

SOUBORNÁ DATABÁZE

KOOPERAČNÍHO SYSTÉMU ČLÁNKOVÉ BIBLIOGRAFIE

- OPTIMALIZACE INTEGRACE A SPRÁVY HETEROGENNÍCH DAT

Souhrnná zpráva za rok 2002

(Redakčně upraveno)

Předkládá: PhDr. Vojtěch Balík, ředitel NK
Zpracovala: PhDr. Ivana Anděrová, hlavní řešitelka



Praha, listopad 2002

OBSAH

A

Konstatační část

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

B

Analytická část

- B.0 Automatická či poloautomatická indexace a topiky ANL FULL - analýza
- B.1 Vlastní řešení
 - B.1.1 Architektura systému zpřístupnění plných textů, funkce systému procesy. Linka TTDE. Systém TOPIC a topiky
 - B.1.2 TOPIC (charakteristika)
 - B.1.3 Charakteristika serveru full.nkp.cz a báze ANL FULL. Periodika na WWW - portál
 - B.1.4 Metadata
 - B.1.5 Tisk, stahování a export plných textů a metadat - přístup z nabídky vyhledávání
 - B.1.6 Administrace
 - B.1.7 Standardizace - UNIMARC, Dublin Core, HTML, XHTML, XML, propojení záznamů s plnými texty
 - B.1.8 Souborná databáze ANL Kooperačního systému článkové bibliografie (KOSABI), metodika a organizace. ČNB
 - B.1.9 Perspektivy Kooperačního systému článkové bibliografie a jeho další vývoj, některé předpoklady a cíl
 - B.1.10 Management Kooperačního systému článkové bibliografie, aplikace MNG KOSABI
 - B.1.11 HW a SW podpora pro správu KOSABI a pro bázi ANL FULL
 - B.1.12 Smluvní ošetření projektu
- B.2 Přínos řešitele
- B.3 Posun znalostí

C

Návrhová část

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

D

Použití finančních prostředků

- D.1 Komentář
- D.2 Tabulky (vynechány)

E

Resumé a klíčová slova

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

F Přílohy

(začleněny do textu)

A Konstatační část

A.1 Rešerše

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.

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

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

CoBRA+ - Progress and Developments, June 1996 to January 1998. [online]. Dostupný z URL: <http://www.bl.uk/gabriel/projects/pages/cobra/cobprog.html>.

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

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

OLSON, Nancy B. Cataloging Internet Resources [online]. Dublin (Ohio, USA) : OCLC, 1997. Dostupný z URL: <http://www.purl.org/oclc/cataloging-internet>.

BRATKOVÁ, Eva. 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 URL: <http://inforum.cz/infomedia98/>.

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

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

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

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 z URL: nkpnew.2b.cz/pages/page.php3?page=oazp_prip.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 z URL: nkpnew.2b.cz/pages/page.php3?page=fond_ann_vf_opr.htm

JONÁK, Z. Reflektuje teorie informace a komunikace dostatečně na zvýšený zájem společenských věd o semiotické a komunikační aspekty života? *Ikaros* [online]. 1999, č. 3 [cit. 1999-03-01]. Dostupný z URL: <http://ikaros.ff.cuni.cz/1999/c03/veda2.htm>.

PAPÍK, R. Trendy v rozvoji informačních služeb. *Ikaros* [online]. 1999, č. 8 [cit. 1999-09-01]. Dostupný z URL: http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_papik.htm.

SVOBODA, Martin. Elektronické publikování. *Ikaros* [online], 1999, č. 3. Dostupný z URL: <http://ikaros.ff.cuni.cz/ikaros/1999/c03/elpubl98/index.htm>.

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

BURGETOVÁ, Jarmila. Právní aspekty poskytování knihovních elektronických a reprografických služeb. *Ikaros* [online], 1999, č. 6. Dostupný z URL: <http://ikaros.ff.cuni.cz/ikaros/1999/c06/repro.htm>.

HEIJTING, Inge. Interconnectivity and the Hybrid Library. *Ikaros* [online], 1999, č. 10. Dostupný z URL: <http://ikaros.ff.cuni.cz/ikaros/1999/c10/ebSCO.htm>.

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

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

TKAČÍKOVÁ, Daniela. Když se řekne digitální knihovna ... *Ikaros* [online], 1999, č. 8. Dostupný z URL: http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_tkacikova.htm.

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

Topic : systém pro inteligentní vyhledávání dokumentů. Praha : Tovek, 199?. JONÁK, Zdeněk. Krize mezilidské komunikace v období komunikační a informační exploze. *Ikaros* [online], 1999, č. 5. Dostupný z URL: <http://ikaros.ff.cuni.cz/ikaros/1999/c05/veda4.htm>.

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

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

KOSEK, J. XML pro každého : podrobný průvodce. 1. vyd. Praha : Grada 2000. 163 s.

ANDĚROVÁ, Ivana. Programový projekt MK ČR "Souborná databáze Kooperativního systému článkové bibliografie - optimalizace integrace a správy heterogenních dat". *Ikaros* [online]. 2000, č. 10 [cit. 2000-12-01]. Dostupný z URL: <http://ikaros.ff.cuni.cz/2000/c10/clanky.htm>. ISSN 1212-5075.

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

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

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

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

HORA, Michal a RICHTER, Vít. Veřejné informační služby knihoven - nový program pro občany a knihovny. *Ikaros* [online], 2000, č. 8. Dostupný z URL: <http://ikaros.ff.cuni.cz/ikaros/2000/c08/visk.htm>.

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

VOJTÁŠEK, Filip a CELBOVÁ, Iva. Helsinská univerzitní knihovna přívětivá vůči každému. *Ikaros* [online], 2000, č. 9. Dostupný z URL: <http://ikaros.ff.cuni.cz/ikaros/2000/c09/helsinky.htm>.

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

VOJTÁŠEK, Filip. Služby iDNES se rozrostly o regionálně členěný katalog odkazů. *Ikaros* [online]. 2001, č. 4 [cit. 2001-02-04]. Dostupný z URL: <http://www.ikaros.cz/Clanek.asp?ID=200208277>. ISSN 1212-5075.

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

Networked European Deposit Library [online]. Hague (Nizozemí) : Koninklijke Bibliotheek, last upd. 11-Mar-2001 [cit. 14. 3. 2001]. Dostupné z URL: <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 URL: http://nkpnew.2b.cz/pages/page.php3?page=oazp_propoj.htm.

ANDĚROVÁ, I. Souborná databáze Kooperačního systému článkové bibliografie - optimalizace integrace a správy heterogenních dat. Souhrnná zpráva za rok # [online]. Dostupný z URL: http://nkpnew.2b.cz/pages/page.php3?page=oazp_grant.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 URL: <http://full.nkp.cz>, <http://full.nkp.cz/nkk/NKKR.html> .

ANDĚROVÁ, Ivana. Metodika popisu článků ve formátu UNIMARC [online]. 2001. Dostupný z URL: http://nkpnew.2b.cz/pages/page.php3?page=oazp_popis1.htm.

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

CELBOVÁ, Ludmila. Informace o projektu registrace domácích internetových zdrojů nově na serveru WebArchiv. *Ikaros* [online]. 2001, č. 5 [cit. 2001-05-01]. Dostupný z URL: <http://ikaros.ff.cuni.cz/2001/c05/webarchiv.htm>. 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. Též dostupný z URL: <http://full.nkp.cz/nkk/NKKR.html>.

SCHWARZ, J. (2001a). Praktické aspekty hodnocení kvality a konzistence indexace. *Ikaros* [online]. 2001, č. 2 [cit. 2001-02-01]. Dostupný z URL: <http://ikaros.ff.cuni.cz/2001/c02/kvalind.pdf>

H.W. Wilson Company Selects Verity to Power the New WilsonWeb Site - the Premier Reference Resource for Librarians and Researchers [online]. Dostupný z URL: <http://www.verity.com/company/press/2001/20010108.html>.

Integration Heterogenous 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 URL: <http://www.aip.cz>, <http://full.nkp.cz>, [Rubrika Co je nového ...](#) .

ANDĚROVÁ, I. Báze ANL FULL v systému TOPIC. Prezentace PPT. Inforum 2002. Dostupný z URL: <http://full.nkp.cz>, [Rubrika Co je nového ...](#) .

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.

Knihovny současnosti 2002, Seč 24-26.9. 2002. PPT prezentace. Dostupný z URL: <http://www.mzk.cz/aktivity/sec.php3> .

ANDĚROVÁ, I. Kooperační systém článkové bibliografie - KOSABI. (Vývoj a současný stav metodiky zpracování, zpřístupnění, organizace kooperace, perspektivy). *Knihovny současnosti 2002*. Sborník z 10. konference, konané ve dnech 24.-26.září 2002 v Seči u Chrudimi. 2002, s. 223-255. Dostupný též z URL: <http://full.nkp.cz>, [Rubrika Co je nového ...](#) .

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 URL: <http://www.mzk.cz/aktivity/sec.php3>, <http://full.nkp.cz>, [Rubrika Co je nového ...](#) .

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. Dostupný též z URL: <http://www.mzk.cz/aktivity/sec.php3>

SVOBODOVÁ, E. Spolupráce paměťových institucí v rámci krajského bibliografického systému - Utopie. Nebo reálná možnost? *Knihovny současnosti 2002*. Sborník z 10. konference, konané ve dnech 24.-26.září 2002 v Seči u Chrudimi. 2002, s. 190-194. Dostupný též z URL: <http://www.mzk.cz/aktivity/sec.php3>

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. Dostupný též z URL: <http://www.mzk.cz/aktivity/sec.php3>

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. Dostupný též z URL: <http://www.mzk.cz/aktivity/sec.php3>

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

IFLA. Dostupný z URL: <http://www.ifla.org/act-serv.htm>

68th IFLA Council and General Conference, August 18-24, 2002. Glasgow. Dostupný z URL: <http://www.ifla.org/IV/ifla68/index.htm>.

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 URL: <http://www.ifla.org>, <http://www.ifla.org/IV/ifla68/papers/069-152e.pdf>.

DAGERSTEDT, S.: Cataloguing and organizing library workflow - New wals. 68th IFLA Council and General Conference, August 18-24, 2002. Glasgow. Dostupný z URL: <http://www.ifla.org>, <http://www.ifla.org/IV/ifla68/papers/067-152e.pdf>.

SMITH, R. The European Library Project: managing bibliographic standards at the European level. 68th IFLA Council and General Conference, August 18-24, 2002. Glasgow. <http://www.ifla.org/IV/ifla68/papers/068-152e.pdf>

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

SCHWARZ, J: Současný stav a trendy automatické indexace dokumentů. Přehledová studie. 2002. Dostupný z URL: <http://full.nkp.cz/nkdb/docs/studie/MAIobsah.html>

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. Dostupný též z URL: <http://www.mzk.cz/aktivity/sec.php3,259-265>.

Dieper : digitised European periodicals [online]. Dostupný z URL : <http://gdz.sub.uni-goettingen.de/dieper/home.htm>.

Serial Item and Contribution Identifier. Dostupný z URL: <http://sunsite.berkeley.edu/SICI/version2.html>.

SICI Generator. Dostupný z URL: <http://www.ep.cs.nott.ac.uk/~sgp/sicisend.html>.

Dublin Core Metadata Initiative Progress Report and Workplan for 2002. Dostupný z URL <http://www.dublincore.org/>.

Praktické výsledky projektů prezentované na www

Seriály (periodika) a analytický popis (články) v České republice, plné texty Propojení bibliografických záznamů s plnými texty [online]. Dostupný z URL: http://nkpnew.2b.cz/pages/page.php3?page=oazp_anal_popis.htm.

Metodika popisu článků ve formátu UNIMARC [online]. Dostupný z URL: http://www.nkp.cz/start/knihcin/OAZ/page.php3?page=oazp_popis1.htm.

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

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

Management Kooperačního systému článkové bibliografie - MNG KOSABI. Dostupný z URL: <http://full.nkp.cz>.

Plné texty v českých novinách a časopisech - přehled. Dostupný z URL: <http://full.nkp.cz>.

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

Seznam seriálů excerpovaných v oddělení analytického zpracování. Dostupný z URL: http://nkpnew.2b.cz/pages/page.php3?page=oazp_Seznam_OAZ.htm.

Knihovny spolupracující v kooperačním systému článkové bibliografie a excerpční základny Dostupný z URL: http://nkpnew.2b.cz/pages/page.php3?page=oazp_kooper_svk.htm.

Báze ANL [online]. Dostupný z URL: <http://sigma.nkp.cz:4505/ALEPH/7315F16RAY35G4NNL4MC7NUGSLSUVSNIYHCAGNR1SR47H9TYVB-01862/file/start-0>.

JIB Caslin. Dostupný z URL: <http://octopus.ruk.cuni.cz/>.

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 URL: <http://portico.bl.uk/gabriel/index.html>

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

Tel Digital deposits state of the art review. Marco de Niet, Koninklijke Bibliotheek. With contribution from Liesbeth Pskamp, Koninklijke Bibliotheek. 18 December 2001. 0.2 (Second draft version).D1.1/R/Report. DEL/007. Hague, Koninklijke Bibliotheek 2001. 84 p. Dostupný z URL: 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 URL: <http://www.ifla.org/VII/s16/pubs/directory.htm>.

A.2 Současný stav ve světě a v ČR

A.2.1 Obecně

Databáze citací článků a další sekundární zdroje informací patří v současné době mezi standardní služby, poskytované uživatelům knihoven spolu s přístupem do katalogů. Dostupné jsou z mnoha zdrojů rozptýlených po síti, z lokálních připojení, ze systémů CD-ROM. Technologie jako Z39.50 umožňují zavádění konzistentních uživatelských rozhraní pro širokou škálu databází přístupných po síti. Většina uživatelů používá databáze sekvenčně (vždy jen jednu), roste potřeba rozhraní, které by slučovalo záznamy, získané z několika databází do logické "souborné" databáze.

Elektronické dokumenty jsou zpřístupňovány prostřednictvím nakladatelství, distributorů, informačních institucí či služeb a jejich produktů, dále pak prostřednictvím digitálních knihoven a služeb vznikajících na základě projektů, konzorcií a licencí. Přístup k plným textům je zajišťován přes různé formy bibliografií a soupisů, obsahů časopisů a plnotextových databází. Vyhledávání v plných textech zvyšuje komfort přístupu uživatelů k informacím. **Při zpřístupňování elektronických informací se stále více prohlubuje spolupráce mezi státním a soukromým sektorem.** Elektronické dokumenty jsou zpřístupňovány v dohodnutých formátech. Služby knihoven jsou založeny na typu služby "document delivery".

Poměrně dobře jsou zpřístupňovány plné texty novin, týdeníků aj. časopisů. Problém vytváření vazeb na primární obsah se v současnosti soustřeďuje na článkové databáze proto, že technologie přístupu k datům v síti celkem dobře umožňuje přístup k článkům v elektronické podobě, zatímco přístup k jiným typům dokumentů je problematictější.

Kromě vyhledávačů typu "search engines" (Alta Vista aj.) nebo předmětových katalogů Internetu (Yahoo! aj.) registrující informační zdroje zatím v nestructurované podobě a u nichž relevance jejich zpětného vyhledávání je zatím problematictější, se na Internetu objevují registrační systémy, které přistupují ke zpracování těchto zdrojů přes strukturované záznamy. Tyto údaje mohou být obsaženy ve zdrojích samotných (metadata). **Je zdůrazňována nutnost průběžné implementace metadat do plných textů.**

Pro popis webovských informačních zdrojů navržen formát **Dublinské jádro (DC)** jako základní soubor údajů pro popis zdrojů. Dublin Core může být vytvářen autorem, vydavatelem, distributorem těchto zdrojů, knihovníkem.

Zpřístupnění relevantních, resp. pertinentních informací uživateli předpokládá jasná pravidla pro vytváření dokumentů jak na úrovni strukturální, tak na úrovni sémantické. Zdá se, že možným nástrojem na úrovni strukturální má největší perspektivy formát či jazyk **XHTML, XML (eXtensible Markup Language)**.

XML považován v současné době za nástupce jazyka HTML. Jeho aplikací je **RDF (Resource Description Framework)**, který má definovanou standardní DTD (Document Type Definition). Implementace souboru metadat DC (i MARC) ve struktuře **RDF/XML** je předpokladem efektivního vyhledávání a využívání digitálních informací, tj. efektivní komunikace na www.

Velká pozornost se věnuje protokolům pro komunikaci a sdílení dat (např. **Z39.50 a Bath Profile**).

Propojují se katalogy knihoven, záznamy s plnými texty dokumentů, "síťové dokumenty", je podporována spolupráce s muzeji, 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. Velká pozornost je věnována standardizaci popisu fondů. **Informační brány a portály** jsou na pořadu dne.

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**.

Propojení mezi dokumenty může být **statické i dynamické**, na základě "**base URL**" a "**open URL**".

A.2.2 Současný stav v zahraničí

(služby, knihovny aj. instituce, projekty, nástroje a strategie týkající se zpřístupnění článků a plných textů v zahraničí)

SLUŽBY, DATABÁZE, PORTÁLY, DIGITÁLNÍ SBÍRKY, KNIHOVNY AJ. INSTITUCE

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. Plné texty jsou zpřístupňovány jak soukromými společnostmi, tak knihovnami.

- **Ingenta** poskytuje přístup k elektronickým verzím časopisů vydávaných v některých předních nakladatelstvích (**Academic Press, Arnold, Blackwell Publishers, Blackwell Science, Elsevier Science** aj.). Její součástí jsou online časopisy dalších nakladatelů zpřístupňované službou **CatchWord**. Uživatelé mají přístup k informacím k cca 13 milionu článků z více než 27 000 vědeckých časopisů 190 nakladatelství z větší části ve formátu pdf. Přístup k plným textům článků je placený na základě uživatelského oprávnění. Je možno prohledávat v online katalogu volně. V roce 2000 získala Ingenta službu **UnCover**. Katalogy obou služeb jsou dnes integrovány do jednoho přístupového bodu pod názvem **UnCoverPlus**.
- **Web of Science: citační** databáze z 8500 periodik (pouze **abstrakty**).
- **ERIC**: databáze plných textů článků z oblasti **vzdělávání**. Od roku 1966 přes milion záznamů z 980 periodik.
- **ProQuest**: databáze plných textů článků z 7250 periodik (viz dále). ([ukázky](#))
- **EBSCO**: databáze plných textů článků ze 3000 periodik všech vědních oborů (viz dále). ([ukázky](#))
- **ScienceDirect**: databáze plných textů více než 1200 odborných časopisů nakladatelství Elsevier Science.
- **Literature Online**: 300 000 plných **literárních textů** z britské a americké literatury.
- **Arts and Humanities Data Service** - služba ve Velké Británii organizovaná King's College London - tvorba a uchovávání digitálních sbírek ze společenských věd.
- **Portály**: Science Gateway (biologie, chemie, vědy o zemi aj.) a Social Science Gateway (sociální vědy).
- Nejkompletnější přístup k elektronickým časopisům nabízí **OCLC FirstSearch Electronic Collection Online**. Interface umožňuje prohledávání periodik i čísel podle různých kritérií, přístup k citacím časopisů je zdarma, přístup k abstraktům a plným textům je možný jen u předplacených titulů, OCLC podporuje konzorciální přístup včetně přístupu kombinovaného s individuálním a "document delivery" (viz dále).
- Ve **Švédsku**, jsou články prezentované v systému **LIBRIS**. Záznam článku je možno zobrazit ve zkrácené podobě i ve struktuře MARC. Formulář obsahuje hypertextové odkazy na knihovny, v jejichž fondu se titul nachází.
- V **DBC** (Dánském knihovnickém centru) se zpracovávají články a recenze v rámci báze **BASIS** (ročně 30000 článků a 20000 recenzí z 9000 dánských periodik). Toto centrum buduje souborný katalog **DANBIB**, přes který lze zaslat objednávku elektronické kopie článku. Je propojený se švédským souborným katalogem LIBRIS a norský souborným katalogem **BIBSYS**.
- Univerzitní knihovna v **Helsinkách** (plní funkci Finské národní knihovny) provozuje centrální knihovnický systém VTLS síť **Linnea**, v rámci které je zpřístupňována databáze článků **ARTO** z tisíce finských periodik. Ve Finsku existuje rozsáhlý projekt pro zpřístupňování elektronických zdrojů - **FinLib**.
- **NORDINFO** - projekt skandinávského souborného virtuálního katalogu zohledňuje aspekty specifické pro severské země - předmětová hesla, klasifikační systémy, národní systémy identifikátorů.
- **The Nordic Metadata project** - kooperační projekt severských zemí jako jeden z prvních řeší problematiku metadat Dublin Core v rámci spolupráce Norska, Dánska, Švédska, Finska a Islandu.

- Systém knihoven **Oxfordské univerzity** zpřístupňuje rozsáhlou sbírku elektronických dokumentů v rámci Electronic Reference Library. Záznamy článků obsahují krátké citace, abstrakty a možnost získání elektronické kopie ve formátu PDF.
- Program **PICA v Holandsku** zajišťuje přístup k centralizovaným bibliografickým databázím a zejména vytváří tzv. Otevřenou síť knihoven (OBN - Open Bibliothek Network), v rámci které je umožněno elektronické propojení knihoven s databází článků
- V **Německu** je vyvinut projekt **JADE** (Journal Articles Database) ve spolupráci s British Library. Obsahuje pouze krátké citace článků (tato báze obsahuje i záznamy z NKČR). JADE je doplněna projektem **JASON** (Journal Article Send On Demand). JASON umožňuje dodávání článků v elektronické podobě z německé databáze časopisů.
- **National Parliamentary Library of Georgia**, Gruzie, Tbilisi - Respublica - databáze novinových článků (1993-).
- **Martynas Mazvydas National Library of Lithuania**, Vilnius, Litva
 - **LIBIS** (Lithuanian Integral Library Information System) - Bibliografická databáze článků (1994-).
 - **TEISE** - bibliografická databáze článků s právní problematikou.
 - **TEISPOL** - bibliografická databáze s právní a oblasti politiky (1997-).
 - **EKONOMIKA** - bibliografická databáze obor ekonomika (1995-).
- **Koninklijke Bibliotheek**, Haag, **Holandsko** - meta katalog obsahující více než 20 databází. Data konvertována do XML, plnotextové vyhledávání. Online Contents KB (OLC) - Catalogue of journal articles (1992- , přes 1.400.000 záznamů článků z holandských periodik).
- **National and University Library**, Ljubljana, **Slovinsko** - Kooperativní databáze COBIB ve Slovinsku obsahuje 1,200 000 záznamů knih, seriálů, článků. Je zároveň suborným katalogem.
- **Consorti de Biblioteques Universitaries de Catalunya**, Barcelona, **Španělsko** - Kooperativní databáze obsahující 2 milionů článků.
- **Analyticky se zpracovávají články také v Lotyšsku a Maďarsku.**
- **H.W.Wilson**

Bibliografické a plnotextové **báze dat americké firmy H.W. Wilson** pro společenskovědní a humanitní obory (WilsonWeb a OmniFile) patří mezi nejlepší na světě. **Tato báze je do určité míry též vzorem pro budování kooperační databáze článkové bibliografie a má s ní mnoho společných metodických prvků** - proto se o ní zmiňujeme podrobněji. (Jistým vzorem pro bázi ANL FULL je také služba ProQuest 5000 - viz dále).

Základní skupina databází - dříve pouze bibliografických, později referátových/abstraktových a dnes i plnotextových - pokrývá různé tematické oblasti.

Excerptovaná periodika/seriály a rovněž tak články v nich obsažené jsou vybírány s ohledem na názory a podněty amerických, zejména veřejných knihoven. Specifická spolupráce při výběru časopisů je realizována také s Odborem referenčních a uživatelských služeb Americké knihovnické asociace (American Library Association Reference & User Services). Každá báze má stanovenou svoji excerptční základnu s tím, že některá periodika se excerptují i pro více bází dat. Cílem firmy je vytvářet báze záznamů/plných textů z klíčových periodik v dané oblasti.

Zpracování záznamů z hlediska formálního i věcného je precizní. Firma Wilson buduje svůj vlastní řízený předmětový heslář (automatizovaný soubor předmětových autorit), který je založen na hesláři Kongresové knihovny LCSH. Heslář je obohacován novými hesly proto, že jsou zpracovávány článkové dokumenty, které obsahují detailnější tematiku, nejnovější poznatky z různých oborů. Firma zaměstnává velké množství profesionálů - knihovníků, katalogizátorů, indexátorů i oborových specialistů. Editoři bází dat kontrolují. Záznamy obsahují v průměru 2-6 předmětových hesel. Vedle předmětového hesláře se buduje soubor jmenných autorit pro záznamy jmen osob a korporací. V případě potřeby jsou při bibliografickém popisu doplňovány málo významové názvy článků o další klíčová slova.

Pozornost je věnována tvorbě abstraktů/referátů. Informace lze vyhledávat a zobrazovat několika způsoby, tisknout a stahovat záznamy i plný text dokumentu, buď ve formátu HTML nebo PDF.

V systému WilsonWeb lze uplatnit v zásadě 3 základní způsoby vyhledávání:

1. jednoduché vyhledávání
2. formulování dotazu prostřednictvím formuláře s rozšířenou nabídkou selekčních údajů/polí
3. vyhledávání prostřednictvím slovníků jednotlivých databází; tento způsob poskytuje možnosti vyhledávání hypertextem

Při formulaci dotazů je možno využít operátory, vyhledávání v polích, zkracování selekčních údajů apod. Ve všech databázích jsou k dispozici seznamy selekčních polí formální a zejména věcné povahy . Velmi důležitým údajem je typ dokumentu (nekrology, recenze divadelních her, oper apod.).

- **GABRIEL (The Gateway to Europe's National Libraries) - údaje o člancích jsou ponechány v originále :**

Souborné katalogy seriálů (noviny a časopisy) nebo databáze přístupné na Internetu ve většině evropských národních knihoven (Belgie, Řecko, Finsko, Německo, Island, Norsko, Polsko, Slovensko, Španělsko, Švýcarsko, Švédsko, Bulharsko, Maďarsko, Holandsko, Norsko, Portugalsko, Slovinsko, Rakousko, Finsko, Francie, Polsko).

Články přístupné ve speciálních databázích či dohromady s ostatními dokumenty, někde jako součást národní bibliografie, digitální knihovny článků, databáze zpřístupňující obsahy periodik ve formě analytického popisu :

- Deutsche Bibliothek Frankfurt am Main, NLG OPAC.
- Nationalbibliografien von Liechtensteinische Landesbibliothek, Liechtenstein, Online-Benutzekatalog (OPAC), ALEPH.
- OPACs of Koninklijke Bibliotheek, Netherlands. KB-catalogue.
- National Union Catalogue of Koninklijke Bibliotheek, Netherlands, Netherlands Central Catalogue NCC, Pica.
- Online Services of Koninklijke Bibliotheek, Netherlands. Bibliotheca Van der Linde-Niemeijeriana, Catalogue of the chess collection. Articles. Databases of the Special Collections Department. Speciální bibliografie: Poortman's Bibliography of Dutch Philosophy, BL Online, The bibliographical database of linguistics.
- OPACs of Narodna i univerzitetska biblioteka Sv. Kliment Ohridski - Skopje, Republic of Macedonia, COBIB (Co-operative Online Bibliographic Macedonian National Bibliography of Macedonian Journal Articles).
- OPACs of Slovenská národná knižnica (SNK), Slovakia, the Slovak National Bibliography. ALEPH.
- National Bibliographies of Narodna biblioteka Sv. Sv. Kiril i Metodii (NBKM), Bulharka, BULNB.
- National Bibliographies of Narodna in univerzitetna knjižnica, Slovenia, Slovenska bibliografija, COBISS/OPAC.
- National Union Catalogue of Slovenian Libraries, COBIB.
- National Union Catalogue of Biblioteca Nacional, Portugal, Porbase.
- Finland, Helsingin Yliopiston Kirjasto - Suomen kansalliskirjasto. Indices to Periodical Contents of Helsingin Yliopiston Kirjasto - Suomen kansalliskirjasto, Finland. ARTO, 470000 záznamů
- Iceland, Landsbókasafn Íslands - Háskólabókasafn (Lbs.). GREINIR, přes 30 000 článků.
- Lithuania, Lietuvos Nacionalinė Martyno Mazvydo Biblioteka . Bibliographic database of articles from Lithuanian periodicals since 1994.
- Norway, Nasjonalbiblioteket. NORART, Norwegian index to periodical articles. 303198 záznamů.
- United Kingdom, The British Library. Indices to Periodical Contents of The British Library, United Kingdom. Inside.
- The Finnish Historical Newspaper Library. 44 titulů.
- France, Bibliothèque nationale de France. Gallica (monografie a periodika)

- Germany, Die Deutsche Bibliothek . Digital Exile Periodicals - Exilpresse Digital.

PROJEKTY, SMĚRY A NÁSTROJE PRO INTEGRACI HETEROGENNÍCH ZDROJŮ

Projekty

Projekt Renardus: Akademický tematický portál konsorcia 12-ti institucí. Řešen v rámci 5 tého rámcového programu EU "Technologie pro informační společnost". Renardus má umožňovat paralelní pohyb uživatele po tematických portálech (metadata DC, Z39.50, DDC).

Architektura pro britskou národní digitální knihovnu UK DNER (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 projektů 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.

Připravují se **nástroje pro převod dat Dublin Core/MARC** (Library of Congress: Dublin Core/MARC/GILS crosswalk), které využívají prvky **DC v katalogizaci**. V Evropě je to projekt **Nordic Metadata** v rámci kterého byl vytvořen konvertor dat, který je schopen generovat záznamy ve formátech MARC severovýchodních zemí a USMARC ze zdrojových údajů DC.

Služba OCLC - **Cooperative Resource Catalog (CORC)** 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.

Nástroje integrace heterogenních dat: XML, identifikace zdrojů, propojování, protokoly, digitální knihovny viz též zpráva z r. 2001

DC (Dublin Core), XML (eXtensible Markup Language), RDF (Resource Description Framework), propojování - FRBR, URL, PURL, URN a DOI, SFX, informační brány - viz podrobně zpráva z r. 2001

Automatická či poloautomatická indexace dokumentů (citace z materiálu SCHWARZ, P. [Současný stav a trendy automatické indexace dokumentů](#). Přehledová studie. 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.

Metody automatické indexace nepřinášejí zatím dostatečné výsledky proto, že automatické porozumění textu je v podstatě v počátcích a jsou používány metody, který celý proces indexace podstatně zjednodušují a tím snižují kvalitu výsledků, nehledě na časovou a finanční náročnost vývoje těchto technologií. Na druhé straně, automatická indexace umožňuje odbourat subjektivní faktory intelektuální indexace, které mají vliv především na konzistenci indexace, a podstatně snížit časovou náročnost indexačního procesu.

V současnosti je výzkum a vývoj systémů automatické indexace ve fázi, kdy nelze hovořit o plně automatické indexaci. Technologie automatické indexace jsou většinou implementovány jako hybridní systémy, ve kterých se uplatňuje automatická indexace coby automatizovaná podpora intelektuální činnosti indexátora. Základní princip takových hybridních systémů spočívá v tom, že pomocí automatizované indexace je připraven seznam termínů, které jsou potenciálními kandidáty pro výběr indexačních termínů. Tento

výběr v závěrečné fázi indexace provádí indexátor. Jak ukazují některé výzkumy, tento způsob indexace může nejen zrychlit a zjednodušit indexační proces, ale i zkvalitnit jeho výsledky, zejména z hlediska míry přesnosti. **Automatická indexace patří do širší oblasti automatizovaného (strojového) zpracování textu (*text processing*), resp. obecně zpracování přirozeného jazyka (*natural language processing*).** Přestože se podařilo vyvinout řadu funkčních systémů pro automatickou indexaci, více než čtyřicetiletá snaha zatím nevedla k vývoji systémů, které by byly plně funkční z hlediska ideálních nároků na úplnou automatizaci procesu indexace a na univerzálnost těchto systémů (většina v současnosti fungujících systémů je specificky oborově zaměřená). Intelektuální indexace přináší ve srovnání s řadou automatických procedur (vč. např. latentního sémantického indexování) ještě stále lepší výsledky. Systémy automatické indexace jsou také účinné pouze částečně proto, že doposud nebyly dostatečně prozkoumány a podrobně popsány všechny intelektuální procesy, které probíhají při indexaci.

Na druhou stranu je potřeba uvést, že řada technologií, jejichž účinnost byla v předchozích letech potvrzena výzkumem a řadou studií, je v současnosti implementována ve formě expertních systémů nebo systémů pracujících na základě umělé inteligence.

Obecně lze konstatovat, že systémy automatické indexace vyvíjejí oborově zaměřené instituce, které zpracovávají velké objemy dokumentů, které je nezbytné kvalitně a konzistentně indexovat. V řadě případů se však jedná o dokumenty, u kterých je dostupný pouze komprimovaný text (např. abstrakt), a které je tudíž žádoucí indexovat.

Příklady systémů automatické indexace

Jedná se o plně funkční systémy, které jsou provozovány rutinně nebo v testovacím provozu. U jednotlivých systémů není označeno, zda se jedná o systémy pracující na bázi automatické extrakce nebo automatického přiřazování, protože většina systémů tyto dva přístupy kombinuje.

Media On Line Project

[\(ukázky\)](#)

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 systémy 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).

Konzorcia

Pro přístup elektronickým zdrojům zejména k plným textům článků se zakládají konzorcia. Výhody: získání zdrojů jinak nedostupných, přístup malým institucím/pracovištím, příznivější ceny na instituci, využití centrálních finančních zdrojů, levnější provoz (administrativní a organizačně-technické výhody), spolupráce.

Členy některých konzorcií jsou i české knihovny a instituce.

V ČR mají uživatelé přístup k mnoha kvalitním zahraničním zdrojům elektronických informací díky **programu MŠCHT "LI - Informační zdroje pro vědu a výzkum"**

- **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 I Collection, obsahující všechny články od prvního čísla časopisu, s retrospektivou do minulého století, až po současnost).

- **ProQuest 5000 International**

[\(ukázky\)](#)

Národní akademická licence pro produkty **ProQuest 5000 International** v rámci grantu financovaného MŠMT umožňuje přístup k online službám ProQuest 5000 a PCI Web. Přístup je na základě registrace prostřednictvím online formuláře. Služba obsahuje bibliografické záznamy článků 7000 zahraničních periodik, plné texty cca 4000 titulů. Licence pokrývá medicínu, humanitní a společenské obory, obchod, aplikované přírodní vědy, výpočetní techniku a telekomunikace. Vyhledávání je jednoduché i pokročilé, pomocí rejstříků a tzv. topiků, tj. témat. Báze ANL FULL se v některých aspektech podobá ProQuestu, **kteřý je do určité míry inspirací pro českou plnotextovou databázi.**

- **EIFL-Direct, EBSCO**

[\(ukázky\)](#)

Zpřístupnění plnotextových databází odborných zahraničních periodik na základě programu Open Society Institute EIFL-Direct - **plošná multilicence.**

Plné texty celkem 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,** nabízené ve 4 dílčích databázích:

- Academic Search Elite (společenské a humanitní vědy)
- Business Source Premier (ekonomie, finance, management, účetnictví, mezinárodní obchod)
- Newspaper Source Plus (přes půl milionů článků z více jak 100 novin v angličtině)
- MasterFILE Premier (obecné zájmové tituly, obchod, zdraví, kultura)

Zahrnuje i databázi Medline z oblasti lékařství a biomedicínského výzkumu.

- **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 12 dalších databází: ArticleFirst (bibliografické citace článků z 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, sympozií, 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.

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"). Pro otázky týkající se **problematiky novin je určen Round Table of Newspapers, z kterého se v r. 2002 stala také samostatná sekce.** Jedním z cílů sekce je spolupráce s nakladateli, autory, dodavateli a institucemi, které se zabývají zpřístupněním informací obsažených v seriálech a podpora aktivit týkajících se akvizice těchto dokumentů.

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 Serials Content. V rámci tohoto průzkumu byly poskytnuty informace o zpracování článků v ČR.

Na půdě **Round Table on Newspapers** se řeší problematika akvizice, zpřístupnění a archivace těchto

dokumentů s důrazem na monitoring nejnovějších technologií, které tyto procesy mohou v knihovnách ovlivnit. Je zdůrazněna kooperace a spolupráce s institucemi, které se zpracováním novin zabývají a bibliografická kontrola těchto zdrojů a informací v nich obsažených.

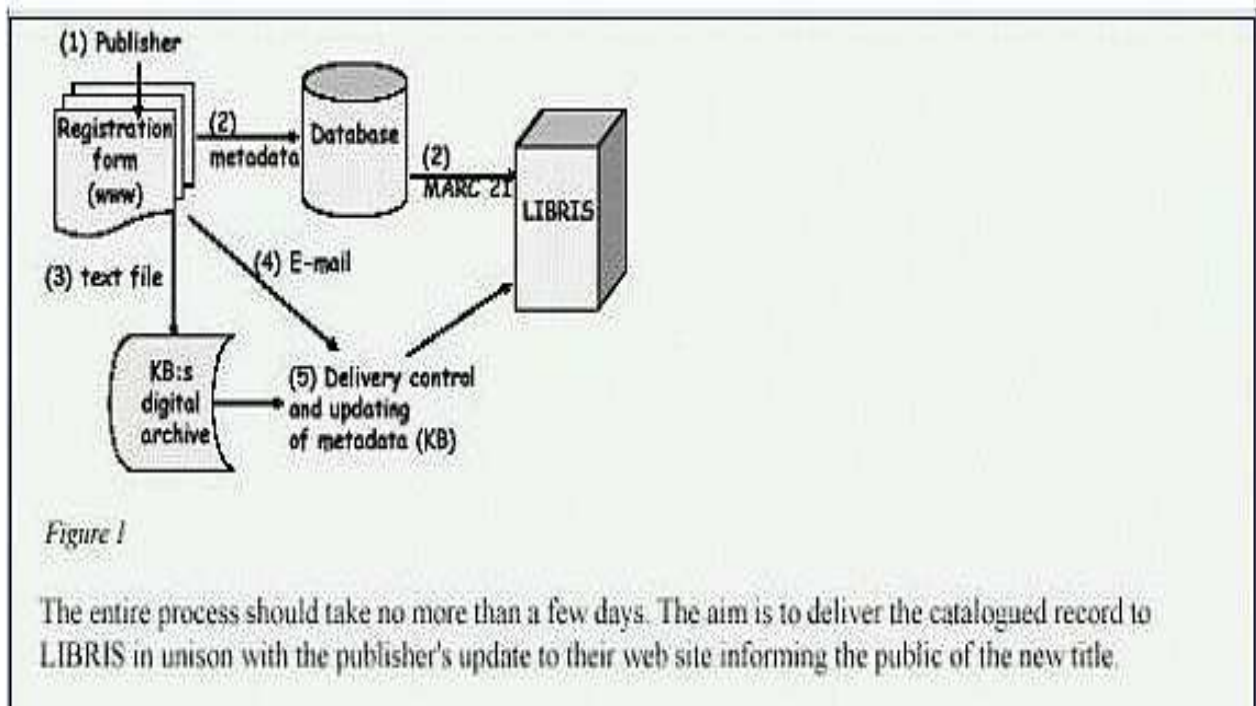
- **Austrálie**

Australské zkušenosti prezentované na konferenci: National Bibliographic Database, de facto australská národní bibliografie, přistupuje k novému modelu nové 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.**

Tento přístup v mnohém připomíná metody, řešené v rámci předkládaného projektu.

- **Švédsko**

Velmi podobný model jako je v předkládaném projektu je také vyvíjen 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.

- **Popis, standardizace**

Další příspěvky konference jsou kromě jiného zaměřeny na automatický sběr dat, na metody věcného popisu a vyhledávání na Internetu (Německo), je zdůrazněna spolupráce s institucemi zabývající se zpřístupněním elektronických dokumentů (Dánsko, Norsko).

V rámci Sekce IFLA pro seriálové publikace je revidován standard pro popis seriálů s názvem **International Standard Bibliographic Description for Serials and Continuing Resources, ISBD (CR).** Tato revize může mít dopad i na dosavadní metodu popisu článků. Podobně je potřeba reflektovat standard pro popis elektronických zdrojů ISBD (ER) a příslušné kapitoly AACR2R.

- **TEL**

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.

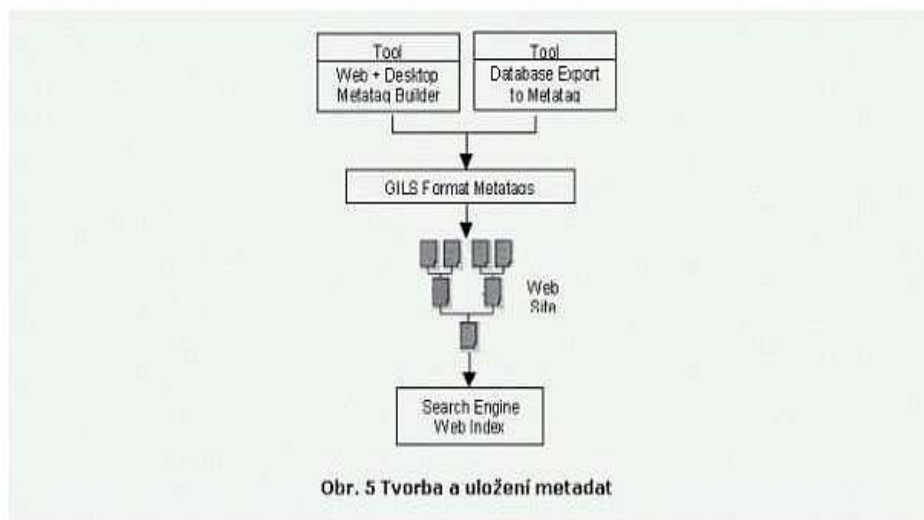
Z dotazníku: pouze 5 národních knihoven denně aktualizuje dokumenty pro digitální knihovnu on-line (Česká republika, Dánsko, Německo, Nizozemí, Velká Británie). Pracovní linky jsou v ČR, Německu, Lotyšsku, Nizozemí, Švédsku. 3 knihovny jsou v kontaktu s IT společnostmi. 4 knihovny automaticky konvertují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

V č. 10 /2002 **News from the IFLA Round Table 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 výstřihků (Norsko, Austrálie, Itálie, Španělsko, Švédsko, Německo).**

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

GILS jako služba a nástroj pro řízení informačních zdrojů z oblasti státní správy USA

GILS jako služba a nástroj pro řízení informačních zdrojů z oblasti státní správy USA



V rámci služby kooperují kromě jiných institucí knihovny v jednotlivých státech. 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 zdrojích.

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. Tento systém by mohl být v mnoha směrech inspirativní i pro nás (je založen na kooperaci a automatizovaném zpracování dat).

Další konference

The National Bibliography: from Print to the Digital Age. Riga, September 12-13, 2002 se zabývala otázkami týkající se bibliografie, povinného výtisku, elektronickými publikacemi. Zazněly též příspěvky týkající se analytické bibliografie lotyšské a maďarské.

Problematice elektronického publikování pro vědu, průmysl a "general public" je věnována **6th International Conference on Electronic Publishing at Karlovy Vary, Czech Republic. ELPUB 2002++ "Technology Interactions" ++ November 06-09th, 2002.**

A.2.3 Současný stav v ČR

Některé zdroje přístupné na českém Internetu, digitální knihovny, služby a projekty zabývající se zpřístupněním sekundárních informací o článcích a zpřístupněním plných textů

Některé knihovny vytvářejí soupisy volných českých zdrojů plných textů na Internetu. Plné texty jsou organizovány na Internetu do tematicky či jinak utříděných portálů.

Relativně velký rozvoj na Internetu nastal v nabídce **českých novinových a časopiseckých elektronických zdrojů** - 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.

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.

Např. nakladatelství **Economia a.s.**, nakladatelství ekonomické literatury vystavuje na Internetu plné texty produkce tohoto nakladatelství. Server IHNEP nabízí pokročilé vyhledávání ve zdrojích i řazení výsledku podle relevance.

Sagit (elektronické nakladatelství - právní texty), **Tigis s.r. o.** (časopisy pro lékařskou odbornou veřejnost a edukační časopisy pro veřejnost), **Nakladatelství Muzikus** - hudební tematika, **Portál**.

Current Contents, abstrakty, citace, plné texty

AVČR - Časopisy vydávané Akademií věd České republiky. Některé časopisy jsou vybaveny abstraktem a plným textem, někde pouze obsahy časopisů.

Nakladatelství Karolinum, Nakladatelství Univerzity Karlovy - vydávání učebních textů, vědeckých monografií, sborníků vědeckých prací, slovníků a vědeckých časopisů - current contents.

Odborná knihovnická periodika

Národní knihovna. Knihovnická revue - samostatná webová prezentace periodika v rámci Projektu propojení analytických záznamů s plnými texty ve formátu html a pdf - r. 1999 - 2002 s rejstříky, do r. 2001 plnotextově přístupná na serveru full.nkp.cz (báze ANL FULL) a www.anopress.cz. V dalších letech se plánuje vytvoření jednotného interface s bází ANL FULL a vytvoření topiku pro tento časopis. **Ikaros** - elektronický časopis o informační společnosti. **Bulletin SKIP, U nás, Knihovní obzor**.

Některé digitální knihovny a databáze ASPI (Automatizovaný systém právních informací). **Digitální knihovna "Český parlament" a "Dokumenty Senátu"**. **Digitální knihovna v NKČR** obsahuje vzácné a ohrožené dokumenty digitalizované v národních programech Memoriae Mundi Series Bohemica (převážně rukopisy, staré tisky a perspektivně další dokumenty) a Kramerius (starší noviny a časopisy a další vzácné dokumenty tištěné na kyselém papíře). V r. 2002 byla zpřístupněna digitální knihovna rukopisů, prvotisků a starých tisků v krajské knihovně [v Olomouci](http://v.Olomouci).

Informační agentury

Albertina icome

Albertina icome Praha je česká soukromá společnost zaměřená na zpřístupnění profesionálních informačních zdrojů v elektronické formě a jejich využití v praxi. AiP nabízí elektronické tituly předních světových vydavatelství. Elektronické vydavatelství spolupracuje na vydávání ČNB na CD-ROM.

Newton I.T.

Elektronická výstřižková služba, elektronický archiv novin a časopisů celostátních a regionálních. Buduje archivy některých deníků a časopisů.

Anopress IT, a.s.

On-line databanka novin a časopisů celostátních a regionálních, monitoring na zakázku, vědomostní databáze, archivy zdrojů. Spolupracuje s NKČR v rámci předkládaného projektu a výzkumného záměru Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů. Anopress IT, a.s. umožňuje on-line přístup do databanky plných textů **TamTam**, na jejíž bázi poskytuje následné služby. Společnost zpřístupňuje informace zákazníkovi na dané téma. Anopress umožňuje přístup do databanky novin on-line na základě licenčních smluv a umožňuje nákup celých titulů periodik. Společnost Anopress je výhradním zpracovatelem elektronické podoby většiny českých regionálních titulů (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. Agentura od roku 1998 průběžně vytváří rozsáhlou databanku, která 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.

Anopress zpracovává zatím cca 35 titulů, které odpovídají excerpční základně Kooperačního systému článkové bibliografie.

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. Agentura plní své dosavadní závazky a smlouvy.

V r. 2002 je podprogram VISK8 - Informační zdroje - linie A zaměřen na zajištění dostupnosti elektronických informačních zdrojů formou multilicencí z domácí, české proveniencí. **V září 2002 je poskytnuta dotace MKČR pro multilicenční zpřístupnění databází TamTam a ČTK v rámci VISK 8.**

Analogicky vyplývá: v budoucnu podobné zpřístupnění báze ANL FULL a licence resp. multilicence pro přístup do báze TamTam pro linku zpracování bibliografických záznamů z plných textů.

Společnost se zúčastnila 12. akvizičního semináře konaného ve Středočeské vědecké krajské knihovně, 12.6.2002.

Projekty, metody a nástroje, související se zpřístupněním plných textů

Národní lékařská knihovna - katalogizace elektronických periodik dostupných v rámci licencí. Některé záznamy elektronických zahraničních časopisů byly předány do STK v rámci projektu Portal STM a měly by být zahrnuty do terciální databáze. Dále by měla být zahrnuta česká elektronická periodika online. V současné době pokus o katalogizaci internetových domácích zdrojů z oblasti lékařství a zdravotnictví.

Státní technická knihovna - elektronické časopisy se katalogizují v terciální databázi v rámci projektu LI01018 z oblasti STM. Uvažuje se o komerčních tematických portálech. Na stránkách knihovny je databáze on-line časopisů podle oborů a databáze českých ISSN.

[Portál STM - Elektronické informační zdroje STM](#)



A&I Databáze

Food Science and Technology Abstracts	Silver Platter
VINITI	Žurnál VINITI
CambridgeScientificAbstracts	Internet Database Service - CSA Materials Science Collection with METADEX
Biological Abstracts	SilverPlatter
Zoological Record	SilverPlatter
MathSci	American Mathematical Society, Silver Platter
EMBASE - Excerpta Medica	Elsevier Science B.V.
MEDLINE (R)+	National Library of Medicine
MICROMEDEX(R)	MICROMEDEX Thomson Healthcare
Web of Science	Institut for Scientific Information
AGRICOLA	EBSCO
INSPEC	DIALOG
Journal Articles DatabasE	British Library
World Bibliographies on CD-ROM	K. G. Saur
ABSEES	The American Bibliography of Slavic and East European Studies
Iron and Steel Works of the World	Metal Bulletin
Continuous Casting Series CD-ROM, Volumes 1-7	Iron and Steel Society
Mezinárodní desetinné třídění na CD ROM	Albertina icome, s.r.o. Praha
CETEDOC Index of Latin Forms	Brepols Publishers

Terciální databáze

A &I Databáze

Elektronické informační zdroje na VŠ

[WebArchiv](#) je vytvářen v rámci programového projektu výzkumu a vývoje "Registrace, ochrana a zpřístupnění domácích elektronických zdrojů v síti Internet". 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.

Do češtiny byla přeložena nejnovější verze standardu Dublin Core Metadata Element Set, Version 1.1 proběhla lokalizace metadatového formuláře převzatého od Helsinské univerzitní knihovny z projektu Nordic Metadata. 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.

[Jednotná informační brána](#)

Projekty JIB:

Jednotná informační brána pro hybridní knihovny (VAV, UK, NK).

Portál STM (STK a 11 institucí).

Jednotná informační brána (VISK 8) - provoz JIB, zpřístupnění zdrojů.

Česko- slovenský virtuální katalog Caslin.

Projekt JIB umožňuje základní prohledávání inf. zdrojů, jejich tematický výběr, stahování záznamů. Zdroje jsou volně přístupné a licencované. Aplikace metody konspektu.

Komunikace v Metalib je na základě protokolu Z39.50 nebo HTTP.

Báze ANL a ANAL (Olomouc) zpřístupněna v JIB jako zdroj. Báze ANL FULL a články krajské knihovny v

Liberici zpřístupněny v JIB jako odkaz. Plné texty báze ANL FULL a přes bázi ANL přístupné v JIB pro externí uživatele na 7 dnů. V řešení jsou připojeny dalších institucí a zdrojů.

Metoda konceptu jako nástroj popisu fondů má přispět k realizaci této brány. Základem metody základem je popis dle věcného třídění na několika úrovních. První obsahuje 24 tematických skupin, druhá 500 kategorií, které se dále člení na 4 000 témat. Vazba na jednotnou informační bránu: výběr věcného oboru v Metalibu podle členění tematických skupin.

Při věcném popisu báze ANL se užívají také předmětové kategorie (od r. 1994) určené pro zařazení článků do hrubých oborů či témat pro zpřehlednění báze. Domníváme se, že tyto kategorie budou podrobnější než ty, aplikované v metodě konceptu. Podobně topiky na serveru full.nkp.cz se používají tři úrovně definice - tematická oblast, skupin témat, detailní témata.

Propojování - Metalib a SFX

SFX je standard pro propojení dokumentů a je využíván v Metalib k vytváření vazeb. V Metalibu je statický způsob propojení nahrazován dynamickým propojováním založeném na open URL.

Autorskoprávní a legislativně právní problematika zpřístupňování elektronických dokumentů - v ČR je třeba aktualizovat zákon o povinném výtisku seriálových publikací, event. autorský zákon.

Povinný výtisk elektronických publikací je předpokladem jejich uchování a zpřístupnění. Dále je třeba v budoucnu uzavírat dohody mezi knihovnami a příslušnými nakladateli a vydavateli, které se budou týkat jednak zpřístupnění elektronických dokumentů, jednak spolupráce.

Automatická či poloautomatická indexace (citace z práce J. Schwarze viz výše)

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 komputační 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, komputační a korpusové lingvistiky pod vedením P. Sgalla; vzniklo specializované pracoviště (dnes Ústav formální a aplikované lingvistiky/Centrum komputační lingvistiky na MFF UK), znalosti a zkušenosti jeho pracovníků byly zúročeny i v současném projektu *Pražského závislostního korpusu*.

Automatická indexace sněmovních tisků v KPS PČR

Parlamentní knihovna jako odbor Kanceláře Poslanecké sněmovny Parlamentu ČR od r. 2000 indexuje v testovacím provozu sněmovní tisky, od 4. volebního období (červen 2002) přešla indexace sněmovních tisků do rutinního provozu. V rámci indexace sněmovních tisků byla ve spolupráci s Odborem informatiky KPS PČR a diplomantem M. Urbanem (VŠE Praha) implementována automatická indexace založená na automatické extrakci slov a sousloví z textu dokumentu za podpory lematizátoru, frekvenční analýzy, rozpoznávání víceslovných výrazů, negativního slovníku a komparace slov z textu s lexikálními jednotkami tezauru EUROVOC. Výsledkem je frekvenčně uspořádaný seznam deskriptorů, který je určen pro další intelektuální zpracování. Automatická indexace je stále pouze v testovacím provozu z důvodů omezené využitelnosti jejích výsledků.

Moderní vyhledávací metody se zkoumají v rámci projektu *Inteligentní vyhledávání v dokumentografických informačních systémech*, který řeší MFF UK ve spolupráci s KPS PČR. Projekt má za cíl pomocí standardních metodik otestovat vyhledávací techniky založené na statistickém přístupu (např. vektorové vyhledávání, latentní sémantické indexování, pojmové shlukování).

Bibliografické zpracování článků v ČR

[\(ukázky\)](#)

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 8 krajských knihoven a MZK, specializované odborné knihovny (STK, ÚZPI, SPKK-ÚIV, ČSAV, experimentálně ČGÚ). Na základě této spolupráce vzniká souborná databáze ANL. V systému LANIUS, resp. Clavius se zpracovávají bibliografické záznamy článků v knihovnách na úrovni okresů. V budoucnu je třeba sladit systém KOSABI a LANIUS tak, aby nedocházelo k duplicitnímu zpracování. Instituce spolupracující v rámci KOSABI přecházejí na nové SW vyšší generace. V krajských knihovnách v Kladně a Liberci se články popisují v systému RAPID, 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, SPKK-ÚIV přetrvává ISIS. V krajské knihovně v Plzni se články zpracovávají v systému KIMS. V nově konstituovaných krajských knihovnách v Pardubicích a Zlíně se používá KP-Sys, v Karlových Varech a Havlíčkově Brodě systém LANIUS, resp. Clavius.

Knihovny pracující v ALEPHu přešly v r. 2002 na verzi 14.2.4. Data jsou zpracovávána ve nebo konvertována do formátu UNIMARC, respektují se pravidla popisu AACR2 s respektováním mezinárodních standardů věcného popisu - MDT-MRF pro oblast systematické indexace. V oblasti verbální věcné indexace se kombinují klíčová slova, věcné obecné kategorie a předmětová hesla. Vyváženost vazby mezi jednotlivými vrstvami popisu je klíčovým momentem. V rámci kooperačního systému byla stanovena pravidla pro výběr titulů k popisu (na základě územní gesce - tituly regionální a celostátní provenience a dále pak na základě odborného zaměření). V rámci KOSABI jsou stanoveny zásady výběru článků co do úplnosti i co typů.

Zpracováváné 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).

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-politické 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).

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.

- Počet záznamů k 15.11.2002: ANL cca 742 600 bibliografických záznamů, ANL FULL cca 100 000 plných textů s metadaty
- počet zpracovávaných titulů: ANL - cca 212 v NKČR, 217 v regionech, 108 v odborných knihovnách, celkem cca 537 titulů; ANL FULL 30 titulů, běžně zpracovávaných zatím 16 v NKČR
- Časové pokrytí: ANL 1990/91 -, ANL FULL 1997- .

Záznamy NKČR tvoří 79,5 procent báze ANL. Záznamy regionů tvoří 14,1 procent. Záznamy specializovaných knihoven tvoří 6,4 procent. (Podíl v procentech souhlasím zhruba s údaji z r.1999).

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

Zpřístupnění výsledků analytického zpracování prostřednictvím Internetu (r. 1998) - projekt řešil zejména konverzi článků do UNIMARCu. V rámci průzkumu Internetu se ukázalo, že postupné propojení článků s některými plnými texty již vystavovanými na Internetu na různých serverech je krajně nespolehlivé (různá retrospektiva a úplnost vystavovaných plných textů, různá strategie vystavovatelů). Výběr spolehlivých zdrojů plných textů je možné řešení.

Výzkumný záměr NK **Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů** (VaV, r. 1999-2003) - je projekt analyticko-koncepční a připravuje půdu pro praktickou realizaci účelového projektu popisovaném v této zprávě a dalších projektech. Cílem výzkumného záměru je optimalizace přístupu uživatelů k plným textům dokumentů domácí proveniencí (nikoli zahraniční). Základem je propojení analytických záznamů o článcích s plnými texty. 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 **periodikum Národní knihovna** v Anopressu převedeno do digitální formy a zpřístupněno na Internetu (v r. 1999 pouze technikou OCR, v r. 2000 se přistoupilo i k prezentaci obrázků). V současné době je zpřístupňováno v podobě html na serveru full.nkp.cz (do r. 2001). Zároveň r. 1999-2002 je vystavován ve formátu html a pdf ve speciální webovské aplikaci pro toto periodikum. Speciální aplikace pro Národní knihovnu se bude integrovat sází ANL FULL pomocí jednotného interface. V rámci projektu je též částečně aktualizováno zadání pro aplikaci pro [management KOSABI](http://management.KOSABI) a provádějí se analyticko-koncepční práce týkající se metod automatické indexace, zpřístupňování plných textů (topiky) a je z části saturován další vývoj aplikace v systému TOPIC a linky zpracování bibliografických záznamů z plných textů (TTDE).

Projekt Západočeský ANAL - Kooperativní zpracování periodické produkce západních Čech se zabývá 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, VaV, r. 2000-2001 řeší problematiku bibliografického zpracování článků v tomto systému.

Velmi významný je z hlediska tvorby a rozvoje regionálních faktografických databází a souborů autorit je projekt řešený **SVK Kladno**.

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

V JIB jsou přístupné plné texty báze ANL FULL (propojení z ANL na ANL FULL) v NKČR, pro externí uživatele pouze na 7 dnů po registraci.

V roce 2002 pokračuje spolupráce v rámci KOSABI, probíhá poloprovoz linky zpracování bibliografických záznamů z plných textů a ladí se aplikace pro správu a údržbu KOSABI, aktualizuje se databáze ANL, ANL FULL. Báze ANL FULL je vybavena metadaty DC ve formátech html, XHTML, XML v kvalifikované i nekvalifikované formě. Pro zpřístupnění plnotextových informací je navržen nový layout serveru full.nkp.cz. Je vytvořena aplikace pro administraci báze ANL FULL a export a stahování záznamů. Jsou zahájeny práce na internetové verzi TTDE. Báze ANL FULL je vybavena dalšími topiky. Je nastíněn další možný vývoj KOSABI s ohledem na moderní metody zpracování.

V rámci KOSABI se řeší otázky zpracování článků s ohledem na nové státoprávní uspořádání tak, aby knihovny zároveň plnily své regionální funkce a byla zajištěna excerptce titulů v rámci systému.

Výsledky práce KOSABI a projektů týkajících se zpřístupnění článků prezentovala řešitelka projektu na konferencích Inforum 2002 a Knihovny současnosti 2002 v Seči u Chrudimi.

Na posledně jmenované konferenci zazněly též příspěvky, týkajících se KOSABI a krajských bibliografických systémů.

Příspěvky z regionů analyzují novou situaci krajských knihoven po reformě státní správy a samosprávy, zabývající se regionálními funkcemi knihoven a krajským bibliografickým systémem. Většina knihoven je

připravena a zdá se i schopna dále své bibliografické povinnosti plnit. Probíhají/proběhly jednání o přerozdělení titulů s ohledem na územní reformu. Zazněly následující příspěvky:

Nádvorníková, M.. Úloha krajských knihoven v národním bibliografickém systému;

Nejezchlebová, J. Systém knihoven z pohledu zajištění regionálních funkcí;

Svobodová, E. Spolupráce paměťových institucí v rámci krajského bibliografického systému - utopie nebo reálná možnost.

Knihovna v Hradci Králové pokrývá plně excerpci na území svého kraje, poskytuje záznamy knihovnam kraje.

Prezentace J. Míky **Regionální bibliografie a faktografie - příklad spojení tradičního a moderního přístupu ke knihovnické práci** se zaměřila na moderní metody zpracování a jejich integraci. Byl zmíněn projekt Rozvoj regionálních databází na základě retrospektivní konverze článkové bibliografie (MKČR 2001-2002), týkající se provázání článkové bibliografie se seriály a faktografickými databázemi REOS, REFA.

Prezentace J. Kaňky **Koncepce krajského bibliografického systému se zabývá vztahem národní a regionální bibliografie**, koncepcí KBS v rámci Zlínského kraje, spoluprací na úrovni národní a krajské. ([ukázka](#)).

Národní úroveň:

"plánované zapojení do kooperačního systému článkové bibliografie - přispívání do báze ANL od roku 2003 proporcionální a obsahové pokrytí území kraje (1 krajský deník, 1 krajská regionální příloha celostátního deníku, 1 týdeník a/nebo 1 měsíčník z každého okresu, významné muzejní časopisy, významné kulturně-vlastivědné časopisy, archivní sborníky, odborné časopisy celostátního významu, sborníky konferencí, kongresů, symposií, seminářů) si pravděpodobně vynutí rozšíření stávající excerpční základny".

Krajská úroveň:

"retrospektivně svazky jednotlivých okresů z edice Bibliografie a prameny k vývoji Moravy souběžně bibliografie buduje každá bývalá okresní knihovna v elektronické podobě (většinou od poloviny 90. let) , v současné době úhrnem přes 66 tisíc záznamů ,celkový roční přírůstek přes 11 tisíc záznamů ,regionální záběr většinou okres, příp. etnografický region ,heterogenní prostředí AKS => nutnost shody na standardech (do budoucna možnost tvorby souborné reálné či virtuální báze), tištěný výstup vychází pouze v jednom okrese (ročenka souběžně regionální článkové bibliografie)".

A. 3 Cíl, vstupní data

A. 3.1 Cíl

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 jsou postupně propojované s elektronickou podobou článku a takto prezentované na Internetu. Obě části souborné databáze - vznikající databáze plných textů a báze bibliografických záznamů ve formátu UNIMARC - vyžadují permanentní kvalitní SW a HW podporu. Budování, doplňování, správu a údržbu plnotextové databáze s možností vyhledávání zajišťuje informační agentura ANOPRESS (systém TOPIC) a její pracovníci. Zároveň jde o vývoj manažerského systému pro příjem a správu dat kooperačního systému. Hlavním cílem projektu je zkvalitnění bibliograficko-informačních služeb a integrace elektronických zdrojů do služeb knihovny.

Cíle

- **Cílem projektu v r. 2002 je aktualizace databáze ANL, ANL FULL a tedy i spolupráce v rámci KOSABI (po stránce faktické, metodické i organizační).**
- **Průběžné doplňování báze ANL FULL plnými texty a metadaty DC ve formátech HTML, XHTML, XML v kvalifikované a nekvalifikované formě.**
- **Aktualizace báze ANL bibliografickými záznamy s propojením na plné texty za současného vydávání báze ANL v rámci ČNB.**
- **Další cílem je poloprovozní zpracování záznamů o člancích z plných textů v rámci linky automatické indexace.**
- **Zkvalitnění pracovního formuláře, který je ještě umístěn na lokálních pracovních stanicích. Probíhají přípravné práce pro umístění formuláře na Internet.**

- Dalším cílem je upgrade a update serveru full.nkp.cz po stránce SW (vyšší verze) i HW (rozšíření kapacity disku).
- Dalším cílem je nový moderní layout a struktura serveru FULL.NKP.CZ.
- Dalším cílem je implementace aplikace pro administraci báze.
- Aplikace pro export a stahování plných textů a metadat.
- Pozornost se soustřeďuje na definici báze ANL FULL - zejména na vyhledávání pomocí topiků.
- Dalším cílem je údržba stávajícího strukturovaného protálu s popisem zdrojů na serveru full.nkp.cz pro volně dostupné zdroje na českém Internetu z hlediska krajů a tematiky.
- Cílem je propojení plného textu se záznamem, které probíhá na základě jednak statických URL (volně přístupné texty na Internetu v rámci báze ANL), jednak dynamických URL (generovaných v rámci linky zpracování bibliografických záznamů z plných textů).
- Cílem je zpřístupnění plných textů v rámci JIB přes bázi ANL, ze které se odkazuje na plné texty v bázi ANL FULL (asp skript) a úprava adres z dělených na adresu v celku.
- Průběžně řešení otázek standardizace jmenného a věcného popisu bibliografických záznamů a elektronických dokumentů a možnosti přebírání metadat do bibliografických záznamů ve formátu UNIMARC a plných textů v podobě Dublin Core a ve struktuře XML je předpokladem funkčnosti celého systému.
- Projekt ovlivní v budoucnu i práce na konverzi UNIMARC a MARC21, prováděné v r. 2002.
- Navazování spolupráce s nově konstituovanými krajskými knihovnami (Zlín, Pardubice, Havlíkův Brod, Karlovy Vary).
- Cílem je ladění a testy vstupů dat do KOSABI a výstupů pro bázi ANL - aplikace pro správu a údržbu KOSABI.
- Naplnění databázové aplikace pro evidenci zpracovávaných titulů v rámci KOSABI základními daty.
- Cílem je navržení jednotného www vstupu k vyvíjeným aplikacím pro management KOSABI.
- Cílem je upgrade systému pro management KOSABI po stránce programové (vyšší verze).
- Řešení pro rok 2002 spočívá v realizaci navržených metod v roce 2001 v poloprovozním stádiu - budování báze ANL FULL, linka zpracování bibliografických záznamů z plných textů (TTDE), zpřístupnění plných textů.
- Cílem je převedení aplikace pro management KOSABI do poloprovozního stádia.
- Dalším cílem jsou praktické závěry pro další vývoj projektu na základě analýzy současného stavu a trendů automatické indexace a selekční efektivnosti topiků v databázi ANL FULL

A. 3.2 Vstupní data

- Vstupními daty pro bázi ANL jsou bibliografické záznamy zpracovávané v r. 2002 v rámci KOSABI spolupracujícími institucemi.
- Vstupními daty pro bázi ANL jsou bibliografické záznamy zpracované v NKČR v ALEPHU a v lince zpracování bibliografických záznamů, resp. metadat z plných textů.
- Vstupními daty pro bázi ANL FULL jsou plné texty s metadaty, které vznikají v rámci linky zpracování bibliografických záznamů z plných textů při generování dynamických URL za současného generování formátů UNIMARC, DC v HTML XHTML a XML.
- V roce 2002 se v lince zpracovává výběr z celostátních deníků a některých časopisů zároveň pro bázi ANL i ANL FULL.
- Bibliografické záznamy zpracované v ALEPHu jsou také propojované na volné www plné texty statickými URL adresami - záznamy především titulů z oboru knihovnictví (U nás, Ikaros, Bulletin SKIP aj.) a oborů dalších (Harmonie, Vesmír, Psychiatrie, Veřejná správa, Učitelství noviny aj.).
- Vstupní data pro strukturovaný portál www periodik jsou volně přístupné zdroje na Internetu.

Tituly volně přístupné na WWW a propojované se záznamy ANL (od r. 1998) - 15 titulů

(Veřejná správa, Vesmír, Lesnická práce, Harmonie, Psychiatrie, Učitelství noviny, Jezuité, Kriminalistika, Národní knihovna, Knihovní obzor, Collection of Czechoslovak Chemical Communication -abstrakta, Inforum #, Ikaros, U nás, Bulletin SKIP).

PERIODIKA VOLNĚ PŘÍSTUPNÁ NA WWW	2002	2001	2000	1999	1998	1997	Suma
Bázi ANL FULL doplňuje portál - Periodika na WWW (struktura oborově a regionálně)							
Volná periodika na WWW - přístup z báze ANL							
Collection of Czechoslovak Chemical Communication			105		35		140
Bulletin SKIP							127
Harmonie			29	63	64		156
Ikaros							427
Inforum...							94
Jezuité			12	26			38
Knihovní obzor	23	29	1	25	27		105
Kriminalistika	17	30	33				80
Lesnická práce	74	54	91	70	1		290
Národní knihovna	244	65	56	69	63		497
Psychiatrie			9	15			24
Psychologie dnes							
Učitelské noviny		24	30				54
U nás							107
Veřejná správa	162	417	411	269	35		1294
Vesmír	145	102	73	86	69		475
Celkem	665	721	850	623	294		3908

V bázi ANL FULL zpřístupněno výběrově cca 37 titulů většinou od r. 1997.

V lince pravidelně zpracováváno 15 titulů (Bankovníctví, Ekonom, Haló noviny, Hospodářské noviny, Kapitál, Lidové noviny, Magazín Práva, Mladá fronta Dnes, Pátek magazín LN, Právo, Profit, Reflex, Respekt, Týden, Večerník Praha, Euro - bude).

NOVINY A ČASOPISY ZPŘÍSTUPŇOVANÉ V NČR A V ANOPRESSU (REGISTRACE, KONSORCIUM)	2002	2001	2000	1999	1998	1997	Suma
Titul, počet článků v ANL FULL vydaných v 1990- 31.10.2002 (zpracovaných v 1999-2002, metadata+text), tučně tituly zpřístupňované v současnosti v ANL FULL (výběrová báze). U titulů je uveden odkaz na Anopress, kde je archiv příslušných titulů							
Archiv titulů a aktuální rok je zpřístupněn v bázi TamTam (Anopress)							

V bázi ANL jsou přístupné tytéž tituly/články jako v ANL FULL, a to přes bibliografický záznam a link do ANL FULL							
Bankovníctví	67	33	73			173	346
Berounský deník				96	154	189	439
České Slovo			310				310
Deník Jablonecka				199	94	444	737
Ekonom	385	998	1070	46	14	385	2898
Haló noviny	1025	1055	983	1340	1327	780	6510
Hanácký a středomoravský den			182				182
Hospodářské noviny	1636	1652	221	3094	2631	2782	14016
Hradecké noviny			23	148	260	389	820
Chebský deník				32	148	156	336
Kapitál	52	3	97	178			330
Kladenský deník				256	19		275
Liberecký den				1		73	74
Lidové noviny	2240	2246	1831	3575	4789	4110	18791
Magazín LN		22	56				78
Magazín Práva	138	165	47				350
Mladá fronta Dnes	2718	3140	2396	2489	2998	4104	17845
Moravskoslezský den			132	202		44	378
Národní knihovna*		37	132	151	164	138	1303
Nedělní noviny			123				123
Pátek magazín LN	147	55					202
Plzeňský deník			21	267	227	343	858
Právo	1494	1800	1198	1431	1936	1671	9530
Pražské Slovo		414					414
Profit	223	162					385

* samostatná aplikace (r. 1999-2002), též Anopress (r. 1990-2001) volně přístupná, z důvodů úspornosti v tabulce implicitně sloupec 1990-1996 pro Národní knihovnu 681 záznamů

Summa summarum:

- Vstupními daty pro plnotextovou bázi ANL FULL jsou plné texty zejména celostátních deníků a některých odborných časopisů za rok 2002 a jim odpovídající bibliografické záznamy v tomto období pro bázi ANL v rámci KOSABI.
- Vstupními daty pro automatizovanou linku zpracování bibliografických záznamů jsou plné texty získané z databáze TAMTAM na základě speciální aplikace. V rámci této linky vznikají vstupní data pro bibliografickou bázi ANL a fultextovou databázi ANL FULL.

- **Vstupními daty pro bibliografickou bázi ANL (ALEPH) záznamy článků excerpovaných z titulové základny systému. Vstupními daty pro bázi ANL jsou dále záznamy spolupracujících institucí v rámci KOSABI.**
- **Vstupními daty pro portál Periodika na www jsou volně přístupná periodika na Internetu.**

B Analytická část

B.0 Automatická či poloautomatická indexace a topiky v bázi ANL FULL - analýza a praktický dopad na další vývoj projektu

Tento bod projektu je též součástí projektu Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů.

B.0.1 Analýza problematiky automatické či poloautomatické indexace a selekční úplnosti topiků v bázi ANL FULL ([citace z materiálu J. Schwarze](#) - viz též výše) - teoretická analýza s praktickým doporučením

Automatická indexace úzce souvisí s vyhledáváním informací (*information retrieval*). Na jedné straně kvalita automatické indexace výrazně ovlivňuje kvalitu vyhledávání, na druhé straně jsou metody automatické indexace a vyhledávání informací z hlediska požadovaného výsledku - vyhledání relevantních dokumentů - zástupné; **někteří autoři např. považují vyhledávání v plném textu dokumentu za nejjednodušší formu automatické indexace.** Podstatnější je však **trend směřující 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 - o těchto tématech se často hovoří společně a někdy se také zaměňují - v případě indexace obecnými třídami se mluví spíše o automatické klasifikaci než automatické indexaci. Prolínají se také problematiky automatického shlukování a klasifikace, protože obě mají stejný účel - rozčlenit dokumenty do základních tříd; každá z metod však používá odlišných postupů. Automatická indexace také souvisí s automatickou tvorbou tezauru, příp. vytvářením sémantických sítí nebo znalostníchází, 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.

Automatické přiřazování je z hlediska praktické aplikace podstatně náročnější, funkční systémy rovněž existují, ale nejsou zcela běžné a často nepřinášejí žádoucí výsledky. V rámci automatického přiřazování jsou aplikovány i metody automatické extrakce, které však slouží pouze jako dílčí podklady pro další zpracování. Jako systémy automatické indexace jsou také někdy nesprávně označovány systémy, které slouží jako automatizovaná podpora indexátora při indexačním procesu. **Tyto systémy umožňují dnes již standardně automatizované procedury, jako je kontrola chyb indexátora (např. zda nezvolil nesprávnou kombinaci termínů), automatický návrh preferovaného termínu místo termínu nepreferovaného, poskytnutí informace, jak byl indexační termín použit, popř. jak vypadá předchozí indexace vybraných dokumentů apod.** V takových případech nelze mluvit o automatické indexaci, ale pouze o automatizované indexaci.

Automatická extrakce

Postup

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.

Hodnocení automatické extrakce

Ve srovnání s intelektuální indexací, je automatická extrakce absolutně konzistentní. Jako podstatné omezení automatické 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.

Postup

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.**

Stále ještě není reálné zcela automatické přiřazování termínů bez intelektuální zásahu. Tato skutečnost se týká zejména systémů používajících velkých řízených slovníků (tisíce až desetitisíce termínů) a obsahujících dlouhé texty pojednávající o komplexních, obecných nebo heterogenních tématech.

Obecně tak metoda automatického přiřazování selhává, pokud je hodnocena z hlediska kvality výsledků intelektuální indexace. Nicméně i v těchto případech mohou být, podobně jako u automatické extrakce, výsledky automatizovaných procedur podkladem pro další intelektuální zpracování indexátorem.

Dlouhodobě neuspokojivé (resp. ne zcela uspokojivé) praktické výsledky metod automatického přiřazování vedly v posledních dvou dekádách k určitému odklonu zájmu od této problematiky; řada autorů se věnuje spíše teoretickým aspektům problematiky (např. pojmovému modelování) než praktickým implementacím ověřených metod, mj. také z toho důvodu, že všeobecná dostupnost plných textů v elektronické podobě a účinných moderních vyhledávacích technik (např. vektorové vyhledávání, pravděpodobnostní modely vyhledávání, latentní sémantické indexování, fuzzi modely vyhledávání, neuronové sítě, bayesovské sítě apod.) význam indexace coby základního předpokladu pro efektivní vyhledávání dokumentů podstatně snižuje. To ovšem nic nemění na skutečnosti, že řada funkčních systémů založených na automatickém přiřazování ve specifických disciplínách existuje a je úspěšně používána.

Intelektuální indexace přináší ve srovnání s řadou automatických procedur (vč. např. latentního sémantického indexování) ještě stále lepší výsledky. Systémy automatické indexace jsou také účinné pouze částečně proto, že doposud nebyly dostatečně prozkoumány a podrobně popsány všechny intelektuální procesy, které probíhají při indexaci.

Na druhé straně řada technologií, jejichž účinnost byla v předchozích letech potvrzena výzkumem a řadou

studii, je v současnosti implementována ve formě expertních systémů nebo systémů pracujících na základě umělé inteligence.

Praktické aspekty implementace automatické indexace Obecně lze konstatovat, že systémy automatické indexace vyvíjejí oborově zaměřené instituce, které zpracovávají velké objemy dokumentů, které je nezbytné kvalitně a konzistentně indexovat. V řadě případů se však jedná o dokumenty, u kterých je dostupný pouze komprimovaný text (např. abstrakt), a které je tudíž žádoucí indexovat. Zároveň lze u nich s úspěchem použít některou z metod automatické indexace, které jsou v případě nasazení na komprimovaný text relativně velice úspěšné. **PŘI DOSTUPNOSTI PLNÝCH TEXTŮ DOKUMENTŮ SE DNES JIŽ SPÍŠE POUŽÍVÁ NĚKTERÉ Z POKROČILÝCH VYHLEDÁVACÍCH METOD, JEJICHŽ VÝVOJ A IMPLEMENTACE JE OBDOBĚ TECHNICKY I FINANČNĚ NÁROČNÝ JAKO U AUTOMATICKÉ INDEXACE, ALE V ŘADĚ PŘÍPADŮ PŘINÁŠÍ ADEKVÁTNĚJŠÍ VÝSLEDKY.**

B.0.2 Selekční úplnost topiků v bázi ANL FULL (na základě analýzy J. Schwarze)

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 (DR) byly srovnány s dokumenty z TMD (DT) 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 RMD.

Výsledkem komparace není jen číselné vyjádření úplnosti výsledků, ale především rozvaha, proč některé dokumenty nebyly nalezeny a jak by bylo možné vylepšit vyhledávání pomocí pokročilé aplikace topiků. 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 kolem 80 %.

Hodnocení výsledků komparace

Nedostatky na straně topiků

Přestože téměř polovina TMD nebyla zcela úplná, není to třeba interpretovat jako selhání příslušných topiků. V zásadě lze neúplnost při vyhledávání vysvětlit těmito důvody: RRD a topik nemusí zcela korespondovat ve svém tematickém zaměření, resp. topik může vyjadřovat pouze jeden aspekt příslušného RRD, což je dobře patrné u duplicitních položek.

Otázkou u jednotlivých topiků zůstává, nakolik správně vyjadřují příslušný aspekt problematiky, Potenciálně problematická jsou také témata z hraničních oborů.

Pouze na základě výsledků komparace nelze stanovit, který z topiků filtruje dokumenty "správně" a který by vyžadoval rozšíření o nová klíčová slova. Tento závěr vyžaduje další podrobnou obsahovou analýzu jednotlivých dokumentů a zejména topiků, která není možná bez znalosti jejich konkrétní struktury a složení.

Nedostatky na straně předmětových hesel

Jako příčinu nenalezení dokumentu topikem nelze vyloučit ani chybné nebo nesprávné přiřazení daného předmětového hesla k dokumentu, který byl zařazen do RMD a srovnáván s výsledky TRD. V tomto případě mohou sloužit výsledky analýzy k případné korekci selekčního obrazu dokumentu, kterou je třeba provést se znalostí příslušných pravidel pro věcné zpracování dokumentů.

Další možnosti rozvoje a využití topiků

V případě topiků je třeba si klást zásadní otázku, jakým způsobem je uživatel může neefektivnějším způsobem využít. Z hlediska uživatele se mohou jevit jako limitující především tyto aspekty:

- problematické zařazení některých dílčích topiků (např. vojenství, vojenská technika se "skrývá" pod spol. a humanitními obory)
- obecně sice jednoduchá struktura, ale pro neznalého uživatele nebo uživatele nedostatečně obeznámeného s funkcí a strukturou topiků může znamenat překážku v nalezení relevantních informací.
- zařazení duplicitních nebo z hlediska uživatele kvaziduplicitních topiků, např. geografie/geografie, zbrojní průmysl/vojenství, vojenská technika, chemie/chemický průmysl, farmaceutický průmysl/farmacie, farmakologie apod. U takových topiků mohou vznikat pochybnosti, který z nich použít, zvláště když řazení dle vědeckých oborů má v případě struktury topiků spíše arbitrární povahu
- nerovnoměrné rozložení topiků: některé přináší několik jednotek nebo desítek záznamů (v dvouleté retrospektivě), některé několik tisíc nebo desítek tisíc záznamů. Tato skutečnost může být samozřejmě dána obsahem databáze, v případě některých specifických témat je nadměrný počet často irelevantních záznamů zřejmě způsoben příliš obecnou definicí topiku (např. topik Kybernetika, robotika). Topik Věcné zpracování přináší z větší části zjevně irelevantní výsledky.
- Naopak topiky mohou být pro uživatele užitečné, protože:
- umožňují řadit záznamy podle relevance, na rozdíl např. při použití pouze předmětových hesel, kdy všechny záznamy dostávají hodnotu relevance 1.00 a jsou řazeny v podstatně podle toho, jak jsou vyhledány z databáze
- omezují počet irelevantních dokumentů

Využití topiků: uživatel nebude zřejmě topiky používat samostatně, protože většina z nich představuje pro vyhledávání příliš široká témata. S vysokou mírou pravděpodobnosti lze předpokládat, že topik nebude využit ani v případě, kdy by bylo žádoucí omezit dotaz na specifickou oblast či obor, a to z důvodu, že procedura je prostě pro řadu laických uživatelů příliš složitá.

Z tohoto hlediska lze spatřovat jako optimální řešení v současnosti neexistující možnost interaktivního využití topiků, kterou lze stručně popsat takto:

Po vyhledání dokumentů podle libovolného selekčního prvku dostane uživatel možnost zúžit nebo rozšířit dotaz pomocí topiků, které mu systém automaticky nabídne - vygenerované na základě aktuální množiny vyhledaných dokumentů. V případě zúžení dotazu by byly topiky s původním dotazem spojeny s operátorem AND, v případě jeho rozšíření pomocí operátoru OR.

Kromě interaktivního využití topiků při vyhledávání připadá do úvahy možnost oboustranné automatické zpětné vazby mezi topikem a např. předmětovým heslem: při věcném zpracování a přiřazení určitého předmětového hesla by systém kontroloval, zda zpracovávaný dokument náleží do stejného topiku jako dokumenty označené stejným předmětovým heslem. Opačně, při zpracování dokumentu může systém na základě podobnosti dokumentu s dalšími, dříve zpracovanými dokumenty informovat o použitých předmětových heslech.

B.0.3 Analýza automatické indexace a selekční úplnosti topiků v ANL FULL a její možný dopad na další praktický vývoj projektu

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 se zprovožňuje 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 fuzzi 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 topiků na základě jejich ladění a zabudování prvků věcného popisu (kategorií a prvků předmětových hesel) přímo do topiků, generování topiků on line a budování znalostní báze, než k automatickému přiřazování.

5. V systému je náběh na řízený slovník tematických kategorií, který pravděpodobně bude sloužit jako základ k tvorbě příslušných topiků. Protože tyto kategorie jsou vázány na MDT, mohly by sloužit také jako základ k automatickému shukování dokumentu podle těchto kategorií (dosud shukování na základě stejných klíčových slov obsažených v dokumentu - tuto funkci nelze realizovat v rámci TOPICU, musela by být řešena mimo - pouze hypotéza). Slovník bude pravděpodobně možno využít také jako pomůcka pro katalogizátora při přiřazování těchto kategorií a vstupních prvků hesel event. k v budoucnu k poloautomatické klasifikaci dokumentů podle těchto kategorií.

6. Interaktivní využití topiků při vyhledávání a interaktivní zpětná vazba mezi topikem a předmětovým heslem při zpracování dokumentů. (hypotetický předpoklad - zatím nekonzultován s pracovníky kolem TOPICu). Souvisí též poměrně s reálnou možností generování věcných topiků online z kategorií a předmětových hesel (předpoklad: redakce těchto prvků již přítomných v systému).

7. V budoucnu zprovoznit interakci systému - hledání dokumentů s podobným obsahem - volný dotaz (Free Text Query), dotaz příkladem (Query By Example).

8. Topiky ladit na konkrétních dokumentech z příslušného oboru - vytypovat jak obory, tak odpovídající dokumenty s kvalitním předmětovým popisem. V úvahu přichází periodikum Národní knihovna (zprístupňované v projektu Propojení analytických záznamů s plnými texty), ke kterému vzniká řízený slovník, dále pak obory a dokumenty, o kterých lze tvrdit, že jsou kvalitně indexovány katalogizátorem (např. ekonomie/ekonomika, některé společensko politické časopisy).

9. K ladění topiků je nutný přístup ke struktuře jednotlivých topiků. Rovněž je třeba zkvalitnit samotnou věcnou indexaci pomocí předmětových hesel indexátory.

B.1 Vlastní řešení - praktická realizace

Stručný popis stávajících produktů TamTam a prací v r. 2002 :

1. **Hlavní linka pro zpracování dokumentů**
 - a. **TamTam Special NK**
 - stažení vybraných článků z TamTamu v Anopressu do NK přes Internet
 - (skripty na serveru v Anopressu)
 - b. **TTDE - TamTam Data Extractor**
 - úprava hlavičkových údajů ve stažených člancích prostřednictvím 3 formulářů
 - výstup Aleph a Full
 - hlavičky Uni, Uni-A, D.C., H1 (zobrazovací), H2 (indexovací)
 - odeslání zpracovaných dat do importní schránky Aleph a k fulltextovému indexování (Topic)
 - (program TTDE.EXE na každé stanici uživatele)
 - **přípravné práce a vývoj internetové verze (nové 2002)**
 - c. **MKINDEX - indexování, generování XML**
 - slouží k on-line indexování upravených článků do fulltextové db (Topic)
 - zároveň se on-line generují XML- a XML+ (nové od 9/2001)
 - generování rejstříků
 - (program MKINDEX.EXE běží na serveru FULL.NKP.CZ)
2. Jednorázové aplikace/moduly
 - a. **TTXMLGEN - TamTam XML generátor**
 - generuje XML- i XML+ z celé DB (nové od 9/2001)
 - (program TTXMLGEN.EXE na severu FULL.NKP.CZ)
 - b. **TTDT - TamTam Data Tester**
 - slouží k testování obsahu DB (nové od 8/2001)

- vyhledá dokumenty, které obsahují nebo naopak neobsahují zadaný řetězec
- (program TTDT.EXE na severu FULL.NKP.CZ)>
- c. **TTDF - TamTam Data Filter**
 - slouží k filtrování a úpravě hlavičkových polí v DB
 - na základě indexovací hlavičky se vygeneruje Dublin
 - s Unimarcem nepracuje
 - poslední verze k 2.6.2001
 - pro použití se současnými hlavičkami je nutná jeho podstatná úprava
 - (program TTDF.EXE na severu FULL.NKP.CZ)
- d. **TTDC - TamTam Data Converter**
 - konverze článků, stažených metodou Standard do fulltextové DB
 - pro použití se současnými hlavičkami je nutná jeho podstatná úprava
 - stahování metodou Standard by se nemělo už používat
 - (program TTDC.EXE na severu FULL.NKP.CZ)
- e. **TTEXT - TamTam Export**
 - export dat z DB Anopress pro NK
 - (program KNIHOVNA.EXE v Anopressu)
- 3. **Internet <http://full.nkp.cz> (kompletní přepracování v asp skriptu v r. 2002)**
 - a. **Obecné základní informace, nápovědy**
 - dynamická horní nabídka, counter
 - úvodní homepage
 - charakteristika DB
 - další informace (báze ANL)
 - jak se registrovat
 - jak vyhledávat
 - b. **Autorizace**
 - registrační formulář a mechanismus (mail + aktivace)
 - autorizační panel (login)
 - c. **Fulltextové vyhledávání**
 - 3 formuláře: základní, rozšířený bez a s tématy
 - vyhledávání dle hlavičkových údajů z metadat prostřednictvím pomocných hlavičkových údajů
 - výsledky jednoduché, se souhrnem, (seskupené v nové verzi do konce r. 2002)
 - Zobrazení výsledků hledání - článku
 - zobrazení textu, hlaviček (bez/s odkazy pro další hledání) - MKTXT.ASP
 - generování a zobrazení UNImarcu MKUNI.ASP
 - generování a zobrazení DC, XHTML- +, XML- + - MKXML.ASP.
 - zobrazení klíčových slov-- MKKEY.ASP
 - zobrazení XML- + - vygenerování při indexování dokumentu - MOFXML.ASP
 - d. **Zobrazení plného textu z ALEPHu**
 - modul MKDOC.ASP
 - e. **Export a stažení plných textů (komplet nové v r. 2002)**
 - modul dnload1.asp, dnload2.asp, dnload3.asp
 - tvar výstupního souboru (formáty): HTML, RTF formát, textový tvar, UNICODE - vše se zvýrazněním klíčových slov a bez, XML formát, HTML formát se souhrnem
 - zobrazení detailního průběhu zpracování
 - f. **Administrace (nové v 2002)**
 - úpravy textu článků
 - statistika přístupů
 - Údržba rejstříků podle data zápisu a počátečních písmen
 - Údržba regionálních periodik - portál
 - Údržba oborových periodik - portál
 - Údržba informačních zdrojů
 - Údržba skupin témat
 - Uživatelé (aktualizace)
 - Změny hesel uživatelů
 - Maily (komu se posílá registrace)
 - g. **Aplikace Regionální periodika - portál**
 - rozdělení dle krajů
 - zobrazení seznamu dle krajů

- zobrazení detailu periodika
- odkaz na www stránku a email periodika
- h. **Aplikace Oborová periodika - portál**
 - rozdělení dle oborů
 - zobrazení seznamu dle oborů
 - zobrazení detailu periodika
 - odkaz na www stránku a email periodika
- i. **Rejstříky**
 - rejstříky dle požadavku zadavatele

ROK 2002 - seznam prací s plánem do budoucna (kurzívou)

- A. **Úprava WWW stránek:**
 - změna vyhledávání v rejstřících
 - období od-do v rejstřících
 - úprava designu full.nkp.cz
 - stránka Co je nového
- B. **Úprava databáze rejstříků**
 - změna struktury databáze rejstříků (číslo, propojení)
 - úprava programu MkIndex (úprava modulu pro generování rejstříků)
- C. **Zavedení nových polí do hlavičky**
 - přeindexování dat z důvodu zavedení nových polí v roce 2001
 - program pro výpis všech souborů z archívních databází
 - znovuinindexování databází
- D. **Konverze XML do UTF-8**
 - modul pro konverzi
 - úprava zobrazování XML přes www
 - úprava indexovacího modulu MkIndex
- E. **Update Serveru I (převod stávajícího stavu do nového prostředí)**
 - instalace Win 2000 server
 - záloha serveru, obnova dat i programů do nového prostředí
 - instalace MS IIS 5 (internet information server)
 - instalace Portal One SE (Topic verze 3.7)
 - konfigurace a nastavení serveru
 - zprovoznění www stránek v novém prostředí
- F. **Update Serveru II (optimalizace a úpravy)**
 - optimalizace stávajících stránek (změna technologie pro nové prostředí, zrychlení zobrazování článků, nahrazení PHP skriptů metodou XMLHTTP a ASP skriptingem)
 - úprava propojení do Aleph
 - úprava modulu MkIndex
 - hromadné stahování článků a export do různých tvarů
 - Interface pro opravy metadat v plných textech
- G. **Rejstřík kategorií**
 - vytvoření www rozhraní pro údržbu rejstříku kategorií
 - provázání nadřazených a podřazených položek (plán)
 - generování topiků za běhu (plán)
- H. **Ostatní rejstříky**
 - vytvoření www rozhraní pro údržbu všech ostatních rejstříků
- I. **Portál**
 - vytvoření www rozhraní pro údržbu regionálních a oborových periodik
- J. **TTDE pro Web (příprava a návrh v r. 2002, částečně plán)**
 - kompletní překlopení linky TTDE do www prostředí
- K. **Úprava TTDE (plán)**
 - stahování jiných zdrojů než Anopress
 - obecná úprava programu pro vstup z více vstupních kanálů
 - přidání filtrů pro každý vstupní kanál
- L. **MARC 21 (plán)**
 - úprava programu TTDE (modul pro konverzi dat do MARC 21)

- úprava www rozhraní (zobrazování MARC 21)
- doplnění MARC 21 do starých dat (konverze)

Popis řešení

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.

Vývoj a údržba systému je financován v rámci předkládaného projektu a podílí se na něm ing. Mattern, dr. Kučera jako soukromé osoby, ing. Mattern a dr. Kučera jako zaměstnanci firmy a společnost Anopress jako firma.

Vývoj technologie (linka, export, administrace, topiky) je též financován v rámci projektu Propojení analytických záznamů s plnými texty (vývojové práce ing. Materna a dr. Kučera).

V rámci poloprovozu v r. 2002 byla dále laděna linka zpracování a ověřována v poloprovozu. V současné době má pracovní formulář, do kterého se extrahují data a ta dále doplňují věcným popisem, tři strany. **V r. 2002 je dále laděn výstup v UNIMARCu.** Automaticky je generována URL adresa, SICI, "provizorní" URN, které sleduje filozofii a strukturu stanovenou pro tvorbu URN.

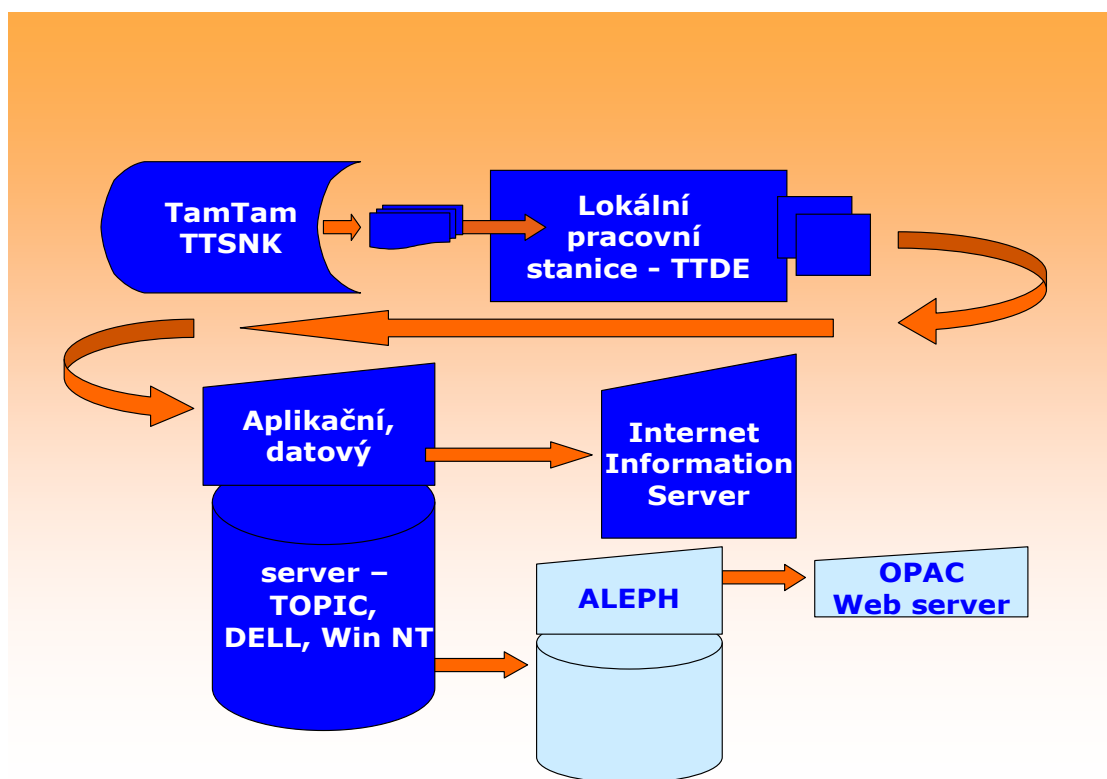
Automaticky lze generovat klíčová slova a extrahovat do formuláře. Na výstupu procesu zpracování zabudovaná funkce kódování výsledného souboru záznamů v Ansel, UNICODu a UTF-8 v r. 2001 je ověřena při rutinním ukládání v r. 2002 (velmi důležitá konverze do UTF-8 vzhledem k nové verzi Alephu). Po odeslání souboru do příslušného adresáře na server full.nkpcz je tento aplikací vyvinutou v NK naimportován do báze ANL (ALEPH), zaindexován a zpřístupněn v systému TOPIC na serveru full.nkp.cz v podobě plného textu vybaveného metadatami (UNIMARC a aplikace Dublin Core 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.

V roce 2002 jsou implementovány další topiky (literatura, archeologie, historie, umění, geografie), implementována sekce Země světa a geografie, aktualizovány stávající topiky. Je vypracován návrh na propojení jednotlivých úrovní některých topiků. Rovněž se plánuje vývoj obecné šablony topiku pro Země světa a šablony pro generování věcných topiků online z kategorií a předmětových hesel.

Nově byla implementována aplikace pro stahování a export záznamů, dále pak aplikace pro administraci. V současné době probíhají přípravné práce a návrh internetové verze linky zpracování TTDE.

B.1.1 Architektura systému zpřístupňování plných textů, funkce systému a procesy. Linka TTDE. Systém TOPIC a topiky



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.

Plné texty báze ANL FULL jsou zatím přístupné interním uživatelům NKČR. Externím uživatelům je běžně k dispozici pouze bibliografický popis (metadata), na dobu 7 dnů je možno získat i přístup k plným textům na základě zkušební registrace. **Vybraní uživatelé mohou po registraci zkusit stahování a export metadat a plných textů v různých formátech. Zpřístupnění plných textů externím uživatelům závisí na vývoji a řešení legislativně právních i finančních otázek souvisejících se zpřístupňováním plných textů, na dohodě s vlastníky autorských práv. V roce 2002 proběhla na půdě NK mailová diskuse ohledně začlenění databáze ANL FULL do konzorcia v rámci VISK 8 (multilicenční zpřístupnění báze TamTam a ČTK). Jednání budou pokračovat. Zatím je báze ANL FULL přístupná stále externím uživatelům pouze na 7 dnů. V rámci konzorcia by měla být používána i linka zpracování bibliografických záznamů z plných textů.**

Inspirací a do jisté míry vzorem pro koncipování báze ANL FULL je databáze **ProQuest 5000** (viz Příloha F 0a).

B.1.1.1 Architektura systému získávání, zpracování a zpřístupnění plných textů v systému TOPIC (báze ANL FULL) a ALEPH (báze ANL) (viz Příloha F2, F3).

- **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 návěšťími označeny příslušné údaje (v budoucnu možná další úprava vstupních filtrů podle charakteru získávaných dokumentů z jiných zdrojů).
- **Zpracování plných textů** (extrakce dat z hlavičky dokumentu, doplnění a generování DC, UNIMARC, URL, SICI, HTML, výstup v ANSEL, UNICODE, UNICODE-UTF 8) - aplikace TTDE (TamTam Data Extractor) zatím na lokálních pracovních stanicích.
- **Konverze, generování dat v XHTML+-, XML+-, indexace, uložení a vyhledání informací** (vyhledávání pomocí textu dotazu, formulářů, topiků, rejstříků) - aplikace a datový server: Portal

One SE - fulltextový vyhledávací systém, server **Dell PoweEdge 6300** (1 GB RAM, 68 GB, Pentium III). **MKINDEX** - indexace, generování XML- a XML+, jednorázové aplikace/moduly (TTEXT- TamTamExport, TTXMLGen - TamTam XML generator, TTDT- TamTam Data Tester, TTDF- TamTam Data filter, TTDC- TamTam Data Convertor, MKDOC.ASP - propojení bibliografického záznamu s plným textem).

- **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 (nové v r. 2002) - formáty a tvary: HTML, RTF, text, XML, UNICODE**
- **Administrace (nové v r. 2002) - úpravy a opravy textů (metadat), statistiky, údržba rejstříků, údržba portálu**

B.1.1.2 Linka TTDE - TamTam Data Extractor (získávání a zpracování bibliografických záznamů z plných textů)

[Linka zpracování bibliografických záznamů z plných textů](#)

Plné texty jsou získávány na základě speciální aplikace **TTSNK (TamTam Special NK) z databáze TamTam (Anopress)**.

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 vlastní **automatickou indexaci článků a plných textů** - pro vytváření bibliografických záznamů v UNIMARCU a metadat Dublin Core v různých jeho aplikacích v HTML, XHTML, XML na základě údajů uložených v plných textech - je určena technologie **TamTam Data Extractor (TTDE)**.

Údaje jsou extrahovány z plného textu a na jejich základě je generováno **5 hlaviček (headers)**. **Data jsou zpracovávána pomocí rozšířeného formuláře pro editaci**.

Rozšířený formulář je nástroj vyvinutý pro editaci a doplnění extrahovaných dat. 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 klíčová slova. 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 (pro stahování je možné využít i verzi TamTam Standard - TTS).

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. Poté je soubor zpracovaných článků odeslán. Tímto způsobem jsou zpracovávány všechny celostátní deníky a některé odborné časopisy. Podrobný popis linky a pracovního formuláře je uveden ve zprávě z r. 2001.

Po doplnění formuláře je po odražce různé je **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.

Linka je použitelná po úpravě vstupním filtrem i na data existující v jiné databázi, event. v komunikaci mezi autorem, nakladatelstvím, bibliografickou agenturou a naopak.

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.

Linka zpracování zabudovaná do předpokládaného možného vývoje (viz Příloha 2).

B.1.1.3 Konverze do XML, XHTML, indexace, propojení, uložení, vyhledávání a zpřístupnění informací v systému TOPIC, topiky

Uživatelský formát: Dublin Core



Celkem nalezeno 1 článků [Tisk seznamu]

1: 0.87 20.03.2002 Lidové noviny str. 6 [13325 B]

Báseň, jenž se národu nezavděčil

Dublin Core Metadata **Stáhnout označené**

Označit všechny články v seznamu

DC/HTML

Title
Báseň, jenž se národu nezavděčil

Creator.personalName
Slomek, Jaronir

Subject
čeští básníci

Subject, scheme=ANL-KATEGORIE
literatura: česká literatura

Subject, scheme=MDI
821.162.3-051

Subject, scheme=MDT
(437.3)

Subject, scheme=PHNK
Machar, Josef Svatopluk, 1864-1942

Subject, scheme=PHNK
básníci-česko-stol. 19.-20.

Coverage.spatial
Česko

Coverage.temporal
stol. 19.-20.

Description.abstract
Před šedesáti lety, ten pár tvdmů předtím.

Pro **propojení plných textů s bibliografickými záznamy v systému ALEPH** (doplnění URL adres do záznamů) byl vytvořen **skript MKDOC.ASP**. Propojení probíhá na základě dynamicky generovaného odkazu na dokument. Program vyhledá požadovaný dokument dle identifikace (identifikační číslo), provede statistiku a

kontrolu autorizace a na jejím základě zobrazuje plný text, abstrakt (souhrn) a citaci dokumentu v rámci databáze ANL FULL.

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

Věcná indexace

Automatická indexace věcná je závislá na vytvoření tezaurů řízených heslářů s vazbou na klasifikační notaci v podobě MDT (UDC) oproti kterým je porovnáván plný text a na základě tohoto porovnání přidělovány příslušné věcné termíny. V rámci projektu **Propojení analytických záznamů s plnými texty je prováděna v roce 2002 analýza problematiky automatické či poloautomatické indexace (viz výše). Z výše uvedené analýzy zřejmě nebude řešena indexace automatických přiřazování věcných termínů. V současné době jsou funkční některé formy automatické extrakce, do budoucna lze uvažovat a podpoře intelektuální věcné indexace a možnosti generování věcných topiků on line, o využití interaktivních vlastností TOPICu.**

Pro vyhledávání v datech ve fulltextové databázi jsou vyvinuty formuláře pro vyhledávání jednoduché, pokročilé, pokročilé s tématy (topiky). Vyhledávat lze též podle rejstříků. Vyhledávání probíhá v systému TOPIC a definice formulářů vychází z jeho filozofie.

V rámci projektu Propojení analytických záznamů s plnými texty a v projektu Souborná databáze jsou v r. 2002 vytvořeny a implementovány připraveny další topiky, a to literatura, archeologie, historie, umění, geografie. Stávající topiky jsou aktualizovány.

Výhledově je možno propojit jednotlivé úrovně topiků. V současné době probíhají práce na vytvoření šablony pro geografické entity a věcná témata. Plánuje se postupné zabudování materiálu získaného věcnou indexací v rámci linky do tematických topiků, které by byly složeny z údajů uvedených v kategoriích a z údajů, uložených ve vstupních prvcích předmětových heslech. V rámci projektu Propojení byl též připraven řízený slovník pro periodikum Národní knihovna. Může sloužit jako základ topiku.

Plné texty a metadata jsou zpřístupňována pomocí uživatelských formátů a údaje DC jsou aplikována ve formátu HTML, XHTML a XML.

B.1.2 TOPIC (charakteristika)

[\(ukázky\)](#)

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 (větších) 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 skutečností, MDT je víceméně taxativní systém jednotlivých oborů, nikoli témat. Proto je třeba k definici topiků přistupovat svébytně.

V oddělení analytického zpracování při věcném popisu článků se používají k indexaci hrubých témat a podtémat předmětové kategorie, které připomínají svou podstatou topiky, resp. témata a skupiny témat v systému TOPIC v databázi Anopress. Je však třeba je sladit obsahově.

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 sblížuje 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 určíme, 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.

B.1.3 Charakteristika serveru full.nkp.cz a báze ANL FULL, Periodika na WWW - portál

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

B.1.3.1 Základní členění stránek - základní a dílčí nabídky

Stránky jsou tvořeny horní 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)

Další informace (seriály-periodika a analytický popis-články v České republice)

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)

B.1.3.2 Charakteristika databáze ANL FULL

[\(obsah, zdroje\)](#)

Databáze ANL FULL - fulltextové vyhledávání ve vybraných člancích 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, dále pak tzv. topiků (definovaných dotazů, resp. témat v rámci jednotlivých oborů). Topiky jsou rozděleny do tří úrovní: tematické oblasti (8), skupiny témat (25), detailní témata (zatím 113).

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 excerptce a typy článků, tituly viz dříve). Výběr titulů - v současné době celostátní deníky a některé časopisy (Bankovníctví, Ekonom, Euro - v přípravě, Kapitál, Magazin Práva, Pátek Magazin LD, Profit, Respekt, Reflex, Týden). O ANL viz Další databáze.

B.1.3.3 Další databáze (ANL) (viz též Kapitola A.2 , B.1 dále, [příloha](#))

Souborná databáze báze ANL obsahuje bibliografické záznamy vybraných článků novin, časopisů a sborníků zpracovávaných v rámci Kooperačního systému článkové bibliografie (KOSABI). Databáze vychází jako řada České národní bibliografie (kategorie titulů, úplnost excerptce a typy článků, tituly viz dříve). Vyhledávání je možné z údajů bibliografického popisu. Záznamy zpracovávané v lince TTDE jsou propojeny na plné texty umístěné v bázi ANL FULL. **Báze ANL obsahuje navíc linky na plné texty některých periodik a plná znění některých zákonů volně dostupných na na WWW (odborná knihovnická periodika - Ikaros, U nás, Bulletin SKIP, Knihovní obzor), další odbornová periodika (Lesnická práce, Psychiatrie, Vesmír, Harmonie, Collection of Czechoslovak Chemical Communications, Veřejná správa, Kriminalistika).**

Výběr titulů a článků ke zpracování.

Obsah:

V analytické bibliografii jsou obsaženy tyto obory: filozofie, psychologie, knihovnictví, náboženství, politika, ekonomika, ekonomie, sociologie, právo, pedagogika matematika, přírodní vědy, lékařství (okrajově), technika a průmysl, hudba, film, výtvarné umění, sport (omezeně), jazykověda a literatura, historie, geografie a všeobecnosti.

B.1.3.4 Periodika na WWW - [portál](#)

[Tituly v bázi ANL FULL. ANL. Anopress. volně přístupné na www](#)

[ukázky](#)

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 uveden oficiální název, www adresa, popis obsahu, vydavatel, tel., fax, e-mail, adresa. V budoucnu bude možno podle těchto údajů vyhledávat. **Nabídka není ještě zdaleka proporcionální, postupně se ladí.**

Dílčí 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 a umění**
 - film, televize
 - hudba
 - divadlo
 - literatura
 - **Odborná periodika**
 - architektura
 - astronomie
 - přírodní vědy
 - filozofie
 - ekonomie a finance
 - historie
 - knihovnictví a informační věda
 - lékařství
 - stavebnictví
 - technika, průmysl
 - **Společnost**
 - náboženství, církev, religionistika
 - životní styl
 - politika, politická scéna, zprávy, komentáře a glosy událostí
 - **Zahraničí**
 - zajímavosti, cestování
 - **Zájmy, hobby**
 - **Ostatní**
-
- **Další zdroje (institucionální členění)**

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

Takto zpřístupněná volně dostupná periodika jsou strukturována do přehledné formy portálu na úrovni krajské, tematické a institucionální.

B.1.3.5 Jak se zaregistrovat (komentář k registraci v databázi ANL FULL)

Bibliografické údaje o člancích, resp. metadata a plné texty jsou k dispozici čtenářům Národní knihovny. Pro externí uživatele jsou běžně k dispozici pouze bibliografické údaje o člancích. Tito uživatelé se musí pro získání plného textu zaregistrovat pomocí formuláře v nabídce Registrace. Po vyplnění povinných údajů bude těmto uživatelům zasláno e-mailem potvrzení registrace s aktivním odkazem a heslem - po odeslání tohoto odkazu je registrace aktivována a prostřednictvím přiděleného hesla zajištěn bezplatný zkušební přístup na dobu sedmi dnů. Rutinní zpřístupňování plných textů bude možné po vyřešení autorskoprávních, popř. ekonomických otázek souvisejících se zpřístupňováním plných textů. **Po registraci a přihlášení mohou také vybraní uživatelé stahovat a exportovat plné texty a metadata (test).**

Nabídka **Přihlášení** slouží k autentifikaci registrovaného uživatele. Ve formuláři pro přihlášení je třeba vyplnit jméno a přidělené heslo.

B.1.3.6 Jak vyhledávat a způsoby zobrazení (stručný přehled)

Báze ANL FULL obsahuje jak bibliografické popisy 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 míry relevance výsledku vyhledávání k položenému dotazu. Vyhledávat lze též pomocí rejstříků. Systém TOPIC navíc umožňuje pojmové vyhledávání podle témat, resp. topiků (viz dále). 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 topikami (viz dále).

Rejstříky - nedefinováno 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

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

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. **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 nedefinovaných topiků (označeny šipkou vlevo). Úrovně topiků se budou postupně propojovat a jednotlivé topiky ladit. **Perspektivy topiků v ANL FULL viz výše B 0.**

Tematický dotaz je vytvořen v podobě nedefinovaného 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 - obsaženo 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 (názevů vyhledaných článků) - viz dále.

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

Seznam výsledků (názevů 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 zprovozuje

Třídění seznamu výsledků (názevů vyhledaných článků) - viz dále.

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

B.1.3.7 Vyhledávání - metodika

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. Operator <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>.

B.1.3.8 Formuláře

Jsou k dispozici tři typy formulářů vzhledem k pokročilosti způsobu vyhledávání a možnostem kombinací při vyhledávání. Jednotlivé funkce formulářů a metodika zápisu údajů pro vyhledávání viz dále.

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 stejné 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. P>

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)
Skóre (relevance, od které zobrazovat články)
Výsl./str. (počet článků na stránku)

Třídění dle:

- skóre (relevance)
- názvu
- zdroje
- data (vydání)
- stran

a to sestupně nebo vzestupně.

Rozšířený formulář s tématy

Nabídky:

Tento formulář má **stejné 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.

B.1.3.9 Metodika vyhledávání

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
- **Kombinací** výše uvedených způsobů :
 - 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
- Pomocí **rejstříků** (které jsou tvořeny údaji v příslušných polích)

Pole dotaz, resp. text dotazu

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, planetky

Vyhledá dokumenty, které obsahují různé gramatické tvary slov "hvězdy" nebo "asteroidy" nebo "planetky" (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.

dac=29.10.2001

dac <contains> 29.10.2001

dat#29.10.2001

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 obsah u 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" .

[Aktualizovaná tabulka návěští pro pole dotaz a Prvky Dublin Core a Anl Core aplikované v bázi ANL FULL](#)

Využito 14 prvků DC, navíc přidán DC.Subtitle (vzhledem k minimálnímu záznamu pro ANL), 14 prvků AC definovaných pro vyhledávání a užívaných v popisu článku

Je-li případně tečková konvence s dc (Dublin Core) použita ještě tečková konvence s anl, používá se konvence s anl (Anl Core). Rovněž nepoužívejte návěští phnk, ale jeho alternativu.

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

Př: zdr=Lidové noviny

Zvýraznění vyhledaných údajů pomocí dotazového pole

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

Údaje, které jsou vyhledávány pomocí návěští (polí) nejsou zvýrazněny červeně, jsou obsaženy v citaci článku.

Spíše než návěští doporučujeme používat formulář s implicitně nastavenými poli.

Formulářová pole s implicitními údaji v rozšířených formulářích.

Přehledná tabulka metodiky pro vyhledávání

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

Ukázka vyhledávání pomocí topiků

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 (25), třetí obsahuje detailní témata ve skupinách témat (zatím 113). Z první úrovně nelze zatím vyhledávat, z druhé a třetí ano - tam, kde je topik nadefinován (označeno šipkou vlevo). Tematická nabídka bude v budoucnu postupně rozšiřována a témata upřesňována a propojována.**

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

Přehled dosud nadefinovaných/navržených topiků strukturovaných do tří úrovní

> označuje možnost vyhledávání z této úrovně topiku v současné době

→ návrh a postupná realizace propojení jednotlivých úrovní

1. úroveň	2. úroveň	3. úroveň	
Tematická oblast	Skupiny témat	Detailní témata	Poznámka
Ekonomika, obchod, finance	Ekonomika, ekonomie →→→→→	→→→ > Ekonomika, ekonomie > Makroekonomika > Pozemky, nemovitosti, byt > Regionální hospodářství	

	Finance →→→→→	> Bankovníctví	
		> Daně	
		> Finance	
		> Investice	
		> Kapitálový trh	
		> Měna	
	> Obchod	Práce	
	> Práce		
Země světa a geografie	Země světa	→→→> Filipíny	Bude rozpracováno
		→→→> Francie	
		→→→> Malajsie	
		→→→> Srí Lanka	
	→→→Geografie		
Hospodářství, výroba	Doprava, spoje →→→	→→→> Doprava →→→> Pošta, filatelie →→→> Telekomunikace	
	Energetika →→→	→→→> Energetika →→→> Hornictví, těžební průmysl →→→> Jaderná energetika →→→> Plynárenství	
	> Metrologie, normy, standardy		
	> Průmysl →→→	→→→> Automobilový →→→> Dřevozpracující →→→> Elektrotechnika →→→> Farmaceutický →→→> Hutnictví →→→> Chemický →→→> Papírenství →→→> Petrochemie →→→> Polygrafie, tiskařský →→→> Stavebnictví →→→> Strojírenství →→→> Textilní →→→> Zbrojní	
	> Výpočetní technika		
	Zemědělství →→→	→→→> Lesnictví, myslivectví →→→> Mlékárenství →→→> Pivovarnictví →→→> Potravinářství →→→> Zemědělství	

Kultura, umění	Kultura		
	Literatura, písemnictví →→→	→→→> Písmo →→→> Literatura →→→> Tisk →→→> Vzácné tisky	
	Umění →→→	→→→> Architektura →→→> Divadlo →→→> Film, kinematografie →→→> Fotografie →→→> Hudba →→→> Urbanismus →→→> Výtvarné umění	
Přírodní a matematické obory	Matematika, fyzika →→→	→→→> Astronomie, astrofyzika →→→> Fyzika →→→> Kybernetika, robotika →→→> Matematika →→→> Statistika →→→? Umělá inteligence	
	Přírodověda →→→	→→→> Biologie →→→> Botanika →→→> Ekologie, životní prostředí →→→> Genetika →→→> Geodézie, kartografie →→→> Geografie →→→> Geologie →→→> Chemie →→→> Mykologie →→→> Příroda obecně →→→> Veterinářství →→→> Zoologie	
Společenské a humanitní obory	Humanitní obory →→→	→→→> Demografie →→→> Etika →→→> Etnografie →→→> Filozofie →→→> Historie →→→> Jazykověda, lingvistika →→→> Psychologie →→→> Sociologie, společnost	
	Knihovnictví a informační věda →→→	→→→> Autority →→→> Bibliografie	Tyto topiky jsou provizorní. Pro periodikum

		<ul style="list-style-type: none"> →→→> Databáze →→→> Dokumentace →→→> Knihovnictví →→→> Knihovny →→→> Informační služby →→→> Jmenný popis →→→> Písmo →→→> Příručky →→→> Seriály →→→> Standardizace →→→> Zpracování dokumentů 	Národní knihovna se počítá s tvorbou tezauru/řízeného hesláře, který bude podkladem topiku pro periodikum Národní knihovna
	Společnost →→→	<ul style="list-style-type: none"> →→→> Feminismus →→→> Náboženství →→→> Politika →→→> Právo →→→> Sociální zabezpečení →→→> Školství →→→> Věda a technika →→→> Veřejná správa →→→> Vojenství, vojenská technika 	
Sport, volný čas	> Společenské a lidové zábavy		
	Sport →→→	<ul style="list-style-type: none"> →→→> Atletika →→→> Basketbal →→→> Cyklistika →→→> Fotbal →→→> Gymnastika →→→> Lední hokej →→→> Motorismus →→→> Tenis →→→> Volejbal 	
	> Volný čas		
Zdravotnictví, lékařství	Zdravotnictví	<ul style="list-style-type: none"> > Terapie > Toxikologie > Zdravotnictví 	Bude rozpracováno
	Farmacie, farmakologie		

Rejstříky

[\(ukázky\)](#)

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

- Údaje obsažené v polích bibliografického popisu doporučujeme vyhledávat spíše pomocí implicitních polí v rozšířených formulářích než vypisováním návštějí 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)

B.1.3.10 Výsledky vyhledávání, zobrazení, tisk (všechny formáty)

B.1.3.10.1 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 zprovoznována - 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)
- názvu
- zdroje
- data (vydání)
- stran

to sestupně nebo vzestupně.

B.1.3.10.2 Zobrazení údajů o článku

Formáty zobrazení

[\(ukázky\)](#)

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 určeny pro běžného uživatele, ale pro budoucí vývoj systému a zpřístupnění plných textů.)

B.1.4 Metadata

[Dublin Core, ANL CORE s vyhledáváním v bázi ANL FULL, metadata](#)
[Kompletní seznam metadat v uživatelském formátu, resp. citace](#)

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

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 Hvízd'ala
Zdroj:	Mladá fronta Dnes
Zdroj-příl.:	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)
	Kohout, Pavel, 1928-
Téma jako předmět:	anglicky psaná literatura
	americká próza
	spisovatelé-Spojené státy americkéo-stol. 20.
Typ dokumentu:	recenze

B.1.5 Tisk, stahování a export plných textů a metadat - přístup z nabídky Vyhledávání (nové)

Tisk seznamu výsledků - z nabídky Tisk seznamu, tisk plného textu - z nabídky Tisk detailu.

V r. 2002 je řešena aplikace **pro stahování a export metadat a plných textů**, která vychází z již vyvinuté technologie v Anopressu. Aplikace umožňuje stáhnutí a export příslušných plných textů a metadat na lokální počítač v zip souboru v několika formátech: 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 metadata mohou zatím jen registrovaní vybraní uživatelé po dobu 7 dní (test).

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

V roce 2002 byla ověřena plná funkčnost aplikace pro registraci, autentifikaci a přihlášení externích uživatelů prozatím na dobu 7 dnů. Otázky zpřístupnění externím uživatelům závisí na vyřešení legislativně právních otázek event. ekonomických. **Externím uživatelům jsou běžně zpřístupňována metadata. Plné texty a metadata jsou zpřístupňována interním uživatelům NK.** **Stahování a export je podle testu také plně funkční.**

B.1.6 Administrace (nové)

[\(ukázky\)](#)

V roce 2002 byla vyvinuta aplikace pro administraci databáze ANL FULL a portálu. Úpravy textů článků - aplikace umožňuje opravovat zdrojové kódy článků, metadata v nich uložená. 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.

B.1.7 Standardizace - UNIMARC, DUBLIN CORE, HTML, XHTML, XML, propojení záznamů s plnými texty

[Dublin Core, ANL CORE s vyhledáváním v bázi ANL FULL, metadata](#)
[Kompletní seznam metadat v uživatelském formátu, resp. citace](#)

Struktura bibliografických dat respektuje formát UNIMARC a knihovnická pravidla AACR2 v oblasti jmenného popisu. **V r. 2002 probíhají práce na konverzi UNIMARC - MARC 21.** 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 sbližovat s tématy systému TOPIC. Předmětové kategorie do jisté míry konvenují metodě konspektu aplikované pro popis a mapování fondů. Dále se používají klíčová slova, která jsou dále částečně řízená a **předmětová hesla**. Automaticky je generován **souhrn článku**, automaticky lze také generovat klíčová slova, která však mají v současném stádiu "počítačovou formu". **O automatické indexaci a perspektivách viz výše B.0.** V budoucnu se předpokládá intenzivnější využívání vznikajících souborů autorit jmenných i věcných.

V záznamech určených k propojení s plným textem se v lince automaticky generuje **dynamická URL adresa, SICI**. Do zpracovávaných článků v ALEPHu jsou doplňovány 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. Pole LKR je v nové verzi ALEPHu nefunkční.

Dublin Core obsahuje 14 z 15 definovaných údajů. Je generován pro formát HTML, XHTML a XML ve kvalifikované i nekvalifikované formě. Do HTML je zabudován LINK tag pro potřeby odkazu na webovský zdroj, v němž se nachází specifikace daného použitého souboru metadat. Bylo nadefinováno 14 údajů **Anl Core** vzhledem k detailnosti popisu a vzhledem k možnostem vyhledávání v současné verzi systému TOPIC.

Formáty zobrazení jsou popsány dříve.

Skutečný stav propojení bibliografických záznamů s plnými texty

Plné texty jsou zpracovávány v lince a zároveň jsou ukládána metadata do těchto textů za současného generování url. Takto vzniklá metadata resp. bibl. záznamy se importují do ALEPHu, odkud se propojují na plný text v systému TOPIC. Zároveň jsou metadata a plné texty indexovány v systému TOPIC. V roce do konce října 2002 bylo takto zpracováno propojeno cca 13 000 metadat a plných textů.

V roce 2002 dále pokračovalo **ruční propojování záznamů zpracovávaných v ALEPHu s volně přístupnými texty na WWW**. Bylo propojeno cca **2000 záznamů**.

Souhrn za r. 2000-2002: na serveru full.nkp.cz jsou přístupné plné texty článků vydané v letech 1990-2002 a získané v rámci tohoto grantu a grantu Propojení analytických záznamů v letech 1999-2002, 15. listopad v počtu cca **100 000**. **Zhruba stejný počet je propojen se záznamy v ANL.**

U malého procenta záznamů nejsou patrně odkazy aktivní (souvisí se změnou adres). V budoucnu bude třeba aplikovat technologii na ověřování hypertextových odkazů.

V lince bylo zpracováno v letech 2001-2002, 15. listopad celkem cca 24000 plných textů s metadaty, v r. 2000-2001, říjen propojeno přes **cca 4000 záznamů s volně dostupnými texty** na Internetu (některá propojení realizovaná 2000-2001 spadla). Jsou propojené také některé záznamy s plnými texty zákonů.

B.1.8 Souborná databáze ANL Kooperačního systému článkové bibliografie (KOSABI), metodika a organizace, ČNB

[\(ukázky\)](#)

B.1.8.1 Kooperace, excerpční základna

V roce 2002 pokračovalo oddělení ve zpracování záznamů pro bázi ANL(ALEPH).

Kooperující instituce - **8 krajských knihoven, MZK a 3 odborné knihovny** - přispívají do

souborné databáze. V budoucnosti se mají součástí KOSABI stát i 4 další krajské knihovny. V roce 2002 byly pravidelně dodávány záznamy z těchto knihoven: Kladno (RAPID), Ostrava (T-Series), SPKK (ISIS), STK (ISIS), ÚZPI (ISIS), Liberec, (RAPID).

Spíše po větších dávkách a nárazově byly dodávány záznamy z krajské knihovny v Českých Budějovicích (souvisí s problematikou Tinlibu resp. T-Series).

Ukládání online do Alephu - Moravská zemská knihovna v Brně.

Tento rok nebyly téměř dodávány záznamy z krajské knihovny v Plzni, kde jsou články ukládány v systému KIMS. Modul pro články a jejich export je patrně ještě ve vývoji. Podle posledních zpráv mají být záznamy dodány k posouzení do konce roku.

V r. 2002 se dále **ladí metodika ukládání v různých systémech a řeší se otázky převoditelnosti do báze ANL a UNIMARCu**. Tato činnost je poměrně náročná v detailech a je založena na úzké spolupráci s příslušnými knihovnami. V r. 2002 v tomto směru intenzivně spolupracuje oddělení článkové bibliografie s krajskou knihovnou v Ostravě a Českých Budějovicích. Jsou navržena některá možná řešení a doporučení týkající se věcného popisu a konverze do UNIMARCu. Plnohodnotné předávání záznamů do národní článkové bibliografie závisí na dotažení konverzního programu Tinlib - UNIMARC. Krajská knihovna v Ústí nad Labem na zkušenosti tinlibovských knihoven navazuje. Knihovny pracující v ALEHu spolupracují dobře. U knihoven pracujících v KP-sysu je třeba dopracovat převod do UNIMARCu. Knihovny pracující v Rapidu spolupracují též bez problémů. Krajská knihovna v Plzni přešla na zpracování v KIMSu. **S nově jmenovanými a konstituovanými krajskými knihovnami** v Pardubicích, Havlíčkově Brodě, Zlíně a Karlových Varech byla navázána dobrá spolupráce. Ladí se převod záznamů do UNIMARCu i metodika a zvažují možnosti zpracování některých titulů.

Kromě **KOSABI** existuje v ČR fungující systém na úrovni městských (okresních) knihoven - **LANius a jeho produkt SKAT - Souborný katalog článků**. Jeho vyšší verzí je systém Clavius. Systém umožňuje sdílenou katalogizaci. Mezi **LANiem a NK** pravděpodobně proběhnou v budoucnu jednání o možné kooperaci týkající se excerptce titulů a omezení duplicit při zpracování. Export do Baze ANL je v zásadě možný. Systém používá ve věcném popisu klíčová slova. Podobná duplicita existuje mezi krajskými knihovnami a LANiem. **Součástí širšího kooperačního systému článkové bibliografie** po linii oborové je zpracování článků lékařské a zdravotnické literatury **Národní lékařskou knihovnou, které nejsou součástí ANL**

V bázi ANL je v r. 2002, k 15.11. 742 379 záznamů, za rok 2002 přibylo do báze cca 54 196 záznamů, z toho cca 30 118 zpracovaných NKČR, ve spolupracujících institucích cca 24 078 zpracovaných záznamů. Počty jsou přibližné.

V roce 2002 je pravidelně aktualizována excerptční základna.

- počet zpracovávaných titulů: ANL - cca 212 v NKČR, 217 v regionech, 108 v odborných knihovnách, celkem cca 537 titulů; ANL FULL 30 titulů, běžně zpracovávaných zatím 16 v NKČR

Excerptční základny jednotlivých kooperujících institucí pro bázi ANL

Národní knihovna České republiky (ALEPH) - Oddělení analytického zpracování: celkem 196 seriálů z toho 6 deníků, 11 týdeníků, 6 čtrnáctideníků, 31 dvouměsíčníků, 4

čtvrtletníky, ostatní periodika s menší periodicitou ; dále nepravá periodika - ročenky, nepravidelně vycházející periodika, sborníky - počet pohyblivý.

NKČR - Odbor knihovnictví: 19 titulů.

Moravská zemská knihovna v Brně (ALEPH): celkem 54 titulů, 2 deníky, vlastivědné časopisy, zpravodaje, sborníky a nepravá periodika, v jednání postupné předání titulů Krajské knihovně Vysočiny.

Jihočeská vědecká knihovna v Českých Budějovicích (Tinlib): celkem 23 titulů, 3 deníky z toho 2 přílohy regionální deníků, periodika s menší periodicitou a nepravá periodika.

Studijní a vědecká knihovna v Hradci Králové (ISIS): celkem 21 titulů, z toho 8 deníků, postupně delimitace titulů po dohodě s Krajskou knihovnou v Pardubicích, Krajskou knihovnou Vysočiny, Krajskou knihovnou v Liberci.

Středočeská vědecká knihovna v Kladně (Rapid): celkem 37 titulů, z toho 12 deníků.

Krajská vědecká knihovna v Liberci (Rapid, resp. Arev): celkem 11 titulů, z toho 2 deníky.

Vědecká knihovna v Olomouci (ALEPH): celkem 17 seriálů: 2 deníky - pouze regionální přílohy, 6 týdeníků, nepravá periodika. Delimitace titulů s MVKOS.

Moravskoslezská vědecká knihovna v Ostravě (Tinlib, resp. T-Series): celkem 16 titulů, z toho 1 deník, 1 týdeník, 4 čtvrtletníky, ostatní nepravá periodika a sborníky.

Studijní a vědecká knihovna v Plzni (KIMS): celkem 22 titulů, z toho 8 deníků, 8 týdeníků, 6 měsíčníků a další periodika.

Severočeská vědecká knihovna v Ústí nad Labem (ISIS, Tinlib): celkem 16 titulů, z toho 7 deníků.

Nově konstituované krajské knihovny - příprava kooperace

Krajská knihovna Františka Josefa Bartoše ve Zlíně (KP-sys): cca 67 periodik, duplicita s celostátními deníky a časopisy, do její báze přispívají městské knihovny okresu Zlín, nutná delimitace s titulů s MZK, VKOL, MVKOS - v říjnu 2002 stanovena pracovníky knihovny excerpcí základna s ohledem na nové státoprávní uspořádání.

Krajská knihovna Karlovy Vary (LANius): celkem cca 26 titulů, zpracovává články pro LANius.

Krajská knihovna v Pardubicích (KP-sys): cca 10 titulů, výběrově vlastivědné a místní zpravodaje.

Krajská knihovna Vysočiny (Havlíčkův v Brod, Clavius): údaje z r. 2001 - celkem cca 30 titulů, duplicita s NK, doporučeno nejdříve zpracovávat tituly týkající se Havlíčkovobrodské, postupně přibrat další, nutná delimitace s MZK (cca 7 titulů), nedostatečné personální obsazení.

Kooperující specializované knihovny

Státní technická knihovna (ISIS): celkem 41 odborných titulů

Ústav zemědělských a potravinářských informací (ISIS) : celkem 24 odborných titulů

Ústav pro informace ve vzdělávání - Státní vědecká knihovna Komenského (ISIS) : celkem cca 43 titulů

Český geologický ústav (ISIS): odborná periodika v rámci experimentu

V roce 2002 probíhala mezi některými knihovnami KOSABI a 4 novými krajskými knihovnami jednání o přerozdělení titulů k excerpci vzhledem k novému státoprávnímu uspořádání.

V rámci regionálních funkcí mají krajské knihovny koordinovat bibliografickou činnost v rámci krajů.

V r. 2002 byla provedena anketa týkající se podchycování regionálních materiálů z ústředního tisku. V odpovědích bylo zaznamenáno, že regionální bibliografové považují ČNB - Články v českých novinách, časopisech a sbornících za důležitý zdroj regionálních informací. Zdůrazňuje se důsledné používání regionálních kódů (viz Nádvořnicková, M. Nové formy a metody práce při poskytování regionálních bibliografických informací).

V r. 2002 byl agentuře Anopress navržen seznam titulů, které by eventuelně mohla plnotextově zpřístupňovat.

B.1.8.2 Popis, standardizace

V roce 2002 se poměrně hodně času věnuje projektu **krajské knihovny v Ostravě** týkající se popisu článků v systému T-Series a jejich zpřístupnění. Jsou dále 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. byl 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. Tento slovník je ve většině knihoven KOSABI používán a do jisté míry modifikován s ohledem na praktické potřeby. Slovník se musí v samotné bázi ANL zredigovat a v současné době již prochází první korekturou. V budoucnu patrně dojde ke sblížení kategorií používaných v KOSABI a těch, používaných v metodě konspektu.

V Ostravě nebylo možno zatím začlenit kategorie ANL do popisu, počítá se s využitím metody konspektu v budoucnu. V Ostravě zatím používají vlastní slovník klíčových slov.

Poměrně problematická situace vznikla v krajské knihovně v Českých Budějovicích, V roce 1998 jsme s paní Strakovou (JVK České Budějovice) minulosti navrhovaly 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 tezurů, 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 jde o rozhodnutí, jak pokračovat v zápisu článků v Českých Budějovicích.

Hlavní řešitelka projektu Souborná databáze navrhla několik variant věcného popisu v Českých Budějovicích.

Pracovnice krajské knihovny se překlony k následujícímu řešení: používat výrazy tezauru, které 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.

V roce 2002 na T-Series přechází i krajská knihovna v Ústí nad Labem, kde bude používána stejná metodika jako v Ostravě.

Záznamy 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**.

V rámci ANL i ANL FULL jsou formulovány základní principy věcného popisu, aplikovaného při zpracování článků v NKČR. **Pořadí kategorií, předmětových 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. Předmětová hesla tematická (pole 606 UNIMARC) se užívají zejména v NK, ve většině knihoven se praktikují klíčová slova (pole 610 UNIMARC). Vazba na autority se objevuje ve větší míře v NK a v krajské knihovně v Kladně. Je třeba důsledněji používat pole 660 Geografický kód.

V současné době se řeší NKČR problematika standardizace věcných selekčních údajů a jejich harmonizace pomocí souboru věcných autorit. Aplikuje se **metoda konspektu** - metoda popisu fondu a všech informačních zdrojů pomocí předmětových kategorií konspektu. Domníváme se, že po důkladné redakci předmětových **kategorií v rámci báze ANL**, které jsou někde detailnější a vyplývají ze specifiky článkových informací, bude moci dojít k částečnému sblížení konspektu a kategorií v ANL, je možno je také užívat paralelně. 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.

V červnu 2002 přechází NK a oddělení analytického zpracování na verzi ALEPH 14.2.4, která umožňuje větší komfort pro opravy uložených dat. Na tuto verzi přecházejí též v r. 2002

(před NK) i knihovny v Brně a Olomouci. Pracovníci MZK ukládají záznamy přímo do báze ANL.

B.1.8.3 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.

B.1.8.4 Česká národní bibliografie - řada Články v českých novinách, časopisech a sbornících na CD-ROM

V roce 2002 byla pravidelně poskytována data KOSABI pro měsíční aktualizace a **CD-ROM ČNB - řada Články v českých novinách , časopisech a sbornících.**

B.1.9 Perspektivy Kooperačního systému článkové bibliografie a jeho další možný vývoj, některé předpoklady a cíle

Organizace spolupráce

- Postupně reálně začlenit nově konstituované krajské knihovny v Pardubicích, Karlových Varech, Havlíčkově Brodě a Zlíně do KOSABI. Přerozdělení některých titulů vzhledem k novému státoprávnímu uspořádání.
- Postupně sladit profil zpracovávaných titulů v KOSABI se Souborným katalogem článků systému LANius, resp. Clavius.
- Spolupráce s institucemi pedagogického a vědeckého zaměření (UK, ČSAV aj.).
- Spolupráce se společností Anopress a event. navázání spolupráce s dalšími nakladateli/vydavateli .

Metody zpracování, standardizace

- Doladit převod výstupů z některých systémů do standardního UNIMARCu a báze ANL.
- Harmonizace popisu článků v rámci KOSABI se soubory autorit zejména v oblasti věcného zpracování (předmětová hesla, předmětové kategorie a metoda konspektu).
- Konverze UNIMARC - MARC 21, pravidla pro popis článků vzhledem k tomuto formátu.
- Postupně testovat nové metody zpracování vyvíjené v projektech na KOSABI (linka automatického zpracování bibliografických záznamů s plných textů a zpřístupnění plných textů v bázi ANL FULL): výhoda jednotného a poměrně jednoduchého a

rychlého zpracování importního souboru bibliografických záznamů pro integrované knihovnické systémy za paralelního zabudování metadat do plných textů a možnosti vyhledávání v plných textech. Pracovníci v současné době pouze kontrolují extrahované jmenné údaje z plného textu a doplňují věcné údaje. Předpokládá se, že v budoucnu dosud doplňované věcné údaje bude možno zabudovat do topiků a tak podle nich vyhledávat. Metody automatické či poloautomatické indexace věcné jsou ve stádiu zkoumání. **Cíl: pouhý výběr titulů ke zpracování a kontrola, doplnění věcným popisem (viz analýza automatické indexace výše).**

- Testovat metody, popř. navázat na metody získávání a zpracování elektronických plných textů (projekt Webarchiv).
- **Postupně přecházet z tradičního zpracování článků obsažených v určených titulech na zpracování pomocí nových technologických metod. Přechod však musí být uvážlivý vzhledem k zachování kontinuity zpracování titulů.**

Organizace zpracování článků vzhledem k excerpovaným titulům z hlediska systémového

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

- I. Zpracování sborníků a nepravých periodik (v rámci integrovaných knihovnických systémů), které zatím nejsou dostupné na Internetu. Většinu pracovních kapacit přesunout na tuto činnost.**
- II. Zpracování deníků a časopisů v rámci linky zpracování bibliografických záznamů z plných textů. Tato linka ušetří pracovní kapacity, zpracování je rychlé, kvalitní se spolehlivým napojením na plné texty.**
- III. 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.

Současné výstupy z KOSABI a jejich možné zpřístupnění v budoucnu

- Báze ANL a ANAL (Olomouc) je zpřístupněna v Metalibu jako zdroj, ANL FULL a článková databáze krajské knihovny v Liberci jako odkaz. Další zpřístupnění článkových databází závisí na strategii Metalibu a příslušných institucí.

- Propojení je realizováno v rámci linky zpracování bibliografických záznamů s plnými texty zatím v rámci NK, tuto metodu lze aplikovat i v dalších institucích KOSABI na dohodnutých titulech.
- Linky na volně přístupné plné texty na Internetu jsou součástí ČNB, Řady články v českých novinách, časopisech a sbornících. Linky na plné texty báze ANL FULL zatím součástí ČNB nejsou.

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.

Diferencované zpřístupnění plných textů

Jasně vymezení a ošetření přístupu k plným textům, které jsou zpřístupňovány volně, přes konzorcía, interním a externím uživatelům institucí a způsob plateb (paušál, kredity, apod.).
V současné době probíhají práce a jednání o zpřístupnění plných textů z báze ANL FULL - stahování a export záznamů pro NKČR, knihovny krajské a MZK.

Personální a finanční zajištění KOSABI

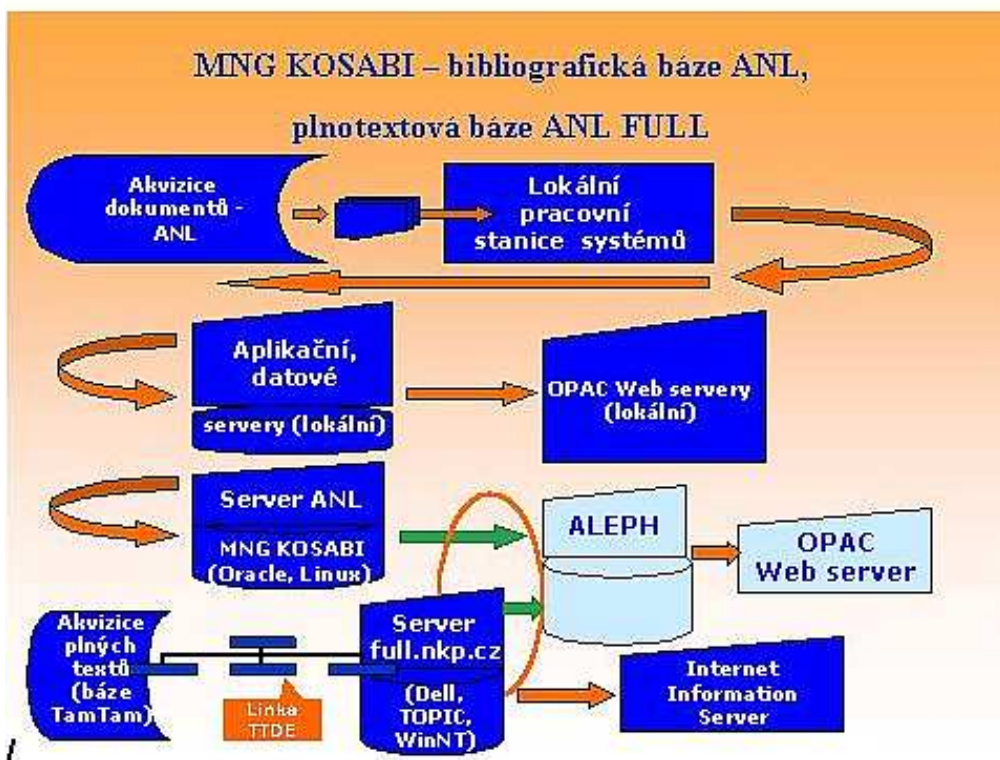
- KOSABI se již tradičně potýká s problémem personálního zajištění činností. Dobrá personální situace je podmínkou převedení systému do nového prostředí a pro aplikaci nových metod. Tyto metody pomohou práci zrychlit a zkvalitnit za menšího počtu pracovníků. Metody je však dobré nejdříve vyvinout a dobře 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 jsou plánovány na cílový počet 11 pracovníků v r. 2003
- Vývoj systému je po finanční stránce zajištěn do roku 2004 díky podpoře MKČR. Je na místě pomalu uvažovat a návazném projektu.

B.1.10 Management Kooperačního systému článkové bibliografie, aplikace MNG KOSABI

[\(ukázky\)](#)

B.1.10.1 Management KOSABI - systémový pohled

Management KOSABI - architektura



Management KOSABI se dělí na správu a údržbu dvouází: **bibliografické báze ANL a plnotextové báze ANL FULL.**

Správa a údržba ANL FULL byla popsána výše, v letošním roce byla pro ni vyvinuta aplikace pro administraci báze. Administrace je funkční a zralá pro poloprovozní stádium.

Aplikace pro správu a údržbu báze ANL je ve stádiu experimentu až poloprovozu.

- **MNG ANL**
báze titulů - volně přístupná
upload dat od kooperujících institucí - registrace
export/import záznamů pro bázi ANL v ALEPHu - registrace
- **ANL**
bibliografická báze - volně přístupná
- **MNG ANL FULL**
TTSNK licence (TamTam Special NK)
linka zpracování TTDE (TamTam Data Extractor) zatím lokálně v NKČR
(zpracování bibliografických záznamů z plných textů, zpřístupnění ve formátech UNIMARC, HTML, XHTML, XML)
báze ANL FULL (import metadat a plných textů z TTDE) - systém TOPIC
báze ANL (import metadat z TTDE) - systém ALEPH
externí uživatelé - metadata volně
externí uživatelé - plné texty na 7 dnů - registrace a přihlášení
vybraní uživatelé - export dat na 7 dnů - registrace a přihlášení - test WWW

**periodika (portál) - volně
administrace báze ANL FULL a portálu - registrace**

B 1.10.2 Báze ANL

Na základě zadání pro řízení a správu kooperačního systému a za využití již vyvinutých řešení v rámci Souborného katalogu CASLIN probíhají dále práce na vývoji aplikace pro KOSABI na serveru ANL(systém LINUX a ORACLE). Ve stádiu experimentu až poloprovozu **vzhledem k nutnosti průběžného náročného ladění je aplikace pro příjem, automatizované zpracování a export dat (přijímání analytických záznamů, integrace stávajících programů pro konverzi analytických záznamů, globální úpravy analytických záznamů, vývoj programů na formálně logické kontroly kooperujících knihoven - test na UNIMARC pro analytické záznamy, test na kritické chyby, test na duplicitu klíčů).** Export záznamů pro ALEPH má být řešen do konce roku 2002. Dále probíhají práce na aplikaci pro bázi titulů, která je v poloprovozním stádiu. Upload záznamů je zatím testován v NKČR, v r. 2003 proběhne test s kooperujícími institucemi.

Seznam prací v roce 2002

1. Testování uploadu souborů získávaných od institucí spolupracujících v KOSABI
2. Specifikace úprav dat na výstupu z aplikace (export pro ALEPH)
3. Vypracování podkladu pro import dat (Excel) pro aplikaci Báze titulů (evidence excerptní základny)
4. Návrh a implementace jednotného www rozhraní pro management KOSABI pro přístup do dílčích aplikací:
 - o Management KOSABI: ([stručná charakteristika - html stránky](#))
 - o Vstup do aplikace:
 - o Báze titulů
 - o Upload předávání dat do KOSABI
 - o Import - statistiky (včetně exportu do Alephu)
 - o Klíče (kontrola na duplicitu záznamů)
5. Upgrade serveru ANL: operační systém Linux SuSE 7.3
 - o Nové verze: Oracle 9.2
 - o Java j2sdk1.4.0-01
 - o Servlety jakarta-tomcat 4.0.4

B.1.10.3 Aplikace pro příjem a zpracování dat, export do ALEPHu

Aplikace pro správu a údržbu KOSABI - přehled dílčích funkcí

1. Příjem dat pro systém

- způsoby příjmu dat: ftp, upload, disketa
- formáty: Aleph, Aleph500, UNIMARC řádkový, UNIMARC ISO 2709, Výměnný formát

- kódování: GIZMO, PC Latin 2, Kamenicky, ISO 8859-2, UTF-8, Ansel, ISO 646
- při importu se u každého záznamu vytvoří klíč na duplicitu záznamů (dále uložení klíčů v bázi a následně zobrazení klíče s počtem záznamů)

2. Zpracování přijímaných dat

- přidělení identifikace pro zpracovávanou dávku (batch_id)
- formát a kódování dle názvových konvencí
- překódování do UTF-8.
- konverze do UNIMARC (u formátu, které to vyžadují)
- převod do interní struktury SQLLCB
- globální úpravy záznamu - 1. etapa (chyby v datech, které lze programově ošetřit, např. tvrdé mezery, tagy ve vazebních polích)
- test UNIMARC
- uložení do baze SKA
- statistika zpracování

Testování uploadu ve všech zmíněných formátech a kódování, vzhledem k poměrně náročnému průběžnému ladění vstupů do ANL a přechodu na novou verzi ALEPHu ve stádiu experimentu až poloprovozu.

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

Export:

- **kritické chyby**
- **chyby na UNIMARC**

Na exportu dat pro ALEPH probíhají práce do konce roku 2002 vzhledem k požadavku diferenciacce mezi kritickými chybami a chybami na UNIMARC.

Aplikace pro správu a údržbu KOSABI - přehled základních funkcí z hlediska správce KOSABI

- Přijímání dat: zpracování probíhá bez zásahu správce KOSABI (kromě kopírování dat z diskety)
- Export dat: správce zadá číslo batch_id, které určuje exportované záznamy
- Statistika: přehled o přijatých datech, statistiky zpracování členěny:
 - dle batch_id
 - přehled o exportovaných datech

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

crerr - test na kritické chyby

B.1.10.4 Aplikace pro evidenci zpracovávaných titulů (doplnění báze základními daty excerptní základny do nadefinovaných polí, které odpovídají SK CASLIN)

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. Pro jednotlivé tituly je definován krátký záznam dle polí UNIMARCu.

Formát pro bázi titulů (modifikovaný formát SHORT SK CASLIN).

Názvové údaje: 200a. 200i [200b] : 200e

Zdroj katalogizace: [801b].[801c]

Sigla vlastníka: [910a].[910a]

Region: [952a].[952a]

Údaje o excerptci: [980a].[980b].[980c].[980k].[980p].[980r]

K této bázi budou nadefinovány přístupové soubory pro možnost hledání a báze bude doplňována a aktualizována. Jsou uvedeny zatím základní údaje o titulech.

B.1.10.5 Test na duplicitu klíčů

Klíče, podle nichž je prováděna kontrola na duplicitu přijímaných dat.

Kontrola na duplicitní záznamy již uložené v ALEPHu není mimo ALEPH možná.

Mezi základní klíče patří název článku, autor, název zdrojového dokumentu, datum vydání.

Ve stádiu řešení a testování, na jehož základě budou klíče upraveny a proveden návrh na řešení duplicit.

B.1.11 HW a SW podpora pro správu KOSABI a pro bázi ANL FULL

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 jednoprocessorový 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.

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

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

B.1.12 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. Všechny smlouvy zůstávají v platnosti.

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 dosavadní závazky vůči projektům (zpráva pro MKČR z května 2002).

B.2 Přínos řešitele

Přínos projektu v r. 2002 spočívá v realizaci metod navržených v r. 2000 a 2001 ve stádiu úspěšného poloprovozu, který má až provozní charakter v případě budování báze ANL FULL v NKČR. V případě aplikace pro příjem a zpracování dat pro bázi ANL je pro projekt v experimentálním až poloprovozním stádiu vzhledem k nutnosti průběžného ladění v závislosti na různých systémech spolupracujících institucí a zařazování nových knihoven do systému.

Přínos projektu v r. 2002 spočívá v praktické integraci elektronických zdrojů do služeb knihovny pomocí technologicky nejvypělejších nástrojů pro zpřístupnění těchto zdrojů při zachování tradičně zpřístupňovaných sekundární informací formou bibliografických záznamů. Jde o integraci heterogenních dat do Kooperačního systému článkové bibliografie, v němž dochází k propojení tradičních knihovnických postupů a fondů s určitými prvky digitální knihovny.

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

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.**

Ukazuje se, že věcný popis dokumentů, v němž jsou zabudovány tzv. **předmětové kategorie konvenuje principu popisu fondů metodou konspektu a koresponduje s trendem budování "subject gateways" pro přístup k informacím. Pro pojmové vyhledávání byly aktualizovány topiky vypracované v roce 2000 a 2002 doplněny o nové definice topiků. Některé topiky jsou navzájem propojeny napříč jejich úrovněmi.**

Další přínos projektu spočívá v rychlém **zpřístupnění analytických záznamů provázaných s plnými texty zdrojových dokumentů v rámci Kooperačního systému článkové bibliografie v postupném propojování záznamů báze ANL se záznamy o zdrojových dokumentech v bázi NKC. Báze ANL je jedním ze zdrojů Metalib, ve kterém jsou plné texty také zpřístupňovány. Báze ANL FULL je do Metalibu připojena jako odkaz.**

Další přínos spočívá v **převedení digitalizované formy periodika Národní knihovna do báze ANL FULL v podobě HTML do báze ANL FULL a vybavení těchto plných textů základními automaticky generovanými metadaty.**

Pro toto zpřístupnění plných textů byl dále aktualizován **moderní layout stránek serveru full.nkp.cz v asp skriptu.**

Velká pozornost byla věnována definici báze ANL.FULL, zejména pak systému vyhledávání.

Byla navržena a realizována kompletní aplikace pro administraci báze (opravy metadat a statistiky).

Byla navržena a realizována aplikace pro stahování a export plných textů s metadaty v různých formátech (text, html, rtf, XML, UNICODE). Stahovat a exportovat plné texty s metadaty mohou pouze vybrané instituce.

V r. 2002 je udržován **portál** pro zpřístupnění volně dostupných elektronických periodik na Internetu, která jsou členěna na principu krajů, tematiky a oblastí, institucionálním.

V roce 2002 byla **posílena kooperace v rámci KOSABI** po stránce metodické i organizační. Dále byla posouzena konverze T-Series UNIMARC vzhledem k systému zavedenému v bázi ANL a navrženy některé korekce. Byla **nastíněna možnost vývoje systému v budoucnu v rámci tří zpracovatelských subsystémů podle typu dokumentů. Agentuře Anopress byl navržen seznam titulů, které by eventuelně mohla plnotextově zpřístupňovat.**

Možnost vzniku **duplicit či multiplicit** při zpracování je minimalizováno delimitací periodik mezi spolupracující instituce. Tím má projekt další i ekonomický význam. Byla navázána spolupráce se čtyřmi novými krajskými knihovnami.

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

otestována aplikace k zajištění registrovaného přístupu externích uživatelů k plným textům. Tito uživatelé se mohou dnů zkušebně zaregistrovat na dobu 7 dnů.

V případě vhodných podmínek přichází v úvahu možnost aplikace vypracovaných metod na některé spolupracující subjekty v Kooperační systému článkové bibliografie v budoucnu a po jisté úpravě aplikace linky zpracování bibliografických záznamů z plných textů v komunikaci mezi autorem, nakladatelem, informační institucí, bibliografickou agenturou.

Pro projekt jsou důležité analýzy, které se týkají automatické indexace a které byly provedeny v rámci projektu Propojení analytických záznamů s plnými texty. Vyplývá z nich, že metody řešené v projektu souvisí spíše s automatickou extrakcí dat. Metoda automatického přiřazování údajů věcného popisu nebude patrně aplikována. Projekt se dále zaměří na využití možnosti inteligentního vyhledávání systému TOPIC a interakci mezi systémem a uživatelem a systémem a indexátorem.

Od spolupráce mezi Národní knihovnou a ANOPRESSEM se očekává **ekonomický efekt** ve smyslu šetření **pracovních kapacit** Národní knihovny. V oddělení je plánována do r. 2003 redukce o 6 pracovních úvazků ve prospěch jiných oddělení NK. **V tomto je realizována další, pátá redukce.**

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. Dále bylo zjištěno, že podobný model zpracování a podobné technologie jsou vyvíjeny např. ve Švédsku, Spojených státech a jsou do jisté míry s nimi oprávněně srovnatelné. Podobný model reorganizace zpracování národní bibliografie se vyvíjí např. v Austrálii.

Aplikace pro management KOSABI byla testována a dále vyvíjena (další specifikace pro opravy dat při jejich exportu ze systému). Daty byla naplněna báze titulů a navrženo a implementováno jednotné rozhraní pro správu KOSABI.

Výsledky projektu byly prezentovány na konferenci Inforum 2002 a Knihovny současnosti 2002.

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**, která umožní jeho přetrvání v budoucnu (metadata, pojmové vyhledávání, propojování informací, vazba na klasický fond, fond elektronických dokumentů a registrace v ČNB, budování digitální knihovny), u báze ANL FULL téměř rutinní provoz v NKČR.
- **Fungující linka zpracování bibliografických záznamů z plných textů.**

- **Propojení bibliografických záznamů s elektronickými zdroji na základě dynamických i statických URL adres, propojení analytických záznamů na zdrojový dokument a naopak, kombinace pojmového vyhledávání s metadaty jako předpoklad pro implementaci kooperačního systému do konceptu metaknihovny, informačních bran a "subject gateways".**
- **Vytvoření předpokladů pro distribuované vyhledávání informací za současného využití tradičního slučování informací do souborné databáze založené na architektuře statických bází.**
- **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.**
- **Byl položen základ pro přehledně strukturovaný portál 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.**
- **Možnost opravy metadat (nové)**
- **Možnost stahování a export plných textů a metadat (nové)**
- **Upgrady a updaty systémů (nové)**
- **Další aktualizace topiků (nové)**
- **Analýza problematiky automatické indexace (nové)**
- **Ověření hypotézy na základě analýzy problematiky automatické indexace: orientace spíše na automatickou extrakci dat, clustering, automatické abstrahování event. klasifikaci a podporu indexátorů a využití dalších možností inteligentního vyhledávání systému TOPIC, než na automatické přiřazování věcných termínů. Zohlednit interakci systému s uživatelem a indexátorem (nové)**
- **Ověření hypotézy, že intelektuální indexace věcná je zatím nezastupitelná**
- **Analýza selekční úplnosti topiků v bázi ANL FULL jako východisko pro ladění topiků (nové)**
- **Posílení vazby v rámci KOSABI v situaci reformy státní správy a samosprávy z hlediska metodického i organizačního - navázání spolupráce s nově vznikajícími krajskými knihovnami**
- **Testování a další vývoj aplikace pro MNG KOSABI, interface pro MNG KOSABI**
- **Naplnění báze titulů (nové)**
- **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 (nové)**
- **Nastínění perspektiv KOSABI a možný vývoj jeho subsystémů z hlediska typu zpracovávaných dokumentů**
- **Zjištění, že výběrová článková bibliografie je potřebná a využívána (anketa)(nové)**
- **Zjištění, že článková bibliografie je zpracovávána v zahraniční řadou knihovnických i komerčních institucí**
- **Zjištění trendu spolupracovat při zpřístupňování informací z novin a dalších seriálů s institucemi(i komerčními), které se touto problematikou zabývají a implementovat nové technologické metody již používané mimo oblast knihovnictví**

(strategický plán IFLA - sekce pro seriály, nově vytvořená sekce pro noviny v r. 2002)
(*nové*)

- **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 v rámci ČR a reference v zahraničí (DCMI, TEL, Gabriel) (*nové*)**

C Návrhovaná část

C.1 Výsledky řešení

- Nový layout www stránek, asp skript (nové)
- 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 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
- Realizace aplikace pro získávání dat TAMTAM SPECIAL NK (TTSNK).
- Realizace aplikace pro automatické generování bibliografických údajů do analytických záznamů a metadat Dublin Core a jejich umístění do plných textů i mimo ně - TamTam Data Extractor (TTDE)
- Aplikace pro indexaci dat do plnotextové databáze ANL FULL
- Aplikace pro vyhledávání a zpřístupnění plných textů formou uživatelských formátů - citací s hypertextovými odkazy i bez nich, citací se souhrnem, v Dublin Core, ale i v pracovních formátech ve formě HTML, XHTML, XML. Rozšíření prvků Dublin Core o prvky Anl Core pro české analytické záznamy pro účely zpracování a vyhledávání
- Definice topiků (aktualizace a nové topiky)
- Definice a automatické generování rejstříků
- Plnotextové a pojmové vyhledávání
- Zkvalitnění věcné indexace ve formě předmětových kategorií, předmětových hesel a klíčových slov
- Statické a dynamické linky na plné texty
- Průběžné doplňování báze ANL bibliografickými záznamy a báze ANL FULL plnými texty (on line)
- Průběžné doplňování databáze ANL FULL, propojování bibliografických záznamů s plnými texty na základě dynamických a statických URL adres.
- Zkvalitnění zpřístupnění periodika Národní knihovna v bázi ANL FULL
- Údržba strukturovaného portálu pro volně přístupné zdroje na WWW v regionální, tematické a institucionální struktuře
- Ověření funkčnosti registrace a autentifikace uživatelů
- Návrh a implementace aplikace pro administraci báze ANL FULL (opravy dat, statistiky) (nové)
- Návrh a implementace aplikace pro stahování a export plných textů a metadat pro vytypované uživatele (nové)
- Testování a vývoj aplikace pro management Kooperačního systému článkové bibliografie (MNG KOSABI), jednotné rozhraní pro aplikaci
- Naplnění Báže titulů na serveru ANL (nové)
- Aplikace pro kontrolu duplicitních záznamů
- Standardizace týkající se popisu článků v UNIMARCu a AACR2R
- Zabezpečení HW a SW podpory kooperačního systému, implementace nových verzí (nové)

- Zabezpečení KOSABI po stránce metodické i organizační
- Návrh akvizice dalších elektronických zdrojů pro bázi TamTam (Anopress) *(nové)*
- Analýza problematiky automatické či poloautomatické indexace věcné a potvrzení hypotézy orientovat se spíše na automatickou extrakci a využití inteligentního vyhledávání v TOPICu, dále pak orientace na skloubení pojmového vyhledávání s intelektuální indexací věcnou a zabudování věcných termínů přímo do topiků, než na automatické přiřazování věcných termínů *(nové)*
- Potvrzení nutnosti intelektuální indexace věcné, která je stále kvalitnější než automatické přiřazování termínů
- Analýza selekční úplnosti topiků v bázi ANL FULL a praktické závěry pro zkvalitnění topiků
- Návrh a vývojové práce na internetové verzi získávání a zpracování bibliografických záznamů z plných textů *(nové)*
- Důkladná analýza zpracování článků v zahraničí (metody, organizace) a ověření, že podobné a srovnatelné metody se vyvíjejí v rámci tohoto projektu *(nové)*
- Důkladný průzkum týkající se systémů zpracování článků v zahraničí (národní knihovny a jiné knihovnické instituce a komerční organizace zpracovávající článkové a plnotextové databáze, národní bibliografie)
- Nastínění perspektiv KOSABI a vznik jeho možných subsystémů z hlediska typu zpracovávaných titulů *(nové)*
- Prezenace výsledků projektu na konferenci Inforum 2002 a Knihovny současnosti 2002 *(nové)*

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 v bázi ANL FULL. 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 je možno transformovat díky finanční podpoře projektů a do budoucna ošetřit tak, abychom články zpracovávaly moderně. O důležitosti informací uveřejněných v člancích nelze pochybovat. Nemyslím tím pouze informace odborné a vědecké povahy, na které je třeba se především zaměřit.

Spolupráce s Anopressem mohla být navázána díky zmíněným projektům a NK mohla navázat na moderní technologii, kterou tato firma užívá a dále rozvíjí. Báze ANL FULL a určité procento báze ANL vzniká díky spolupráci s touto firmou.

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.

Jedním ze základních východisek nového knihovního zákona je, že **veřejné knihovny pracují v celostátně koordinovaném systému**. Koncepce **státní informační politiky ve vzdělávání** (usnesení vlády ze dne 10. 4. 2000, č. 351) stanoví důležitou úlohu veřejných knihoven při zajištění bezbariérového přístupu k informacím všech typů v procesu výchovy a vzdělávání. Vybudování České digitální knihovny je jeden z hlavních úkolů. Základním cílem

státní informační politiky je vybudovat a rozvíjet informační společnost a tím vytvořit předpoklady zejména pro zlepšení kvality života jednotlivých občanů, zefektivnění státní správy a samosprávy a zkvalitnění podpory rozvoje podnikání. Toto je i cílem institucí spolupracujících v KOSABI.

V návrhu nové "Strategie rozvoje knihoven 2003-2005" je formulován hlavní cíl rozvoje knihoven: "**Prostřednictvím národního knihovního systému umožnit občanům rovný přístup k publikovaným dokumentům a informačním zdrojům v jakékoli formě, vytvářet informační zázemí pro výchovu a celoživotní vzdělávání, pro uspokojování kulturních zájmů občanů, pro výzkumnou a vývojovou činnost, pro ekonomické aktivity a pro nezávislé rozhodování jedince.**"

Z dílčích cílů: vytvořit integrovaný národní systém knihoven a informačních institucí, který s využitím informačních technologií umožní jejich propojení, kooperaci a zapojení do mezinárodní spolupráce; zajistit realizaci regionálních funkcí knihoven a formovat krajské systémy knihoven; podporovat oborovou organizaci systému knihoven; dosáhnout maximální kompatibility a unifikace knihovnických činností a systémů v národním a mezinárodním měřítku s cílem zlepšit kvalitu služeb, zvýšit efektivitu činnosti knihoven, odstranit duplicitu činností.

C.3 Návrhy opatření

- **Zajištění financování projektu v roce 2002 a dále zajištění souvisejícího projektu Propojení analytických záznamů s plnými texty - optimalizace zpřístupnění plných textů, který tento projekt saturuje koncepčně a doplňuje finančně. Koordinace s ostatními stávajícími i budoucími projekty v NK i jinde**
- **Posílení vazeb stávajícího KOSABI na existující kooperační systémy na nižší úrovni z hlediska správního, zejména systém LANIUS v budoucnu**
- **Rozvoj spolupráce s novými čtyřmi krajskými knihovnami**
- **Personální zajištění projektu z hlediska počtu pracovníků oddělení analytického zpracování v NK**
- **Existence souborů autorit v oblasti jmenného a věcného popisu jsou nutným předpokladem kvalitního zpracování a vyhledávání informací. Přesunuté kapacity do příslušných oddělení tuto situaci pomohou řešit**
- **Přísnější výběr článků k indexaci deníků s cílem vyloučení subjektivního faktoru při excerpci deníků**
- **Funkčnost propojovacích vazeb v systému ALEPH a Metalib a možnosti expanze a spolehlivosti systému v tomto ohledu. Moderní informační systémy jsou založeny na propojování sekundárních informací s primárními jak klasickými tak elektronickými, ale také na vzájemném propojování sekundárních informací o různých typech dokumentů. Nejde pouze o propojení záznamů s plnými texty, ale také o provázání seriálů a jednotlivých čísel na analytický rozpis článků obsažených v seriálu v rámci báze NKC či Souborného katalogu CASLIN, dále pak připojení článků - recenzí k recenzovaným dokumentům v rámci těchto bází. Navigační systémy na úplné obsahy čísel seriálů mimo záběr abstraktových databází jsou v zahraničí zcela běžné.**

- Ošetření autorsko právních aspektů, event. ekonomických aspektů zpřístupňování plných textů externím a vzdáleným uživatelům nejenom na úrovni projektu
- Zapojení báze ANL FULL do Konsorcia Anopress a v budoucnu i multilicence na používání linky automatického zpracování pro další knihovny (*nové*)
- Pokračování Konzorcia Anopress
- Eventuelní další zpřístupnění elektronických zdrojů v Anopressu (*nové*)
- Vstřícnost vydavatelů, nakladatelů, autorů (*nové*)
- Strukturované údaje v plných textech (*nové*)
- 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 kategorií (*nové*)
- Opravy báze ANL FULL, ANL (*nové*)
- Další zkoumání možností automatické či poloautomatické indexace zejména ve smyslu extrakce údajů (*nové*), generování topiků on line
- Nutnost intelektuální indexace věcné
- Internetová verze linky získávání a zpracování bibliografických záznamů z plných textů (*nové*)
- Dostatečné informace o možnosti podání návazného projektu a možnostech financování z domácích a zahraničních zdrojů (*nové*)
- Jednání s nakladateli/vydavateli o spolupráci týkající se elektronických zdrojů by mělo být vedeno a zaštitěno z vyšší úrovně managementu institucí, které mohou ovlivnit další vývoj zpřístupnění elektronických zdrojů (*nové*)

D Použití finančních prostředků

D.1 Komentář

Použití finančních prostředků je zohledněno v následujících tabulkách. První z nich zachycuje využití neinvestičních prostředků. V rámci neinvestičních prostředků jsou odděleny placené služby, mzdy. Druhá zachycuje použití investic. V tomto roce opět nebyla realizována plánovaná zahraniční stáž z důvodu pracovního vytížení řešitelského týmu a prospěšnosti investovat ušetřené prostředky do jiných typů služeb. Plánované čerpání finančních prostředků do konce roku 2002 je zachyceno kurzívou.

Prostředky z podnikových zdrojů a jiných zdrojů činí podle doplňku ke smlouvě mezi NKČR a MKKČR 320 000 Kč. Z toho 120 000 Kč je plánovaného jako vklad Anopressu. Společnost umožňuje trvalé zpřístupnění 1 licence Tam Tam Professional (cca 15 000 měsíčně) pro další vývoj systému v NKČR. Pracovníci oddělení mají vložit do projektu v tomto roce 200 000 Kč. Úplný vklad do projektu bude vyčíslen v konečném zúčtování projektu po jeho dokončení v tomto roce. Vzhledem k tomu, že projekt je velmi náročný koncepčně i realizačně zároveň, je vklad hlavní řešitelky poměrně velký.

D.2 Tabulky

(vynechány)

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 a metadata jsou uložena ve zdrojových kódech plných textů.

Výsledkem řešení projektu v r. 2002 je poloprovoz aplikace pro získávání a automatickou extrakci/indexaci bibliografických záznamů z plných textů (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.

V r. 2002 byl vytvořen nový design serveru full.nkp.cz, www stránky byly zprovozněny v novém prostředí. Pozornost je průběžně věnována definici báze ANL FULL a definici vyhledávání v bázi (prostý dotaz, formulářový dotaz, topiky, rejstříky). Indexace plnotextové databáze v systému TOPIC umožňuje pojmové vyhledávání informací na základě strukturovaných dotazů - topiků, které byly dále aktualizovány. V r. 2002 je nově zprovozněna aplikace pro administraci databáze ANL FULL (statistiky, opravy, evidence uživatelů). Dále je zprovozněna aplikace pro stahování a export metadat a plných textů (formát text, rdf, html, UNICODE, XML). V lince zpracování

bibliografických záznamů z plných textů (TTDE) bylo zpracováno a následně naimportováno do báze ANL a ANL FULL cca 13 000 metadat (bibliografických záznamů) a plných textů. Průběžně jsou staticky propojovány záznamy s plnými texty z volně dostupných stabilních titulů na Internetu v bázi ANL (cca 2000 propojení). V r. 2002 je udržován portál pro zpřístupnění volně dostupných textů na Internetu ve struktuře tematické a regionální s popisem zdrojů. V r. 2002 byly zahájeny práce na www verzi linky zpracování bibliografických záznamů z plných textů (TTDE). V roce 2002 byla průběžně aktualizována souborná databáze kooperačního systému ANL a vydáván CD-ROM s Českou národní bibliografií.

V r. 2002 je dále laděna a testována aplikace pro správu a údržbu bibliografické báze dat KOSABI včetně Báze pro evidenci titulů. Je implementován jednotný interface pro vstup do MNG KOSABI.

V r. 2002 proběhl update i upgrade SW i HW serveru full.nkp.cz, upgrade SW serveru anl.nkp.cz pro správu a údržbu KOSABI.

Na základě analytických prací týkajících se problematiky automatické indexace je potvrzena hypotéza orientovat se spíše na automatickou extrakci dat s využitím možností inteligentního vyhledávacího systému TOPIC a jeho interaktivních vlastností. Byla provedena analýza selekční úplnosti topiků a navrženy metody jejich zkvalitnění.

Klíčová slova:

Plné texty; TOPIC; topiky; analytická indexace; záznam; seriály; články; zpřístupnění; souborná databáze; 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; Dublin Core; metadata ; HTML; XHTML; XML

E.2 ABSTRACT AND KEY WORDS IN ENGLISH

Abstract:

The contents of this project is optimization of integration and management of heterogenous data which are involved in union 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 and matadata are involved in source documents of full texts.

The 2002 has resulted in pilot system operation of acquisition and workflow of automated exctraxion indexing of bibliographical entries from fulltexts (TTDE) and creating of imported file for bibliographical database ANL and full texts database ANL FULL in NLCR with metadata embadement in full texts including automatically generated DC in HTML, XHTML, XML (qualified, unqualified).

New design of server full.nkp.cz has been realized, new www pages have been implemented. Definition of ANL FULL database and search method have been specified regularly(simple search, advanced search, topics, index). Topics in database ANL FULL in TOPIC system (concept based retrieval) have been updated. In 2002 an application for administration (statistics, correstions, users) has been put into operation as well as application for download and export of full texts with metadata in several formats (text,

rtf, html, UNICODE, XML). In the workflow of automatic extraction indexing have been produced and then imported to ANL and ANL FULL database ca 13 000 bibliographic records/metadata records and full texts. Statics links have been done between bibliographic records and free Internet full texts regularly (ca 2000 links). Portal for free texts on Internet has been maintained (subject and regional structure) with resource description. In 2002 works on www version of workflow (TTDE) has started.

The union database of co-operative system has been regularly updated and published on CD-ROM as Czech National Bibliography.

In 2002 an application for management of COSABI is tested and optimised including Titles' database. Interface for MNG COSABI has been done.

In 2002 update and upgrade of SW and HW of full.nkp.cz server has been realized, SW upgrade of anl.nkp. cz server as well.

Orientation on automatic extraction indexing and data generating with possibilities of concept and intelligent retrieval of TOPIC system has been confirmed and interactive properties of TOPIC system as well. Analysis of recall of topics has been done and suggestions for improvement of their quality as well.

Key words:

Full texts; TOPIC; topics; analytical indexing; entries; serials; articles; access; union database; linking; 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; Dublin Core; metadata; HTML, XHTML, XML

14. listopadu 2002

PhDr. Vojtěch Balík, ředitel NK

PhDr. Ivana Anděrová, hlavní řešitelka