků v ANL FULL je nepřístupné (různé důvody v řetězci od získání ke zpracování plného textu)

- **Počet propojení z báze ANL do báze ANL FULL – cca 125 042**
- **Z báze ANL propojeno na volné plné texty cca 24 280 viz tabulka Příloha F7, z toho cca 400 nefunkčních a cca 1 500 netestovaných** (též v rámci dalších dvou projektu VaV: Propojení analytických záznamů ... a Budování vzájemně kompatibilních informačních systémů ...)
- **Z báze ANL celkem cca 149 322 linků na plné texty**
- **Portál [Periodika na WWW](#) obsahuje linky na cca 90 periodik oborových, cca 90 periodik regionálních.**
- **Počet běžně zpracovávaných titulů v lince TTDE: 17**
- **Propojovaných titulů v bázi ANL na [www.periodika](#): cca 44**
- **Počet propojovaných titulů na volné [www](#) zdroje: cca 59**

V lince Windows TTDE lokální + WWW TTDE lince zpracováno a propojeno v letech 2001, 2. pololetí - 2004: cca 53496 záznamů (bez simulované linky v letech 2000, 2001 1. pololetí)

(malý rozdíl v importech dán: do báze ANL FULL se importují záznamy po 10 minutách průběžně, tj. téměř ihned po zpracování, do báze ANL dávkově jednou večer; proto je lepší výsledek u báze ANL FULL; výsledek rozdíl v importech je dán poruchami sítě v době dávkového importu do ANL a výpadky proudu, špatnou volbou kódu katalogizátorem aj.)

Tabulka zpracování záznamů v lince TTDE

Rok	Počet stažených, tj. zpracovaných záznamů v lince	Poznámka	Průměr na den zpracovaných záznamů
2000	6398*	Příprava k propojení	50
2001	13657*	Windows TTDE od poloviny r. 2001	53
2002	16005		63
2003	15877		62.5
2004	13000 **	Těž WWW TTDE	55.4
Celkem	64937	Celkem v TTDE 53496	Průměrně ca 56.7 tj cca 57

* simulovaná linka TTDE, dále specifikováno v následující tabulce, tyto texty musely být zakoupeny extra, jinak pro texty zpracovávané v TTDE byl z projektu placen speciální přístup do TamTam

*

Rok	Zpracování a příprava k propojení záznamů (ALEPH)
2000	6398
2001	5043
Celkem	11441

**

Počet záznamů zpracovaných v WWW TTDE v prosinci 2004. Rozdíl v importech do obouází je dán většinou tím, že katalogizátoři nevolili správný kód exportu z linky. V době zavádění do provozu WWW linky byly také časté výpadky Internetu a sítě v NKČR. Přístupy do TamTam hrazeny v r. 2004, říjen – prosinec z projektu Budování vzájemně kompatibilních informačních systémů ... (za tuto dobu popsáno cca 3000 plných textů, z toho cca 900 v WWW TTDE) .

Datum 2004	Import do ANL	Import do ANL FULL
29.11.	46	69
30.11.	42	91
1.12.	87	92
2.12.	57	59
3.12.	17	29
6.12.	82	88
7.12.	92	92

8.12.	89	89
9.12.	31	31
10.12.	44	45
13.12.	49	54
14.1.	53	66
15.1.	32	46
16.1.	49	49
17.1.	21	22
20.1.	56	58
21.1.	37	38
23.1.	23	23
28.1.	38	39
Celkem	945	1080

- **Počet záznamů k prosinci 2004: [ANL](#) - poslední sysno v r. 2004 886 121 , v roce 1999 poslední sysno 572 904**
- **V kooperaci za léta 2000-2004 zpracováno a zasláno do systému cca 134 627 záznamů:** počty jsou přibližné, ne vždy je možné vzhledem k vývoji báze identifikovat záznamy dle zpracovatelské provenience a adekvátně vyhledat; dále pak počet naimportovaných záznamů může být menší než zaslanych (neprošly konverzí nebo nebyly naimportovány – vyvinutá aplikace pro upload a zejména statistiky tento nedostatek předchozích let eliminuje (lze přesně vidět, co je uploadováno, lze sledovat počet importovaných záznamů a chyb). Za výchozí se bere poslední sysno v bázi v r.2004.

Tabulka stav kooperace – r. 2004

Knihovna	Rok 1999 záznamy	Rok 2004 záznamy	Za léta 2000- 2004 záznamy zasláno cca	NKČR
NKČR	443766	622325		
ÚZPI	22929	37529	14600	
NPKK	7552	16367	8812	
STK	26715	42755	16040	
MZK	11885	26323	14138	
Plzeň	7498	9818	2320	
České Budějovice	10736	35964	25228	
Hradec Králové	11162	14644	3482	
Liberec	4276	12442	8166	
Ústí nad Labem	4287	14350	10063	
Olomouc	9746	23307	13561	
Ostrava	5472	12009	6537	
Kladno	6880	18560	11680	
Celkem	572904	886120	134627	178559*

- *** od čísla je třeba odečíst cca 5000 - nejsou zohledněny experimentující knihovny nebo nově přístupující knihovny, jejichž záznamy byly v rámci testů naimportovány do báze.**
- **V kooperaci zasláno uploadem v listopadu a prosinci 2004 do báze ANL cca: 4 959 záznamů**

Upload

Datum 2004	Sysno event. počet	Instituce Počet záznamů č.dávky				
30.11.		189+997 z.	OLA00 + ABA013 KLG001 OSA001 ABA009			
1.12.	880316-881090	CBA001 650 z. Dávka 624	ULG001 112 z. Dávka 609	HKA001		
3.12.	881333-881427-8	ABA 012 13 z. Dávka 616	ABA012 85 z. Dávka 618			
6.12.	881625-882339	LIA 01 521 z. Dávka 626	ZLG001 52 z.	HBG001 142 z.		
1.12 Zkouška	ABC135	835 z Dávka 622				
14.12.	OSA001	81 z. Dávka 653				
	CBA001	404 z. Dávka 654				
	KLG001	261z. Dávka 655				
	ABA009	226 z. Dávka 656				
	HKA001	78 z. Dávka 658				
15.12.	OLA001	217 z. Dávka 664				
	ABA 013	300 z. Dávka 665				
21.12.20 04	KLG001	261 z. Dávka 678				
	ULG001	48 z. Dávka 679				
4.1.2005	ABC135	Dávka 689 673 z.				
Celkem upload		3773 z.				

- Časové pokrytí titulů zahrnutých v: ANL 1990/91 -, ANL FULL 1997- .
- Statistiky využívání báze ANL

Statistiky vstupů do báze ANL (novější k dispozici řešitelka nemá):

Období 2003	ANL
září	3629
říjen	5878
listopad	5339
prosinec	4042

- **Statistiky využívání báze ANL FULL**
Sumarizace přístupů na plné texty z báze ANL a ANL FULL za r. 2004, 4.12 – viz originál výpisy z administrace Příloha F7

Měsíc 2004	Počet přístupů na plné texty ANL, ANL FULL
12, 4.12	211
11	2504
10	1781
9	1296
8	1419
7	935
6	1155
5	3315
4	3440
3	4155
2	2186
1	1555
Celkem	23952

- **Aktuální titulová základna kooperačního systému - viz Příloha F4**

Knihovna	Cca seriály	Sborníky vedené extra r. 2000-2004 cca
NKČR	195	90
ÚZPI	25	112
NPKK	54	
STK	47	
MZK	61	z toho cca 10 budou zpracovávat jiné instituce (přístupující)
Plzeň	43	zatím přechází na ALEPH, cca 10 budou zpracovávat jiné (přístupující) instituce
České Budějovice	24	
Hradec Králové	21	z toho cca 4 budou zpracovávat jiné instituce (přístupující)
Liberec	14	
Ústí nad Labem	31	
Olomouc	23	

Ostrava	18	
Kladno	37	
Celkem*	593	

***Nejsou započítány tituly nově přístupujících institucí, v rámci kterých jsou započítány i tituly přecházející – cca 74 titulů**

B.2 Přínos řešitele

Přínos projektu v spočívá v realizaci metod navržených v zadání projektu.

Do rutinního provozu byla uvedena Windows TTDE linka poloautomatické extrakce/indexace bibliografických záznamů z plných textů s konverzí dat do DC a UNIMARC, do stádia ověřovacím provozu byla uvedena WWW linka s napojením na soubory autorit a konverzí dat do M21.

Přínos projektu je v rutinní v aplikaci systému TOPIC (pojmové a fuzzy vyhledávání), ve zpřístupnění plných textů v plnotextové databázi ANL FULL, ve vytvoření portálu Periodika na WWW.

Do rutinního provozu byla uvedena aplikace pro MNG Kooperačního systému článkové bibliografie zahrnující aplikaci pro evidenci a aktualizaci Báže Titulů, pro automatizovaný upload, zpracování a export dat spolupracujících institucí do báze ANL, aplikaci pro kontrolu URL adres a duplicit.

Přínos v spočívá v revidování tradičních postupů při zpracování české národní bibliografie v oblasti jmenného i věcného popisu a ve funkční realizaci linky automatické indexace bibliografických záznamů. Pro popis elektronických informačních zdrojů je aplikován formát Dublin Core v jeho dosud možných aplikacích v jazyku HTML, XHTML a XML.

Přínos spočívá v experimentální aplikaci metody konspektu v NKČR a v návrhu aplikace konspektu v rámci celého systému. Vytvoření předpokladu pro zapojení systému do oborových informačních bran.

Předmětové kategorie používané v kooperačním systému budou možno po důkladné redakci a revizi doplnit metodu konspektu.

Přínos spočívá v pravidelné aktualizaci excerptní základny systému, zpřehlednění její evidence s možností aktualizace, v jejím zkvalitnění směrem k odborným zdrojům, v eliminaci duplicit.

Důležité jsou analýzy, které se týkají automatické indexace a které byly provedeny v rámci projektu Propojení analytických záznamů s plnými texty a ze kterých kooperační projekt vychází. Potvrdily nutnost doplnění autotmatizované indexace zejména intelektuální indexací věcnou.

V případě zájmu přichází v úvahu možnost aplikace linky TTDE po určitých úpravách na zpracování článků v rámci některých kooperujících institucí, dále pak, po jisté úpravě, aplikace linky zpracování bibliografických záznamů z plných textů v komunikaci mezi autorem, nakladatelem, informační institucí, bibliografickou agenturou.

V případě vhodných podmínek a zájmu přichází v úvahu možnost aplikovat systém **TOPIC** po jisté modifikaci na zpracování a zpřístupnění **odborných zdrojů vědecko technického charakteru**.

Přínos projektu spočívá v praktické integraci elektronických zdrojů do služeb NKČR pomocí technologicky nejvyspělejších nástrojů pro zpřístupnění těchto zdrojů při zachování tradičně zpřístupňovaných sekundární informací formou bibliografických záznamů.

Přínos spočívá v integraci heterogenních dat (dat různých SW a plných textů) do Kooperačního systému článkové bibliografie, v němž dochází k propojení tradičních knihovnických postupů a fondů s určitými prvky digitální knihovny.

Přínos projektu spočívá v organizačním a metodickém vedení Kooperačního systému článkové bibliografie a vytváření předpokladů pro zapojení kooperujících institucí do JIB.

Spolupráce NKČR a Anopress, IT byla oboustranně přínosná a vytvořila předpoklady pro integraci produktů Anopress do JIB.

Z analýzy zahraničních materiálů vyplývá, že články se zpracovávají jednak soukromými společnostmi, jednak národními knihovnami a jsou zpřístupňovány online, ve formě digitálních knihoven a někde jsou součástí národní bibliografie. Zpracování odpovídá současným světovým trendům, v kooperaci i metodice je do jisté míry předchází.

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.

Výsledky projektu byly prezentovány na konferencích a seminářích.

B.3 Posun znalostí

K významnému posunu znalostí došlo především v těchto oblastech:

- Poměrně dobrá **orientace v nových trendech zpracování** a zpřístupňování informací.
- Praktická **realizace těchto trendů na konkrétní fungující systém**
- **Fungující linka poloautomatické extrakce/indexace TTDE**
- **Propojení internetové verze linky TTDE na authority**
- **Propojení bibliografických záznamů s elektronickými zdroji na základě dynamických i statických URL adres, kombinace pojmového vyhledávání s metadaty**
- **Zpřístupnění plných textů v rámci moderně koncipované báze ANL FULL na základě vyspělých technologií pro zpřístupňování elektronických dokumentů, praktická implementace Dublin Core v rámci HTML, XHTML, XML v kvalifikované a nekvalifikované formě. Definování Anl Core pro zpřístupnění jemnějších údajů analytického popisu**
- **Údržba strukturovaného portálu volně dostupných zdrojů na Internetu s popisem těchto zdrojů**
- **Možnost využití metody vypracované pro linku automatické indexace pro zpracování informací a toku informací mezi autorem, nakladatelem, informační agenturou, bibliografickou agenturou/knihovnou**
- **Analýza problematiky automatické indexace a selekční úplnosti topiků**
- **Ověření hypotézy na základě analýzy problematiky automatické indexace: orientace spíše na automatickou extrakci dat, clustering, automatické abstrahování event. klasifikaci a podporu indexátorů a automatizovanou kategorizaci s využití dalších možností inteligentního vyhledávání systému TOPIC, než na automatické přiřazování věcných termínů, zohlednit interakci systému s uživatelem a indexátorem**
- **Ověření hypotézy, že intelektuální indexace věcná je zatím nezastupitelná**

- Indexace a vyhledávání v klíčových slovech plných textů je velmi nepřesné a zavádějící. U běžných plnotextových systémů nepostihuje pojmy, koncepty. **Možná, že řešením je integrace řízených slovníků do pojmového vyhledávání podpořená lematizátorem a automatickou kategorizací textu a tvorba ontologií. Tedy kombinace všech výše zmíněných nástrojů pro věcné zpřístupnění**
- **Posílení vazby v rámci KOSABI a připravení podmínek pro vstup nových institucí do kooperace**
- **Automatizovaný management kooperačního systému a koordinovaná excerpční základna s výraznějším směřováním k odborným informačním zdrojům**
- **Ověření, že podobné metody zpracování, organizace zpracování i bibliografického systému se ve světě vyvíjejí a jsou s nimi srovnatelné a do jisté míry je i předcházejí z hlediska kooperace a metodiky**
- **Nová forma poskytování informačních služeb za spolupráce knihovnické a moderní informační instituce. Bez kvalitního know how těchto společností není možný vývoj v oblasti knihoven**
- **Prezentace výsledků projektu v rámci ČR a reference v zahraničí (řešitelce projektu nabídnuto členství v IFLA Section for Newspapers, což je určitým uznáním práce, kterou tým pracovníků, kolem projektu, vykonává)**

C Navrhovaná část

C.1 Výsledky řešení

- **Praktická realizace nástrojů pro optimalizaci integrace a správy heterogenních dat v rámci Kooperačního systému článkové bibliografie.** Vznikl tak nástroj pro poloautomatickou přípravu dat pro bibliografickou databázi založenou na UNIMARCu/M21 a plnotextovou databázi založenou na pojmovém vyhledávání systému TOPIC a metadatech zpřístupňovaných v Dublin Core a jeho aplikacích v HTML, XHTML, XML v kvalifikované a nekvalifikované formě
- **Rutinní provoz linky** zpracování bibliografických záznamů z plných textů Windows TTDE pro lokální pracovní stanice (generování UNIMARCu, DC) a **zavedení ověřovacího provozu WWW TTDE** (navíc generování M21) s možností stahování autorit přes Z39.50 Toolkit Klienta.
- **Rutinní provoz serveru full.nkp.cz, systému TOPIC, báze ANL FULL a portálu Periodika na WWW včetně administrace**
- **Zpřístupnění českých plných textů**
- **Rutinní provoz Aplikace pro Management Kooperačního systému článkové bibliografie** (koordinovaná titulová základna, upload a automatizované zpracování dat systému, export dat do Alephu)
- **Standardizace týkající se popisu článků v UNIMARCu a AACR2R**
- **Zabezpečení kooperačního systému článků po stránce metodické i organizační**
- **Analýza problematiky automatické či poloautomatické a selekční úplnosti aplikovaných topiků**
- **Důkladná analýza zpracování článků v zahraničí (metody, organizace)**
- **Nastínění perspektiv KOSABI a vznik jeho možných subsystémů z hlediska typu zpracovávaných titulů**
- **Prezentace výsledků projektu**

C.2 Závěr

Zpracování českých článků prochází transformací, jde o inovaci jak po stránce technické i technologické, ale i koncepční, jde o zkvalitnění báze ANL a diferencované plné zpřístupnění plných textů. Tyto změny však musí být postupné a citlivé při zachování toho, na co je třeba navázat, pokud nechceme systém zlikvidovat. Systém zpracování a zpřístupnění článků bylo možno částečně transformovat díky finanční podpoře projektů. O důležitosti informací uveřejněných v člancích nelze pochybovat. Nemyslím tím pouze informace odborné a vědecké povahy, na které je třeba se především zaměřit v budoucnu.

Spolupráce s Anopressem mohla být navázána díky zmíněným projektům a NK mohla navázat na moderní technologii, kterou tato firma užívá a dále rozvíjí. Spolupráce byla přínosná jak pro NKČR, tak pro Anopress.

Na zpřístupňování českých plných textů mají vliv nejen vyvíjené technologie, ale i koncepce a strategie zainteresovaných subjektů, jejich smysl a citlivost pro dobrý odhad, znalost věci, transparentnost řešení a schopnost kompromisů, respekt k domácímu terénu a světovým trendům.

C.3 Návrhy opatření

- **Zajištění financování navazujícího projektu zejména z hlediska rozvoje plnotextové báze a systému TOPIC, spolupráce v oblasti moderních metod strukturace a analýzy textu**
- **Koordinace s ostatními stávajícími i budoucími projekty v ČR i v zahraničí**
- **Posílení vazeb stávajícího KOSABI na existující kooperační systémy, souborné katalogy aj. systémy**
- **Rozvoj spolupráce s dalšími institucemi**

- Užší spolupráce s Národní lékařskou knihovnou (citace)
- Personální zajištění projektu z hlediska počtu pracovníků oddělení analytického zpracování v NK i v rámci kooperačního systému
- Průběžné sledování vývojových trendů ve standardizaci a jejich praktická aplikace
- Přísnější výběr článků k indexaci deníků s cílem vyloučení subjektivního faktoru při excerpce deníků, orientace na odborné texty
- Propojování údajů přes SFX
- Ošetření autorsko právních aspektů, event. ekonomických aspektů zpřístupňování plných textů externím a vzdáleným uživatelům nejenom na úrovni projektu
- Smluvní zajištění zpřístupňování báze ANL FULL interním i externím uživatelům, uživatelům v rámci konzorcia Anopress
- Zapojení báze ANL FULL do Konzorcia Anopress
- Zpřístupnění báze externím uživatelům
- Pokračování Konzorcia Anopress
- Eventuelní další zpřístupnění elektronických zdrojů v Anopressu
- Vstřícnost vydavatelů, nakladatelů, autorů
- Strukturované údaje v plných textech
- Další vývoj topiků a jejich ladění
- Zkvalitnění možností vyhledávání v systému TOPIC a integrace termínů věcného popisu do topiků, interakce systému s uživatelem a katalogizátorem
- Práce na sjednocování rejstříku předmětových kategorií a rutinní aplikace metody konspektu
- Opravy báze ANL a ANL FULL. V ANL zdrojových dokumentů a ISSN
- Zkvalitněné ukládání údajů v rámci NKČR i kooperujících institucí
- Doladění existující konverze UNIMARC M21 pro články
- Automatizovaná kategorizace a možnosti koexistence „ručně tvořených předmětových kategorií“ a automatizovaně vzniklých
- Další zkoumání možností automatické či poloautomatické indexace zejména ve smyslu extrakce údajů
- Průzkum možností uplatnění vyvinutých linek poloautomatické indexace/extrakce TTDE event. Jejich doladění dle potřeby
- Pokračování aplikace pro automatizovaný management Kooperačního systému článkové bibliografie event. jeho doladění dle potřeby nebo modifikace
- Retrokonverze záznamů článků v tištěné formě (ČNB i jiné)
- Zkoumání možností sdílené katakogizace pro články v rámci ALEPhu
- Vývoj linky zpracování plných textů pod ALEPHem (záležitost výzkumného záměru Budování vzájemně kompatibilních systémů ...)
- Propojování údajů přes SFX
- Napojení zpracování článků na oborové brány
- Zkoumání problematiky ontologií a sémantického WWW
- Spolupráce knihovní, akademické a podnikové informační sféry

D Použití finančních prostředků - institucionální podpory (tabulky, komentář)

Financování v jednotlivých letech bylo částečně přizpůsobeno vývoji zpřístupňování informací o člancích v NKČR, potažmo v ČR a výzkumnému záměru Propojení analytických záznamů s plnými texty ... , který je s účelovým projektem komplementární.

Byly podepsány 3 doplňky ke smlouvě, týkající se přesunu finančních prostředků. Oproti původnímu rozpočtu byl snížen vklad NKČR do projektu v letech 2001-2004. V době řešení projektu probíhaly redukce počtu pracovníků oddělení a výše původně kalkulovaného vkladu nebylo možno dodržet. Čerpání celkové výše státní dotace bylo dodrženo.

Rok 2000

Náklady na projekt celkem: 2 896 000.-

<u>Výše státní dotace:</u>	2 320 000.-	
<u>Neinvestiční prostředky:</u>	1 025 000.-	
Služby	850 000.-	
Materiál	48 000.-	
Licence	36 000.-	
Mzdové prostředky (OON)	91 000.-	
<u>Investiční prostředky:</u>	1 295 000.-	
Dell Computer	588 955.-	
Search '97 - TOPIC	523 979.-	
PC-PIII-7000 Mhz	77 958.-	
ORACLE 8i	42 941,60	
ORACLE 8i. Upgrade	2 520.-	
PC AMD K7-800 MhZ	58 646,40	
Vklad do projektu:	576 000,-	
<u>Vlastní zdroje</u>		
Národní knihovna ČR – odd. analytického zpracování	456 000.-	
<u>Finanční prostředky získané z jiných zdrojů</u>		
Anopress, s.r. o.	120 000,-	

Tabulky

Neinvestiční prostředky

	A	B	C	D
1	Neinvest.	Souborná	databáze Kooperáčního systému článkové bibliografie	limit
2	Datum	Čís. fakt.	Popis	Částka
3			Služby	
4	18.7.2000	2000-088	I. Mattern - úprava hlaviček, interface pro export, program TAMTAM	48 500,00 Kč
5	18.7.2000	806/2000	Anopress - úprava hlaviček, interface pro export, program TAMTAM	50 925,00 Kč
6	18.7.2000	0508/2000	Anopress - nákup plných textů článků I.	19 440,80 Kč
7	18.7.2000	0509/2000	Anopress - nákup plných textů článků II.	19 446,00 Kč
8	18.7.2000	658/2000	Anopress - nákup plných textů článků III.	19 440,80 Kč
9	18.7.2000	657/2000	Anopress - nákup plných textů článků IV.	19 446,00 Kč
10	21.8.2000	0975/2000	Anopress - úpravy programu TAMTAM, párování hlaviček, interface	32 550,00 Kč
11	22.8.2000	2000-090	I. Mattern - úpravy programu TAMTAM, párování hlaviček, interface	31 000,00 Kč
12	22.8.2000	6816	Dell Computer - instalace serveru Dell	3 050,00 Kč
13	5.10.2000	2000-093	I. Mattern - instalace produktů na server NK	39 500,00 Kč
14	31.10.2000	1325/2000	Anopress - indexace a konverze dat a server NK	41 475,00 Kč
15	31.10.2000	2000-095	I.Mattern - indexace, konverze a vyhledávání dat	39 000,00 Kč
16	7.11.2000	36	Koktan - instalace Linux a Oracle na server ANL, definice databaze	40 000,00 Kč
17	7.11.2000	2000-087	I. Mattern - činnosti v rámci propojení s plnými texty	8 000,00 Kč
18	7.11.2000	410008	F. Traksl - činnosti v rámci propojení s plnými texty, opt. Programu	22 798,90 Kč
19	7.11.2000	0507/2000	Anopress - nákup plných textů článků V.	16 359,00 Kč
20	15.11.2000	1490/2000	Služby Anopress - činnosti v rámci propojení s plnými texty	41 475,00 Kč
21	23.11.2000	1489/2000	Anopress-převod periodika Národní knihovna do elektronické podoby	20 422,50 Kč
22	11.12.2000	41	J. Koktan - aplikace pro management kooperačního systému	80 000,00 Kč
23	13.12.2000	42	J. Koktan - aplikace pro management kooperačního systému	80 000,00 Kč
24	1.12.2000	1507/2000	Anopress - údržba databáze - Topic	105 000,00 Kč
25	1.12.2000	1508/2000	Anopress- knihovnicko-bibliografické práce	17 969,90 Kč
26	1.12.2000	0659/2000	Anopress - nákup plných textů článků VI.	16 359,00 Kč
27	1.12.2000	2000-097	I. Mattern	37 000,00 Kč
28			Materiál	
29	4.9.2000	1363/2000	CDR, ddiskety, pásy do tiskáren	9 142,70 Kč
30	7.11.2000	1763/2000	CDR	907,40 Kč
31	23.11.2000	20020123	Gaia - PC AMD Duron	37 943,00 Kč
32			Licence	
33	22.8.2000	6821	Win NT 0,4	36 849,00 Kč
34			Mzdy	
35	31.8./6.11.2000		Chocholová, Levá, Kopp, Rumlenová, Holá, Magúthová, Střítecká,	91 000,00 Kč
36			Štochlová, Molitorisová - stahování textů, příprava k propojení	
37				
38				
39				
40	zůstatek			
41	čerpáno			1 025 000,00 Kč

Mzdové prostředky (OON) - detailní čerpání jednotlivých pracovníků oddělení analytického zpracování NKČR

	A	B	C	D	E	F	G
1	Mzdy	Souborná databáze Kooperačního systému článkové bibliografie				limit	91 000
2	Datum	Jméno	Popis činnosti	Hodiny	Kusy	Částka	Zůstatek
3	k 30.11.2000	Střítecká M.	Příprava k propojení	70	700	10 500	80 500
4	k 30.11.2000	Molitorisová D.	Příprava k propojení	100	1000	15 000	65 500
5	k 30.11.2000	Čápková V.	Příprava k propojení	66	622	10 500	55 000
6			Korektury	4	67		
7	k 30.11.2000	Magúthová J.	Příprava k propojení	70	700	10 500	44 500
8	k 30.11.2000	Chocholová J.	Příprava k propojení	100	1000	15 000	29 500
9	k 30.9.2000	Holá E.	Příprava k propojení	13	130	3 750	25 750
10			Korektury	12	240		
11	k 30.11.2000	Levá J.	Příprava k propojení	100	1000	15 000	10 750
12	k 30.9.2000	Rumlenová A.	Příprava k propojení	26	260	3 900	6 850
13	k 30.9.2000	Kopp P.	Příprava k propojení	21	210	3 150	3 700
14	k 31.12.2000	Štochlová H. x	Korektury	43	731	3700	0
15	Suma			625	6660	91 000	
16	z toho:		korektury	59	1038	3 700	
17			propojení	566	5622	87 300	
18	Suma			625	6660	91 000	
19	Komentář:	Pracovní normy: korektury		20 korektur/hod.			
20		příprava propojení, propojení		10 propojení/hod.			
21		Částka za hodinu:		150 Kč/hod.			
22		x U této pracovnice norma:		17 korektur/hod.			
23		Částka za hodinu:		86 Kč/hod.			

Investiční prostředky

	A	B	C	D
1	Investice	Souborná databáze Kooperačního systému článkové bibliografie		limit
2	Datum	Čís. fakt.	Popis	Částka
3				
4	18.7.2000	4468	Dell Computer	588 955,00 Kč
5	18.7.2000	2000022	Search '97 - TOPIC	523 979,00 Kč
6	22.9.2000	2000200096	PC-PIII-7000 Mhz	77 958,00 Kč
7	18.9.2000	200182	ORACLE 8i.	42 941,60 Kč
8	26.10.2000	200209	ORACLE 8i. Upgrade	2 520,00 Kč
9	23.11.2000	20020122	PC AMD K7 - 800 MhZ	58 646,40 Kč
10	čerpáno			1 295 000,00 Kč

Vklad do projektu

Vklad pracovníků odd. analytického zpracování NKČR

	A	B	C	D	E	F
1	Vklad oddělení	Souborná databáze Kooperačního systému článků		limit vkladu		456000
2	Datum	Jméno	Popis činnosti	Hodiny	Kusy	Vklad v Kč
3	10.6./31.12.2000	Střítecká M.	Korektury	140	2801	21000
4			Příprava propojení	65	650	9750
5	10.6./31.12.2000	Molitorisová D.	Korektury	155	3101	23250
6	1.2./31.12.2000		Prezentace WWW stránek	80		12000
7	10.6./31.12.2000	Čápková, V.	Ukládání záznamů	84	253	12600
8	10.6./31.12.2000	Magúthová J.	Korektury	51	1012	7650
9			Příprava propojení	26	264	3900
10	10.6./31.12.2000	Chocholová J.	Korektury	260	5200	39000
11			Příprava propojení	30	300	4500
12	10.6./31.12.2000	Holá E.	Korektury	114	2281	17100
13	10.6./31.12.2000	Levá J.	Korektury	21	420	3150
14			Příprava propojení	13	132	1950
15	10.6./31.12.2000	Rumlenová A.	Korektury	45	893	6750
16			Příprava propojení	10	100	1500
17	10.6./31.12.2000	Kopp P.	Korektury	59	1190	8850
18			Příprava propojení	26	260	3900
19	10.6./31.12.2000	Brejníková D.	Ukládání záznamů	24	72	3600
20	10.6./31.12.2000	Kindl D.	Korektury	84	1680	12600
21			Propojení	20	200	3000
22	10.6./31.12.2000	Pašník M.	Korektury	49	984	7350
23	10.6./31.12.2000	Macková R.	Korektury	225	4500	33750
24	10.6./31.12.2000	Vrbičanová L.	Korektury	73	1457	10950
25	10.6./31.12.2000	Kozáková A.	Korektury	53	1070	7950
26	Suma			1707	28820	256050
27						
28	z toho:		korektury	1329	26589	199350
29			propojení	190	1906	28500
30			záznamy	108	325	16200
31			prezentace WWW stránek	80		12000
32	Suma			1707	28820	256050
33	Zůstatek vkladu - mezisoučet					199950
34	Vklad hl.řešitelky					
35	Datum	Jméno	Popis činnosti	Hodiny	Kusy	Vklad v Kč
36	1.1./31.12.2000	Anděrová I.	Korektury	495	9900	74250
37	1.6./31.12.2000		Management projektu	100		15000
38	1.5./31.12.2000		Příprava smluv	98		14700
39	1.6./31.12.2000		Podklady pro programátorské práce	120		18000
40	1.8./31.12.2000		Testování programů	120		18000
41	1.5./31.12.2000		Studium materiálů, práce s Internetem	150		22500
42	1.5./31.12.2000		Správa databáze kooperačního systému	150		22500
43	1.6./31.12.2000		Podklady pro dodávky plných textů			
44			a částečná správa url adres	100		15000
45	Suma			1333		199950

53		V případě hlavní řešitelky byla práce vykonávána mimo pracovní dobu - cca 20 pracovních sobot/180 hod.,				
54		od dubna 2000 práce do 20 hodin, tj. 3 hodiny navíc denně, tj. 175 dnů/525 hod.				
55		Cca 522 hodin bylo odpracováno doma (3 hodiny denně, 174 dnů).				
56						
57		Pracovní normy: korektury	20 korektur/hod.			
58		příprava propojení, propojení	10 propojení/hod.			
59		záznamy	3 záznamy/hod.			
60		Částka za hodinu:	150 Kč/hod.			
61						

Porovnání mzdových prostředků a činností v rámci projektů VaV Propojení analytických záznamů s plnými texty ... a Souborná databáze Kooperačního systému článkové bibliografie

2	Porovnání činností oddělení analytického zpracování NKČR v rámci řešených projektů VAV													
3														
4	Celkové porovnání činností oddělení analytického zpracování NKČR (korektury, propojení záznamu													
5	s plným textem, tvorba nových záznamů) v rámci projektu Propojení analytických záznamů s plnými													
6	texty a optimalizace zpřístupnění plných textů, dále pak projektu Souborná databáze Kooperačního													
7	systému článkové bibliografie - optimalizace integrace a správy heterogenních dat (v tomto projektu													
8	oddělen vklad oddělení a vklad řešitelky pro názornost). Oba projekty se doplňují. První z nich je													
9	především koncepční, druhý pak realizační.													
10														
11														
12														
13	Činnost	Propojení analyt.zázn			Soub. databáze Koop.s			Vklad oddělení			Vklad hl. řeš.			Celkem
14		Hodin	Kusů	Částka	Hodin	Kusů	Částka	Hodin	Kusů	Částka	Hodin	Kusů	Částka	kusů
15	Korektury													
16	Propojení	188	2 596	28 200	59	1 038	3 700	1 329	26 589	199 350	495	9 900	74 250	40 123
17	Záznamy	246	2 460	35 800	566	5 622	87 300	190	1906	28 500				9 988
18	Jiné činnosti	40	120	6 000				108	325	16 200				445
19								80		12 000	838		125 700	
20	Celkem	474		70 000	625		91 000	1 707		256 050	1 333		199 950	

Vklad Anopress

	A	B	C	D	E	F	G
1			Vklad Anopress.s.r.o.				
2							
3	Datum	Jméno	Popis produktu	Hodiny	Kusy	Vklad v Kč	Kč/měsíc
4	1.5./31.12.2000	Anopress	TAMTAM Profesional		1	120 000	15 000

Komentář – účelové použití dotace Neinvestiční prostředky

Služby

Návrh a programová realizace linky automatického získávání plných textů, indexace bibliografických záznamů a plných textů, propojování záznamů s plnými texty on-line a zpřístupňování plných textů. údržba databáze Topic - Anopress, I. Mattern.

Instalace produktů (NT, TOPIC) na server full.nkp.cz (server Dell) - I. Mattern.

Indexace a konverze plných textů na server, údržba databáze - I. Mattern.

Zpřístupnění databáze plných textů na WWW. Vyhledávání v plnotextové databázi a vytvoření formulářů pro vyhledávání - I. Mattern.

Nákup plných textů odpovídajících profilu zpracovávaných článků za účelem propojení analytických bibliografických záznamů s těmito texty- cca 17930 textů – Anopress.

Převod periodika Národní knihovna do elektronické podoby a její zpřístupnění na Internetu - Anopress.

Aplikace pro řízení, správu a údržbu Kooperačního systému článkové bibliografie (přijímání záznamů, globální úpravy, formálně-logické kontroly) - J. Koktan

Instalace produktů (Linux, ORACLE) na server ANL - J. Koktan.

Optimalizace programu pro off-line doplňování adres k plným textům do bibliografických záznamů - F. Traksl.

Materiál

Nákup CD-R, disket a pásek do tiskáren pro zálohování a provoz oddělení, PC AMD Duron.

Licence

Win NT 0.4 (20 licencí).

Mzdové prostředky (OON)

Pracovníci oddělení analytického zpracování uváděli částečné adresy plných textů a zároveň tyto texty stahovali v rámci Konzorcia Anopress (příprava k propojení s plným textem). Dále byly propojovány bibliografické záznamy zpracované v rámci oddělení s plnými texty vystavenými na Internetu. Byly prováděny korektury bibliografických záznamů.

Investiční prostředky:

Server Dell (full.nkp.cz) pro plné texty a systém TOPIC (pojmové vyhledávání), PC-PIII-7000 Mhz (server anl.nkp.cz) pro management kooperačního systému a systém ORACLE 8i.
PC AMD K7-800 Mhz pro práci v oddělení analytického zpracování.

Vklad do projektu

Finanční prostředky z vlastních zdrojů

NKČR - vklad spočívá ve formě práce navíc a nebyl honorován finančně (kromě minimální částky OON a odměn). V rámci této práce byly prováděny **korektury** báze ANL (oprava chyb, které byly z velké části dány dvojí existencí báze v ISIS, kde se pracovalo a báze v ALEPHu, kam se zároveň konvertovaly ještě nezredigované záznamy kvůli rychlému zpřístupnění na Internetu v minulosti). Korektury se týkají také záznamů zpracovaných jinými institucemi.

Příprava propojení a propojení záznamů s plnými texty. Pracovníci oddělení analytického zpracování uváděli částečné adresy plných textů a zároveň tyto texty stahovali v rámci Konzorcia Anopress (příprava k propojení s plným textem). Dále byly propojovány bibliografické záznamy zpracované v rámci oddělení s plnými texty vystavenými na Internetu.

Ukládání záznamů - vzhledem k převedení 2 pracovních úvazků do jiných oddělení bylo nutno zachovat kontinuitu zpracování některých dokumentů.

Tvorba WWW stránek je velmi důležitá pro prezentaci kooperačního systému na WWW.

Vklad hlavní řešitelky představuje přes polovinu pracovního úvazku .

Finanční prostředky z jiných zdrojů

Anopress:

Zpřístupnění TamTam Professional k testování v NKČR a vývoji programových produktů v rámci grantu - hlavní řešitelce projektu k dispozici přibližně od května 2000.

Rok 2001

Tabulka

<u>Náklady na projekt celkem:</u>	1 376 000.-
<u>Výše státní dotace:</u>	1 056 000.-
<u>Neinvestiční prostředky:</u>	1 015 000.-
Služby	924 000.-
Mzdové prostředky	91 000.-
<u>Investiční prostředky:</u>	41 000.- (část plánovaných prostředků přesunuta do r. 2002)

Vklad do projektu: 320 000.-

Vlastní zdroje:

Národní knihovna ČR - odd. analytického zpracování 200 000.-

Finanční prostředky získané z jiných zdrojů (vklad):

Anopress, s.r.o. 120 000.-

Neinvestiční prostředky

	A	B	C	D
1	Neinvest.		Souborná databáze kooperačního systému článkové bibliografie	limit
2	Datum	Čís. fakt.	Popis	Částka
3			Služby (924 000 Kč)	
4	5.5.2001	2001-101	I. Mattern-plnotextové databáze-údržba, ladění (leden)-úprava hlaviček, linka aut.index.	10 000,00 Kč
5	5.5.2001	2001-103	I. Mattern-plnotextová databáze-údržba, ladění (únor)-úprava hlaviček, linka aut.index.	10 000,00 Kč
6	5.5.2001	2001-105	I. Mattern-plnotextové databáze-údržba, ladění (březen)-úprava hlaviček, linka aut.index.	10 000,00 Kč
7	5.5.2001	2001-108	I. Mattern-plnotextové databáze-údržba, ladění (duben)-úprava hlaviček, linka aut.index.	10 000,00 Kč
8	5.5.2001	2001-109	I. Mattern-dohledání plných textů, vytvoření aplikace pro jejich výběr (r. 97, 98)	12 000,00 Kč
9	5.5.2001	1134/2000	Anopress-přístup do databanky TAMTAM (srpen 2000, dluh)	15 120,00 Kč
10	5.5.2001	1667/2000	Anopress-přípr.práce pro plnotext.datab. (list.2000, dluh), nákup plných textů článků VII.	44 415,00 Kč
11	5.5.2001	0204/2001	Anopress-plnotextová databáze-údržba, ladění (leden)-nákup plných textů článků I.	52 500,00 Kč
12	5.5.2001	01/10344	Anopress-plnotextová databáze-údržba, ladění (únor)-nákup plných textů článků II.	52 500,00 Kč
13	5.5.2001	01/10495	Anopress-plnotextová databáze-údržba, ladění (březen)-nákup plných textů článků III.	52 500,00 Kč
14	19.5.2001	01/10666	Anopress-plnotextová databáze-údržba, ladění (1.pol.dubna)-převod NK do html, CD I.	19 250,00 Kč
15	19.5.2001	01/10135	Anopress-plnotextová databáze-údržba, ladění (2.pol.dubna)-převod NK do html, CD II.	33 250,10 Kč
16	28.5.2001	2001/001	D. Molitorisová-příprava a údržba www stránek s plnými texty na Internetu	3 000,00 Kč
17	9.6.2001	2001-111	I. Mattern - plnotextová databáze-údržba, ladění (květen)-úprava hlaviček, linka aut.index.	10 000,00 Kč
18	9.6.2001	2001-112	I.Mattern - filtr pro konverzi databáze do nového tvaru, další doplnění hlaviček, indexace	22 500,00 Kč
19	2.7.2001	01/10343	Anopress - plnotextová databáze-údržba, ladění (květen)-převod NK do html, CD III.	52 500,00 Kč
20	2.7.2001	2001-106	I. Mattern - optimalizace zpřístupnění plných textů - registrace uživatelů	30 000,00 Kč
21	2.7.2001	2001-116	I. Mattern - plnotextová databáze-údržba, ladění (červen)-úprava hlaviček, linka aut.index.	10 000,00 Kč
22	24.7.2001	2001/002	D. Molitorisová - www zpřístupnění metodických materiálů k popisu článků v UNIMARCu	8 000,00 Kč
23	9.8.2001	01/10507	Anopress - plnotextová databáze-správa, údržba, ladění (červen)-linka aut.indexace	52 500,00 Kč
24	9.8.2001	10100023	TOVEK - rešerše internetových zdrojů k jejich integraci do služeb NKČR	9 000,00 Kč
25	9.8.2001	20120069	Symantec Notron Antivirus 2001	1 672,70 Kč
26	9.8.2001	2001-118	I. Mattern - plnotextová databáze-údržba, ladění (červenec)-úprava hlaviček, linka aut.ind.	10 000,00 Kč
27	18.9.2001	01/10673	Anopress - plnotextová databáze-správa, údržba, ladění (červenec)-linka aut. indexace	52 500,00 Kč
28	18.9.2001	2001-120	I. Mattern - plnotextová databáze-údržba, ladění (srpen)-úprava hlaviček, linka aut.ind.	10 000,00 Kč
29	18.9.2001	01/10839	Anopress - plnotextová databáze-správa, údržba, ladění (srpen)-linka aut.indexace	52 500,00 Kč
30	23.10.2001	21110995	Anopress - plnotextová databáze-správa, údržba, ladění (září)-linka aut. indexace	52 500,00 Kč
31	10.11.2001	2001-124	I. Mattern - plnotextová databáze-údržba, ladění (září)-úprava hlaviček, linka aut.ind.	10 000,00 Kč
32	10.11.2001	2001-127	I. Mattern - plnotextová databáze-údržba, ladění (říjen)-úprava hlaviček, linka aut.ind.	10 000,00 Kč
33	21.11.2001	2111166	Anopress - plnotextová databáze-správa, údržba, ladění (říjen)-linka aut. indexace	52 500,00 Kč
34	27.11.2001	200121	J. Koltan-správa serveru ANL	20 000,00 Kč
35	30.11.2001	2001-130	I. Mattern - plnotextová databáze-údržba, ladění (listopad)-úprava hlaviček, linka aut.ind.	10 000,00 Kč
36	30.11.2001	2001-131	I. Mattern - plnotextová databáze-údržba, ladění (prosinec)-úprava hlaviček, linka aut.ind.	10 000,00 Kč
37	3.12.2001	21111195	Anopress - plnotextová databáze-správa, údržba, ladění (listopad)-linka aut. indexace	52 500,00 Kč
38	3.12.2001	21111196	Anopress - plnotextová databáze-správa, údržba, ladění (prosinec)-linka aut. indexace	52 500,00 Kč
39	4.12.2001	2001/1	I.Anděrová - vedení projektu, koncepce, organizace, podklady pro programy	8 292,20 Kč
40			Mzdové prostředky (91 000 Kč)	
41	31.8./6.11.2000		Střítecká, Kotrba, Molitorisová, Pašník, Magúthová, Chocholová, Holá, Levá, Rumlenová,	91 000,00 Kč
42			Kopp, Anděrová - stahování textů, příprava k propojení, testování linky, korektury, program	
43				
44	zůstatek			
45	čerpáno			1 015 000,00 Kč

Mzdové prostředky (OON) - detailní čerpání jednotlivých pracovníků oddělení analytického zpracování NKČR

2	Detailní čerpání u jednotlivých pracovníků:					
3			Souborná databáze kooperačního systému článkové bibliografie			
4	Datum vystavení	Jméno	Popis činnosti	Hodiny	Kusy	Částka
5	31.5.2001	Střítecká M.	Příprava k propojení, kontrola propojení	52	1040	7 800
6	12.6.2001		Korektury	48	1440	7 200
7	12.6.2001	Kotrba P.	Program na import do Alephu z TTDE	20		6 000
8	31.5.2001	Molitorisová D.	Příprava k propojení, kontrola propojení	20	400	3 300
9			Testování linky	2		
10	28.6.2001		Testování linky	8		1 200
11						
12	31.5.2001	Pašník, M.	Korektury	26	780	3 900
13	28.6.2201		Korektruy	22	660	3 300
14						
15	31.5.2001	Magúthová J.	Příprava k propojení, kontrola propojení	47	940	7 050
16	28.6.2001		Příprava k propojení, kontrola propojení, linka	10	200	2 250
17			Testování linky	5		
18						
19	31.5.2001	Chocholová J.	Příprava k propojení, kontrola propojení	47	940	7 050
20	28.6.2001		Příprava k propojení, kontrola propojení, linka	13	260	3 150
21			Testování linky	8		
22	31.5.2001	Holá E.	Příprava k propojení, kontrola propojení	28	560	4 200
23	28.6.2001		Příprava k propojení, kontrola propojení, linka	11	220	3150

24			Testování linky	10		
25	31.5.2001	Levá J.	Příprava k propojení, kontrola propojení	34	680	5 100
26	28.6.2201		Příprava k propojení, kontrola propojení, linka	21	420	3 150
27						
28	31.5.2001	Rumlenová A.	Příprava k propojení, kontrola propojení	35	700	5 250
29	28.6.2001		Příprava k propojení, kontrola propojení, linka	6	120	3 150
30			Testování linky	15		
31	26.11.2001		Příprava k propojení, kontrola propojení, linka	33	660	4 950
32	31.5.2001	Kopp P.	Příprava k propojení, kontrola propojení	31	620	7 350
33			Korektury	18	540	
34	28.6.2001		Příprava k propojení, kontrola propojení, linka	11	220	1 650
35	27.11.2001	Anděrová	Konrola linky	5		850
36	Suma			586		91 000
37						
38			Pracovní normy			
39			Příprava a kontrola propojení, linka aut. ind.+kont	20 propojení/hod.		150 Kč/hod.
40			Korektury	30 korektur/hod.		150 Kč/hod.
41			Testování linky automatické indexace			150 Kč/hod.
42			Kontrola linky automatické indexace			170 Kč/hod.
43			Program na off-line propojení			300 Kč/hod.

45		Sumarizace:	Hodiny	Kusy	Částka
46		Příprava a kontrola propojení, linka aut. ind.+kont	399	7980 x	59 850
47		Korektury	114	3420	17 100
48		Testování linky automatické indexace	48		7 200
49		Kontrola linky automatické indexace	5		850
50		Program na off-line propojení	20		6 000
51			586		91000
52		x z toho linka a kontrola	4749,	příprava a kontrola propojení	3231
53					
54		V případě pracovníků oddělení analytického zpracování šlo o experiment a ověření funkčnosti linky automatické indexace.			
55		Práce musela být částečně prováděna v pracovní době, protože šlo o ověřovací ologie. Kontroly propojení byly prováděny mimo pracovní dobu.			
56					
57					

Investiční prostředky

	A	B	C	D
1	Investice		Souborná databáze kooperačního systému článkové bibliografie	limit
2	Datum	Čís. fakt.	Popis	Částka
3				
4	23.5.2001	20/20041	PC DURON 750 Mhz - Gaia	41 000,00 Kč
5				
6				
7				
8				
9	zůstatek			
10	čerpano			41 000,00 Kč
11				

Vklad do projektu

Vklad pracovníků odd. analytického zpracování NKČR

	A	B	C	D	E	F
1			Souborná databáze Kooperačního systému článkové bibliografie			
2	Vklad pracovníků oddělení:			1 hod./150 Kč		200 000 Kč
3	Datum	Jméno	Popis činnosti	Hodiny	Kusy	Vklad v Kč
4						
5	31.8.2001	Brejniková D.	Záznamy	124	372	18 600 Kč
6	30.12.2001	Molitorisová, D.	www stránky, překlad zemí ke geograf. kódům	50		7 500 Kč
7	30.12.2001	Kindl	Propojení záznamů s volně dostup. texty www	70	1400	10 500 Kč
8						
9	Vklad	- mezisoučet		244		36 600,00 Kč
10						
11	Vklad hl.řešitelky			1 hod./200 Kč.		
12	Datum	Jméno	Popis činnosti	Hodiny	Kusy	Vklad v Kč
13	1.1./31.12.2001	Anděrová I.	Koncepční práce	100		20 000 Kč
14	1.1./31.12.2001		Redaktorské práce, metodická činnost	45		9 000 Kč
15	1.6./31.12.2001		Management projektu	80		16 000 Kč
16	1.5./31.12.2001		Částečná správa url adres	34		6 800 Kč
17	1.5./31.12.2001		Podklady pro programátorské práce	150		30 000 Kč
18	1.8./31.12.2001		Testování programů	140		28 000 Kč
19	1.1./31.12.2001		Studium materiálů, práce s Internetem	150		30 000 Kč
20	1.1./31.12.2001		Správa databáze kooperačního systému	88		17 600 Kč
21	1.1./30.5.2001		Podklady pro dodávky plných textů	20		6 000 Kč
22	Suma			807		163 400 Kč
23	Vklad	celkem				200 000,00 Kč

24			Pracovní normy			
25			propojení s volně dostupnými texty na www			
26			záznamy	20 propojení/hod.		150 Kč/hod.
27			testování linky automatické indexace	3 záznamy/hod.		150 Kč/hod.
28			kontrola linky automatické indexace			150 Kč/hod.
29			www stránky, překlad zemí ke geograf. kódům			170 Kč/hod.
30			činnosti hlavní řešitelky			150 Kč/hod.
31						200 Kč/hod.
32						
33			Vklad NK do projektu je ve formě práce navíc pracovníků oddělení analytického zpracování a hlavní řešitelky.			
34			V případě pracovníků oddělení byla práce dělána mimo prac. dobu, částečně v ušetřeném čase v prac. době.			
35			V případě hlavní .			
36			225 hod./celkem.			
37			Cca 150 běžných pracovních dnů bylo navýšeno v průměru o 2,5tj. 375 hodin celkem.			
38			Dále hlavní řešitelka vykonávala práci doma - cca 217 hod.			

Porovnání mzdových prostředků a činností v rámci řešených projektů VaV

	A	B	C	D	E	F	G	H	I	J	K	L	M
2	Porovnání činností oddělení analytického zpracování NKČR v rámci řešených projektů VaV												
3	r. 2001												
4	Celkové porovnání činností oddělení analytického zpracování NKČR (korektury, propojení záznamu												
5	s plným textem, tvorba nových záznamů) v rámci projektu Propojení analytických záznamů s plnými												
6	texty a optimalizace zpřístupnění plných textů, dále pak projektu Souborná databáze Kooperativního												
7	systému článkové bibliografie - optimalizace integrace a správy heterogenních dat (v tomto projektu												
8	oddělení vklad oddělení a vklad řešitelky pro názornost). Oba projekty se doplňují. První z nich je												
9	především koncepční, druhý pak realizační.												
10	V tabulce jsou porovnány mzdové prostředky v rámci projektů VaV a vklad oddělení analytického zpracování												
11													
12	Celkem činnosti v oddělení v r. 2001												
13	Činnost	Propojení analyt.zázn.			Soub. databáze Ko			Vklad oddělení			Vklad hl. řeš.		Celkem
14		Hodin	Kusů	Částka	Hodí	Kusů	Částka	Hodin	Kusů	Částka	Hodin	Částka	kusů
15													
16	Korektury				114	3 420	17 100						3 420 K
17	Propojení, linka	206	4 120	30 900	399	7 980	59 850	70	1400	10 500 Kč			13 50
18	Záznamy-popis	100	300	15 000				124	372	18 600 Kč			67
19	www stránky, překlad			2 100				50		7 500 Kč			
20	NKČR na WWW	65		13 000									
21	Testování linky				48		7 200						
22	Kontrola linky				5		850						
23	Progr.na off-line propojeníx	30		9 000	20		6 000						
24	Koncepční práce, management, údržba bá-										807	163 400	
25	ze, podklady pro program, práce aj.												
26	Celkem	401		70 000	586		91 000	244		36 600	807	163 400	
27	Poměry mezi linkou aut. indexace a pouhým propojením:												
28	Projekt Propojení anl.zázn.	4 120 linka a kontrola (on-line propojení)											
29	Projekt Spuborná dat.	7980 - z toho linka a kontrola 4749 (on-line propojení), příprava a kontrola propojení 3231 (off-line)											
30	Sumarizace:	celkem v lince automat. Indexace zpracováno v 2001, květen-listopad 8869 záznamů											
31		celkem propojeno off-line 3231 záznamů v 2001, leden-květen + 1400 volně dostup.textů											
32		x pracovník pracuje v jiném oddělení											

Vklad Anopress

	A	B	C	D	E	F	G
1							
2			Vklad Anopress, s.r.o	2001			
3	Datum	Jméno	Popis produktu	Hodiny	Kusy	Vklad v Kč	Kč/měsíc
4	r. 2001(6 měsíců)	Anopresss	TAMTAM Profesional		1	90 000	15 000
5	list. a pros. 2001	Anopresss	Konzorcium -licence		18	30 000	15 000
6	Celkem					120 000	

Komentář – účelové použití dotace

Neinvestiční prostředky

Služby

Ladění a další vývoj programové aplikace linky automatické indexace - automatického získávání plných textů, indexace bibliografických záznamů a plných textů, propojování záznamů s plnými texty on-line a zpřístupňování plných textů (rozšíření pracovního formuláře, úprava hlaviček, vyhledání – formuláře, rejstříky, topiky, zpřístupnění – uživatelské formáty, pracovní formáty – DC, XHTML, XML. Údržba serveru full.nkp.cz a báze anl.full (mj. indexace a konverze plných textů), aplikace pro přístup uživatelů (registrace, přihlášení) - I. Mattern.

Nákup plných textů odpovídajících profilu zpracovávaných článků za účelem propojení analytických bibliografických záznamů s těmito texty - Anopress, s.r.o. (cca 24 429).

Převod periodika Národní knihovna do html a převod na CD (1991-2000) - Anopress, s.r.o.

Anopress – údržba, ladění, linka (6 přístupů do databáze TamTam) v 2. polétí 2001.

Tvorba a údržba www stránek Kooperativního systému článkové bibliografie – D. Molitorisová.

Aplikace k off-line propojení záznamů s plnými texty – P. Kotrba.

Správa serveru ANL, aplikace pro řízení, správu a údržbu Kooperativního systému článkové bibliografie – experiment - ing. J. Koktan.

Byl zakoupen Northon Antivirus.

Mzdové prostředky

Pracovníci oddělení analytického zpracování uváděli částečné adresy plných textů a zároveň tyto texty stahovali v rámci Konzorcia Anopress (příprava k propojení s plným textem) 1.pololetí 2001, květen – prosinec 2001 popis v rámci linky automatické indexace, její testování a kontrola propojení. Dále byly propojovány bibliografické záznamy zpracované v rámci oddělení s plnými texty vystavenými na Internetu. Byly prováděny korektury bibliografických záznamů.

Investiční prostředky:

PC DURON 750 Mhz.

Vklad do projektu

Finanční prostředky z vlastních zdrojů

NKČR -vklad hlavní řešitelky spočívá ve formě práce navíc a nebyl honorován finančně.

Byly propojovány bibliografické záznamy zpracované v rámci oddělení s plnými texty vystavenými na Internetu.

Ukládání záznamů - vzhledem k převedení 2 pracovních úvazků do jiných oddělení bylo nutno zachovat kontinuitu zpracování některých dokumentů (obor knihovnictví).

Tvorba WWW stránek je velmi důležitá pro prezentaci kooperačního systému na WWW.

Příslušné počty, propojení, záznamů jsou ověřeny v databázi ANL.

Finanční prostředky z jiných zdrojů

Anopress - k dispozici přístup do TamTam Professional jako záložní zdroj, dále pak k testování a vývoji programových produktů v rámci projektu.

Rok 2002

Náklady na projekt celkem:

1 565 000.-

Výše státní dotace:

1 245 000.-

Neinvestiční prostředky:

1 015 000.-

Služby

924 000.-

Mzdové prostředky

91 000.-

Investiční prostředky:

230 000.-

Vklad do projektu: 320 000.-

Vlastní zdroje:

Národní knihovna ČR - odd. analytického zpracování

200 000.-

Finanční prostředky získané z jiných zdrojů (vklad):

Anopress, s.r.o.

120 000.-

Tabulky

Neinvestiční prostředky

1	Neinvestice	Souborná databáze Kooperativního systému článkové bibliografie	limit	1 015 000,00 Kč	
2	Datum	Čís. fakt.	Neinvestice	Částka	Zůstatek
3			Služby (924 000 Kč)		
4	9.5.2002	22110141	Anopress - ANL FULL - údržba, ladění, linka (leden)	52 500,00 Kč	962 500,00 Kč
5	9.5.2002	22110299	Anopress - ANL FULL - údržba, ladění, linka (únor)	52 500,00 Kč	910 000,00 Kč
6	9.5.2002	22110447	Anopress - ANL FULL - údržba, ladění, linka (březen)	52 500,00 Kč	857 500,00 Kč
7	9.5.2002	2002-136	I. Mattern - ANL FULL - údržba, ladění, linka (leden)	10 000,00 Kč	847 500,00 Kč
8	9.5.2002	2002-138	I. Mattern - ANL FULL - údržba, ladění, linka (únor)	10 000,00 Kč	837 500,00 Kč
9	9.5.2002	2002-141	I. Mattern - ANL FULL - údržba, ladění, linka (březen)	10 000,00 Kč	827 500,00 Kč
10	9.5.2002	2002-143	I. Mattern - ANL FULL - údržba, ladění, linka (duben)	10 000,00 Kč	817 500,00 Kč
11	17.5.2002	22110600	Anopress - ANL FULL - údržba, ladění, linka (duben)	52 500,00 Kč	765 000,00 Kč
12	11.6.2002	22110754	Anopress - ANL FULL - údržba, ladění, linka (květen)	52 500,00 Kč	712 500,00 Kč
13	24.6.2002	2002-145	I. Mattern - ANL FULL - údržba, ladění, linka (květen)	10 000,00 Kč	702 500,00 Kč
14	24.6.2002	2002-147	I. Mattern - ANL FULL - údržba, ladění, linka (červen)	10 000,00 Kč	692 500,00 Kč
15	9.7.2002	2241120	Anopress - ANL FULL - údržba, ladění, linka (červen)	52 500,00 Kč	640 000,00 Kč
16	16.8.2002	22410273	Anopress - ANL FULL - údržba, ladění, linka (červenec)	52 500,00 Kč	587 500,00 Kč
17	16.8.2002	2002-155	I. Mattern - ANL FULL - údržba, ladění, linka (červenec)	10 000,00 Kč	577 500,00 Kč
18	5.9.2002	22410438	Anopress - ANL FULL - údržba, ladění, linka (srpen)	52 500,00 Kč	525 000,00 Kč
19	5.9.2002	2002-158	I. Mattern - ANL FULL - údržba, ladění, linka (srpen)	10 000,00 Kč	515 000,00 Kč
20	31.10.2002	22410559	Anopress - ANL FULL - údržba, ladění, linka (září)	52 500,00 Kč	462 500,00 Kč
21	31.10.2002	2002-162	I. Mattern - ANL FULL - údržba, ladění, linka (září)	10 000,00 Kč	452 500,00 Kč
22	31.10.2002	2002-164	I. Mattern - ANL FULL - údržba, ladění, linka (říjen)	10 000,00 Kč	442 500,00 Kč

23	8.11.2002	22410729	Anopress - ANL FULL - údržba, ladění, linka (říjen)	52 500,00 Kč	390 000,00 Kč
24	8.11.2002	20220070	Upgrade PC Duron	20 986,50 Kč	369 013,50 Kč
25	8.11.2002	7/2002	J. Schwarz - analýza topiků	15 000,00 Kč	354 013,50 Kč
26	12.11.2002	20223845	Adobe Acrobat - 2 licence	8 862,00 Kč	345 151,50 Kč
27	18.11.2002	200221	J. Koktan - upgrade serveru ANL	59 000,00 Kč	286 151,50 Kč
28	18.11.2002	200224	J. Koktan - testování a ladění aplikace MNG, Báze titulu	38 000,00 Kč	248 151,50 Kč
29	18.11.2002	2002-166	I. Mattern - ANL FULL - údržba, ladění, linka (listopad)	10 000,00 Kč	238 151,50 Kč
30	18.11.2002	2002-168	I. Mattern - ANL FULL - údržba, ladění, linka (prosinec)	10 000,00 Kč	228 151,50 Kč
31	21.11.2002	účet	Copy general - kopie zprávy o projektu a vazby	1 330,50 Kč	226 821,00 Kč
32	25.11.2002	20220079	AVG - antivirus	3 969,00 Kč	222 852,00 Kč
33	26.11.2002	22410747	Anopress - ANL FULL - údržba, ladění, linka (list.,pros.)	105 000,00 Kč	117 852,00 Kč
34	9.12.2002	2002/2	I. Anděrová - management projektu	14 331,95 Kč	103 520,05 Kč
35	10.12.2002	A009225	Oracle JDeveloper	7 934,20 Kč	95 585,85 Kč
36	10.12.2002	A009226	Oracle JDeveloper - technická podpora	4 585,85 Kč	91 000,00 Kč
37					
38			Mzdové prostředky (91 000 Kč)		
39			Holá, Chocholová, Kindl, Kopp, Kozáková, Levá,		
40	8.6.2002		Magúthová, Molitorisová, Mráková, Rumlenová,	91 000,00 Kč	0,00 Kč
41			Vrbičanová		
42	zůstatek				0,00 Kč
43	čerpáno			1 015 000,00 Kč	

Mzdové prostředky (OON) - detailní čerpání jednotlivých pracovníků oddělení analytického zpracování NKČR

	A	B	C	D	E	F
1	Mzdové prostředky - detailní čerpání jednotlivých pracovníků					
2	Souborná databáze Kooperačního systému článkové bibliografie					
3	Datum vystavení	Jméno	Popis činnosti	Hodiny	Kusy	Částka v Kč
4	1.6.2002	Holá E.	Korektury jmenné (ANL)	35	700	
5			Kontrola záznamů ALEPH/TOPIC (linka)	20		5 500
6	1.6.2002	Chocholová J.	Korektury jmenné (ANL)	36	720	
7			Kontroly záznamů ALEPH/TOPIC (linka)	20		5 600
8	1.6.2002	Kindl D.	Globální korektury (ANL)	20		2 000
9	1.6.2002	Kopp P.	Kontrola záznamů ALEPH/TOPIC (linka)	35		3 500
10	1.6.2002	Levá J.	Kontroly záznamů ALEPH/TOPIC (linka)	30		3 000
11	1.6.2002	Magúthová J.	Kontroly záznamů ALEPH/TOPIC (linka)	35		3 500
12	1.6.2002	Molitorisová D.	www stránky - údržba (kontrola)	35		3 500
13	1.6.2002	Rumlenová A.	Kontrola záznamů ALEPH/TOPIC (linka)	20		
14			Korektury jmenné (ANL)	43	860	6 300
15	1.6.2002	Vrbičanová L.	Korektury jmenné (ANL)	24	480	2 400
16	22.11.2002	Holá E.	Korektury jmenné, věcné hromadné (ANL)	30	900	
17			Kontrola záznamů ALEPH/TOPIC (systém)	15		9 000
18	22.11.2002	Chocholová J.	Korektury jmenné, věcné hromadné (ANL)	30	900	
19			Kontrola záznamů ALEPH/TOPIC (systém)	14		8 800
20	22.11.2002	Kopp P.	Korektury jmenné, věcné hromadné (ANL)	24	729	
21			Kontrola záznamů ALEPH/TOPIC (systém)	4		5 600
22	22.11.2002	Kozáková A.	Korektury jmenné, věcné hromadné (ANL)	13	390	
23	22.11.2002	Kozáková A.	Autority	25		5 700

24	22.11.2002	Levá J.	Korektury jmenné, věcné hromadné (ANL)	19	570	
25			Kontrola záznamů ALEPH/TOPIC (systém)	1		4 000
26	22.11.2002	Magúthová J.	Korektury jmenné, věcné hromadné (ANL)	19	570	
27			Kontrola záznamů ALEPH/TOPIC (systém)	12		6 200
28	22.11.2002	Mráková D.	Práce pro progr. import záznamů, glob. opravy	36		9 000
29	22.11.2002	Rumlenová A.	Korektury jmenné, věcné hromadné (ANL)	33	990	
30			Kontrola záznamů ALEPH/TOPIC (systém)	4		7 400
31	Suma			632		91 000
32						
33			Pracovní normy			
34			Kontrola záznamů ALEPH/TOPIC (linka)			100 Kč/hod.
35			Korektury jmenné (ANL)	20 korektur/hod.		100 Kč/hod.
36			Kontrola záznamů ALEPH/TOPIC (systém)			200 Kč/hod.
37			Korektury jmenné, věcné hromadné (ANL)	30 korektur/hod.		200 Kč/hod.
38			Globální korektury (ANL)			100 Kč/hod.
39			Práce pro progr. import záznamů, glob. opravy			250 Kč/hod.
40			Korektury hromadné (ANL)			200 Kč/hod.
41			www stránky - údržba (kontrola)			100 Kč/hod.
42			www stránky - nové			200 Kč/hod.
43			Autority (sjednocování, podklady)			124 Kč/hod.

Investiční prostředky

1	Investice	Souborná databáze Kooperačního systému čl. bibliografie		limit	230 000,00 Kč
2	Datum	Čís. fakt.	Popis	Částka	Zůstatek
3	22.7.2002	207047	Upgrade serveru Dell - disk, instalace	13 664,00 Kč	216 336,00 Kč
4	22.7.2002	206825	Upgrade serveru Dell - Win2000, RAM	61 083,00 Kč	155 253,00 Kč
5	31.10.2002	211210	Dell Latitude (laptop)	65 624,00 Kč	89 629,00 Kč
6	31.10.2002	211209	Dell Workstation (PC)	70 015,00 Kč	19 614,00 Kč
7	12.11.2002	20220074	Tiskárna LaserJet	14 124,00 Kč	5 490,00 Kč
8	19.11.2002	20220075	UPS	5 490,00 Kč	0,00 Kč
9	zůstatek				0,00 Kč
10	čerpano			230 000,00 Kč	

Vklad do projektu

Vklad pracovníků odd. analytického zpracování NKČR

1	Vklad pracovníků oddělení			limit	200 000, 00 Kč
2	Souborná databáze Kooperačního systému článkové bibliografie				
3	Datum	Jméno	Popis činnosti	Hodiny	Vklad v Kč
4	1.1./31.12.2002	Anděrová I.	Koncepční práce	119	35 700 Kč
5	1.1./31.12.2002		Redaktorské práce, metodická činnost	45	13 500 Kč
6	1.6./31.12.2002		Management projektu	50	15 000 Kč
7	1.5./31.12.2002		Údržba portálu	16	4 800 Kč
8	1.5./31.12.2002		Podklady pro programátorské práce	150	45 000 Kč
9	1.8./31.12.2002		Testování programů	100	30 000 Kč
10	1.1./31.12.2002		Studium materiálů, práce s Internetem	100	30 000 Kč
11	1.1./31.12.2002		Správa databáze kooperačního systému	47	14 100 Kč
12	1.1./20.9.2002		Příprava prezentací (Inforum, Knihovny současnosti)	39,6	11 900 Kč
13	Suma			666,6	200 000 Kč
14					
15					
16					
17					
18			V případě hlavní řešitelky byla práce vykonávána mimo pracovní dobu , cca 9 pracovních sobot, 9 hod./ den,		
19			81 hodin celkem.		
20			Cca 150 běžných pracovních dnů bylo navýšeno v průměru 2,5 hod./den, tj. 375 hodin celkem.		
21			Cca 84,5 dnů vykonávala řešitelka práci doma v průměru 2,5 hod./den, tj. cca 211 hodin.		
22			Cca suma 667 hodin práce navíc, 300 Kč/hod.		

Vklad Anopress

	A	B	C	D	E	F	G
1	Vklad Anopressu	IT, a.s.	Souborná databáze Kooperačního systému článkové bibliografie				
2	Datum	Jméno	Popis produktu	Hodiny	Kusy	Vklad v Kč	Kč/měsíc
3	r. 2002(8 měsíců)	Anopress	TamTam Profesional		1	120 000 Kč	15 000 Kč
4							
5	Celkem					120 000 Kč	

Komentář - účelové použití dotace

Neinvestiční prostředky

Služby

Anopress – údržba, ladění, linka (6 přístupů do databáze TamTam).

I. Mattern - ladění a další vývoj programové aplikace linky automatické indexace/extrakce - automatického získávání plných textů, indexace bibliografických záznamů a plných textů, propojování záznamů s plnými texty on-line a zpřístupňování plných textů (úpravy pracovního formuláře, úprava hlaviček, vyhledání - formuláře, rejstříky, topiky, zpřístupnění - uživatelské formáty, pracovní formáty - DC, XHTML, XML). Částečně interface pro administraci báze ANL FULL a strukturovaného portálu pro volná www periodika (aktualizace dat, statistiky), převod linky automatické indexace/extrakce (TTDE) do www prostředí (hrazeno též z projektu Propojení analytických záznamů s plnými texty).V rámci údržby serveru full.nkp.cz a báze anl.full – částečně též hrazen upgrade a update serveru full.nkp.cz (instalace WIN 2000, nové verze TOPIC - verze 3.7 Portal One, zvýšení kapacity disku na 68 GB, RAM na 1 GB).

Ing. J. Koktan - upgrade serveru ANL (operační systém Linux SuSE 7.3, verze Oracle 9.2), ladění aplikace pro management KOSABI. J. Schwarz - analýza selekční úplnosti stávajících topiků na serveru full.nkp.cz.. I. Anděrová - koncepce a management projektu.

Z neinvestičních prostředků byl dále zakoupen upgrade PC, Adobe Acrobat (2 licence), AVG - antivirus, OracleJDeveloper - SW pro server ANL pro management Kooperačního systému článkové bibliografie.

Mzdové prostředky (OON)

Pracovníci oddělení analytického zpracování vykonávali tyto činnosti: opravy údajů jmenného a věcného popisu v rámci báze ANL (i záznamů kooperujících institucí) - překlepy, sjednocování popisu, kontrola oproti souborům autorit aj., M. Molitorisová navíc udržovala www stránky týkající

se projektu. Další pracovníci: D. Mráková (NKČR) – globální opravy, aktualizace podkladů pro program pro import záznamů do ALEPHu.

Investiční prostředky (tabulka Investice 2000):

Pro update a upgrade serveru Dell, na kterém je provozován TOPIC, bylo zakoupeno: 18 G disk, Win2000, 256 MB RAM. Byla hrazena instalace HW. Dále byl zakoupen laptop a PC (Dell Workstation). Jako navýšení investice byla zakoupena k PC tiskárna. K serveru ANL (zakoupenému v r. 2000 z investic) byl zakoupen záložní zdroj (UPS).

Vklad do projektu

Finanční prostředky z vlastních zdrojů

NKČR - vklad oddělení analytického zpracování - vklad hlavní řešitelky – koncepce, podklady pro programy a jejich testování, příprava prezentací a redaktorské práce, studium materiálů zejména na Internetu, správa databáze ANL kooperačního systému aj.

Finanční prostředky získané z jiných zdrojů

Anopress – k dispozici přístup do TamTam Professional jako záložní zdroj, dále pak k testování a vývoji programových produktů v rámci projektu.

Rok 2003

<u>Náklady na projekt celkem:</u>	1 235 000,-
<u>Výše státní dotace:</u>	915 000,-
Neinvestiční prostředky:	915 000,-
Služby:	824 000,-
Mzdové prostředky:	91 000,-
Vklad do projektu:	320 000,-
Vlastní zdroje:	
Národní knihovna ČR – odd. analytického zpracování	200 000,-
Finanční prostředky získané z jiných zdrojů:	
Anoress IT, s.r.o.	120 000,-

Tabulky

Neinvestiční prostředky

	A	B	C	D
1	Neinvestice	Souborná databáze Kooperacího systému článkové bibliografie		limit
2	Datum	Čís. fakt.	Neinvestice	Částka
3			Služby (824 000 Kč)	
4	23.4.2003	23410076	Anopress - ANL FULL - údržba, ladění, linka (leden)	52 500,00 Kč
5	23.4.2003	23410182	Anopress - ANL FULL - údržba, ladění, linka (únor)	52 500,00 Kč
6	23.4.2003	23410353	Anopress - ANL FULL - údržba, ladění, linka (březen)	52 500,00 Kč
7	7.5.2003	23410516	Anopress - ANL FULL - údržba, ladění, linka (duben)	52 500,00 Kč
8	15.5.2003	2003-174	I. Matterm - ANL FULL - údržba, ladění, linka (01/05)	37 500,00 Kč
9	3.6.2003	23410671	Anopress - ANL FULL - údržba, ladění, linka (květen)	52 500,00 Kč
10	16.6.2003	200311	J. Koktan - program na kontrolu URL adres	12 000,00 Kč
11	1.7.2003	2003-178	I. Matterm - ANL FULL - údržba, ladění, linka (červen)	7 500,00 Kč
12	2.7.2003	23410845	Anopress - ANL FULL - údržba, ladění, linka (červen)	52 500,00 Kč
13	3.7.2003	200315	J. Koktan - záloha ANL FULL (zrcadlo), export dat pro ANL	5 000,00 Kč
14	31.7.2003	2003-181	I. Matterm - ANL FULL - údržba, ladění, linka (červenec)	7 500,00 Kč
15	31.7.2003	2435	Instalace serveru full.nkp.cz po havárii serveru a IIS	4 685,00 Kč
16	20.8.2003	23410996	Anopress - ANL FULL - údržba, ladění, linka (červenec)	52 500,00 Kč
17	1.10.2003	23411210	Anopress - ANL FULL - údržba, ladění, linka (srpen)	52 500,00 Kč
18	6.10.2003	2003-183	I. Matterm - ANL FULL - údržba, ladění, linka (srpen)	7 500,00 Kč
19	6.10.2003	2003-187	I. Matterm - ANL FULL - údržba, ladění, linka (září)	7 500,00 Kč
20	1.10.2003	23411334	Anopress - ANL FULL - údržba, ladění, linka (září)	52 500,00 Kč
21	19.11.2003	2003-189	I. Matterm - ANL FULL - údržba, ladění, linka (říjen)	7 500,00 Kč
22	19.11.2003	23411467	Anopress - ANL FULL - údržba, ladění, linka (říjen)	52 500,00 Kč
23	24.11.2003	2003-192	I. Matterm - ANL FULL - údržba, ladění, linka (listopad)	7 500,00 Kč
24	24.11.2003	2003-193	I. Matterm - ANL FULL - údržba, ladění, linka (prosinec)	7 500,00 Kč
25	26.11.2003	23411541	Anopress - ANL FULL - údržba, ladění, linka (listopad)	52 500,00 Kč
26	26.11.2003	23411542	Anopress - ANL FULL - údržba, ladění, linka (prosinec)	52 500,00 Kč
27	2.12.2003	200326	J. Koktan - správa serveru ANL pro MNG KOSABI	20 000,00 Kč
28	2.12.2003	200327	J. Koktan - úpravy pro gramů (báze titul, opravy, export, import)	33 000,00 Kč
29	3.12.2003	2003-188	I. Matterm - konverze Unimarc M21	29 315,00 Kč
30			Mzdové prostředky (91 000 Kč)	
31	30./24.11.03		Chocholová, Kopp, Kozáková, Levá,	
32			Magúthová, Molitorisová, Rumlenová,	91 000,00 Kč
33			Vrbičanová, Anděrová	
34	zůstatek			
35	čerpáno			915 000,00 Kč
36				

Mzdové prostředky (OON) – detailní čerpání jednotlivých pracovníků oddělení analytického zpracování NKČR

	A	B	C	D	E	F	
2		Souborná databáze Kooperačního systému článkové bibliografie			2003		
3	Datum	Jméno	Popis činnosti	Hodiny	Kusy	Částka v Kč	
4	30.6.2003	Holá E.	Korektury jmenné, věcné, hromadné (ANL)	30	900	6 000	
5			Kontrola záznamů ALEPH/TOPIC (linka)	5		1 000	
6	30.6.2003	Chocholová J.	Korektury jmenné, věcné, hromadné (ANL)	38	1140	7 600	
7			Kontroly záznamů ALEP/TOPIC (linka)	5		1 000	
8	30.6.2003	Kopp P.	Korektury hromadné (ANL)	15		3 000	
9			Kontrola záznamů ALEPH/TOPIC (linka) 688 z.	15		3 000	
10	30.6.2003	Levá J.	Kontroly záznamů ALEPH/TOPIC (linka)	10		2 000	
11	30.6.2003	Magúthová J.	Korektury hromadné (ANL)	15		3 000	
12			Kontroly záznamů ALEPH/TOPIC (linka) ?	15		3 000	
13	30.6.2003	Molitorisová D.	Kontroly URL adres, věcné korektury	28	url+769	5 600	
14			www stránky - údržba (kontrola), tech. Redakce	15		3 000	
15	30.6.2003	Rumlenová A.	Kontrola záznamů ALEPH/TOPIC (linka)	15		3 000	
16			Korektury věcné, hromadné (ANL)	28	840	5 600	
17	30.6.2003	Vrbičanová L.	Korektury jmenné, věcné (ANL)	30	900	6 000	
18	24.11.2003	Holá E.	Korektury jmenné, věcné hromadné (ANL)	15	450	3 000	
19			Kontrola záznamů ALEPH/TOPIC (systém)	20		4 000	
20	24.11.2003	Chocholová J.	Korektury jmenné, věcné, hromadné (ANL)	17		3 400	
21			Kontrola záznamů ALEPH/TOPIC (systém).	15		3 000	
22	24.11.2003	Kopp P.	Korektury jmenné, věcné, hromadné (ANL)	15		3 000	
23			Kontrola záznamů ALEPH/TOPIC (systém)	15		3 000	
24	24.11.2003	Kozáková A.	Korektury jmenné, věcné, hromadné (ANL)	10	260	2 000	

Vklad do projektu

Vklad pracovníků odd. analytického zpracování NKČR

	A	B	C	D	E	F
1	Vklad pracovníků oddělení				limit	200 000, 00 Kč
2		Souborná databáze Kooperačního systému článkové bibliografie				
3	Datum	Jméno	Popis činnosti	Hodiny		Vklad v Kč
4	1.1./31.12.2003	Anděrová I.	Koncepční práce	119		35 700 Kč
5	1.1./31.12.2003		Redaktorské práce, metodická činnost	45		13 500 Kč
6	1.6./31.12.2003		Management projektu	50		15 000 Kč
7	1.5./31.12.2003		Údržba portálu volných www periodik	16		4 800 Kč
8	1.5./31.12.2003		Podklady pro programátorské práce	150		45 000 Kč
9	1.8./31.12.2003		Testování programů	100		30 000 Kč
10	1.1./31.12.2003		Studium materiálů, práce s Internetem	100		30 000 Kč
11	1.1./31.12.2003		Správa databáze kooperačního systému	47		14 100 Kč
12	1.1./20.9.2003		Příprava prezentací (Infos, SDRUK)	39,6		11 900 Kč
13	Suma			666,6		200 000 Kč
14						
15						
16						
17						
18		V případě hlavní řešitelky byla práce vykonávána mimo pracovní dobu , cca 9 pracovních sobot, 9 hod./ den,				
19		(většinou doma), tj. 81 hodin.				
20		Cca 150 běžných pracovních dnů bylo navýšeno v průměru 2,5 hod./den, tj. 375 hodin celkem.				
21		Cca 84,5 dnů vykonávala řešitelka práci doma v průměru 2,5 hod./den, tj. cca 211 hodin.				
22		Cca suma 667 hodin práce navíc, 300 Kč/hod.				

Vklad Anopress

	A	B	C	D	E	F	G
1	Vklad Anopressu	IT, a.s.	Soub. databáze Koop. systému čl. bibliografie 2003				
2	Datum	Jméno	Popis produktu		Kusy	Vklad v Kč	Kč/měsíc
3	r. 2003, (7,06) měsíců)	Anopress	TamTam Profesional - k dispozici přístup do báze TamTam		1	120 000 Kč	17 000 Kč
4							
5	Celkem					120 000 Kč	

Komentář – účelové použití dotace

Neinvestiční prostředky

Služby

Anopress, IT.- správa, údržba, linka (6 přístupů do databáze TamTam).

I. Mattern – částečné ladění a další vývoj programové aplikace linky automatické indexace/extrakce (lokální a částečně její internetové verze s propojením na rejstříky ANL FULL a národní autority s možností oprav metadat) - tj. automatické získávání plných textů, indexace bibliografických záznamů a plných textů, propojování záznamů s plnými texty on-line a zpřístupňování plných textů (úpravy pracovního formuláře, úprava hlaviček, vyhledání - formuláře, rejstříky, topiky, zpřístupnění - uživatelské formáty, pracovní formáty - DC, XHTML, XML) - vývoj internetové verze hrazen též z projektu Propojení analytických záznamů s plnými texty. Údržba a aktualizace portálu volně přístupných www periodik. V rámci údržby serveru full.nkp.cz a báze anl.full hrazena nová instalace serveru full.nkp.cz a databáze ANL FULL po havárii serveru (porucha IIS z dosud nezjištěných příčin a následně havárie disku – I. Mattern, fy Dell a Getronics). I. Mattern - částečně konverze do UNIMARC – M21.

Ing. J. Koktan – údržba serveru ANL a jištění báze ANL FULL (zrcadlení na serveru ANL a archivace na DVD), ladění aplikace pro management KOSABI (příjem, zpracování a export dat.), dialog programy pro editaci báze titulů kooperačního a bibliografických záznamů, aplikace pro kontrolu URL adres.

Mzdové prostředky (OON)

Pracovníci oddělení analytického zpracování vykonávali tyto činnosti: opravy údajů jmenného a věcného popisu v rámci báze ANL (i záznamů kooperujících institucí) - překlepy, sjednocování popisu, kontrola oproti souborům autorit, hromadné opravy, kontrola záznamů produkovaných v lince automatické indexace v Alephu a Topicu. D. Molitorisová navíc opravovala URL adresy v bázi ANL a spravovala www stránky v rámci technické redakce.

Finanční prostředky z vlastních zdrojů

NKČR - vklad oddělení analytického zpracování - vklad hlavní řešitelky – koncepce a podklady pro programy (též v projektu Propojení analytický záznamů s plnými texty), testování nových produktů, aktualizace portálu volných www periodik, příprava prezentací a redaktorské práce, studium materiálů zejména na Internetu, správa databáze ANL kooperačního systému aj.

Finanční prostředky získané z jiných zdrojů

Anopress – k dispozici přístup do TamTam Professional jako záložní zdroj, dále pak k testování a vývoji programových produktů v rámci projektu.

Rok 2004

<u>Náklady na projekt celkem:</u>	1 235 000,-
<u>Výše státní dotace:</u>	915 000,-
<u>Neinvestiční prostředky:</u>	915 000,-
<u>Mzdové prostředky:</u>	10 000,-
<u>Vklad do projektu:</u>	320 000,-
<u>Vlastní zdroje:</u>	
Národní knihovna ČR – odd. analytického zpracování	200 000,-
<u>Finanční prostředky získané z jiných zdrojů:</u>	
Anopress IT, s.r.o	120 000,-

Tabulky

Neinvestiční prostředky

	A	B	C	D
1	Neinvestice	Souborná databáze Kooperčního systému článkové bibliografie		limit
2	Datum	Čís. fakt.	Neinvestice	Částka
3			Služby (915 000 Kč)	
4	6.4.2004	24410063	Anopress - ANL FULL - údržba, ladění, linka, plné texty (leden)	61 000,00 Kč
5	6.4.2004	24410197	Anopress - ANL FULL - údržba, ladění, linka, plné texty (únor)	61 000,00 Kč
6	6.4.2004	24410322	Anopress - ANL FULL - údržba, ladění, linka, plné texty (březen)	61 000,00 Kč
7	14.4.2004	2004-201	I. Mattern - ANL FULL - údržba, ladění, linka (leden-duben)	30 000,00 Kč
8	10.5.2004	24410448	Anopress - ANL FULL - údržba, ladění, linka (duben)	61 000,00 Kč
9	21.5.2004	2004-204	I. Mattern - ANL FULL - údržba, ladění, linka (květen)	7 500,00 Kč
10	21.5.2004	2004-205	WWWTTDE - ladění konverze M21 v lince -část 1/2	39 500,00 Kč
11	17.6.2004	24410584	Anopress - ANL FULL - údržba, ladění, linka (květen)	59 500,00 Kč
12	29.6.2004	2004-207	I. Mattern - ANL FULL - údržba, ladění, linka (červen)	7 500,00 Kč
13	7.7.2004	24410806	Anopress - ANL FULL - údržba, ladění, linka (červen)	59 500,00 Kč
14	14.7.2004	20040701	Cosmotron - Toolkit klient pro linku zpracování	92 649,90 Kč
15	23.7.2004	2004-209	I. Mattern - ANL FULL - údržba, ladění, linka (červenec)	7 500,00 Kč
16	12.8.2004	24410915	Anopress - ANL FULL - údržba, ladění, linka (červenec)	59 500,00 Kč
17	3.8.2004	2004-211	I. Mattern - ANL FULL - údržba, ladění, linka (srpen)	7 500,00 Kč
18	9.9.2004	24411071	Anopress - ANL FULL - údržba, ladění, linka (srpen)	59 500,00 Kč
19	8.10.2004	2004-212	I. Mattern - ANL FULL - údržba, ladění, linka (září)	7 500,00 Kč
20	8.10.2004	24411233	Anopress - ANL FULL - údržba, ladění, linka (září)	59 500,00 Kč
21	27.10.2004	2004-215	I. Mattern - ANL FULL - údržba, ladění, linka (říjen)	7 500,00 Kč
22	8.11.2004	2004-216	WWWTTDE - ladění konverze M21 v lince, další úpravy -část 2/2	39 500,00 Kč
23	6.12.2004	200420	J. Koktan - správa serveru ANL , ladění aplikace MNG KOSABI	87 000,00 Kč
24	6.12.2004	200422	J. Koktan - jištění plných textů báze ANL FULL	3 000,10 Kč
25	6.12.2004	2004-218	I. Mattern - ANL FULL - údržba, ladění, linka (listopad-prosinec)	15 000,00 Kč
26	6.12.2004	2004-219	WWWTTDE -ladění linky - authority, konverze	11 850,00 Kč
27			Mzdové prostředky (91 000 Kč) - OON	
28	29.10.2004		J.Kotrba - program pro import dat z aplikace pro MNG KOSABI	10 000,00 Kč
29				
30				
31	zůstatek			
32	čerpáno			915 000,00 Kč

Vklad do projektu

Vklad pracovníků odd. analytického zpracování NKČR aj.

	A	B	C	D	E
1	Vklad pracovníků odd. analytického zpracování NKČR aj.				200 000, 00 Kč
2	Souborná databáze Kooperačního systému článkové bibliografie				
3	Datum	Jméno	Popis činnosti	Hodiny	Vklad v Kč
4	2004	Anděrová I.	Koncep. a analytické práce (napojení www linky na autority, upload)	100	35 700 Kč
5	2004		Redaktorské práce, metodická činnost	45	13 500 Kč
6	2004		Management projektu	50	15 000 Kč
7	2004		Údržba portálu volných www periodik	16	4 800 Kč
8	2004		Podklady pro program. práce a vývoj apl.www linky a MNG KOSABI	150	45 000 Kč
9	2004		Testování konverze M21, uploadu dat, uvedení do provozu www linky	100	30 000 Kč
10	2004		Studium materiálů, práce s Internetem	80	30 000 Kč
11	2004		Správa databáze kooperačního systému	47	14 100 Kč
12			mezisoučet	588	176 400 Kč
13		prac. odd.	testování www linky a uvedení do provozu	39,54	7 909 Kč
14	20.5.2004	odd. autom.	Výměna HDD - 18 G		15 691 Kč
15	Suma				200 000 Kč
16		V případě hlavní řešitelky byla práce vykonávána mimo prac. dobu, 11 prac. sobot po 10 hodinách,			
17		tj. 110 hodin , tj. 33 000 Kč.			
18		Cca 82 pracovních dnů bylo navýšeno v průměru o 2,5 hod. denně, tj. 205 hod., tj. 61 500 Kč.			
19		Cca 91 dnů vykonávala řešitelka práci doma, cca 3 hod. denně, tj. 273 hodin, tj. 81 900 Kč.			
20		Celkem vklad hlavní řešitelky je cca 176 400 Kč.			
21		Pracovníci	oddělení anl. zpracování testovali linku zpracování bibl. záznamů z plných textů	39,54 hod.,	
22		200 Kč/hod.	tj. 7 909 Kč.		

Vklad Anopress

	A	B	C	D	E	F	G
1	Vklad Anopressu	IT, a.s.	Soub. databáze Koop. systému čl. bibliografie				2004
2	Datum	Jméno	Popis produktu	Kusy	Vklad v Kč		Kč/měsíc
3	r. 2004 (7,06) měsíců	Anopress	TamTam Profesional - (jako náhradní zdroj)	1	120 000 Kč		17 000 Kč
4							
5	Celkem				120 000 Kč		

Komentář – účelové použití dotace

Neinvestiční prostředky

Služby

Anopress IT, a.s. – údržba, ladění linka (6 přístupů pro linku – leden-září 2004, říjen-prosinec nefinancován z tohoto projektu). I. Mattern - údržba a správa databáze serveru full.nkp.cz a báze ANL FULL, částečné ladění a další vývoj programové aplikace www linky automatické indexace/extrakce s napojením na národní autority(přes Z39.50 Toolkit Klienta) a s konverzí zpracovávaných dat do DC, UNIMARCu, M21, XHTML a XML. Údržba a aktualizace portálu volně přístupných www periodik Firma Cosmotron – Toolkit Klient pro linku automatické indexace/extrakce (linka automatického zpracování bibliografických záznamů z plných textů).

J. Koktan – údržba serveru ANL a jištění báze ANL FULL (zrcadlení na serveru ANL), ladění aplikace pro management KOSABI (příjem, zpracování a export dat.), dialog programy pro editaci báze titulů kooperačního a bibliografických záznamů, aplikace pro kontrolu URL adres.

Mzdové prostředky (OON)

P. Kotrba (NKČR) – program pro import dat z aplikace MNG KOSABI.

Vklad do projektu

Finanční prostředky z vlastních zdrojů

NKČR - vklad oddělení analytického zpracování - vklad hlavní řešitelky – koncepce a podklady pro programy, redaktorské práce a metodická činnost, management projektu, podklady pro program. práce a vývoj aplikace www linky a aplikace pro management Kooperačního systému článkové bibliografie, testování nových produktů a jejich uvedení do provozu (www linka, upload záznamů do báze ANL), aktualizace portálu volných www periodik, studium materiálů zejména na Internetu, správa databáze

ANL kooperačního systému aj. Testování www linky a uvedení do provozu – pracovníci odd. článkové bibliografie. Výměna hard disku (18 G) – firma Dell (financováno NKČR).

Finanční prostředky získané z jiných zdrojů

Anopress – k dispozici přístup do TamTam professional jako náhradní zdroj (započítáno 7,6 měsíce).

Finanční náklady za r. 2000-2004

Rok	Dotace	Investice	Neinvestice (z toho mzdy, materiál, licence, údržba)	Vklad odd. analytické ho zpracován í NKČR	Vklad Anopress	Celkem náklady na projekt
2004	915 000		915 000 (z toho mzdy 91 000)	200 000	120 000	1 235 000
2003	915 000		915 000 (z toho mzdy 91 000)	200 000	120 000	1 235 000
2002	1 245 000	230 000	1 015 000 (z toho mzdy 91 000)	200 000	120 000	1 565 000
2001	1 056 000	41 000	1 015 000 (z toho mzdy 91 000)	200 000	120 000	1 376 000
2000	2 320 000	1 295 000	1 025 000 (z toho materiál 48 000, licence 36 000, mzdy 91 000)	456 000	120 000	2 896 000
	6 451 000	1 566 000	4 885 000	536 000	600 000	8 307 000

E Resumé a klíčová slova

E.1 Resumé a klíčová slova v češtině

Náplní projektu je optimalizace integrace a správy heterogenních dat souborné databáze Kooperačního systému článkové bibliografie (KOSABI) – bibliografické báze ANL a plnotextové báze ANL FULL. Bibliografické záznamy článků, publikovaných v českém periodickém tisku jsou postupně propojované s elektronickou podobou článku (staticky z báze ANL, dynamicky v bázi ANL FULL) a metadata jsou uložena ve zdrojových kódech plných textů.

Výsledkem řešení projektu je rutinní provoz aplikace pro získávání a automatickou extrakci/indexaci bibliografických záznamů z plných textů ve formátu TXT získávaných z databáze TamTam (Anopress) pro lokální pracovní stanice Windows TTDE (TamTam Data Extractor) a následné vytvoření importního souboru záznamů v UNIMARCu pro databázi bibliografických záznamů ANL a plných textů pro bázi ANL FULL s vygenerovanými metadaty Dublin Core, UNIMARC v rámci linky. Automaticky se generuje URL a SICI. Ve stádiu ověřovacího provozu je aplikace WWW TTDE se stejnými funkcemi jako Windows TTDE navíc se vstupem ve formátu HTML z libovolného URL na WWW, s konverzí extrahovaných a editovaných dat do M21 (vyžaduje drobně ladit) a s funkcí stahování autoritních tvarů přes Z39.50 Tookit Klient, s výstupem v UNIMARCu (provozně ověřeno)/M21 pro bibliografickou bázi.

Plné texty z databáze TamTam pro linky zpracování TTDE jsou zajišťovány přístupem TTSNK (TamTam Special NK). Plné texty se také doplňují do báze ANL FULL v počátečních stádiích projektu off-line.

Výsledkem řešení je rutinní provoz serveru full. nkp.cz, aplikace v systému TOPIC v podobě báze ANL FULL s možností kombinace vyhledávání dle řízených termínů, klíčových slov z plného textu a předem definovaných strukturovaných dotazů - topiků (pojmu) v třístupňové hierarchii. Z uložených dat v TTDE se generuje nekvalifikovaný a kvalifikovaný formát XHTML a XML, dále pak 17 rejstříků.

V rutinním provozu je aplikace pro administraci databáze ANL FULL (statistiky, opravy, evidence uživatelů), aplikace pro registraci a přihlášení uživatelů, aplikace pro stahování a export metadat a plných textů (formát TXT, RDF, HTML, UNICODE, XML).

V rutinním provozu je aplikace Periodika na WWW včetně portálu pro zpřístupnění volně dostupných textů na Internetu ve struktuře tematické a regionální s částečným popisem zdrojů. Souborná databáze kooperačního systému ANL byla převedena do formátu M21. Je pravidelně aktualizována a je vydáván CD-ROM s Českou národní bibliografií též v M21.

Je vypracována metodika popisu článků ve formátu UNIMARC i M21, schválen minimální záznam pro soubornou databázi M21, ve věcném popise se koncem r. 2004 přistoupilo k postupné aplikaci metody konspektu v rámci systému. Při zpracování článků jsou stále více užívány soubory autorit.

Do báze pravidelně přispívají instituce spolupracující v Kooperačním systému článkové bibliografie, který je veden po stránce organizační i metodické. Je aktualizována koordinovaná titulová základna systému, která, ve srovnání s r. 2000, více zohledňuje zpracování odborných informačních zdrojů. Byly vytvořeny předpoklady pro zapojení nových institucí do systému v r. 2005.

V provozu je Aplikace pro management Kooperačního systému článkové bibliografie – server anl.nkp.cz, aplikace pro příjem, zpracování a export dat do ALEPHu. Součástí aplikace je automatizovaný upload dat do pomocné báze v několika formátech (Výměnný formát, UNIMARC, M21), formách a kódech, kontroly na UNIMARC a M21, statistiky chyb a kontrolovaný export do báze ANL. V provozu je i aplikace pro vedení Báže titulů - aktuálně zpracovávaných seriálů (zdroje sborníkového charakteru jsou vedeny zvlášť), aplikace pro kontrolu duplicit a platných URL adres. Nevylučuje se drobné ladění Aplikace pro Management KOSABI ad hoc.

V případech importů z linek TTDE a z Aplikace pro MNG KOSABI do báze ANL je zajištěna jejich kontrola automatizovaným hlášením o provedeném importu.

Projekt je koncepčně finančně průběžně saturován též projektu VaV Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů. V posledních třech měsících r. 2004 je přístup na plné texty do databáze TamTam hrazen z projektu Budování vzájemně kompatibilních informačních systémů k heterogenním informačním zdrojům a jejich zastřešení prostřednictvím JIB. Koncem r. 2004 byly vytvořeny předpoklady pro převedení zpracování zdrojů Anopress pod ALEPH a zapojení článků do oborových bran.

Uvedené nástroje by mohly v budoucnu částečně umožnit optimalizovat zpřístupnění i českých odborných textů, v případě WWW TTDE je nutné ověřit dlouhodobé větší zatížení v provozu.

Klíčová slova:

Plné texty; TOPIC; topiky; analytická indexace; záznam; seriály; články; zpřístupnění; souborná databáze; upload; propojování; Kooperační systém článkové bibliografie; management; Česká národní bibliografie; vyhledávání; automatická indexace; automatická extrakce; automatické shlukování; automatické abstrahování; ANOPRESS; KOSABI; ANL; ANL FULL; plnotextová databáze; pojmové vyhledávání; CD-ROM; UNIMARC; M21; Dublin Core; metadata ; HTML; XHTML; XML; digitální knihovna

E.2 ABSTRACT AND KEY WORDS IN ENGLISH

Abstract:

The contents of this project is optimization of integration and management of heterogeneous data which are involved in union bibliographic database ANL of the Co-operative system of Article Bibliography (COSABI) and full texts database ANL FULL. Bibliographical entries of articles published in Czech periodicals are linked with electronical form (static links in ANL database, dynamic links in ANL FULL database) and metadata are involved in source documents of full texts.

The project has resulted in routine system operation of acquisition and workflow of automated extraction indexing of bibliographical entries from Anopress fulltexts in TXT format designed for local workstations Windows TTDE (TamTam Data Extractor) and creating of imported file (in UNIMARC) for bibliographical database ANL and full texts database ANL FULL in NLCR with metadata embadement in full texts including automatically generated DC and UNIMARC in TTDE. URL and SICI is automatically generated, too. In pilot operation is a WWW application of acquisition and workflow of automated extraction indexing of bibliographical entries from fulltexts - WWW TTDE (with the same function as Windows TTDE) including sources in HTML format directly from their WWW URL (little improvement needed), plus conversion to M21 (little improvement needed), linking to authority files via Z39.50 Toolkit Klient is possible, output in UNIMARC (in operation)/M21,too.

TTSNK (TamTamSpecial NK) is enabling access to full text of TamTam Database. In the first stages of project have been acquired full text off-line, too.

Routine system operation of full.nkp.cz server is a reality, searchin in full text database ANL FULL via controlled terms, forms, key words of full text and topics - concepts (predefined structured queries in three hierarchical levels) is possible. From TTDE extracted and edited data are presented in ANL FULL in HTML, XHTML, XML (qualified, unqualified) formats and in 17 indexes, as well.

Application for ANL FULL administration has been in routine operation (statistics, corrections, evidence of users), application of user registration and authentication, download and export full texts is possible (formats: TXT, RDF, HTML, UNICODE, XML).

Application WWW Periodiclas including portal of www free texts in regional and thematic structure has been in a routine operation, too.

Methodic materials for article description in UNIMARC and M21 have been prepared, minimal data for description of articles in a union database have been approved. Gradual application of Conspectus method has been started, authority files have been used more.

Union Database of Article Bibliography has been converted to M21 and has been actualized regularly, CD-ROM with Czech National Bibliography has been published (in Marc 21, too).

Institutions have been participating in Co-operation system of Article Bibliography in a regular way (data imports). System has been managed methodically and as to the organization. Title database has been co-ordinated (orientation on branch titles), conditions for access of new institutions in 2005 has been worked out.

Application for Co-operation system of Article Bibliography has been in a routine operation – anl.nkp.cz. server, upload of data, data process and export to ALEPH. Data upload in several formats (Exchange format, UNIMARC, MARC 21), codes, controls and statistics, controlled export to ANL is possible. Application for administration of title database has been in operation (collection works separately). Application for URL valid control has been in function, too. A little improvement of Application for management of COSABI may be possible in the future ad hoc. In the case of imports from TTDE and Application for MNG COSABI to ANL database regular reports have been ensured.

This project has been supported as to the conception and partially financially from R&D project Analytical Bibliographic Records as a Gateway to Full Texts of Documents (1999-2003). In last three months 2004 access to TamTam full text database has been financed from the project Building of Mutually Compatible Information System for Access to Heterogeneous Information Resources under the Umbrella of the Uniform Information Gateway (2004-2010). At the end of 2004 conditions for transfer of cataloguing of Anopress full texts via ALEPH have been worked out, conditions for accessing articles into subject gateways, too.

Referred means may help in the future partially realize access to branch texts. WWW TTDE should be verified under higher upload in the future.

Key words:

Full texts; TOPIC; topics; analytical indexing; entries; serials; articles; access; union database; linking; upload; Co-operative system of Article Bibliography; COSABI; management; Czech National Bibliography; searching; machine-aided indexing; automatic extraction indexing; abstracting; clustering; ANOPRESS; COSABI; ANL; ANL FULL; fulltext database; concept based retrieval; CD-ROM; UNIMARC; M21; Dublin Core; metadata; HTML, XHTML, XML; digital library

31. ledna 2005

Mgr. Vlastimil Ježek, ředitel NKČR

PhDr. Ivana Anděrová, hlavní řešitelka

F Přílohy

- F.1 Báze ANL v NKČR (ALEPHU), JIB CASLIN a ANL/ANL FULL, ČNB- vývoj báze ANL od UNIMARCu k MARCu 21 a aplikaci konspektu pro zapojení do oborových informačních bran I**
- F.2 Linka zpracování bibliografických záznamů z plných textů pro lokální pracovní stanice – Windows TamTam Data Extractor (Windows TTDE) - bibliografická metadata pro bázi ANL (ALEPH ve formátu UNIMARC) a plné texty s Dublin Core pro bázi ANL FULL (TOPIC) XI**
- F.3 Linka zpracování bibliografických záznamů z plných textů – WWW TamTam Data Extractor (WWW TTDE) XXII**
- F.4 Zpřístupnění plných textů v systému TOPIC. Dokumenty. Architektura systému získávání, zpracování a zpřístupnění plných textů v systému TOPIC (báze ANL FULL a portál Periodika na WWW) – detailní ukázky formulářů, formátů, výstupů. Administrace (opravy a statistiky) XLVI**
- F.5 Aplikace pro správu a údržbu Kooperačního systému článkové bibliografie (architektura systému, jednotlivé aplikace pro příjem a zpracování dat, báze titulů) – Management Kooperačního systému článkové bibliografie LXX**
- F.6 Excerpční základna Kooperačního systému článkové bibliografie LXXXV**
- F.7 Statistiky XCVIII**
- F.8 Windows TTDE a WWW TTDE , TOPIC – dokumentace – hlavní programy a funkce s plánem do budoucna CIX**
- F.9 Konverze UNIMARC, DC, M21; Dublin Core a ANL Core - tabulka návěští pro vyhledávání v poli dotaz; tabulka formulářových polí, topiky, citace článku CXII**