

# SOUBORNÁ DATABÁZE KOOPERAČNÍHO SYSTÉMU ČLÁNKOVÉ BIBLIOGRAFIE - OPTIMALIZACE INTEGRACE A SPRÁVY HETEROGENNÍCH DAT

Souhrnná zpráva za rok 2003

Předkládá: PhDr. Vojtěch Balík, ředitel NK

Zpracovala: PhDr. Ivana Anděrová, hlavní řešitelka



Praha, listopad 2003

## 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, resp. WWW 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 bibliografie (KOSABI)
  - 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

### Navrhovaná část

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

## D

### Použití finančních prostředků

- D.1 Komentář
- D.2 Tabulky

## E

### Resumé a klíčová slova

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

### Přílohy

## F

- F.1 Příloha F 1 : Báze ANL v NKČR (ALEPHU), JIB CASLIN a ANL/ANL FULL, ČNB
- F.2 Příloha F 2 : Linka zpracování bibliografických záznamů z plných

textů

- TamTam Data Extractor (TTDE) - bibliografická metadata  
pro bázi ANL (ALEPH) a plné texty s Dublin Core  
pro bázi ANL FULL (TOPIC)

- F.3** Příloha F 3 : Architektura systému získávání,  
zpracování a zpřístupnění plných textů v systému TOPIC  
(báze ANL FULL a portál Periodika na WWW)  
- detailní ukázky formulářů, formátů, výstupů.  
Administrace (opravy a statistiky)
- F.4** Příloha F 4 : Tituly v bázi ANL FULL, ANL, volně na WWW,  
Dublin Core a ANL Core - tabulka návštěví pro vyhledávání v poli  
dotaz,  
tabulka formulářových polí, topiky, citace článku
- F.5** Příloha F 5 : Krajský bibliografický systém.  
Aplikace pro správu a údržbu KOSABI  
(architektura systému, jednotlivé aplikace pro příjem  
a zpracování dat, báze titulů)
- F.6** Příloha F 6 : Linka zpracování bibliografických záznamů  
z plných textů - internetová verze - WWW TamTam Data Extractor  
(WWW TTDE)  
- bibliografická metadata pro bázi ANL (ALEPH)  
a plné texty s Dublin Core pro bázi ANL FULL (TOPIC)

## A Konstatační část

### A.1 Rešerše

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) : Helsinki University,

1996, last updated 21 February 2000. Dostupný z: <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+ : Computerised Bibliographic Record Actions [online]. Boston Spa (Velká Británie) : COBRA+, 1997. Dostupný z: <http://www.ddb.de/gabriel/cobra>.

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.htm](http://www.lub.lu.se/metadata/DC_creator.htm).

Nordic Countries URN-generator : provided by the Nordic Libraries [online]. Lund (Švédsko) : Lund universitetsbibliotek, 1997. Dostupný z 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.oclc.org/support/documentation/worldcat/cataloging/internetguide/>.

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://www.inforum.cz/infomedia98/pdf/wilson.htm>.

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>

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

BRATKOVÁ, Eva. Metadata jako nový nástroj pro komunikaci webovských informačních zdrojů. *Národní knihovna : knihovnická revue*, 1999, roč. 10, č. 4, s. 178-195. Dostupný též z URL: <http://full.nkp.cz/nkkr/Nkkr9904/9904178.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/nkkr/Nkkr9901/9901006.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: [http://www.nkp.cz/pages/page.php3?page=fond\\_anal\\_unim\\_opr.htm](http://www.nkp.cz/pages/page.php3?page=fond_anal_unim_opr.htm)

Záznam pro soubornou databázi : Výměnný formát. Fyzicky nesamostatné části dokumentů. Tištěné monografie a seriály. Pracovní skupina pro analytické zpracování, Rada pro katalogizační politiku. 1. vyd. Praha : Národní knihovna České republiky, 1999. 39 s. (Standardizace ; č. 20). Určeno k připomínkám. Dostupný z URL: [http://www.nkp.cz/pages/page.php3?page=fond\\_ann\\_vf\\_opr.htm](http://www.nkp.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>. Pozn.: nepřístupný

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

PAPÍK, R. Trendy v rozvoji informačních služeb. *Ikaros* [online]. 1999, č. 8 [cit. 1999-09-01]. Dostupný z URL: <http://www.ikaros.cz/Clanek.asp?ID=200208571>.

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

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://www.ikaros.cz/Clanek.asp?ID=200205087>.

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

Projects at the Royal Library in Stockholm, Sweden [online]. Stockholm : Royal Library, updated July 1, 1999. Dostupný z 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://www.ikaros.cz/Clanek.asp?ID=200208578>.

Topic : systém pro inteligentní vyhledávání dokumentů. Praha : Tovek, 199?.

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

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

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

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

ANDĚROVÁ, Ivana. Programový projekt MK ČR "Souborná databáze Kooperačního systému článkové bibliografie - optimalizace integrace a správy heterogenních dat". *Ikaros* [online]. 2000, č. 10 [cit. 2000-12-01]. Dostupný z URL: <http://www.ikaros.cz/Clanek.asp?ID=200301003>. ISSN 1212-5075.

Biblink [online]. Bath (Anglie) : UKOLN, last updated 12-Jul-2000 [cit. 14. 3.

2001]. Dostupné 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://www.ikaros.cz/Clanek.asp?ID=200208144>.

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>. Pozn.: nepřístupný

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 přivítává vůči každému. *Ikaros* [online], 2000, č. 9. Dostupný z URL: <http://www.ikaros.cz/Clanek.asp?ID=200208199>.

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>.

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://www.nkp.cz/pages/page.php3?page=oazp\\_granty.htm](http://www.nkp.cz/pages/page.php3?page=oazp_granty.htm).

ANDĚROVÁ, I. Souborná databáze Kooperačního systému článkové bibliografie - optimalizace integrace a správy heterogenních dat. Souhrnná zpráva za rok # [online]. Dostupný z URL: [http://www.nkp.cz/pages/page.php3?page=oazp\\_granty.htm](http://www.nkp.cz/pages/page.php3?page=oazp_granty.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/nkkr/NKKR0101/0101026.html>.

PAPÍK, R. Competitive Intelligence, informační služby, Internet a informační profese. *Ikaros* [online]. 2001, č. 04 [cit. 2001-04-01]. Dostupný z:

<http://www.ikaros.cz/Clanek.asp?ID=200208281>. ISSN 1212-5075.

ANDĚROVÁ, Ivana. Metodika popisu článků ve formátu UNIMARC [online]. 2001. Dostupný z URL: [http://www.nkp.cz/pages/page.php3?page=oazp\\_popis1.htm](http://www.nkp.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://www.ikaros.cz/Clanek.asp?ID=200208297>. ISSN 1212-5075.

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

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

H.W. Wilson Company Selects Verity to Power the New WilsonWeb Site - the Premier Reference Resource for Librarians and Researchers [online]. Dostupný z 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://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.zář 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.zář 2002 v Seči u Chrudimi. 2002, s. 190-194.

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.zář 2002 v Seči u Chrudimi. 2002, s. 195-200.

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

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

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

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

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

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

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

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>

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

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

SKLENÁK, V. Sémantický web. Knihovny v elektronické záplavě. Automatizace knihovnických procesů .9. ročník. Liberec. 2003. Dostupný z:

<http://knihovny.cvut.cz/akp2003/>.

ŽABIČKA, P. OAI-PMH: Protokol pro metadatovou interoperabilitu. Automatizace knihovnických procesů . 9. ročník. Liberec. 2003. Dostupný z:

<http://knihovny.cvut.cz/akp2003/> .

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

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

BARTOŠEK, M. Aktuální oblasti výzkumu digitálních knihoven Infos 2003: zborník z 32. medzinárodného informatického sympózia, ktoré se konalo v dňoch 7.-10 apríla 2003 v Starej Lesnej. Sest. Alojz Androvič, Judita Kopáčiková. Bratislava, Centrum VTISR 2003. S. 84-90.

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

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

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

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

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

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

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

SVOBODA, M. PQNext: Nová verze vyhledávacího prostředí služby ProQuest

5000. Ikaros [online]. 2003, č. 08 [cit. 2003-08-01]. Dostupný z: <http://www.ikaros.cz/Clanek.asp?ID=200308004>. ISSN 1212-5075.

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

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

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

BRATKOVÁ, E. Citace odborné literatury jako nástroj rozvoje služeb a integrace digitálních knihoven. Dostupný z: <http://platan.vc.cvut.cz/akp/clanky/12.pdf>.

Van de SOMPEL, H.- HOCHSTENBACH, P. Reference Linking in a Hybrid Library Environment. Part 1-3. D-Lib Magazine [online]. 2000, vol 5, no. 4, no. 10. Dostupný z: <http://www.dlib.org>.

BALÍKOVÁ, M.: Soubor věcných autorit. Předmětová kategorizace pro potřeby konspektu. Knihovny současnosti 2003 11. konference konaná ve dnech 16.-18.září 2003 v Seči u Chrudimi. Dostupný z: <http://jib-info.cuni.cz/dokumenty/sec2003/1>.

PAVLÍK, J. Linkování na relevantní přidané služby v rámci Jednotné informační brány Knihovny současnosti 2003. 11. konference konaná ve dnech 16.-18.září 2003 v Seči u Chrudimi. Dostupný z: [http://jib-info.cuni.cz/dokumenty/sec2003/sec2003\\_sfx.doc](http://jib-info.cuni.cz/dokumenty/sec2003/sec2003_sfx.doc).

SKLENÁK, V. Vyhledávací stroje v prostředí Internetu - a co bude dál? Dostupný z: [http://platan.vc.cvut.cz/akp2003/sbornik/03\\_sklenak.pdf](http://platan.vc.cvut.cz/akp2003/sbornik/03_sklenak.pdf).

ANDĚROVÁ, I. Aktuální informace o Kooperačním systému článkové bibliografie SDRUK , 13. zasedání Sekce pro bibliografii. Hradec Králové, 5.-6.5. 2003. Dostupný z: <http://full.nkp.cz>, Rubrika Co je nového ... .

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

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

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

Dublin Core Metadata Initiative Progress Report and Workplan for 2002. Dostupný z: <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://www.nkp.cz/pages/page.php3?page=oazp\\_odd\\_anal\\_zprac.htm](http://www.nkp.cz/pages/page.php3?page=oazp_odd_anal_zprac.htm).

Metodika popisu článků ve formátu UNIMARC [online].

Dostupný z: [http://www.nkp.cz/pages/page.php3?page=oazp\\_popis1.htm](http://www.nkp.cz/pages/page.php3?page=oazp_popis1.htm).

Server FULL.NKP.CZ .

Dostupný z: <http://full.nkp.cz>

Báze ANL FULL.

Dostupný z: <http://full.nkp.cz>

Management Kooperačního systému článkové bibliografie - MNG KOSABI.

Dostupný z: <http://full.nkp.cz>

Plné texty v českých novinách a časopisech - přehled.

Dostupný z: <http://full.nkp.cz>.

Výsledky práce společnosti ANOPRESS IT.

Dostupný z: <http://www.anopress.cz>

Seznam seriálů excerpovaných v oddělení analytického zpracování.

Dostupný z [http://www.nkp.cz/pages/page.php3?page=oazp\\_Seznam\\_OAZ.htm](http://www.nkp.cz/pages/page.php3?page=oazp_Seznam_OAZ.htm) , <http://full.nkp.cz>

Knihovny spolupracující v kooperačním systému článkové bibliografie a excerpční základny.

Dostupný z [http://www.nkp.cz/pages/page.php3?page=oazp\\_kooper\\_svk.htm](http://www.nkp.cz/pages/page.php3?page=oazp_kooper_svk.htm), <http://full.nkp.cz>.

Báze ANL [online].

Dostupný z: [http://sigma.nkp.cz/F/?func=file&file\\_name=base-list](http://sigma.nkp.cz/F/?func=file&file_name=base-list)

Národní knihovna. Knihovnická revue [online].

Dostupný z: <http://full.nkp.cz/nkkr/NKKR.html>.

JIB Caslin.

Dostupný z: <http://www.jib.cz>.

## **Zahraníční reference**

General information about ANL - Articles in Czech newspapers, magazines and collections of works. Login procedures to ANL [FULL] - Articles published in Czech journals and newspapers (full texts). *Gabriel*.

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

ANL FULL. *Dublin Core Metadata Initiative*.

Dostupný z: <http://dublincore.org/projects/europe.shtml#denmark>.

Tel Digital deposits state of the art review. Marco de Niet, Koninklijke 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: [www.europeanlibrary.org/pdf/tel\\_results\\_d11\\_v02.pdf](http://www.europeanlibrary.org/pdf/tel_results_d11_v02.pdf).

IFLA-Directory of Serials Content Databases and Current-Awareness Services for Serials Content. [V přípravě]. *IFLANET*.

Dostupný z: <http://www.ifla.org/l/whatsnew/new2002.htm>

## **A.2 Současný stav ve světě a v ČR**

### **A.2.1 Obecně**

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

Elektronické dokumenty jsou zpřístupňovány prostřednictvím nakladatelství, distributorských firem, informačních institucí či služeb a jejich produktů, dále pak prostřednictvím digitálních knihoven a služeb vznikajících na základě projektů, konzorcií a licencí. **Při zpřístupňování elektronických informací se stále více prohlubuje spolupráce mezi státním a soukromým sektorem.**

**Předpokladem plnohodnotného zpřístupnění plných textů je implementace metadat do plných textů. Tato metadata (DC, Marc) mohou** může být vytvářena autorem, vydavatelem, distributorem, knihovníkem a zpřístupňována na webu pomocí XML/RDF s definovanou standardní strukturou DTD. Existují iniciativy, které se zabývají konverzí mezi DC, MARC a XML. DC je určen primárně pro otevřený web, lze ho užít i pro databáze tzv. hlubokého webu. Významný je OAI-PHP (v. 2) protokol - protokol pro metadatovou interoperabilitu, umožňující automatizované získávání metadat a vzájemnou komunikaci archívů, digitálních knihoven, je založený na DC a XML. V současné době se velká pozornost věnuje protokolům pro komunikaci a sdílení dat - **Z39.50 a Bath Profile a tzv. otevřenému nebo dynamickému propojování pomocí tzv. OpenURL.**

Jednoznačná indentifikace plných textů je jednou z podmínek zpřístupnění plných textů. K identifikaci služeb, zdrojů a objektů na internetu slouží nestabilní **URL (Uniform Resource Locator), PURL (Persistent URL), Uniform Resource Name URN (Uniform Resource Name), DOI**

**(Document Object Identifier), SICI (Seriál Item and Contribution Identifier) aj.**

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

Předkládaná zpráva je poměrně podrobná, obsahuje výsledky řešení předchozích let, na které navazuje a bez kterých by se ztratila kontinuita řešení projektu.

## **A.2.2 Současný stav v zahraničí**

### **Obecně**

**Některé systémy a služby zpřístupňující informace o člancích v zahraničí jsou důkladně popsány ve zprávě podané v roce 2002.**

Lze shrnout, že kooperační systémy zabývající se zpřístupňováním článkových informací a článkové bibliografie se vyvíjely a existují zejména v zemích bývalého východního bloku. Články jsou zpřístupňovány zejména v severovýchodních zemích (Švédsko-Libris, Dánsko - Basis, Norsko-Bibsys, Finsko - Arto). Velké článkové databáze se budují v Holandsku, Španělsku, Německu, Gruzii, Litvě, Rusku, Slovinsku, Makedonii aj.). Poměrně kompletní přístup k elektronickým časopisům nabízí OCLC First Search Electronic Collection Online. Kvalitně zpřístupňují plné texty Ingenta, ProQuest, EBSCO, Wilson Web aj. K dispozici jsou časopisecké zdroje na nakladatelských serverech, tituly vědeckých časopisů s volným přístupem k obsahům a abstraktům, někde i k plným textům, denní tisk a časopisy populární, popularizační i odborné včetně jednotlivých článků. Plné texty jsou zpřístupňovány jak soukromými společnostmi, tak knihovnami v rámci portálů, digitálních knihoven, archívů, multioborových i oborových databází. Jde o databáze konkrétních vydavatelů, agregátorů nebo servery konkrétních titulů.

Uživatelé v České republice mají dnes přístup k zahraničním informacím o člancích různého typu z hlediska formy i obsahu díky programu MŠMT "Informační zdroje pro výzkum a vývoj" (LI), který byl vyhlášen v září 1999 na čtyřleté období (2000-2003). Cílem bylo zajistit systematickou podporu financování oborových a

polytematických informačních zdrojů, které si většinou jednotlivé instituce nemohou dovolit. Přehled možné najít na www stránkách [MŠMT Portálu STM](#) (projekt LI01018) a v příslušném [dokumentu](#) STK z r. 2002.

Národní knihovna ČR se koncem roku 2002 zapojila do projektu Univerzitní knihovny v Regensburgu. Elektronische Zeitschriftenbibliothek (EZB). Tituly jsou rozděleny podle toho, zda jsou volně přístupné, licencované pro určitou knihovnu a nepřístupné pro danou knihovnu.

Nová online česká služba [Infozdroje.cz](#) obsahuje údaje o projektech zahrnujících nákup elektronických informačních zdrojů realizovaných v letech 2000-2003 v rámci grantového [programu LI "Informační zdroje pro výzkum a vývoj"](#). [Infozdroje.cz](#) tak nyní obsahují informace o všech produktech a službách dostupných díky financování v rámci tohoto programu, a je zde možno zjistit, které instituce k nim mají přístup.

## **Trendy, nástroje, metody a projekty integrace získávání, zpracování a zpřístupnění heterogenních informačních zdrojů**

### **1. Trendy**

**Základní trend:** kooperace v rámci interoperability systémů na základě spolupracujících komponent v celém procesu získávání, zpracování a zpřístupňování informací, tj. přechod od explicitní pevně svázané kooperace ke kooperaci nezávislé na použitých SW a HW za účelem zajištění pružné integrace dat a jejich zpřístupnění v rámci jednotného interface za využití moderních vyhledávacích metod umožňující interakci uživatele se systémem.

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

**2. Některé nástroje, metody, projekty odpovídající současným trendům** (pro definice některých pojmů jsem použila databázi KTD - Česká terminologická databáze knihovnictví a informační vědy -- - TDKIV)

### **Pravidla popisu - jejich zjednodušení a zefektivnění, formáty**

#### **Jmenný popis**

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

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



## **Hybridní knihovna, vznik digitálních knihoven, interoperabilita, identifikátory, propojování, legislativa, služby**

### **Hybridní knihovna**

Knihovna integrující klasickou knihovnu představovanou především tištěnými dokumenty a digitální knihovnu, obvykle s cílem zkvalitnění služeb uživatelům. Většina knihoven má tuto povahu. Pro současnost je charakteristický přechod od klasické k digitální knihovně.

**Digitální knihovna** je „integrováný systém zahrnující soubor elektronických informačních zdrojů a služeb umožňující získávání, zpracovávání, vyhledávání a využívání informací v tomto systému uložených. Digitální knihovny jsou zpřístupňovány prostřednictvím počítačových sítí. Účelem budování digitální knihovny je poskytnout uživatelům možnost jednotného přístupu k digitálním anebo digitalizovaným dokumentům, případně i k sekundárním informacím o tištěných primárních zdrojích, uložených ve fondu knihovny i mimo fond knihovny.

### **Nejdůležitější komponenty digitální knihovny: metadata, interoperabilita, globální vyhledávání zdrojů.**

Metadata jsou "strukturovaná data, která nesou informace o primárních datech. Pojem metadat je používán především v souvislosti s elektronickými zdroji a vztahuje se k datům v nejširším smyslu slova (datové soubory, textové informace, obrazové informace, hudba aj.). Funkce metadat je popisná, selekční a archiváční. V souvislosti s těmito funkcemi se rozlišují metadata pro účely popisu, správy, právních nároků, technické funkčnosti, užití a archivace. Údaje se obvykle vkládají přímo do zdroje (umísťují se např. v záhlaví dokumentu HTML)".

**Dublin Core (DC)** je "standardizované metadatové schéma pro popis informačních zdrojů, zejména elektronických a internetových. Dublin Core je tvořen souborem patnácti základních prvků (jednoduchý Dublin Core), které lze specifikovat kvalifikátorem (kvalifikovaný Dublin Core). Dublin Core nepředepisuje žádnou závaznou syntaxi (jedná se o sémantický standard), je základem řady dalších metadatových standardů". Vyjadřuje sémantiku dokumentu.

### **Z DC vychází např. metadatový standard OAI (Open Archives Initiative) primárně vyvinutý pro vědecké a akademické komunity**

Projekty, které využívají DC, je možno najít na adrese <http://dublincore.org/projects>.

### **XML (eXtensible Markup Language)**

Jazyk XML je, podobně jako jazyk HTML, prostředek sloužící k zapsání **strukturovaného** textu, zvláště pak textu určeného k šíření v prostoru www. XML odděluje popis struktury dat od jejich prezentace (pomocí tzv. style sheetů). To umožňuje snadnou konverzi do jiných formátů, možnost prezentace dat různými způsoby (HTML, postscript, UNIMARC apod. textový formát). Každý dokument má definovanou svoji strukturu prostřednictvím tzv. DTD (Document Type Definition). Velký potenciál XML se skrývá v novém způsobu odkazování (oběma směry, na více dokumentů najednou či dokonce v rámci hierarchické struktury) pomocí speciálních jazyků XLink, XPointer a XPath. Totéž lze říci o stylovém jazyku XSL, který doplňuje a nahrazuje tzv. kaskádové styly (CSS).

### **RDF (Resource Description Framework)**

„Obecný rámec pro popis jakéhokoli elektronického zdroje, resp. webové stránky a jejího obsahu, tedy pro vyjádření sémantiky a pro podporu sémantického webu. Popisná metadata mohou zahrnovat údaje o autorovi zdroje, datu vytvoření nebo aktualizace, organizaci stránek (sitemap), klíčová slova, předmětové kategorie aj. Jazyk **RDF** poskytuje robustní flexibilní architekturu pro zpracování metadat na internetu; umožňuje komukoli

definovat a používat metadatové schéma, které slouží nejlépe jeho potřebám, a současně umožňuje interoperabilní výměnu metadat. **RDF** je aplikací formátu XML a je vyvíjen konsorciem W3C (World Wide Web).“

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

V současnosti aktivity v oblasti mapování formátu DC do MARC a opačně a převod do XML.

**Interoperabilita** je schopnost dynamické spolupráce mezi technicky různorodými a nezávislými komponentami z hlediska syntaktického, strukturálního a sémantického. Interoperabilitu kromě jiných umožňují i zde jmenované nástroje. Z hlediska použitých prostředků lze stanovit různé úrovně interoperability (Z39.50 a OpenURL, XML aj.)

Významné jsou aktivity v oblasti ontologií (systém konceptů a vztahů mezi nimi). Ontologie - metoda získávání znalostí. Dílčí ontologie souvisí s konceptualizací jednotlivých oblastí. Na základě interoperability mohou vznikat snadněji elektronické archívy, souborné katalogy virtuální i reálné, brány a portály.

### **Elektronický archív**

„Organizovaná sbírka digitálních dokumentů shromážděná za účelem jejich dlouhodobého uchování. Může se jednat o digitalizované dokumenty, tj. tištěné druhy dokumentů převedených do digitální podoby, nebo o dokumenty vytvořené již jako digitální.“

### **Reálný souborný katalog**

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

### **Virtuální souborný katalog**

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

### **Informační brána**

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

**Informační brány** řeší přístup k různým zdrojům z jednotného prostředí. Nejpoužívanějším standardem pro tvorbu metadat je [Dublin Core \(DC\)](#). Jejich fungování závisí na existenci pokud možno homogenního nástroje pro věcnou indexaci harmonizací řízených slovníků a tezurů a hledají se cesty k řešení vícejazyčnosti. Jednou z řešených metod pro zpřístupnění elektronických informačních zdrojů je metoda konspektu, předmětová kategorizace pro popis informačních zdrojů.

Spočívá hierarchickém uspořádání předmětových kategorií, na nejvyšší úrovni není propojena se systematickou klasifikací. Cílem je jednotný tematicky strukturovaný popis heterogenních informačních zdrojů pro potřeby koordinovaného budování knihovních sbírek a knihovních fondů a pro tvorbu nástroje určeného ke zpřístupnění kvalitních (zhodnocených) heterogenních informačních zdrojů v síťovém prostředí, tj. tematických bran. Slouží k tomu údaj o předmětové kategorii spolu s vybraným znakem MDT. Pro mezinárodní srozumitelnost je potřebná konkordance MDT a DDC.

### **Portál**



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

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

## **Propojování informací**

### **Druhy propojování (aplikované v komerčních a nekomerčních aktivitách)**

Propojovat je možno uzavřeně (closed linking), otevřeně (open linking) staticky (static linking), dynamicky (dynamic linking).

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

### **OpenURL a SFX**

SFX (Special Effects) je technologie založená na OpenURL. Je založena na oddělení popisu zdroje od služeb, které jsou poskytovány. Mezi popisem zdroje a službami stojí tzv. servisní služba (linking service) která propojuje metadata obsazená ve výchozím zdroji, které našel uživatel s vhodným cílovým zdrojem pro uživatele na základě jeho registrace u této servisní služby. Zaručuje tzv. propojení citlivé na kontext (context sensitive linking). Pro připojení informačního či knihovního systému k SFX je nutné, aby váš systém podporoval OpenURL.

### **Některé aplikace OpenURL**

**Ovid** - služba **Ovid OpenLinks** („universal link resolver“). Bibliografické záznamy přístupné v databázích od firmy **H.W.Wilson** jsou propojeny pomocí funkce WilsonLink (OpenURL) s plnými texty článků v elektronické knihovně JSTOR, která obsahuje více než 320 časopisů zaměřených na humanitní a společenské obory. OpenURL podporuje i nová verze vyhledávacího prostředí **ProQuest**.

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

### **Konzorcia a licence**

Pro přístup elektronickým zdrojům zejména k plným textům článků se zakládají konzorcia a uzavírají licenční smlouvy (národní, plošné, individuální pro instituci) s agregátorem event. přímo s nakladatelem. Výhody: získání zdrojů jinak nedostupných, přístupů malým institucím/pracovištím, příznivější ceny, využití centrálních finančních zdrojů, levnější provoz (administrativní a organizačně-technické výhody), spolupráce.

### **Cenová politika**

Poplatky, které se váží na proces vyhledávání, jsou účtovány databázovým centrem na měsíční/roční bázi nebo časově jinak. Částky mohou být fixní nebo variabilní. Nejčastější způsoby úhrady:

- předplatné (*flat-fee, fixed-fee, all-you-can-eat*, subscription) - neomezené využívání databáze nebo skupiny databází za fixní periodické platby. Předplatné není obvyklé u koncových uživatelů, ale spíše u korporativních uživatelů

- platby jen za uskutečněné operace (*pay-as-you-go*) .

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

Problematika získávání, zpracování, zpřístupňování a využívání oblasti). Hlavním cílem těchto technologií je zdokonalení **znalostního managementu** v institucích, firmách V důsledku je efektivní vazba tzv. knowledge managementu na rozhodovací a plánovací procesy.

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

Tato témata jsou předmětem pravidelně se konajících konferencí [Znalosti](#) a [DataKon](#) s mezinárodní účastí (dobývání znalostí z databází, získávání znalostí od experta, vyhledávání, indexace a extrakce informací z textu, aplikovaná lingvistika, implementace znalostních a expertních systémů, znalostní management).

### **Poloautomatická indexace textu (J. Schwarz, 2002)**

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

**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.**

**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 .**

### **Vyhledávání**

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

### **Metody zefektivnění vyhledávacích algoritmů**

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

### **Interaktivní pojetí vyhledávání**

Uživatel systému prohlíží seznam vyhledaných záznamů (často i s plnými texty dokumentů) a jednoduše označuje relevantní záznamy. Z relevantních záznamů pak systém automaticky vybere podstatné výrazy, jimiž upraví původní dotaz. Na základě upraveného dotazu systém vyhledá více relevantních záznamů. "Dotaz příkladem" (anglicky query-by-example, more like this, find similar či similar pages) fungují na stejném principu. Nový dotaz je však vytvořen pouze na základě jediného vybraného dokumentu.

Koncový uživatel nemusí znát vyhledávací metody a strategie, daný vyhledávací systém, uspořádání sbírky či strukturu záznamů. Aplikace interaktivních technik (re)formulace dotazu tak představuje poměrně jednoduchý způsob, jak nezkušeným uživatelům usnadnit vyhledávání.

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

### **Vyhledávací systémy 3. generace**

Jedním z nejdokonalejších systémů pro vyhledávání plnotextových dokumentů je systém TOPIC americké firmy Verity. Systémy třetí generace jsou založeny na principech:

1. rozkladu pojmu na podpojmy
2. vážení jednotlivých podpojmů (větví pojmového stromu)
3. neostrého vyhodnocování dotazů

Dotaz v systému 3. generace reprezentuje pojem, respektive ideu vyhledávaného tématu. Jádrem dotazu je stromová hierarchická struktura, která rozkládá hledané téma na podtémata a přiřazuje jednotlivým částem váhy, které vyjadřují do jaké míry příslušné podtéma přispívá k celkovému určení tématu. Systém je pak schopen vypočítat míru relevance (nejčastěji udávanou v % nebo hodnotou v intervalu 0,1), podle které řadí vyhledané dokumenty.

Systém TOPIC eliminuje jeden z nedostatků booleovských vyhledávacích systémů, kterým je přílišná ostrost operátoru AND, jenž nevyhledá dokument, pokud neobsahuje všechna slova tímto operátorem spojená, zavedením o perátoru ACCRUE.

Klíčovým aspektem úspěšnosti je vlastní vyvážená definice topiků. Je zřejmé, že je to práce pro specialistu, srovnatelná s tvorbou expertních systémů, neboť dobře nadefinovaná báze topiků představuje vlastně bázi znalostí.

### **4. Některé významné databáze, které jsou inspirací pro projekt**

#### **JSTOR - Journal Storage**

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

#### **IBZ - Internationale Bibliographie der Zeitschriftenliteratur**

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

## ProQuest 5000 International

Národní akademická licence pro produkty [ProQuest 5000 International](#) v rámci grantu financovaného MŠMT umožňuje přístup k online službám ProQuest 5000 a PCI Web. Služba obsahuje bibliografické záznamy článků cca 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](#)

Zpřístupnění plnotextových databází odborných zahraničních periodik na základě programu Open Society Institute EIFL-Direct - **plošná multilicence**. Plné texty celkem cca 3300 časopisů od r. 1990 a další inf. zdroje (abstrakty, zpravodajství, příručky) především z oblasti sociálních a humanitních věd od **EBSCO Publishing, jednoho z předních světových dodavatelů el. a tištěných časopisů**, 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 miliónů článků z více jak 100 novin v angličtině)
- MasterFILE Premier (obecně 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 cca 12 dalších databází: ArticleFirst (bibliografické citace článků z cca 13.000 periodik), ContentFirst (seznamy obsahů periodik), NetFirst (bibliografické údaje o odborných zdrojích na Internetu, včetně abstraktů a klasifikace), PapersFirst (referáty ze světových konferencí, kongresů, symposií, výstav a workshopů od 1983), ProceedingsFirst (seznamy obsahů sborníků z vědeckých konferencí), UnionLists (souborný katalog periodik s uvedenými lokacemi), WilsonSelect (plné texty článků z 800 periodik), WorldAlmanac MEDLINE (medicínská informace), ERIC (bibliografie literatury z oblasti vzdělávání) a další.

## [GILS](#)

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

V rámci systému GILS existuje trojí způsob tvorby záznamů: **klasický způsob (ruční), konverze (převod záznamů z jiných bází 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).**

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

V rámci projektu [Nordic Metadata](#) byl vytvořen [konvertor dat](#) který je schopen generovat záznamy ve formátech MARC severských zemí a USMARC ze zdrojových údajů DC.

Služba OCLC [CORC Cooperative Resource Catalog](#) poskytuje nástroj pro automatickou katalogizaci elektronických zdrojů přímo na webu (vyhledávání, vytváření a editace záznamů) ve formátech MARC a DC. Na jeho testování se podílelo více než 450 knihoven z celého světa. Databáze vznikla ze záznamů původně uložených v bázích OCLC InterCat a NetFirst.

**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 vybudováním vazeb mezi národními bibliografickými agenturami a vydavateli elektronických zdrojů s cílem společně vytvořit informace o těchto dokumentech využitelné v obou oblastech.

#### **Příklady systémů automatické indexace**

**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**

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).

#### **5. Mezinárodní akce a související problematika s projektem**

**IFLA a informace z některých příspěvků na 68th IFLA Council and General Conference, August 18-24, 2002. Glasgow**

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

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

## Austrálie

**Australské zkušenosti prezentované na konferenci: [National Bibliographic Database](#)**, de facto australská národní bibliografie, přistupuje k novému modelu bibliografické služby, která **kombinuje tradičně produkovaná bibliografická data s dodáváním metadat od autora či vydavatele**. Koncept předpokládá získávat data z těchto zdrojů: National Bibliographic Database, National Discovery Service, národního repozitáře metadat. **National Bibliographic Database obsahuje: katalogizovaná data, katalogizační záznamy vzdálených zdrojů archivované v National Library, digitalizované dizertace, katalogizační záznamy od prodejců („vendors“, kteří mohou stát mezi producentem a kupujícím) elektronických zdrojů a služeb. The Rource Discovery Service obsahuje: metadata z elektronických archivů, metadata ze „subjekt gateways“, metadata elektronických služeb - výchova a vzdělání, kultura. Uživatel se tak bude moci vybrat tištěnou i elektronickou formu dokumentů.**

**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 peri odika 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.

## 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.**

Problematika Tel na [Elag 2003](#)

Řeší se problematika distribuovaného vyhledávání v různých zdrojích. Snaha najít vazbu mezi skrytým a otevřeným webem. V pro otevřený web je vhodný pro popis zdrojů DC, OPACy lze také převést na DC (možno



doplnit). Lze tak dostat sourodý výsledek vyhledávání a vyřešit problém „dvou“ webů. Jednotlivé sbírky by měly být popsány na této úrovni, takže vyhledávání v OPACích by probíhalo nejprve na této úrovni.

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

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

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

K těmto informacím připojuji informaci o zpřístupňování novin, která byla předmětem semináře [Newspapers for Libraries. Newspapers and the press in Central and Eastern Europe: access and preservation](#) (Berlin-Brandenburgische Akademie der Wissenschaften, 9.-10. 8. 2003) , kterého jsem se zúčastnila (cesta hrazena ze souvisejícího projektu Propojení analytických záznamů ..). V řadě zemí střední a východní Evropy existují specializovaná pracoviště zabývající se uchováváním a zpřístupněním novin (tj. deníků, týdeníků a čtrnáctideníků) v návaznosti na projekty týkající se digitalizace a zpřístupnění těchto materiálů v celé šíři problematiky (technika pro převod tištěných dokumentů na různá média včetně automatické strukturace textu a automatického zpracování metadat). Na mnoha serverech je možno najít portály, které zpřístupňují tyto dokumenty často podle regionálního hlediska ze všech konců světa. V současné době se věnuje velká pozornost samizdatové a exilové literatuře, jejímu shromažďování, uchovávání a zpřístupnění. Stále potřebnější je kooperace na národní i mezinárodní úrovni. Noviny jsou zpřístupňovány v samostatných sbírkách nebo spolu se seriály, v rámci portálů a virtuálních knihoven, jsou budovány jejich rozsáhlé archívy. V některých zemích existuje stanovená strategie budování sbírek těchto informačních zdrojů.

Informace v novinách mapují ekonomický, sociální, kulturní a politický vývoj společnosti z různých hledisek a jsou někdy podceňovány. Jejich využití je možné i pro vědeckovýzkumné účely v těchto oblastech. O nutnosti zabývat se problematikou novin svědčí i fakt, že byla v r. 2002 založena v rámci IFLY Sekce pro noviny - [Newspapers Section](#). Sekce se zabývá strategií, popisem, digitalizací, novými technologiemi pro zpracování a zpřístupnění novin. Na stránkách IFLY je vystavena doporučení pro popis novin [International Guidelines for the Cataloguing of Newspapers](#).

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

Na internetu existuje množství **portálů a serverů** zpřístupňujících noviny často doplněné aktuálním zpravodajstvím. Společnost [Worldpress.com](#) nabízí 1117 deníků publikovaných v 192 zemích. [Newspapers from around the world](#) poskytuje linky na tituly uspořádané regionálně v rámci USA, v různých oblastech světa, Kanadě. V [Actualidad.com](#) jsou noviny vyhledatelné podle kontinentů. [Online newspapers.com](#) nabízí tisíce světových novin vyhledatelných podle regionu. News and Newspapers on-line je služba na [University of North Carolina in Greenboro](#) (vstup přes individuální titul, region, zemi).

**Služby ve východní Evropě.** V [ABYZ News Links](#) je obsažen výběr titulů z evropských zemí aj. oblastí světa. [NewsDirectory.com](#) obsahuje evropské zdroje, [swnewsherald.com](#) zahrnuje zdroje východní Evropy a pobaltských zemí, [megamallandmall.com](#) zahrnuje střední a východní Evropu. [Inkpot Newspapers Link](#) obsahuje noviny 17 východoevropských zemí a Ruska aj. regionů.

[Integrum](#) je nejrozsáhlejší databázová služba poskytující služby on-line v Rusku - obsahuje 4000 databází (přes 140 000 000 dokumentů, 15 000 nových dokumentů denně, archívy národních a regionálních novin, časopisů, TV a rozhlasových pořadů, zprávy a archívy hlavních národních a mezinárodních informačních agentur, plné texty ruských klasiků, dokumenty audiovizuální) - plnotextové vyhledávání, media monitoring service, placená služba, automatické překlady.

Zajímavé byly příspěvky z **ruských knihoven**. Konstantin M. Suchorukov (Head of National Bibliography Department in the Russian Book Chamber) and A. Dzingo (Deputy Director of the Russian Book Chamber): Work with newspapers at the Russian book Chamber: results, problems and prospects - příspěvek podrobně analyzující situaci ve vydávání novin v Rusku, problematiku povinného výtisku a **zpracování článkové bibliografie** - Letopis' gazetnych statej - s týdenní periodicitou (cca 50 titulů novin). Dále vychází Letopis' žurnal'nych statej a Letopis' recenzij. Bibliografie vydává Rossiskaja knižnaja palata . Elektronické bibliografické báze obsahují novinové články od r. 1988, časopisecké články od r. 1991.

Některá z další vystoupení se týkala zpřístupnění **sbírek alternativní a samizdatové literatury** a dalších novinových sbírek. Claus Gravenhorst (Chief Product Management CCS Hamburg): Automated retroconversion of newspapers into fully tagged XML. Tento zajímavý příspěvek se týkal problematiky mikrofilmování, digitalizace včetně metod strukturace textu , extrakce a generování metadat na základě metody vyvinuté [CCS - Content Conversion Specialist](#) . Analyzoval technologii umožňující vytváření a archivování strukturovaných dat během procesu retrokonverze. Tato technologie je použita v projektu [METAe](#) .

Z dalších vystoupení na kongresu, která se týkala problematiky novin. Denise Rosemary Nicholson (Copyright Services Librarian, University of the Witwatersrand, Johannesburg, South Africa): [What has copyright got to do with newspapers? A South African Perspective](#). Charles Oppenheim (Loughborough University, UK): [Newspaper copyright developments: a EU and UK perspective](#).

V obou vystoupeních byly konstatovány malé pokroky týkající se oblasti zpřístupňování novin s ohledem na copyright, platné zákonné normy komplikují zpřístupňování článků z novin, které de facto podléhají dvojnásobné kontrole z hlediska autorských práv: autorská práva vydavatele novin a autorská práva jednotlivých autorů článků. Potěšitelná je aktivita ve Velké Británii směrem k elektronickému povinnému výtisku.

Sekci pro noviny jsem poskytla informace o projektech týkajících se zpracování článků a zpřístupnění článků v rámci oddělení analytického zpracování NKČR, Kooperativního systému článkové bibliografie, báze ANL a ANL FULL V této sekci mně bylo nabídnuto členství.

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

### A.2.3 Současný stav v ČR

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

O některých zdrojích se zmíním podrobněji.

Plné texty jsou organizovány na Internetu do tematicky, regionálně nebo jinak utříděných služeb. Jsou přístupné přes vyhledavače jako samostatné tituly nebo jsou utříděny do kategorií, zpřístupněny na různých portálech a zpravodajských serverech.

Některé knihovny vytvářejí soupisy volných českých zdrojů plných textů na Internetu.

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



Na českém 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 Vydavatelství [Economia](#) nakladatelství ekonomické literatury vystavuje na Internetu plné texty produkce tohoto nakladatelství. Server IHNED nabízí pokročilé vyhledávání ve zdrojích i řazení výsledku podle relevance. Pro předplatitele periodik je možný zdarma přístup do archivů. Čtenáři mají také možnost zaplatit si přístup k online archivům tištěných periodik vydavatelství ECONOMIA nebo si mohou zakoupit prostřednictvím tzv. iKreditu pouze jednotlivé články .

[Sagit](#) (elektronické nakladatelství - právní texty), [Tigis](#) ( časopisy pro lékařskou odbornou veřejnost a edukační časopisy pro veřejnost), [Nakladatelství Muzikus](#) - hudební tematika, [Nakladatelství Portál](#) (pedagogika, psychologie).

### **Current Contents, abstrakty, citace, plné texty**

[Akademie věd ČR](#) - 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 webovská prezentace periodika v rámci Projektu propojení analytických záznamů s plnými texty ve formátu html a pdf - r. 1999 - 2003 s tematickými a autorskými rejstříky, do r. 2001 plnotextově přístupná na serveru [full.nkp.cz](#) (báze ANL FULL) a [www.anopress.cz](#). V dalších letech [Ikaros](#) - elektronický časopis o informační společnosti. [Bulletin SKIP](#), [U nás](#) aj.

### **Některé digitální knihovny a databáze**

[Právní informační systém 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). Pro projekt Memoriae mundi series Bohemica bylo vytvořeno DTD na základě struktury Master. V listopadu 2003 je spuštěna aplikace [Manuscriptorium](#) , Aplikace Kramerius má být uvedena do provozu v koncem r. 2003. Sám digitální archiv nebude po internetu přístupný.

V NKČR se dále používá **RetrievalWare Excalibur k indexaci** (umí překonat chybovost technologie OCR)

Excalibur je intuitivní vyhledávací systém společnosti Excalibur Technologies založený na technologii APRP a sémantické síti (slovní výrazy, spojení idiomy) Convera RetrievalWare.

Adaptive Pattern Recognition Processing je technologie založena na indexování na bipolární úrovni, jazykově nezávislá, má toleranci chybovosti dat na vstupu. Pomocí tohoto systému se indexují www stránky NKČR , krajské knihovny v Olomouci.

Tento systém používá i [AVČR](#). Katalog elektronických zdrojů (KEZ) obahuje zdroje vznikající či využívané v rámci AV ČR (automatické generování deskriptorů i manuální). Je určen k dispozici on-line vědecké obci i široké odborné veřejnosti a to nejen pro vyhledávání, ale i pro pořizování záznamů. KEZ využívá systém

vyvinutý firmou Incad. Akademie věd ČR zpřístupňuje publikační činnost pracovníků v [Ústavní bibliografii ASEP](#) do r. 1993 a [Ústavní bibliografii ASEP](#) od r. 1993.

V r. 2002 byla zpřístupněna digitální knihovna rukopisů, prvotisků a starých tisků v krajské knihovně 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. V tomto roce byl zpřístupněn projekt [Infozdroje.cz](#).

Cíle projektu: zřehlednit přístup k informačním zdrojům z jednotlivých institucí, usnadnit zveřejňování informací o konzorciálních projektech, informovat uživatele o novinkách a pomůckách k jednotlivým informačním zdrojům, usnadnit správu rozsáhlých konzorciálních licencí, sdružit informace o IP adresách, usnadnit zpětnou vazbu mezi uživateli, koordinátory na úrovni jednotlivých institucí, technickou podporou AiP a vydavateli.

### **ČTK**

ČTK nabízí kromě vlastního zpravodajství i plné texty několika desítek nejvýznamnějších tuzemských novin, týdenků a jiných odborných titulů. V elektronické formě dává k dispozici plné texty bez reklamních a inzertních stránek, bez fotografií a grafů. Stejně tak jsou k dispozici přepisy nejdůležitějších zpravodajských a publicistických pořadů významných domácích rozhlasových a TV stanic. Infobanka poskytuje informace vybrané informace o zemích světa včetně politických, makroekonomických, historických a geografických údajů o každém státě. Databáze přináší podrobný přehled o domácím politickém, hospodářském a kulturním životě, včetně vývojových materiálů, jež průběžně sledují všechny klíčové události ve společnosti. INFOBANKA ČTK obsahuje také biografie známých osobností z celého světa, sportovní databázi s výsledky a životopisnými údaji významných sportovců a přehledy výročí a událostí očekávaných doma i v zahraničí v příštích měsících a letech.

### **Newton I.T.**

Agentura vytváří elektronický archiv zpráv z domácích i zahraničních tištěných médií, plných přepisů zpravodajských, publicistických a diskusních pořadů z televize a rozhlasu, agenturního zpravodajství a zpravodajských internetových serverů (až od roku 1993). Z toho archivu poskytuje monitoring podle individuálních potřeb. Nabízí jednotlivým redakcím tvorbu elektronických archivů. Elektronické archivy dostupné přes standardní webové rozhraní či na CD jsou určeny pro vnitřní potřeby samotného vydavatele, tak pro nejširší veřejnost (Právo, Mladá fronta Dnes, 100+1 zahraničních zajímavostí, Respekt (po registraci dostupné i volně).

### **Anopress IT, a.s.**

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

Hlavními produkty jsou: on-line databanka novin a časopisů celostátních a regionálních, monitoring na zakázku, vědomostní databáze, archívy zdrojů. Anopress IT, a.s. umožňuje on-line přístup do databanky plných textů **TamTam**, na jejíž bázi poskytuje následné služby. Společnost Anopress je výhradním zpracovatelem

elektronické podoby většiny českých regionálních titulů (nakladatelství Bohemia). Pro zpřístupnění plných textů ve veřejných knihovnách bylo založeno v r. 2000 **Konzorcium Anopress**. Společnost je výhradním zástupcem slovenské firmy SLOVAKIA ONLINE v ČR, která zpracovává elektronickou podobu slovenských tištěných médií. Kromě mediální části obsahuje databanka TAMTAM i část vědomostní, v níž jsou k dispozici pro fulltextové vyhledávání různé encyklopedie, příručky a další knihy referenčního charakteru. Databank a 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 databance TamTam je obsaženo cca 200 informačních zdrojů.

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 byla schválena dotace MKČR pro multilicenční zpřístupnění databází TamTam a ČTK v rámci VISK 8 v r. 2003. 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.

### [Národní lékařská knihovna](#)

Vydává BMČ a CR-CZ (viz dále), též katalogizuje elektronická periodika dostupná v rámci licencí. Některé záznamy elektronických zahraničních časopisů byly předány do STK v rámci projektu Portál STM a měly by být zahrnuty do terciální databáze. Dále by měla být zahrnuta česká elektronická periodika online. Knihovna by měla také katalogizovat internetové domácí zdroje z oblasti lékařství a zdravotnictví. NLK řeší projekt Medicínská virtuální knihovna - [MEDVIK](#), sdílení informačních zdrojů pro vědu a výzkum.

### [Státní technická knihovna](#)

elektronické časopisy se katalogizují v terciální databázi v rámci projektu LI01018 z oblasti STM. Na stránkách knihovny je databáze on-line časopisů podle oborů a databáze českých ISSN. Provozuje [Portál STM](#) (viz dále).

### [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. V roce 2003 byla zpřístupněna báze bibliografická WEB, obsahující některé on line elektronické zdroje. Cílem je, aby ze záznamů v databázi byl současně umožňován přístup do digitálního archivu (WebArchiv). Protože platná česká legislativa neumožňuje v

současné době zdroje zpřístupňovat, řeší se situace [uzavíráním smluv s autory/vydavateli](#) internetových zdrojů. Na projektu Registrace, ochrana a zpřístupnění domácích elektronických zdrojů v síti Internet spolupracuje Ústav výpočetní techniky Masarykovy univerzity v Brně. Předběžně byla dohodnuta spolupráce při testování využití metadatového schématu **Dublin Core** s několika informačními a dalšími institucemi, které působí současně jako vydavatelé elektronických zdrojů.

**Dublin Core Metadata Generator** je nástroj umožňující autorům webových stránek poloautomaticky nebo ručně vytvořit, editovat, konvertovat a ve zvolené syntaxi uložit metadata respektující pravidla kvalifikovaného Dublin Core. Přidělení jednoznačného identifikátoru je umožněno propojením Dublin Core generátoru s generátorem URN. Přípravuje se úprava systému přidělování URN tak, aby program přiděluje URN fungoval jako samostatný URN server s možností integrace této funkce do publikačních systémů vydavatelů online zdrojů. Díky tomu by se přidělování URN mělo stát zcela automatickým procesem.

V rámci projektu byl lokalizován nástroj **Harvester** z projektu NEDLIB, a kalkulátor **MD5** pro výpočet kontrolního součtu dokumentu. Slizená data se budou ukládat pomocí páskového robota.

Řešení právních otázek je dlouhodobou záležitostí. Je třeba připravit podklady pro změnu zákona obsahujícího ustanovení o povinném výtisku seriálových publikací ("tiskový zákon"), resp. výklad a směrnice k zákonu týkajícímu se povinného výtisku neperiodických publikací a k autorskému zákonu

Zpřístupnění archivu se má realizovat pomocí technologie fulltextového indexování a automatizované extrakce autorem vytvořených metadat. Na MFF UK vypsán ročníkový týmový projekt na vytvoření indexační a vyhledávací aplikace pro Webarchiv. Tato aplikace by měla zpřístupnit stažené dokumenty v jejich kontextu, tedy s vloženou grafikou ze stejné doby a s odkazy vedoucími primárně opět do archivu. Vyhledávání v archivu by mělo být umožněno nejen na základě URL nebo kontrolního součtu dokumentu, ale i na základě z dokumentu extrahovaných metadat nebo fulltextového vyhledávání. Tato aplikace by měla být navržena tak, aby bylo možné k ní kdykoli připojit moduly pro indexování jiných, než textových typů souborů.

Jedním z dalších cílů projektu bude proto pokus o využití systému Convera Retrievalware pro indexování některých typů souborů obsažených v archivu.

Z dokumentů, které obsahují metadata podle DC by se mohla vybudovat menší bibliografická báze. Aplikace vyvíjená v rámci projektu Souborná databáze by se eventuálně dala upravit pro přijímání dokumentů z Webarchivu.

### [Jednotná informační brána](#)

#### **Projekty JIB:**

Jednotná informační brána pro hybridní knihovny (VAV, UK, NK). Součástí projektu JIB je:

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 Liberci zpřístupněny v JIB jako odkaz. Plné texty báze ANL FULL a přes bázi ANL přístupné v JIB pro externí uživatele na 1 den. V řešení jsou připojeny dalších institucí a zdrojů

**Metoda konspektu** jako nástroj popisu fondů má přispět k realizaci této brány. Základem metody 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 konspektu. 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 za loženém na open URL.

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

Povinný výtisk elektronických publikací je předpokladem jejich uchovávání a zpřístupnění. Prozatím je třeba v budoucnu uzavírat dohody mezi knihovnami a příslušnými nakladateli a vydavateli, které se týkají jednak zpřístupnění elektronických dokumentů, jednak spolupráce.

### **Automatická či poloautomatická indexace (J. Schwarz, 2002)**

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

Již od začátku 70. let byly vyvíjeny i systémy SEMAN (V. Smetáček) a MOZAIKA (Z. Kirschner). 70. a 80. léta představovala také období rozvoje matematické, počítačové a korpusové lingvistiky pod vedením P. Sgalla; vzniklo specializované pracoviště (dnes Ústav formální a aplikované lingvistiky/Centrum počítačové 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 (viz [Příloha F 5, F 1, F 3, F 4](#))**

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

**Národní knihovna ČR** zpracovává výběrově bibliografické záznamy článků ze všech druhů seriálů (noviny, časopisy, odborná periodika, sborníky) v rámci **Kooperačního systému článkové bibliografie (KOSABI)**, ve kterém spolupracuje v současnosti 8, resp. 7 krajských knihoven (Plzeň momentálně nedodává záznamy) a MZK, specializované odborné knihovny (STK, ÚZPI, SPKK-ÚIV, ČSAV, experimentálně ČGÚ). Spolupráce je navázána se 4 novými krajskými knihovnami. 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 tento rok instaloval ALEPH (články dříve zpracovávány v 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 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. V roce 2003 probíhaly v NKČR práce na konverzi UNIMARC MARC 21.

**Zpracovávané typy titulů a výběr článků z hlediska úplnosti:** v kooperačním systému jsou zpracovávány články a statě české novinové, časopisecké a periodické sborníkové produkce (sborníky vysokých škol, muzeí, archivů, materiály z konferencí, ročenky, odborné i polytematické časopisy, kulturně-politické časopisy a populárně-naučné časopisy).

**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 11.11.2003: anl cca 807 000 bibliografických záznamů, anl full cca 115 000 plných textů s metadaty
- počet zpracovávaných titulů: ANL - cca 212 v NKČR, 2 08 v regionech, 122 v odborných knihovnách, celkem cca 542 titulů; ANL FULL 30 titulů, běžně zpracovávaných zatím 17 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 souhlasí 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 je r. 1999-2003 vystavován ve formátu html a pdf ve speciální webovské aplikaci pro toto periodikum. Speciální aplikace pro Národní knihovnu se bude integrovat s bází ANL FULL pomocí jednotného interface. V rámci projektu je též částečně koncepčně řešena aplikace pro 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 excerpce 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 **Rozvoj regionálních databází na základě retrospektivní konverze článkové bibliografie** řešený v letech 2000-2002 v **SVK Kladno**.

Analytické záznamy zpracovávané v rámci **KOSABI** jsou zpřístupňované také na **CD-ROM** vydávaném **AIP icode** v rámci ČNB jako **řada Články v českých novinách, časopisech a sbornících**, od června v 2000 v UNIMARCu. CD-ROM je vydáván ve čtvrtletních aktualizacích, každý měsíc je bibliografie aktualizována na Internetu. 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 1 den po registraci.

**V roce 2003 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. Probíhají práce na internetové verzi linky s napojením na rejstříky ANL FULL a autority prostřednictvím komponenty Ole Z39.50. Báze ANL FULL je vybavena dalšími topiky. Je nastíněn další možný vývoj KOSABI s ohledem na moderní metody zpracování. Je vypracována aplikace pro kontrolu platnosti URL adres. Výsledky práce KOSABI a projektů týkajících se zpřístupnění článků prezentovala řešitelka projektu na konferencích Infos 2003, též v IFLA Section for Newspapers a na zasedání bibliografické sekce SDRUK.**

**V r. 2003 vypracován nový pokračující projekt týkající se zpřístupnění odborných zdrojů v kooperačním systému a zapojení báze ANL FULL do JIB v rámci programu 1N MŠMT „Informační zdroje pro vědu a výzkum“ (Anděrová) a komplexní záměr v rámci NKČR (pro MKČR), který se týká zejména věcného zpracování v rámci kooperačního systému.**

Citace z projektu:

V **NK ČR** v rámci běžících projektů - bude zajištěna internetová verze linky a vazba linky na autority, použitelnost linky nejen pro vstupy z Anopressu (TXT, HTML) - na základě vývoje linky v rámci běžícího projektu týkajícího se souborné databáze bude možno již v r. 2004 zpracovávat odborné plné texty již vystavené na internetu.

V navrhovaném projektu projektu pro 1N je třeba linku modifikovat pro více vstupních formátů (DTP formáty - Word, PDF, OCR), doladit zpracování v lince na zpravování i bez plného textu (pouze metadata) a doladit (tj. plně zprovoznit) propojení na bázi autorit při zpracování v lince - Z39.50 klient Database Update), případně otestovat linku v rámci spolupracujících institucí. V projektu je třeba doladit konverzi do M21, doplnit stávající DTD pro články o potřebná data administrativní a specifická, další doplnění DTD vzhledem podrobnější strukturaci textu. Báze ANL FULL je provozována v systému TOPIC. V projektu je třeba zajistit migraci systému TOPIC na systém K2 (automatizovaná tvorba kategorií, integrace odborných tezurů do topiků, implementovat lingvistický analyzátor pro český jazyk). V rámci konceptu sociálních sítí zajistit interakci, adaptivní vyhledávání, doporučování dokumentů vzhledem k uživatelskému profilu a využít vlastností K2 pro mapování stěžejních výzkumných oblastí v oborech, implementovat SQL databázi pro záznamy DC (XML) - (dosud jsou strukturované údaje uloženy v bázi Access, která již nedostačuje), zapojit ANL FULL do JIB (Z39.50 s rozšířenou funkcí Database Update a modul pro podporu Open URL) - zajistit integraci fulltextového vyhledávání do systému JIB“.

Projekt je zaměřen na zpracování odborných informačních zdrojů, a spolupráci odborných knihoven. Týká se i spolupráce s NLK, která je součástí širšího kooperačního systému.

V **ČR NLK** se zpracovává databáze BMČ. Je třeba optimalizovat dosavadní způsob zpracování zdrojové báze Bibliografie Medica Československa (BMČ) a Českého citačního rejstříku (CR-CZ) základě technologické linky s využitím formátu XML, zprovoznit databázi BMČ a CZ-CR v systému WEB Tornado, zapojit obě databáze přes Z39.50 do JIB, zajistit propojení na plné texty. Součástí řešení je i parametrizace výstupních formátů umožňující vstup do dalších databází (PubMed dle požadavků NML Bethesda, RIV) a konverze pro zpracování výstupů z databází Embase a MedLine.



V [ÚZPI](#) se zpracovává Zemědělská a potravinářská bibliografie. V budoucnu je třeba převést také zde zpracování článků do formátu UNIMARC/M21. Má být umožněno propojení záznamů na plný text v rámci technologické linky zpracování bibliografických záznamů z plných textů přes formát XML v návaznosti na existující redakční systém za účelem importu do bibliografické báze a též za účelem kooperace s NLM Bethesda.

V [ČGs](#) jsou záznamy zpracovávány v systému Win ISIS ve struktuře MAKS, vystavovány jako Česká geologická bibliografie. Zpracování článků je třeba převést na strukturu dle aktuální specifikace pro analytický popis z r. 1993 (aktual. 1997) s respektováním standardů a autorit.

V [ÚIV SPKK](#) jsou články zpracovávány v systému CDS ISIS. Zpracování článků je třeba převést na zpracování v systému KP Win (UNIMARC, event M ARC 21). Na internetu je vystavována Pedagogická bibliografická databáze.

Ve [STK](#) jsou články zpracovávány v systému CDS ISIS. Zpracování článků je třeba převést na zpracování v systému KP Win (UNIMARC, event MARC 21).

## 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 zpracováváné 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.

**Tento systém se bude v budoucnu pravděpodobně modifikovat tak, jak budou postupně aplikovány prostředky pro interoperabilitu dat - zejména Z39.50 a OpenUrl. Tyto záležitosti jsou předmětem podaného projektu v rámci programu 1N MŠMT.** Hlavním cílem projektu je zkvalitnění bibliograficko-informačních služeb a integrace elektronických zdrojů do služeb knihovny.

### Cíle

- Cílem projektu v r. 2003 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 rutinní zpracování záznamů o člancích z plných textů v rámci linky automatické indexace (TTDE) - TTDE umístěné na lokálních pracovních stanicích.
- Zkvalitnění ukládání v pracovním formuláři, který je ještě umístěn na lokálních pracovních stanicích v rámci linky automatické indexace.
- Převedení linky automatické indexace na Internet (WWW TTDE), napojení na rejstříky ANL FULL a ověření možnosti napojení na soubory autorit přes komponentu Z39.50.
- Dalším cílem je implementace aplikace pro administraci báze a v rámci ní aplikace pro kontrolu platnosti url adres v rámci báze ANL a ANL FULL.
- Ověření aplikace pro export a stahování plných textů a metadat, zatím běžně nepřístupné s ohledem na autorskoprávní problematiku.
- Pozornost se soustřeďuje na definici báze ANL FULL - zejména na vyhledávání pomocí topiků, jejich ladění (tvorba nových topiků, návrhy na zlepšení relevance vyhledávání podle topiků jsou

hrazeny ze souvisejícího projektu Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů).

- **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, doplnění dalších zdrojů .**
- **Cílem je propojení plného textu se záznamem, které probíhá na základě jednak statických URL vznikajících v bázi ANL (volně přístupné texty na Internetu v rámci báze ANL), jednak dynamických URL (generovaných v rámci linky zpracování bibliografických záznamů z plných textů), které jsou do záznamů generovaných v lince a importovaných do Alephu vkládaných jako statické. V rámci ANL FULL jsou metadata uložena přímo v plných textech.**
- Průběžné 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.**
- Dalším cílem je **návrh minimálního záznamu v M21 pro články.**
- **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 a poloprovozní stádium v NKČR.**
- **Údržba databázové aplikace pro evidenci zpracovávaných titulů v rámci KOSABI.**
- **Cílem je dialog program aplikace pro opravy dat kooperujících institucí pomocí JDeveloperu.**
- **Dalším cílem je implementace aplikace pro kontrolu platnosti url adres v bázi ANL v aplikaci MNG KOSABI.**
- **Dalším cílem jsou praktické závěry pro další vývoj projektu na základě analýzy současného stavu a trendů automatické indexace a selekční efektivity topiků v databázi ANL FULL, které byly podkladem pro navazující projekt v rámci 1 N, tj. migrace systému TOPIC na platformu K 2 s možností automatické kategorizace, vyhledávání na základě podpory jazykového analyzátoru a možnostmi interakce uživatele se systémem a ostatními uživateli v rámci tzv. konceptu sociálních sítí.**

### A. 3.2 Vstupní data

- **Vstupními daty pro bázi ANL jsou bibliografické záznamy zpracovávané v r. 2003 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 a jejich zpřístupnění v HTML XHTML a XML v bázi ANL FULL**
- **V r. 2003 se i nadále 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. 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) - cca 26 titulů** (Veřejná správa, Vesmír, Lesnická práce, Literární noviny, Harmonie, Psychiatrie, Psychologie Dnes, Jezuité, Kriminalistika, Národní knihovna, Knihovní obzor, Collection of Czechoslovak Chemical Communication - abstrakta, Kriminalistika, Bulletin advokacie, Souvislosti, Moderní řízení, Československý časopis pro fyziku, Chemické listy, Země světa, Revue církevního práva, Obec a finance, Věstník (příloha Obec a finance), Právní rádce ..., Ikaros, U nás, Bulletin SKIP ) (**viz Příloha F 4a**). Toto propojování hrazeno z prostředků souvisejícího projektu Propojení analytických záznamů s plnými texty.

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

**V lince pravidelně zpracováváno 17 titulů** (Bankovníctví, Ekonom, Haló noviny, Hospodářské noviny, Kapitál, Lidové noviny, Magazín Práva, Mladá fronta Dnes, Pátek magazín LN, Právo, Profit, Reflex, Respekt, Týden, Kapitál, Večerník Praha, Euro viz **Příloha F 4a**).

**Summa summarum:**

- Vstupními daty pro plnotextovou bázi ANL FULL jsou plné texty zejména celostátních deníků a některých odborných časopisů za rok 2003 a jim odpovídající bibliografické záznamy v tomtéž 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 fulltextovou 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. Pokládám ho za velmi důležitý vzhledem k řešené problematice. Ponechávám informace předložené již ve zprávě z r. 2002, na které navazuje informace o řešení problematiky v roce 2003.

#### B.0.1 Analýza problematiky automatické či poloautomatické indexace a selekční úplnosti topiků v bázi ANL FULL (J. Schwarz, 2002) - 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 nespécifických 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, fuzzy 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 studií, 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 obdobně 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

(J. Schwarz, 2002)

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

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

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

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

**Dokumenty z RMD (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 ne obsahuje všechny dokumenty obsažené v TMD.

**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 přes 80 % (84 %).**

### 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 nejefektivnějším 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 pře kážku 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 Vliv topiků na úplnost vyhledávání v bázi ANL FULL (J. Schwarz, 2003)**

Hlavním cílem analýzy bylo zjistit, jak topiky, tj. předdefinované komplexní dotazy založené na strukturovaných a vážených klíčových slovech, ovlivňují úplnost vyhledávání v plnotextové článkové databázi ANL FULL, a navrhnout případné úpravy topiků, popř. jiných selekčních prvků databáze, které povedou ke zvýšení úplnosti vyhledávání.

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

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

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

Podarilo se **však identifikovat hlavní důvody**, které **mohou vést k nevyhledání dokumentu** (přičemž k nevyhledání konkrétního dokumentu může vést každý z důvodů samostatně stejně jako jejich kombinace). Byly identifikovány topiky při vyhledávání přinášejí z větší části zcela irelevantní záznamy (bude v souhrnné zprávě Propojení analytických záznamů s plnými texty za léta 1999-2003).

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

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

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

### **B.0.4 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 je zprovozněna v nové verzi TOPICu i funkce shlukování (clustering) článků do skupin podle stejných klíčových slov.
- 2. V TTDE je funkční metoda automatického vkládání vybraných klíčových slov do formuláře TTDE a záznamu na základě požadavku při vyhledávání dokumentů v bázi TamTam. Tato slova by se musela indexátorem redigovat. Tyto metoda není při práci v lince využívána - není příliš kvalitní, dosud je věcný popis dokumentu zcela vytvářen katalogizátorem.
- 3. Systém TOPIC generuje také automatický souhrn k článku - zatím začátek textu dokumentu. Zjistit další možnosti v tomto směru.
- 4. Vzhledem k fuzzy vyhledávání v systému TOPIC a ještě nerealizovaným a neobjeveným dalším možnostem, které TOPIC má, se kloním spíše k využití funkcí dosud v projektu nezohledněných, ke zkvalitnění vyhledávání pomocí topiku na základě jejich ladění a zabudování prvků věcného popisu (kategorií a prvků předmětových hesel) přímo do topiků, než k zcela automatickému přiřazování věcných termínů. Užívají se metody např. automatizované kategorizace či klasifikace.
- 5. Hypoteticky: v systému je náběh na řízený slovník tematických kategorií, který by mohl být použit i v rámci topiků (a jejich generování ?) spolu se vstupním prvkem předmětového hesla.
- 6. Slovník bude pravděpodobně možno využít také jako pomůcku pro katalogizátora při přiřazování těchto kategorií.
- 7. Jak mohou být tyto kategorie spojeny s automatizovanou kategorizací nebo do jaké míry mohou být pomůckou při automatizované kategorizaci informací ukáže budoucnost.
- 8. V budoucnu zprovoznit hledání dokumentů s podobným obsahem - volný dotaz (Free Text Query), dotaz příkladem (Query By Example).
- 9. Topiky ladit na konkrétních dokumentech z příslušného oboru - určit jak obory, tak odpovídající dokumenty s kvalitním předmětovým popisem. V úvahu přichází periodikum Národní knihovna (zprístupňované v projektu Propojení analytických záznamů s plnými texty a ke kterému vzniká řízený slovník) aj. odborná plnotextová periodika, která budeme mít pravděpodobně k dispozici, dále pak obory a dokumenty, o kterých lze tvrdit, že jsou kvalitně indexovány katalogizátorem (např. ekonomie/ekonomika, některé společensko politické časopisy).

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

Analytické práce týkající se topiků vyústily v řadu doporučení, které budou posuzovány popř. realizovány v následujících letech.

**V r. 2003 byl podán projekt v rámci 1N programu MŠMT „Informační zdroje pro vědu a výzkum. V tomto projektu je akcentován přechod TOPICu na platformu K2 s možností automatizované kategorizace s využitím technologie LCR (Logistic Regression Classification), která umožňuje vytvořit taxonomii založenou na analýze obsahu textu (Thematic Mapping), vyhledávání za podpory jazykového analyzátoru a zavedení konceptu tzv. sociálních sítí s využitím technologie TSA (Tensor Space Analysis). Technologie umožňuje analyzovat jak uživatelé vytvářejí, modifikují, vyhledávají a přijímají informace. Na základě toho pak provádí doporučování dokumentů a autorit, které mají vztah k dotazu, resp. profilu, uživatele (adaptivní řazení dokumentů - založené na obsahu i na tom, jak často byly dokumenty vyhledány a prohlédnuty v minulosti, doporučování dokumentů vzhledem k uživatelovu kontextu, objevování/nacházení expertů, tj. experti jsou rozpoznáni na základě informací, které vytvořili, modifikovali a/nebo ke kterým přistupovali; na základě společných znaků vznikají komunity uživatelů).**

V projektu je dále akcentována tvorba a ladění topiků pro příslušné odborné zdroje v bázi ANL FULL s ohledem na oborové tezaury, využití systému K2 pro sledování citačních vazeb.

## **B.1 Vlastní řešení - praktická realizace**

Pro zachování kontinuity řešení uvádím výsledky i z předešlých let.

### **Stručný popis stávajících produktů TamTam a prací v r. 2002 a 2003:**

#### **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 (linka poloautomatické indexace) - verze pro lokální stanice**
  - vstup - data Anopress (txt)
  - úprava hlavičkových údajů ve stažených člancích (3 strany formuláře pro editaci - hlavičky)
  - výstup pro Aleph a Full
  - hlavičky Uni, Uni-A, DC, 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 pro vývoj internetové verze TTDE( nové 2002)**

- c. **WWW TTDE - WWW TamTam Data Extractor (linka poloautomatické indexace) - verze pro Internet (nové v r. 2003)**  
- vstupy: formáty txt, html z libovolné URL  
- rozšíření formuláře pro editaci (4 strany - hlavičky)  
- aplikace pro opravy rejstříků a metadat v lince (hlavička opravy)
- d. **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 serveru 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 Convertor**  
- 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. **TTEXPORT - 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
- <http://full.nkp.cz/nkdb/mkdoc.asp?p1=s120001019000064>

#### 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 r. 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 hlaviček**

- 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. WWW TTDE pro Web (příprava a návrh v r. 2002, částečně plán)**

- *kompletní překlopení linky TTDE do www prostředí (plán)*
- *stahování jiných zdrojů než Anopress*
- *obecná úprava programu pro vstup z více vstupních kanálů*
- *přidání filtrů pro každý vstupní kanál*

## **K. 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)*

## **ROK 2003 - seznam prací s plánem do budoucna (kurzívou)**

### **LINKA pro zpracování bibliografických záznamů z plných textů a její převedení na www - WWWTTDE**

#### **A. Design linky zpracování bibliografických záznamů z plných textů - WWWTTDE**

- vstupní strana: načíst soubory (uložené na disk), načíst URL, spustit Linku
- dohromady 13 hlaviček formuláře pro editaci (oproti TTDE 4 strany formuláře pro editaci, hlavička Opravy)
- začlenění funkce Opravy (původně v aplikaci Administrace)

#### **B. Úpravy linky zpracování pro Internet -1**

- úpravy nabídky ... Různé pro stahování www dokumentů obecně s možností extrakce metadat z - plných textů , jejich doplnění a generování ( analogie lokální linky pro texty Anopress i s možností vytvoření pouze hlavičky s url na plný text bez stažení plného textu)



### **C. Úpravy linky pro Internet -2**

- doplnění údajů do formuláře, úpravy v hlavičkách

### **D. Vazba na rejstříky ANL FULL a na Bázi autorit - AUT**

- stahování údajů z rejstříků ANL FULL podle úprav ve formuláři  
- stahování údajů z báze autorit

### **E. Administrace**

- doplnění nových údajů do tabulek (formulářů) k příslušným položkám, např. údržba www periodik

### **F. Konverze z UNIMARu do MARC 21**

- *plan na konec r. 2003 a r. 2004*

#### **Další vstupní formáty pro linku (pdf, Word aj.)**

- *plán do budoucna v pokračujícím projektu*

### **G. Doplnění DTD pro články o data administrativní aj.**

- *plán do budoucna v pokračujícím proje ktu*

### **H. Možné úpravy ve vyhledávání , generování topiků z dat uložených v kategoriích a heslech**

- *plán do budoucna v navrhovaném pokračujícím projektu 1N*

## **Popis řešení**

Opět pro zachování kontinuity informací jsou obsaženy i výsledky předešlého roku.

**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, dr. Kučera, J. Schwarz).**

**Linka zpracování bibliografických záznamů z plných textů má v současné době dvě podoby - linka TTDE pro práci ve formuláři na lokálních pracovních stanicích (pouze stahování textů přes Internet) a WWW TTDE - kompletní překlopení TTDE na Internet, které je realizováno v r. 2003.**

**V rámci poloprovozu v r. 2002 byla laděna linka TTDE a ověřována v poloprovozu .** V současné době má pracovní formulář, do kterého se extrahují data a ta dále doplňují věcným popisem , 3 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.Z uložených a editovaných dat se generuje DC, UNIMARC pro ALEPH a klasický UNIMARC, hlavička zobrazovací a indexovací.

**Automaticky lze generovat klíčová slova a extrahovat do formuláře.** Na výstupu procesu zpracování zabudovaná **funkce kódování výsledného souboru záznamů v Ansel, UNICODu a UTF-8 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.nkp .cz 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 metadaty s generovaným UNIMARCEm, DC v podobě HTML, XHTML, XML v kvalifikované a nekvalifikované formě.**

**Vyhledání je možno z plného textu, z metadat v něm uložených - tj. ze strukturovaných údajů (vyhledávání pomocí polí, formulářů, topiků a rejstříků). Zobrazení je možno pomocí tří uživatelských formátů.** Je možno tisknout vyhledaný jak seznam článků , tak jednotlivé plné texty .

Zároveň je importován do bibliografické báze ANL UNIMARC pro Aleph.

Záznamy ANL jsou propojeny s plným textem automaticky generovanou adresou v lince zpracování, která je uložena jako statický link v ANL záznamech.

V roce **2002 byly 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 (zatím rozhodnuto nerealizovat). **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**. Probíhaly přípravné práce a návrh **internetové verze linky zpracování TTDE**.

V roce **2003** byla uzpůsobena linka zpracování bibliografických záznamů z plných textů TTDE (TamTam Data Extractor) pro práci na Internetu - **WWW TTDE** a **rozšířeny její funkce: možnost zpracovávat plné texty na Internetu s uploadem metadat i bez stažení plného textu, vstupní formáty zip, txt, html, propojení linky na rejstříky báze ANL FULL a na autority (zatím ve formě komponenty O LE ), zabudování funkce Opravy přímo do linky. Formulář pro editování dat byl rozšířen o další pole/podpole UNIMARCU zejména s ohledem na propojení se soubory autorit.**

**Byly implementovány další topiky, např. země Evropy a ostatní země (49 nových topiků).**

**V rámci tohoto projektu bude linka v r. 2004 testována.**

B.1.1 Architektura systému zpřístupňování plných textů, funkce systému a procesy. Linka TTDE, resp. WWW TTDE. Systém TOPIC a topiky (viz Příloha F2; F 3; F4; F6)

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 bylo možno získat v r. 2002 i přístup k plným textům na základě zkušební registrace. **V r. 2003 byl omezen vstup pro externí uživatele pouze na 1 den z ohledem na možné komplikace s autorskými právy. Vybraní uživatelé mohou po registraci zkoušet 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. V rámci konzorcia by měla být používána i linka zpracování bibliografických záznamů z plných textů.**

**Možnost zpřístupnění ANL FULL externím uživatelům a v rámci konzorcia se bude řešit v r. 2004.**

Inspirací a do jisté míry vzorem pro koncipování báze ANL FULL je databáze **ProQuest 5000 a PQ NEXT rozšíření o služby pomocí OpenURL.**

**B.1.1.1 Architektura systému získávání, zpracování a zpřístupnění plných textů textů v systému TOPIC (báze ANL FULL) a ALEPH (báze ANL) (viz Příloha F2; F3; F6).**

**On-line získávání dokumentů z databáze TamTam přes vstupní filtr** pro linku zpracování - **aplikace TT Special NK (TTSNK)**, textové dokumenty vstupují do linky se strukturovanou hlavičkou, kde jsou dohodnutými znaky a návěštými označeny příslušné údaje . **V r. 2003 byla aplikace doplněna o vstup HTML dokumentů z obecné URL (v budoucnu možná další úprava vstupních filtrů podle charakteru získávaných dokumentů z jiných zdrojů a v dalších formátech - pdf, word).**

**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. V roce 2003 vytvořena varianta pro použití na WWW - WWW TTDE.

**Konverze, generování dat v XHTML+-, XML+-, indexace, uložení a vyhledání informací** (vyhledávání pomocí textu dotazu, formulářů, topiků, rejstříků) - **aplikační a datový server: Portal One SE** - fulltextový vyhledávací systém, server Dell PoweEdge 6300 (1 GB RAM, 68 GB, Pentium III). MKINDEX - indexace, generování XML- a XML+, jednorázové aplikace/moduly (TTEXTEXPORT-TamTamExport, TTXMLGen -TamTam XML generator, TTDT-TamTam Data Tester, TTDF-TamTam Data filter, TTDC- TamTam Data Converter, MKDOC.ASP - propojení bibliografického záznamu s plným textem). Strukturované údaje jsou ukládány v Accessu, v budoucnu nutná SQL databáze pro metadata DC a XML, migrace TOPICu na platformu K2.

**Zobrazení a zpřístupnění informací - Internet Information Server Win 2000** - komunikace v síti, zobrazování dat a metadat v uživatelských formátech (citace bez/s odkazy, výsledky jednoduché, se souhrnem, seskupené) a pracovních formátech (klíčová slova, DC v HTML, XHTML+ -, XML+ -), tisk. Obecné základní informace, nápovědy, autorizace, statistika zobrazených informací přes Internet aj.

**Stahování 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.** V r. 2003 zabudována aplikace pro opravy a údržbu u rejstříků/plných textů přímo do linky.

**B.1.1.2 Linka TTDE (TamTam Data Extractor) a WWW TTDE (TamTam Data Extractor pro WWW)**

( získávání a zpracování bibliografických záznamů z plných textů )

**Linka zpracování bibliografických záznamů z plných textů (viz Příloha F2; F6)**

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) pro lokální stanice, WWW Tam Tam Data Extractor (WWW TTDE) pro práci na Internetu.**

## TTDE

Údaje jsou extrahovány z plného textu a na jejich základě je generováno 5 **hlaviček (headers)**. **Data jsou zpracovávána pomocí rozšířeného formuláře pro editaci (detailně popsány ve zprávě z r. 2001).**

**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 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.

Po doplnění formuláře je po odrážce 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 A LEPHu (program vyvinutý v NK), import na server full.nkp.cz.

## TTDE WWW

**Li nka pro zpracování na WWW se liší designem, možností více typů vstupů do formátů, většími možnostmi editace, její formulář pro editaci obsahuje 4 hlavičky detailně strukturované zejména s ohledem na strukturu autoritních**

údajů (z formuláře vyřazeno SICI, SICI se generuje a automaticky vkládá do příslušných metadat, aniž by ho katalogizátor musel vidět ve formuláři). Lze stahovat a doplňovat metadata s para le lním uložením těchto metadat do plného textu a jeho stažením na server. Lze natahovat a editovat metadata určitého plného textu bez stažení tohoto textu do báze.

Z jednotlivých údajů je možno se napojit na rejstříky báze ANL FULL, v současné době se testuje varanta napojení linky na soubory autorit. Je implemenován modul pro opravy v podobě další hlavičky. Původní základní funkce TTDE byly zachovány.

**LINKA WWW TTDE - popis:**

**Vstupní strana - hlavní funkce: načíst soubory, načíst URL, spustit Linku**

- **Načíst soubory (stažené na lokální počítač):**  
Načíst soubory pro zpracování: pouze zip, txt, html a to: z databáze TamTam, z databáze ANL FULL, z jakéhokoli URL - až 10 souborů.  
Poslat soubory na full.nkp.cz a spustit TTDE (otevřít formulář pro zpracování).
- **Načíst URL:**  
Stáhnout zadaná URL (přímo z WWW) - až 10 souborů.  
  
Spustit TTDE (otevřít formulář pro zpracování).

- **Spustit Linku**  
Extrakce metadat, editace pomocí 4 hlaviček formuláře, hlavička pro zobrazení textu, hlavička UNIMARC - generovaný UNIMARC, hlavička UNIMARC A - generovaný UNIMARC pro ALEPH, v budoucnu MARC 21, zobrazovací a indexovací hlavička, hlavička Opravy, hlavička ... (tj. různé)

**Hlavička Opravy:** umožňuje opravy rejstříků ANL FULL a metadat v plných textech - komplet nová funkce (nahrazuje opravy původně připravené v Administraci)

Po vyhledání údaje v příslušném rejstříku lze údaj zrušit, opravit nebo vypsát všechny články databáze, které tento údaj obsahují, tj. lze opravit záznamy a následně znovu přenést do databáze ANL FULL nebo nalezené články úplně zrušit. Funkce se ještě bude ladit.

**Hlavička ... (tj. různé)** - zachovává stejné funkce jako ve TTDE, navíc možnost stahovat z URL.

- **Formulář pro editaci (tj. hlavičky pro editaci aplikace WWW TTDE) a stahování z rejstříků ANL FULL a souborů autorit.**  
Možnost stahování údajů do formuláře a přepisování údajů zanesených do formuláře z rejstříků ANL FULL v rámci příslušných položek formuláře pro editaci.  
Údaje se mohou přenášet najednou do jednoho podpole příslušného pole

UNIMARC nebo se mohou po označení příslušné funkce, natahovat do příslušných podpolí.

- **Možnost stahování údajů do formuláře a přepisování údajů zanesených do formuláře ze souborů autorit v rámci příslušných položek formuláře pro editaci.**

Tato část aplikace je v řešení. Bude na podobném principu jako popsané stahování z rejstříků ANL FULL. V současné době probíhá zprovoznění komponenty OLE v rámci Z.39 50 klienta pro propojení linky na autority. Pokud se osvědčí (a pokud bude schválen navrhovaný projekt), bude aplikováno plné propojení linky na autority v rámci Z39.50 klienta .

- **Aplikace WWW TTDE se bude v roce 2004 ladit a testovat.**

### **LINKA WWW TTDE - přehled funkcí a podporovaných formátů:**

#### **Funkce TTDE WWW pro spuštění aplikace:**

- a) zobrazit panel pro zadání až 10ti vstupních souborů v daném formátu
- b) zobrazit panel pro zadání až 10ti různých url adres (http://
- c) spustit linku

#### **Podporované vstupní formáty a extenze, interní a externí URL, extrakce metadat**

⋮

- 1) vstup článků získaných v rámci přístupu do báze TamTam (Anopress) pomocí TTSNK
- 2) vstup článků z mediální databáze Anopress
- 3) vstup z báze ANL FULL
- 4) vstup článku na základě obecné URL (soubor na disku nebo adresa na webu)

Soubory mohou mít extenzi htm, html nebo txt.

Používají se dva typy adres URL v poli 856: interní (URL v rámci báze ANL FULL) a externí (jinde na WWW).

**Možnost extrakce metadat s paralelním stažením plného textu nebo bez, s příslušnou URL.**

Při načtení URL je možné v hlavičce Text zrušit text a ponechat jen externí URL, nebo ponechat text, interní i externí URL, je možno přenášet údaje ze zdrojového kódu stránky do údajů ve formuláři. Je možno nakonec vytvořit metadatový záznam pouze s propojením na plný text jinde než v bázi ANL FULL.

**Je možné pracovat s obecnou URL adresou buď z menu různé nebo při startu možnost použít formuláře pro zadání až 10ti různých URL.**

## Možné použití TTDE a WWW TTDE

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

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

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

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

**Linka zpracování 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 (viz Příloha F 3)**

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 tezurů 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 vyplývá, že zřejmě nebude řešena indexace automatických přiřazování věcných termínů. V současné době jsou funkční některé formy automatické extrakce, do budoucna lze uvažovat a podpořit intelektuální věcné indexace a možnosti generování věcných topiků on line, o zabudování výrazů odborných tezurů do topiků, o využití interaktivních vlastností TOPICu. Migrace TOPICu na platformu K2 uvažovaná v předkládaném projektu v rámci 1N programu MŠMT by umožnila o automatizovanou tvorbu kategorií.****



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 Pro pojení 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. **V r. 2003 bylo vytvořeno např. 49 nových detailních topiků pro tematickou oblast Geografie ( v rámci ní pro skupinu témat Evropa a Ostatní země světa, celkový počet 49 topiků ).**

**Stávající topiky jsou průběžně aktualizovány.**

**Výhledově je existuje hypotetická možnost propojení jednotlivých úrovní topiků. Je možno uvažovat o postupném 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 hesel.**

V rámci projektu Propojení analytických záznamů probíhají v roce 2003 též další práce na řízeném slovníku pro periodikum Národní knihovna. V uvažovaném novém projektu je plánovaná spolupráce s odbornými knihovnami v rámci kooperačního systému a o rozšíření báze ANL FULL o některá odborná periodika. Na jejich základě by bylo možné dále rozvíjet topiky pro příslušná periodika.

**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) ( Příloha F 3, F4)**

**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 skutečnosti, MDT je víceméně taxativní systém jednotlivých oborů, nikoli témat. Proto je třeba k definici topiků přistupovat svébytně.**

### **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, C CONTAINS, STARTS, ENDS**

**ACCRUE sblíží operátor AND a OR: “čím více různých klíčových slov nalezeno, tím je dokument důležitější“, řeší dilema mezi přesností a úplností.**

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

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

### **B.1.3 Charakteristika serveru full.nkp.cz a báze ANL FULL, Periodika na WWW - portál (viz Příloha F3, F4)**

Server full.nkp.cz je určen primárně ke zpřístupňování plnotextových dokumentů, 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 horním pruhem a základní částí (hlavním prostorem). Horní pruh (frame) je přístupný stále, obsahuje v horní části dynamické rozbalovací menu s jednotlivými nabídkami základních a dílčích funkcí. V základní části se zobrazují požadované funkce.

#### **Obecné (základní) informace**

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

**Další databáze** (databáze NK - charakteristika báze ANL)

**Periodika na WWW** (volně dostupná periodika na www)

**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) (viz Příloha F4a)**

**Databáze ANL FULL** - fulltextové vyhledávání ve vybraných článcích z novin a časopisů zatím s retrospektivou od r. 1997 (periodikum Národní knihovna od r. 1990). Data báze 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 v r. 2002 jsou rozděleny do tří úrovní: tematické oblasti (8), skupiny témat (25), detailní témata (zatím 113). Topiky v r. 2003: tematické oblasti (8), skupiny témat (28), detailní témata (zatím cca 197).**

**Báze obsahuje následující tematické oblasti (8):**

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

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

### **B.1.3.3 Další databáze (ANL) (viz též A.2, B.1 dále, Příloha F1)**

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 zpracováváné 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 WWW** (odborná knihovnická periodika - **Ikaros, U nás, Bulletin SKIP, Knihovní obzor, Národní knihovna** - celkem 5), další odborová periodika (**Veřejná správa, Vesmír, Lesnická práce, Literární noviny, Psychologie Dnes, Harmonie, Psychiatrie, Jezuité, Collection of Czechoslovak Chemical Communications, Kriminálnístika, Bulletin advokacie, Souvislosti, Moderní řízení, Československý časopis pro fyziku, Chemické listy, HOST, Odpady, Tvar, Země Světa, Revue církevního práva, Obec a finance, Právní rádce** - u tohoto časopisu připojené plné texty některých zákonů, **Národní knihovna, U nás, Ikaros, Bulletin SKIP** - celkem počet 22).

**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 literatúra, historie, geografie a všeobecnosti.

#### **B.1.3.4 Periodika na WWW - portál (viz Příloha F3)**

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. V roce 2003 je v portálu zahrnuto 98 regionálních periodik (toho 15 momentálně nepřístupných), 76 oborových periodik.**

**Dílejší nabídky :**

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

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

- **Kultura, umění**  
- film, televize

- hudba
- divadlo
- literatura

- **Odborná periodika**

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

- **Společnost**

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

- **Zahraničí, zajímavosti, cestování**

- **Zájmy (hobby)**

- **Ostatní**

- 

#### **Další zdroje**

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

#### **B.1.3.5 Jak se zaregistrovat (komentář k registraci v databázi ANL FULL) (viz Příloha F3)**

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 aktivačním odkazem a heslem - po odeslání tohoto odkazu je registrace aktivována a prostřednictvím přiděleného hesla zajištěn bezplatný zkušební přístup na dobu sedmi dnů v roce 2002.

**V roce 2003 bylo toto zpřístupnění omezeno na 1 den z důvodu možné reakce některých společností s ohledem na copyright elektronických forem periodik.**

**Řešení je třeba hledat. 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.**

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) (viz Příloha F3)**

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

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

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

#### **Prostý dotaz**

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 nadefinovaných topiků (označeny šipkou vlevo). Úrovně topiků je možno 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** - obsažena v rozšířeném formuláři a v rozšířeném formuláři s tématy .

### Vazby mezi polem dotaz, dalšími poli a tématy

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

**Způsoby zobrazení seznamu výsledků (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 zprovožňuje

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

**Názvy článků je možno zobrazit od určité hranice skóre relevance, v určitém počtu na stránku a třídit 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 (viz příloha F3)

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

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

**Období pro vyhledání od do**

**Implicitní pole:**

název, autor (s nabídkami pro jednotlivé údaje)

číselné údaje (s nabídkami pro jednotlivé údaje)  
předmět (s nabídkami pro jednotlivé údaje)

další nabídky: typ článku (s nabídkami pro jednotlivé typy článků)  
zdrojový dokument (s nabídkami zdrojových dokumentů)

**Způsob vyhledání řetězců (v implicitních polích):**

<contains> (postačí výskyt řetězce v poli)  
= je požadováno přesné zadání i porovnání řetězce

**Typ seznamu výsledků (článků):**

jednoduchý (pouze název s nabídkou zobrazovacích formátů)  
se souhrnem (plus souhrn resp., který tvořený počátečními větami článku)  
seskupený (skupiny článků uspořádané podle shody v klíčových slovech) viz též výše  
Skóre (relevance, od které zobrazovat články)  
Výsl./str. (počet článků na stránku)

**Třídění:**dle skóre (relevance)

dle názvu  
zdroje  
data (vydání)  
stran  
a to sestupně nebo vzestupně.

**Rozšířený formulář s tématy**

**Nabídky:**

Tento formulář má **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 stronomie, 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, planety

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

Př.: 3

(hvězdy, asteroidy) and komety

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

Př. 4

(hvězdy and asteroidy) not komety

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

Př.: 5

komety <near> kolize

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

Př. : 6

"meteorický roj" or "padající hvězdy".

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

Př.: 7

meteo\*

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

Př. : 8

\*stvo

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

Př.: 9

??běr

Otazníková konvence: Vyhledá dokumenty, které obsahují slova končící na "běr" a sestávající z pěti znaků (výběr, záběr, ap

**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 jednoduše i navolit v nabídce období od do ve všech formulářích. Zde je možno označit den či interval.

Vyhledá všechny dokumenty zpracované v tomto dni

Př.:11

zdr=Respekt and naz=Rafinovaný odraz skutečnosti

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

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

Př. 12

zdr <contains>Respekt and naz <contains>odraz  
src<contains>Respekt and ti<contains>odraz  
dc.source<contains>Respekt and dc.title <contains>odraz

Operátor <contains> lze nahradit #:

zdr#Respekt and naz#odraz  
src#Respekt and ti#odraz  
dc.source#Respekt and dc.title#odraz

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

**Tabulka návěští pro pole dotaz a Prvky Dublin Core a Anl Core aplikované v bázi ANL FULL (viz Příloha F4b)**  
**Využito 14 prvků DC, navíc přidán DC.Subtitle (vzhledem k minimálnímu záznamu pro ANL), 14 prvků AC (ANL CORE) definovaných pro vyhledávání a užívaných v popisu článku . Vzhledem k WWW TTDE a rozšíření formuláře pro editaci o některé prvky bude zřejmě potřeba rozšířit v budoucnu ANL Core.**

Je-li případě tečkové 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 návěstí (polí) nejsou zvýrazněny červeně, jsou obsaženy v citaci článku.

**Spíše než návěstí doporučujeme používat formulář s implicitně nastavenými poli.**

### **Formulářová pole s implicitními údaji v rozšířených formulářích. Přehledná tabulka metodiky pro vyhledávání (viz Příloha F 4c)**

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 (viz Příloha F3, F4d)**

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

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

**(viz Příloha F 4 d)**

*Přehled dosud nadefinovaných/navržených topiků strukturovaných do tří úrovní*

> označuje možnost vyhledávání z této úrovně topiku v současné době

→ návrh propojení jednotlivých úrovní

**Rejstříky (viz Příloha F3)**



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**

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

### **B.1.3.10 Výsledky vyhledávání, zobrazení, tisk (všechny formuláře) (viz Příloha F3)**

#### **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 (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**) - v současné verzi TOPICu zprovožňována - viz též výše

### Třídění seznamu výsledků (názevů vyhledaných článků)

**Skóre** (relevance, od které z obrazovat názvy článků)

**Vysl./str.** . (počet článků na stranu)

**Třídění:** dle skóre (relevance)

dle názvu

zdroje

data (vydání)

stran


a to sestupně nebo vzestupně.

### **B.1.3.10.2 Zobrazení údajů o článku (viz Příloha F3)**

#### Formáty zobrazení

#### Uživatelské formáty

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

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

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

 **DC/HTML - Dublin Core Metadata**

**Název článku hypertextově aktivní** - zobrazení citace a plného textu bez odkazů

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

#### Pracovní formáty

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

#### B.1.4 Metadata (viz Příloha F4 b,e)

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

#### Příklady uložených metadat v plném textu

##### Př.1

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

##### Př. 2

<b>Název:</b>	Rafinovaný odraz skutečnosti
<b>Podnázev:</b>	Na pultech se objevil další titul singerovské řady
<b>Hlavní autor:</b>	Hana Ulmanová
<b>Zdroj:</b>	Respekt
<b>ISSN:</b>	0862-6545
<b>Roč.:</b>	12, č. 41 (8.10.2001), s. 23
<b>Rubrika:</b>	KULTURA

Předmět. ktg.:	literatura: americká literatura
MDT:	821.111(73)-31, (070.447)
Osoba jako předmět:	Singer, Isaac Bashevis, 1904-1991
Dílo jako předmět:	Stíny nad Hudsonem (kniha)
Téma jako předmět:	anglicky psaná literatura
	americká próza
	spisovatelé-Spojené státy-stol. 20.
Typ dokumentu:	recenze

### B.1.5 Tisk, stahování a export plných textů a metadat - přístup

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

V r. 2002 byla vyvinuta aplikace pro stahování a export metadat a plných textů, která vychází z již vyvinuté technologie v Anopressu. Aplikace umožňuje stáhnout a export příslušných plných textů a metadat na lokální počítač v zip souboru v několika formách: HTML, RTF formát, textový tvar, UNICODE - vše se zvýrazněním klíčových slov a bez, XML formát, HTML formát se souhrnem. Je možno zobrazit detailní průběh zpracování . Exportovat plné texty a metadata mohou zatím jen registrovaní vybraní uživatelé. V r. 2003 byly ověřena její funkčnost.

Postup:

1. Vyhledané články nejprve označit v Seznamu výsledků (levé okno)
2. Odrážka „Stáhnout označené“
3. Volba formátu výstupního souboru“
4. Odrážka „Přípravit 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.

V roce 2003 omezen přístup pro externí uživatele na 1 den (důvod viz výše).

Otázky zpřístupnění externím uživatelům závisí na vyřešení legislativně právních otázek event. ekonomických (viz výše). **Externím uživatelů jsou běžně zpřístupňována metadata. Plné texty a metadata jsou zpřístupňována interním uživatelům NK**

**Stahování a export metadat a plných textů je také plně funkční v rozsahu testování několika uživateli.**

### **B.1.6 Administrace (viz Příloha F 3)**

**V roce 2002 byla vyvinuta aplikace pro administraci databáze ANL FULL a portálu.**

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

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

**V r. 2003 byla částečně implementována zcela nová aplikace pro opravy hesel a metadat v plných textech, která je zabudovaná přímo do linky WWW TTDE (viz též výše). Opravy dat jsou tak mnohem pružnější než v administraci. Aplikace umožňuje vyhledání plných textů z rejstříků, jejich zobrazení a zatím částečné načtení do pracovního listu k opravě. Po korekci metadat je plný text odeslán do báze, plný text je přepsán a zrušen. Aplikace se dále ladí.**

### **B.1.7 Standardizace - UNIMARC, DUBLIN CORE, HTML, XHTML, XML, propojení záznamů s plnými texty (viz Příloha F3, F4b,e)**

**Struktura bibliografických dat respektuje formát UNIMARC a knihovnická pravidla AACR2 v oblasti jmenného popisu.** V r. 2002, 2003 probíhají práce na konverzi UNIMARC - MARC 21. Byl nadefinován návrh minimálního záznamu pro M21. **Jmenný popis bude muset do budoucna reflektovat pohyb v oblasti standardizace ( metodika popisu článků ve formátu UNIMARC - vychází z AACR2R, ISBD /CP; v budoucnu větší harmonizace AACR2R a ISBD, ISBD (ER), ISBD(S) - ISBD (CR)).** Jak se dotkne popisu článků koncept FRBR - Funkční požadavky na bibliografické záznamy (**Functional Requirements for Bibliographic Records**) - **4 úroňový model manifestace díla - dílo, exprese, manifestace, jednotka) ukáže budoucnost.**

V oblasti **věcného popisu** se používá aktualizovaná verze **MDT-MRF** . Verbální věcný popis obsahuje **předmětové kategorie**, které zasazují dokument do širších souvislostí v rámci databáze z hlediska obecných témat, jež by se měla sblížovat s tématy systému TOPIC eventuelně s automatizovanou tvorbou kategorií. Předmětové kategorie do jisté míry konvenují **metodě konspektu** aplikované pro popis a mapování fondů. Od konspektu budou odlišeny identifikátorem v rámci stejného pole 615. Dále se používají **klíčová slova**, která jsou dále částečně řízena a **předmětová hesla**. V rámci linky je zajištěno také propojení na soubory autorit (ne zatím zcela pro

katalogizátora pohodlné, pracuje se na propojení on-line - viz dříve) Automaticky je generován **souhrn článku**, automaticky lze také generovat klíčová slova, která však mají v současném stádiu "počítačovou formu". **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 doplňované **statické adresy** volně dostupných dokumentů a WWW. Dále se automaticky generuje **URN** (do kterého je zabudováno identifikační číslo plného textu přidělené v Anopressu) a z velké části pole **LKR** určené k propojení se zdrojovým dokumentem.

**Dublin Core** obsahuje 14 z 15 definovaných údajů **viz Příloha F 4 b**. 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 a prvky ANL Core užívané pro vyhledávání.

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. bibliografické záznamy se importují do ALEPHu, jsou propojeny na plný text v systému TOPIC. Zároveň jsou metadata a plné texty indexovány v systému TOPIC. Navíc se propojují záznamy báze ANL s volně dostupnými texty na internetu.**

### **Souhrn za r. 2000-2003**

#### **Linka bibliografického zpracování z plných textů (TTDE)**

**Souhrn za r. 2000-2003** : na serveru **full.nkp.cz** jsou přístupné **plné texty článků** vydané v letech 1990-2003 a získané v rámci tohoto grantu a grantu Propojení analytických záznamů v letech 1999-2003, 11. listopad v počtu cca **115 000**. **Se záznamy v ANL je propojeno cca stejný počet. Možný rozdíl** je dán tím, že v počátcích projektu byly do báze importovány zpětně plné texty a následně programově nebo ručně propojovány s bibliografickými záznamy. Ne vše chna metadata uložená v záznamech ANL byla totožná s metadaty uloženými v plných textech z Anopressu a nešla tudíž programově propojit.

**(v lince bylo zpracováno v letech 2001-2002, 15. listopad celkem cca 24000 plných textů s metadaty , v r. 2003, listopad 13 000 plných textů) .**

Průměrně měsíčně v lince uloženo: 1000 záznamů.

**Statické linky v báze ANL (přibližná a pohyblivá čísla, protože některé adresy jsou nestálé - např. Psychologie Dnes a Učitelské noviny)**

V r. r. 2000-2002, říjen propojeno cca 3000 záznamů s volně dostupnými texty na Internetu. Některá propojení spadla. Do 21.10. 2003 bylo propojeno cca 2181 záznamů Jsou propojené také některé záznamy s plnými texty zákonů.

V Odboru knihovnictví jsou ještě propojována knihovnická periodika.

**V současné době je staticky propojeno cca 6 000 URL. Podle posledního URL testu cca 300 neplatných.**

**B.1.8 Souborná databáze ANL Kooperativního systému článkové bibliografie (KOSABI), metodika a organizace, ČNB (viz Příloha F 1, F5)**

#### **B.1.8.1 Kooperace, excerpční základna**

V roce 2003 pokračovalo oddělení ve zpracování záznamů pro bázi ANL(ALEPH). Kooperující instituce - **8, resp. 7 krajských knihoven, MZK a 3 odborné knihovny** - přispívají do souborné databáze (do báze **nepřispívá krajská knihovna v Plzni**, přechází ze systému KIMS na ALEPH, knihovně byla poskytnuta definice báze ALEPH pro ANL). V budoucnosti se mají součástí KOSABI stát i 4 další krajské knihovny, se kterými byl navázán kontakt (Pardubice, Zlín - systém KP-sys, Karlovy Vary a Havlíčkův Brod systém Lanisus, resp. Clavius). Do báze ANL přispěla dávkově v rámci experimentu i ČGS.

V roce 2003 byly pravidelně dodávány záznamy z těchto knihoven: Kladno (RAPID), Ostrava (T-Series), SPKK (ISIS), STK (ISIS), ÚZPI (ISIS), Liberec, (RAPID), České Budějovice (T -Series).

Ukládání online do Alephu - Moravská zemská knihovna v Brně.

V r. 2003 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 spolupracovalo oddělení článkové bibliografie s krajskou knihovnou v Ostravě a Českých Budějovicích. Byla 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** bylo třeba dopracovat převod do UNIMARCu (téměř dotažen je UNIMARC pro krajskou knihovnu ve Zlíně, kterou následuje krajská knihovna v Pardubicích; v listopadu 2003 ale krajská knihovně ve Zlíně zaslala vzorek záznamů v KP- Winu, které se patrně a od téměř dotaženého KP-sysu budou lišit ). Knihovny pracující v Rapidu spolupracují též bez problémů. Téměř dotažen je UNIMARC pro **Lanisus** (Karlovy Vary). U systému **Clavius** je třeba UNIMARC ještě dopracovat. **S nově jmenovanými a konstituovanými krajskými knihovnami v**

Pardubicích, Havlíčkově Brodě, Zlíně a Karlových Varech byla navázána **dobrá spolupráce**. Byla dohodnuta budoucí excerpční základna.

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 proběhla jednání o možnostech** delimitace duplicit ve zpracování o možné kooperaci týkající se excerpce titulů a omezení duplicit při zpracování zejména s NKČR. Zatím jsou určeny tituly, které by přicházely v úvahu. Systém používá ve věcném popisu klíčová slova, bylo zajištěno propojení na soubory autorit.

**Součástí širšího kooperačního systému článkové bibliografie** po linii oborové je zpracování článků lékařské a zdravotnické literatury **Národní lékařskou knihovnou, které nejsou součástí ANL.**

**V navrhovaném pokračujícím projektu v rámci programu 1N MŠMT je zpracován přechod odborných knihoven z ISIS na UNIMARC a úzká spolupráce s Národní lékařskou knihovnou.**

Tyto knihovny poskytly pro projekt 1 N některé plné texty, na kterých je možno dále vyvíjet pojmové vyhledávání v systému TOPIC, resp. K2. V projektu je také zpracován vývoj a realizace linky zpracování bibliografických záznamů z plných textů v Národní lékařské knihovně a nový způsob produkce Bibliografia Medica Československa a báze citační rejstřík CR-CZ. Podobně ÚZPI má realizovat zpracování bibliografických záznamů článků některých titulů v návaznosti na již existující redakční systém. ČGS má v tomto projektu postupně přejít na pravidelné přispívání do báze ANL ze systému WIN ISIS.

**V bázi ANL bylo je v r. 2002, k 15.11. cca 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 bázi ANL bylo je v r. 2003, k 15.11. cca 807 000 záznamů, za rok 2003 přibylo do báze cca 60 000 záznamů, z toho cca 35 000 zpracovaných NKČR, ve spolupracujících institucích cca 25 000 zpracovaných záznamů. Počty jsou přibližné.**

**V roce 2003 je pravidelně aktualizována excerpční základna.**

**Počet zpracovávaných titulů: ANL - cca 212 v NKČR, 263 v regionech (všechny krajské knihovny, reálně kooperující -208), 122 v odborných knihovnách, celkem cca 597 titulů (zatím reálně 542); ANL FULL 30 titulů, běžně zpracovávaných zatím 17 v NKČR. Došlo k přesunu některých titulů na nově konstituované krajské knihovny, které začnou pravděpodobně přispívat do báze v r. 2004.**

#### **Excerpční základny jednotlivých kooperujících institucí pro bázi ANL**

Excerpční základna byla tento rok v pohybu vzhledem k novému státoprávnímu uspořádání.



Řada titulů přešla na krajskou knihovnu v Havlíčkově Brodě (z Brna, Budějovic, Hradce Králové).

Krajská knihovna Karlovy Vary se zavázala excerpovat bývalé plzeňské tituly. Nově konstitované knihovny budou excerpovat regionální přílohy deníků Mladá Fronta Dnes a Právo.

Postupně přesouvané tituly se přestávají postupně zpracovávat v původních knihovnách a přecházejí na nově konstituované knihovny.

**Národní knihovna České republiky (ALEPH) - Oddělení analytického zpracování:** celkem 195 seriálů z toho 6 deníků, cca 11 týdeníků, 6 čtrnáctideníků, 31 dvouměsíčníků, 4 čtvrtletníky, ostatní periodika s menší periodicitou ; dále nepravá periodika - ročenky, nepravidelně vycházející periodika, sborníky - počet pohyblivý (5 titulů převzala STK).

NKČR - Odbor knihovnictví: 17 titulů. **Moravská zemská knihovna v Brně (ALEPH):** celkem cca 49 titulů, 1 deník, vlastivědné časopisy, zpravodaje, sborníky a nepravá periodika, část excerpce převzala Krajská knihovna Vysočiny (cca 4 tituly postupně přebere krajská knihovna Vysočiny, Brno přestalo excerpovat).

**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 cca 17 titulů, z toho 7 deníků, postupně delimitace titulů po dohodě s Krajskou knihovnou v Pardubicích, Krajskou knihovnou Vysočiny, Krajskou knihovnou v Liberci (4 tituly přešly na Krajskou knihovnu Vysočiny).

**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 (ALEPH):** celkem 22 titulů, z toho 5 deníků, 5 týdeníků, 6 měsíčníků a další periodika (zatím neposílá).

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

**Nově konstituované krajské knihovny - příprava kooperace Krajská knihovna Františka Josefa Bartoše ve Zlíně (KP-sys): pro ANL cca 15 titulů, jinak 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 . Stanovena excerpční základna s ohledem na nové státoprávní uspořádání: cca 2 deníky, 1 týdeník, 2 měsíčníky, 2 ročenky, 7 dalších seriálů.**

**Krajská knihovna Karlovy Vary (LANius):** celkem cca 16 titulů, 5 deníků, 3 týdeníky, další seriály, zpracovává články pro LANius.

**Krajská knihovna v Pardubicích (KP-sys):** celkem cca 9 titulů, výběrově vlastivědné a místní zpravodaje.

**Krajská knihovna Vysočiny (Havlíčkův v Brod, Clavius):** celkem cca 15 titulů, 4 deníky, doporučeno nejdříve zpracovávat tituly týkající se Havlíčkobrodsko, postupně přibrat další, nutná delimitace s MZK, krajskou knihovnou v Budějovicích a Hradci Králové, nedostatečné personální obsazení.

#### **Kooperující specializované knihovny**

**Státní technická knihovna (ISIS):** celkem 47 odborných titulů (některé tituly převzala od NKČR, která po redukcích nemůže zpracovávat)

**Ústav zemědělských a potravinářských informací (ISIS) :** celkem 25 odborných titulů

**Ústav pro informace ve vzdělávání - Státní vědecká knihovna Komenského (ISIS)**  
: celkem cca 50 titulů

**Česká geologická služba (ISIS):** odborná periodika v rámci experimentu

**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).**

#### **B.1.8.2 Popis, standardizace (viz Příloha F1, F3, F4)**

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

Byly řešena také problematika řízeného slovníku pro věcný popis. V kooperačním systému článkové bibliografie se používal ještě za dob ISISu tzv. částečně řízený slovník klíčových slov, tj. byly řízeny nadřazené pojmy a většinou první podřazené slovo k nim. Kromě toho se klíčová slova vyvinula z předmětových hesel a zachovávala částečně i jejich strukturu, tj. 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 v r. 2003 KOSABI používán a do jisté míry modifikován s ohledem na praktické potřeby. Slovník se musí v samotné bázi ANL zredigovat. V budoucnu budou tyto kategorie převedeny do pole 615, k odlišení od Konspektu s 2. indikátorem 9.

## **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 tezaurů, v posledním období pouze pole předmětové kategorie v terminologii Tinlibu, resp. T-Series (něco jiného jsou předmětové kategorie UNIMARCu, pole 615). V roce 2002 se přistupuje k popisu do jemněji strukturovaného formuláře, podobně jako v Ostravě.

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

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

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

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

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

**V r. 2003 se ladila struktura záznamů vzhledem k formátu UNIMAR ve spolupráci s krajskými knihovnami v Pardubicích a Zlíně (KP-sys), v Karlových Varech a Havlíčkově Brodě (Lanius, resp. Clavius). Zároveň byla delimitována excerpční základna vzhledem k novému státoprávnímu uspořádání a vstupu těchto knihoven do reálné kooperace (viz výše).**

**V r. 2003 záznamy ANL respektují metodický materiál Záznam pro soubornou databázi : UNIMARC a Záznam pro soubornou databázi : Výměnný formát. Dle možností je aktualizována pracovní verze příručky pro zpracování článků v UNIMARCu - Metodika popisu článků ve formátu UNIMARC a vystavena na**

**Internetu s názornými příklady.** Většina knihoven vybavuje záznamy předmětovými kategoriemi, které jsou jednotícím prvkem souborné databáze. Oddělení analytického zpracování přistupuje k **intenzivnějšímu využívání souborů autorit v NK , korekturám báze ANL a připravuje data pro soubory jmenných a věcných autorit. Také kooperující instituce postupně používají národní soubory autorit.**

V rámci ANL i ANL FULL jsou formulovány základní principy věcného popisu, aplikovaného při zpracování článků v NKČR. **Pořadí kategorií, 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, Olomouci, Brně, Liberci. 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 je možno je užívat paralelně s konspektem.

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.

### **Propojení na zdrojový dokument**

U 5 seriálů (Týden, Ekonom, Reflex, Respekt, Právo) zpracovávaných v lince je generována v azba LKR (nahoru) k propojení se zdrojovým dokumentem v bázi NKC. Tato vazba je funkční.

Do budoucna lze uvažovat o propojení na zdrojový dokument přes ISSN pomocí OpenUrl. Názvy zdrojových dokumentů v bázi ANL je třeba sjednotit (někdy je problém určit jednoznačně název, protože pod určitým názvem je seriál uložen v regionální knihovně, pod určitým v NKC apod.). Řešením je vazba na ISSN. Existuje báze ISSN budovaná v STK, hypoteticky se mohou lišit i ISSN v NKC, SK ČR a ISSN. Někdy problematický je zápis zdrojových dokumentů tzv. nepravých periodik a

jejich sjednocení v bázi ANL. V bázi ANL probíhají v současné době korektury.

### **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.

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

### **B.1.8.4 Česká národní bibliografie - řada Články v českých novinách, časopisech a sbornících na CD-ROM (viz Příloha F1)**

V roce 2003 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 proběhlo, bude se zřejmě ještě ladit.
- 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.) - zjistit možnosti.

Spolupráce se společností Anopress a event. navázání spolupráce s dalšími

- nakladateli/vydavateli. V r. 2003 jednáno předběžně s nakladatelstvím **Economia**, v rámci excerptní základny je možno určit skupinu titulů z hlediska této spolupráce. Je možné uvažovat o spolupráci s vydavateli titulů, které jsou dotovány MKČR. Pro návrh projektu rámci 1N MŠMT byly poskytnuty některé plné texty z oblasti zemědělství, ekonomie, náboženství, lékařství.

#### **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).
- Korektury báze ANL, zejména zdrojových dokumentů.
- Konverze UNIMARC - MARC 21 (ladit), pravidla pro popis článků vzhledem k tomuto formátu.
- Případně otestovat linku v institucích spolupracujících v KOSABI.
- Testovat metody, popř. navázat na metody získávání a zpracování elektronických plných textů (projekt Webarchiv).
- Sledovat vývoj harmonizace AACR2R, ISBD, aktualizované a nové standardy
- **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 aj. odborných seriálových informačních zdrojů (v rámci integrovaných knihovnických systémů), které zatím nejsou dostupné na Internetu.**

**II. Zpracování odborných seriálových informačních zdrojů, které jsou dostupné na Internetu v lince aut omatické extrakce/indexace.**

**II. Zpracování článků z deníků a časopisů na základě akvizice z databanky TamTam (Anopress) v lince automatické extrakce/indexace.**

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

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

**Dodržování zásad výběru článků.**

#### **Organizace zpracování článků z hlediska typů institucí:**

Vzájemná kooperace mezi následujícími sítěmi knihoven:

- NK, krajské knihovny, základní knihovny
- Specializované knihovny

- Vysokoškolské knihovny
- Vědecké instituce

### **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.
- V r. 2003 podán projekt týkající se kooperačního systému a zpřístupnění odborných zdrojů pro VaV v rámci programu 1N, zapojení báze ANL FULL do JIB přes Z39.50 a Open URL, migrace TOPICu na systém K2.

### **Legislativně právní otázky zpřístupnění plných textů**

Vyřešení autorskoprávních 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. Koncem roku 2002 se zdálo, že zpřístupnění ANL FULL v rámci konzorcia knihoven zdarma nebude problematické. Anopress souhlasil. Později se však ukázalo, že by mohly nastat některé problémy. Jasně je zpřístupnění ANL FULL interním uživatelům knihovny. Národní knihovny je koncový uživatel pro Anopress.

Otázky zpřístupnění ANL FULL zdarma v rámci konzorcia bude předmětem jednání v roce 2004, podobně jako zpřístupnění báze externím uživatelům.

V r. 2004, v posledním roce projektu by měla být uzavřena smlouva mezi NKČR a Anopress o zpřístupňování báze ANL FULL interním a pokud to půjde, i externím uživatelům. Ke stávající smlouvě týkající se Konzorcia Anopress by měl být učiněn dodatek.

### **Diferencované zpřístupnění plných textů**

Jasně vymezení a ošetření přístupu k plným textům, které jsou zpřístupňovány volně, přes konzorcia, interním a externím uživatelům institucí a způsob plateb (paušál, kredity, apod.).

### **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 byly plánovány na cílový počet 11 pracovníků v r. 2003. Tento rok poslední redukce místa zatím neproběhla, současný stav pracovníků je 12.

- Vývoj systému je po finanční stránce zajištěn do roku 2004 díky podpoře MKČR. V r. 2003 je podán návrh projektu v rámci programu 1N MŠMT, který navazuje v plné míře na stávající projekt týkající se KOSABI. Týká se zejména systému TOPIC a jeho vývoje, zapojení báze ANL FULL do JIB přes Z39.50 a OpenUrl, převedení odborných knihoven z ISISu do UNIMARCU (kromě ČGS - zde WIN ISIS). Další projekt - výzkumný záměr podává Národní knihovna jako komplexní projekt, je v něm částečně zahrnuto i analytické zpracování a týká se zejména věcného popisu a práce v ALPEHu.

### **B.1.10 Management Kooperačního systému článkové bibliografie, aplikace MNG KOSABI (viz Příloha F5)**

#### **B.1.10.1 Management KOSABI - systémový pohled**

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

**Vznik, správa a údržba ANL FULL byla popsána částečně výše, lze říci, že je stádiu rutinního provozu v NKČR.**

**Aplikace pro správu a údržbu báze ANL je ve stádiu poloprovozu v NKČR.**

- **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), v budoucnu WWW TTDE**  
**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 1 den - registrace a přihlášení**  
**vybraní uživatelé - export dat a stahování dat - registrace a přihlášení**



**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 poloprovozu pouze v NKČR (zatím) 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)**. Dále probíhají práce na **aplikaci pro bázi titulů z hlediska možného dialogového zpracování této báze**.

Upload záznamů je zatím testován v NKČR, v r. 2004 proběhne test s kooperujícími institucemi.

**Zbývá vyřešit přebírání dat získaných od kooperujících institucí z příslušného adresáře do ALEPHu. Rovněž tak je třeba, v případě potřeby, ještě doladit importy pro KOSABI u 4 nově vzniklých krajských knihoven.**

**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 excerpční základny)**
- 4) Návrh a implementace jednotného www rozhraní pro management KOSABI pro přístup do dílčích aplikací.**

**Management KOSABI (stručná charakteristika - html stránky):**

**Vstup do aplikace**

**Báze titulů**

**Upload předávání dat do KOSABI**

**Import - statistiky (včetně exportu do Alephu)**

**Klíče (kontrola na duplicitu záznamů)**

- 5) Upgrade serveru ANL: operační systém Linux SuSE 7.3**

**Nová verze: Oracle 9.2**

**Java j2sdk1.4.0-01**

**Servlety jakarta-tomcat 4.0.4**

**Seznam prací v r. 2003**

- 1. Testování uploadu souborů získávaných od institucí spolupracujících v KOSABI**
- 2. Specifikace úprav dat na výstupu z aplikace (export pro ALEPH)**
- 3. Test URL adres**

Byla vypracována nová aplikace pro testování platnosti URL adres.

  - a. Převzetí dat z Alephu a naplnění sql tabulek pro test a statistiky.  
Provádí se automaticky, momentálně 1x týdně.
  - b. Vlastní test URL adres. Spuštění testu prováděno ručně.  
Při spuštění možno definovat množinu testovaných adres:  
pouze chybné adresy  
pouze ok adresy  
adresy vyjma full  
pouze adresy full
  - c. Statistiky  
Přehled aktualizace báze adres, ( tj. přebírání dat z Alephu )  
Výsledky provedených testů, včetně historie.
- 4. Test na duplicitu**

Pro každý záznam v bázi SKA je vytvořen klíč.  
Vznikne přehled o klíčích, které se opakovaně vyskytují v různých záznamech.
- 5. Jištění serveru full.nkp.cz**

Vzhledem k tomu, že jištění serveru pouze na pásky je nedostatečné, je zajištěno zrcadlení báze na serveru ANL. Na serveru anl byly vytvořeny skripty pro automatické jištění vybraných adresářů na serveru full.
- 6. Import**

Program pro import upraven: - do baze SKA (KOSABI ANL) ukládány všechny záznamy, nezávisle na výsledku vstupních testů  
- test na UNIMARC doplněn o test na kritické chyby (u každého záznamu je rozlišeno:  
test UNIMARC OK nebo ERROR  
test na kritické chyby OK nebo ERROR  
- úpravy řešící chyby ve vstupních datech  
Např. doplňování pole 005
- 7. Export**

Program pro export doplněn o různé úpravy dat vzhledem k následnému importu do Alephu.  
Např. Pole FMT, LDR, 100, 660, 461/463, 7xx, 600, 801c

## **8. Báze titulů opravy**

Vytvořen dialog program pro úpravy záznamů v bázi titulů aplikace v JDeveloperu)

## **9. Opravy dat určených pro import do Alephu**

Vytvořen dialog program pro úpravy záznamů (aplikace v JDeveloperu)

V roce 2004 se budou programy testovat. V možném návazném projektu v rámci 1N je uvažován vývoj aplikace založený na JDeveloperu pro opravu dat před importem do Alephu.

### **B.1.10.3 MNG KOSABI - 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 záznamů - 1. etapa ( chyby v datech, které lze programově ošetřit, např. tvrdé mezery, tagy ve vazebních polích )  
test UNIMARC
- uložení do baze SKA
- statistika zpracování

#### **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**

### **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
- Statistiky: 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). Názevové údaje:  
200a. 200i [200b] : 200e  
Zdroj katalogizace: [801b].[801c]  
Sigla vlastníka: [910a].[910a]  
Region: [952a].[952a]  
Údaje o excerptu: [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.

#### **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 u pgrade serveru ANL: operační systém Linux SuSE 7.3, verze Oracle 9.2.**

**V r. 2003 nedošlo k žádné inovaci SW ani HW. Došlo k havárii serveru, Předcházela porucha služby publikování - příčina dosud neobjasněna, nebyly shledány problémy ze strany instalace serveru.**

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

V návrhu projektu pro 1N je požadavek na na disk pro zrcadlení báze ANL full, upgrade procesoru, desky, paměti. DVD jednotka pro jištění báze, neomezená licence JDeveloperu - tyto nástroje jsou také předmětem návrhu projektu v rámci 1N.

#### **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. 2003 spočívá v realizaci metod navržených v r. 2000, 2001, 2002 ve stádiu rutinního provozu v případě budování báze ANL FULL v NKČR. Aplikace pro příjem a zpracování dat pro bázi ANL je v poloprovozním stádiu pouze v NKČR vzhledem k nutnosti průběžného ladění v závislosti na různých systémech spolupracujících institucí a zařazování nových knihoven do systému.**

**Přínos projektu v r. 2003 spočívá v praktické integraci elektronických zdrojů do služeb knihovny pomocí technologicky nejvyspělejších 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.**

**Zdá se, že předmětové kategorie používané v kooperačním systému budou možno po důkladné redakci a revizi doplnit metodu konspektu. Pro pojmové vyhledávání byly aktualizovány topiky vypracované v roce 2000, 2002 doplněny o nové definice topiků.**

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

Koncem roku mají ještě probíhat práce na **konverzi UNIMARC M21**. Zatím je navržen minimální záznam pro soubornou databázi.

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

Byla navržena a částečně realizována internetová podoba linky pro zpracování bibliografických záznamů z plných textů s propojením této linky na rejstříky báze ANL FULL.

V současné době probíhají práce na zprovoznění komponenty pro propojení linky na autority v rámci Z39.50.

V r. 2003 byl navázán poměrně intenzivní kontakt s nově konstituovanými krajskými knihovnami, ladil se UNIMARC a excerpční základna těchto institucí.

Aplikace pro MNG KOSABI byla rozšířena o významný test URL adres, probíhaly práce na zkvalitnění www stránek aplikace, dále se ladily programy pro příjem, zpracování a export dat do ALEPHu. Byl vytvořen dialog program pro editaci báze titulů kooperačního systému a bibliografických záznamů přispívajících institucí v JDeveloperu.

V r. 2003 je udržován **portál** pro zpřístupnění volně dostupných elektronických periodik na Internetu, která jsou členěna na principu regionů a oborů.

Stále realističtější je **možnost vývoje systému v budoucnu v rámci několika zpracovatelských subsystémů podle typu dokumentů a přístupnosti plných textů**. Agentura 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 i ekonomický význam.

Další přínos spočívá ve zvýšení **uživatelského komfortu** - v nalezení příslušného článku z novin či časopisu (v budoucnu snad i statě ze sborníku) v elektronické formě. Navigace k primárním dokumentům patří k základním trendům. Byla otestována aplikace k zajištění registrovaného přístupu externích uživatelů k plným textům. Tito uživatelé se mohou zkušebně zaregistrovat na dobu 1 dne a vyhledávat v plných textech.

Pro uživatele NK jsou plné texty přístupné. Uživatelé je mohou prohlížet, tisknout, nikoli však zatím kopírovat na disketu.

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

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. V úvahu přichází automatizovaná kategorizace dokumentů. Projekt se dále zaměří na využití možnosti inteligentního vyhledávání systému TOPIC a interakci mezi systémem a uživatelem a systémem a indexátorem.

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 (metodicky) reorganizace zpracování národní bibliografie se vyvíjí např. v Austrálii.

Aplikace pro management KOSABI se dále ladí, byl navržen test na URL adres.

Výsledky projektu byly prezentovány na konferenci Infos 2003 a zasedání sekce pro bibliografii SDRUK, informace byly poskytnuty sekci IFLA, zabývající se problematikou novin. Aktuální tematikou je zpřístupňování informací, obsažených v novinách.

### 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 CNB, budování digitální knihovny), u báze ANL FULL **rutinní provoz v NKČR**
- **Fungující linka zpracování bibliografických záznamů z plných textů na lokálních stanicích v NKČR - TTDE, návrh a částečná realizace internetové verze linky - WWWTTDE (nové)**
- **Propojení internetové verze linky na autority a rejstříky ANL FULL (nové)**
- **Propojení bibliografických záznamů s elektronickými zdroji na základě dynamických i statických URL adres, kombinace pojmového vyhledávání s metadatami 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**
- **Zpřístupnění plných textů v rámci moderně koncipované báze ANL FULL na základě vyspělých technologií pro zpřístupňování elektronických dokumentů, praktická implementace Dublin Core v rámci HTML, XHTML, XML v kvalifikované a nekvalifikované formě. Definování Anl Core pro zpřístupnění jemnějších údajů analytického popisu**
- **Údržba strukturovaného portálu volně dostupných zdrojů na Internetu s popisem těchto zdrojů**
- **Možnost využití metody vypracované pro linku automatické indexace pro zpracování informací a toku informací mezi autorem, nakladatelem, informační agenturou, bibliografickou agenturou/knihovnou**
- **Možnost opravy metadat v rámci WWWTTDE (*komplet nové*)**
- **Možnost stahování a export plných textů a metadat**
- **Další aktualizace topiků (*nové*)**
- **Analýza problematiky automatické indexace a selekční úplnosti topiků**

**Hypotézy na základě analýzy problematiky automatické indexace: orientace spíše na**

- **automatickou extrakci dat, clustering, automatické abstrahování event. klasifikaci a podporu indexátorů a automatizovanou kategorizaci s využití dalších možností inteligentního vyhledávání systému TOPIC, než na automatické přiřazování věcných termínů. Zohlednit inter akci systému s uživatelem a indexátorem**
- **Ověření hypotézy, že intelektuální indexace věcná je zatím nezastupitelná**
- **Posílení vazby v rámci KOSABI se 4 nově konstituovanými knihovnami - ladění UNIMARCu a excepční základny**
- **Testování a další vývoj aplikace pro MNG KOSABI, zlepšení interface pro MNG KOSABI  
Údržba báze titulů**
- **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**
- **Pravděpodobný vývoj KOSABI - nastínění perspektiv KOSABI a možný vývoj jeho subsystémů z hlediska typu zpracovávaných dokumentů**
- **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)
- **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** ( řešitelce projektu nabídnuto členství v **IFLA Section for Newspapers**, což je určitým uznáním práce, kterou tým pracovníků, kolem projektu, vykonává)

## C Navrhovaná část

### C.1 Výsledky řešení

- **Praktická realizace nástrojů pro optimalizaci integrace a správy heterogenních dat v rámci Kooperačního systému článkové bibliografie. Vznikl tak nástroj pro poloautomatickou přípravu dat pro bibliografickou databázi založenou na UNIMARCu 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 a rutinní provoz **aplikace pro získávání dat TAMTAM SPECIAL NK (TTSNK).**
- Realizace a **rutinní provoz 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) - lokální pracovní stanice**
- Návrh a částečná realizace aplikace pro automatické generování bibliografických údajů do analytických záznamů a **metadat Dublin Core a jejich umístění do plných textů i mimo ně - WWW TamTam Data Extractor (WWW TTDE) - internetová verze TTDE (vývoj aplikace hrazen z projektu Propojení analytických záznamů s plnými texty, koncepce , zadání a testování v rámci tohoto projektu) s propojení linky WWW TTDE na rejstříky báze ANL FULL, příprava a aplikace na propojení WWW TTDE se soubory autorit). Vývojové práce budou probíhat ještě do konce roku, podobně jako testování aplikace , které probíhá v rámci popisovaného projektu.**
- Aplikace **pro indexaci dat do plnotextové databáze ANL FULL - rutinní provoz**
- 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 - rutinní provoz**
- **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**
- **Údržba a doplňování strukturovaného portálu pro volně přístupné zdroje na WWW v regionální, tematické a institucionální struktuře - rutinní provoz**
- **Registrace a autentifikace uživatelů - rutinní provoz**

- Aplikace pro administraci báze ANL FULL (opravy dat, statistiky) - rutinní provoz
- Návrh a implementace aplikace pro stahování a export plných textů a metadat pro vytypované uživatele - rutinní provoz v rámci možného rámce s ohledem na autorská práva
- Testování, ladění a vývoj aplikace pro management Kooperačního systému článkové bibliografie (MNG KOSABI), jednotné rozhraní pro aplikaci - poloprovozně funkční upload a testování dat kooperujících institucí, export dat do adresáře pro ALEPH a adresáře určeného k dalším opravám - odzkoušeno v NKČR.
- Nová aplikace pro test URL adres.
- Údržba Báže titulů na serveru ANL
- Standardizace týkající se popisu článků v UNIMARCu a AACR2R
- Zabezpečení KOSABI po stránce metodické i organizační, navázání poměrně úzké spolupráce se 4 novými krajskými knihovnami - ladění UNIMARCu a delimitace excerpční základny vzhledem k novému státoprávnímu uspořádání
- Další studium a analýza problematiky automatické či poloautomatické indexace věcné a potvrzení hypotézy orientovat se spíše na automatickou extrakci a využití inteligentního vyhledávání v TOPICu s automatizovanou tvorbou předmětových kategorií, dále pak orientace na skloubení pojmového vyhledávání s intelektuální indexací věcnou a zabudování věcných termínů přímo do topiků, než na automatické přiřazování věcných termínů
- 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 (hrazeno z projektu Propojení analytických záznamů s plnými texty)
- Důkladná analýza zpracování článků v zahraničí (metody, organizace) a ověření, že podobné a srovnatelné metody se vyvíjejí v rámci tohoto projektu
- 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
- Prezentace výsledků projektu na konferenci Infos 2003 (Slovensko), zasedání sekce pro bibliografii SDRUK (Hradec Králové), informace pro IFLA Newspaper Section, Berlín 2003 (nabídka členství v této sekci)

## 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 plně 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 v budoucnu.

Spolupráce s Anopressem mohla být navázána díky zmíněným projektům a NK mohla navázat na moderní technologii, kterou tato firma užívá a dále rozvíjí. 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** 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.

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 2004 a dále zajištění návazného projektu v rámci 1 N programu MŠMT na léta 2004-2008 (projekt se týká zpracování a zpřístupnění odborných informacích zdrojů včetně plnotextových moderními metodami a kooperace zejména s odborně orientovanými institucemi v Kooperačním systému článkové bibliografie, zapojení serveru full.nkp.cz do JIB prostřednictvím Z39.50 a OpenUrl)
- Projekt Souborná databáze Kooperačního systému článkové bibliografie končí v r. 2004. Bude třeba zajistit financování zdrojů získávaných z Anopressu pro provoz linky buď z rozpočtu NK, nebo v rámci Konzorcia Anopress nebo dalším projektem
- 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
- Další rozvoj spolupráce s novými čtyřmi krajskými knihovnami a reálná kooperace v rámci Kooperačního systému článkové bibliografie
- Užší spolupráce s Národní lékařskou knihovnou v rámci na základě návrhu projektu v programu 1N a nové metody zpracování BMC a CR-CZ (Český citační rejstřík)

- **Personální zajištění projektu z hlediska počtu pracovníků oddělení analytického zpracování v NK i v rámci kooperačního systému**
- **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
- **Opravy báze ANL a ANL FULL. V ANL zejména sjednocení názvů zdrojových dokumentů a ISSN**
- **Zkvalitněné ukládání údajů v rámci NKČR i kooperujících institucí**
- **Důkladné posouzení a další ladění dosud existující konverze UNIMARC M21 pro články**
- **Průběžné sledování vývojových trendů ve standardizaci a jejich praktická aplikace**
- **Přísnější výběr článků k indexaci deníků s cílem vyloučení subjektivního faktoru při excerpci deníků**
- **V budoucnu zaměření na zpřístupnění odborných plných textů**
- **Funkčnost propojovacích vazeb a možnosti propojování přes OpenUrl v systému ALEPH a Metalib, resp. JIB 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 NKČR či Souborného katalogu CASLIN, dále pak připojení článků - recenzí k recenzovaným dokumentům v rámci těchtoází. Navigační systémy na úplné obsahy čísel seriálů mimo záběr abstraktních databází jsou v zahraničí zcela běžné. Jedním z předpokladů tohoto propojení jsou i kvalitní, pokud možno bezchybná data uložená v bázi ANL. Oprava dat uložených v databázi je aktuálním problémem.
- **Ošetření autorskoprávních aspektů, event. ekonomických aspektů zpřístupňování plných textů externím a vzdáleným uživatelům nejenom na úrovni projektu**
- **Smluvní zajištění zpřístupňování báze ANL FULL interním i externím uživatelům, uživatelům v rámci konzorcia Anopress**
- **Zapojení báze ANL FULL do Konzorcia Anopress a v budoucnu eventuelně i multilicence na používání linky automatického zpracování pro další knihovny**
- **Pokračování Konzorcia Anopress**
- **Eventuelní další zpřístupnění elektronických zdrojů v Anopressu**
- **Vstřícnost vydavatelů, nakladatelů, autorů**
- **Strukturované údaje v plných textech**
- **Další vývoj topiků a jejich ladění**
- **Zkvalitnění možností vyhledávání v systému TOPIC a integrace termínů věcného popisu do topiků, interakce systému s uživatelem a katalogizátorem**
- **Práce na sjednocování rejstříku předmětových kategorií**
- **Automatizovaná kategorizace a možnosti koexistence „ručně tvořených předmětových kategorií“ a automatizovaně vzniklých**
- **Další zkoumání možností automatické či poloautomatické indexace zejména ve smyslu extrakce údajů, generování topiků on line**

- Nutnost intelektuální indexace věcné
- Ověření funkčnosti internetová verze linky získávání a zpracování bibliografických záznamů z plných textů (WWW TTDE)
- Jednání s nakladateli/vydavateli o spolupráci týkající se elektronických zdrojů by mělo být koncepčně vedeno a zaštitěno z vyšší úrovně managementu institucí rovněž tak iniciativy ohledně povinného výtisku elektronických publikací event. autorského práva.
- Smluvní ošetření týkající se zpřístupnění plných textů.

## D Použití finančních prostředků

### D.1 Komentář

Použití finančních prostředků je zohledněno v následující tabulce. V rámci neinvestičních prostředků jsou odděleny placené služby, mzdy.

Plánované čerpání finančních prostředků do konce roku 2003 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

Neinvestice		Soub. databáze Kooper. systému čl. bibliogr. 2002	Limit	915 000,00 Kč
Datum	Čís. fakt.	Neinvestice	Částka	Zůstatek
		Služby ( plán 824 000 Kč)		
23.4.2003	23410076	Anopress - ANL FULL - údržba, ladění, linka (leden)	52 500,0 0 Kč	862 500,00 Kč
23.4.2003	23410182	Anopress - ANL FULL - údržba, ladění, linka (únor)	52 500,00 Kč	810 000,00 Kč
23.4.2003	23410353	Anopress - ANL FULL - údržba, ladění, linka (březen)	52 500,0 0 Kč	757 500,00 Kč
7.5.2003	23410516	Anopress - ANL FULL - údržba, ladění, linka (duben)	52 500,00 Kč	705 000,00 Kč

15.5.2003	2003-174	I. Mattern - ANL FULL - údržba, ladění, linka (01-05)	37 500,00 Kč	667 500,00 Kč
3.6.2003	23410671	Anopress - ANL FULL - údržba, ladění, linka (květen)	52 500,00 Kč	615 000,00 Kč
16.6.2003	20031	J. Koktan - program na kontrolu URL adres	12 000,00 Kč	603 000,00 Kč
1.7.2003	2003-178	I. Mattern - ANL FULL - údržba, ladění, linka (červen)	7 500,00 Kč	595 500,00 Kč
2.7.2003	23410845	Anopress - ANL FULL - údržba, ladění, linka (červen)	52 500,00 Kč	543 000,00 Kč
3.7.2003	200315	J. Koktan - záloha ANL FULL (zrcadlo), export dat pro ANL	5 000,00 Kč	538 000,00 Kč
31.7.2003	2003-181	I. Mattern - ANL FULL - údržba, ladění, linka (červenec)	7 500,00 Kč	530 500,00 Kč
31.7.2003	2435	Instalace serveru full.nkp.cz po havárii serveru a IIS	4 685,00 Kč	525 815,00 Kč
20.8.2003	23410996	Anopress - ANL FULL - údržba, ladění, linka (červenec)	52 500,00 Kč	473 315,00 Kč
1.10.2003	23411210	Anopress - ANL FULL - údržba, ladění, linka (srpen)	52 500,00 Kč	420 815,00 Kč
6.10.2003	2003-183	I. Mattern - ANL FULL - údržba, ladění, linka (srpen)	7 500,00 Kč	413 315,00 Kč
6.10.2003	2003-187	I. Mattern - ANL FULL - údržba, ladění, linka (září)	7 500,00 Kč	405 815,00 Kč
1.10.2003	23411334	Anopress - ANL FULL - údržba, ladění, linka (září)	52 500,00 Kč	353 315,00 Kč
		<i>I. Mattern - ANL FULL - údržba, ladění, linka (říjen)</i>		
		<i>Anopress - ANL FULL - údržba, ladění, linka (říjen)</i>		
		<i>I. Mattern - ANL FULL - údržba, ladění, linka (listopad)</i>		
		<i>Anopress - ANL FULL - údržba, ladění, linka (listopad)</i>		
		<i>I. Mattern - ANL FULL - údržba, ladění, linka (prosinec)</i>		
		<i>I. Mattern - ANL &gt; FULL - údržba, ladění, linka (prosinec)</i>		
		<i>J. Koktan - testování a ladění aplikace MNG, Báze titulů</i>		
		<b>Mzdové prostředky (plán 91 000 Kč)</b>		
30.6.2003		Magúthová, Molitorisová, Mráková, Rumlenová, Vrbičanová	52 800,00 Kč	300 515,00 Kč
<b>zůstatek</b>				<b>300 515,00 Kč</b>
<b>čerpáno</b>			<b>614 485,00 Kč</b>	



## **E Resumé a klíčová slova**

### **E.1 Resumé a klíčová slova v češtině**

Náplní projektu je optimalizace integrace a správy heterogenních dat souborné databáze Kooperčního 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. 2003 je rutinní aplikace pro získávání a automatickou extrakci/indexaci bibliografických záznamů z plných textů (TTDE) a následně vytvoření importního souboru záznamů pro databázi bibliografických záznamů ANL a databázi plných textů ANL FULL v NK s implementací metadat DC v plných textech včetně automaticky generovaných formátů HTML, XHTML, XML v kvalifikované i nekvalifikované formě - aplikace pro lokální stanice v NKČR. Automaticky se generuje URL.

V r. 2003 byla navržena a částečně realizována internetová aplikace pro získávání a automatickou extrakci/indexaci bibliografických záznamů z plných textů (WWW TTDE) rozšířená o vstupní formáty HTML z libovolné URL adresy s možností ukládat metadata do báze ANL FULL (ANL) i bez uložení plných textů, dále s možností propojení na rejstříky báze ANL FULL. V současné době probíhají práce na propojení linky se soubory autorit - v tomto projektu na základě zprovoznění komponenty pro asp rozhraní. V této lince je zakomponován i modul pro opravy rejstříků a metadat v plných textech báze ANL FULL (částečně hrazeno z projektu Propojení analytických záznamů).

V r. 2003 je průběžně věnována pozornost 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. Navrženy a implementovány nové definice topiků (hrazeno z projektu Propojení analytických záznamů).

V r. 2003 je v rutinním provozu aplikace pro administraci databáze ANL FULL (statistiky, opravy, evidence uživatelů) a 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 2181 propojení - říjen 2003, hrazeno z projektu Propojení analytických záznamů).

V r. 2003 je udržován portál pro zpřístupnění volně dostupných textů na Internetu ve struktuře tematické a regionální s popisem zdrojů.

V roce 2003 byla průběžně aktualizována souborná databáze kooperčního systému ANL a vydáván CD-ROM s Českou národní bibliografií.

V r. 2003 je dále laděna a testována aplikace pro správu a údržbu bibliografické báze dat KOSABI včetně Báze pro evidenci titulů - částečně poloprovozní stádium v NKČR.

Hypotéza orientovat se spíše na automatickou extrakci dat s využitím možností inteligentního vyhledávání systému TOPIC a jeho interaktivních vlastností je doplněna o možnost automatické kategorizace plný >ch textů za předpokladu migrace systému TOPIC na platformu K2. Byla provedena analýza selekční úplnosti topiků a navrženy metody jejich zkvalitnění (hrazeno z projektu Propojení analytických záznamů).

V r. 2003 navázán pracovní kontakt se 4 nově vzniklými krajskými knihovnami - laděn UNIMARC a excerpní základna.

#### **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; digitální knihovna

## **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 2003 has resulted in running system operation of acquisition and workflow of automated extraction 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) - TTDE on lokal workstations in NLCR. Automatic URL generation.

The 2003 has resulted in partial realization of WWW aplication of acquisition and workflow of automated extraction indexing of bibliographical entries from fulltexts (WWW TTDE) including sources in HTML format directly from their URL. It is possible include metadata to ANL FULL (ANL) database without fulltexts. Linking to ANL FULL indexes is possible, too. Activities concernig linking to authority files are done - in this project on the base of component for asp interface. A modul for correction of indexes and metadata is a part of this application (partially funded from the rearch project Analytical Bibliographic Records as a Gateway of Documents).

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 (funded from the rearch project Analytical Bibliographic Records as a Gateway of Documents).

In 2003 an application for administration (statistics, corrections, users) has been put into full 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. Static links have been done between bibliographic records and free Internet full texts regularly (ca 2181 links - August 2003, funded from the research project Analytical Bibliographic Records as a Gateway of Documents). Portal for free texts on Internet has been maintained (subject and regional structure) with resource description.

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

In 2003 an application for management of COSABI is tested and optimised including Titles' database - partial pilot operational system in NLCR .

Orientation on automatic extraction indexing and data generating with possibilities of concept and intelligent retrieval of TOPIC system has been confirmed, interactive properties of TOPIC system and automatic categorization of documents as well.. Analysis of recall of topics has been done and suggestions for improvement of their quality, too (funded from the research project Analytical Bibliographic Records as a Gateway of Documents).

#### **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; digital library

PhDr. Vojtěch Balík, ředitel NKČR

PhDr. Ivana Anděrová, hlavní řešitelka

## **F Přílohy**

- F.1** [Příloha F 1](#) : Báze ANL v NKČR (ALEPHU), JIB CASLIN a ANL/ANL FULL, ČNB
- F.2** [Příloha F 2](#) : Linka zpracování bibliografických záznamů z plných textů - TamTam Data Extractor (TTDE) - bibliografická metadata pro bázi ANL (ALEPH) a plné texty s Dublin Core pro bázi ANL FULL (TOPIC)
- F.3** [Příloha F 3](#) : Architektura systému získávání, zpracování a

zpřístupnění plných textů v systému TOPIC (báze ANL FULL a portál Periodika na WWW) - detailní ukázky formulářů, formátů, výstupů. Administrace (opravy a statistiky)

- F.4** [Příloha F 4](#) : Tituly v bázi ANL FULL, ANL, volně na WWW, Dublin Core a ANL Core - tabulka návěští pro vyhledávání v poli dotaz, tabulka formulářových polí, topiky, citace článku
- F.5** [Příloha F 5](#) : Krajský bibliografický systém. Aplikace pro správu a údržbu KOSABI (architektura systému, jednotlivé aplikace pro příjem a zpracování dat, báze titulů)
- F.6** [Příloha F 6](#) : Linka zpracování bibliografických záznamů z plných textů - internetová verze - WWW TamTam Data Extractor (WWW TTDE) - bibliografická metadata pro bázi ANL (ALEPH) a plné texty s Dublin Core pro bázi ANL FULL (TOPIC)

## Příloha F 1

### Báze ANL v NKČR (ALEPHU), JIB CASLIN a ANL/ANL FULL, ČNB Prezentace báze ANL (ALEPH)

#### Příloha F 1 : Báze ANL v NKČR (ALEPHU), JIB CASLIN a ANL/ANL FULL, ČNB Prezentace báze ANL (ALEPH)

Aktuální báze: ANL

Konec | Čtenář | Databáze | Dílčí báze | Nastavení | Otázky | Nápověda  
Rejstříky | Vyhledávání | Výsledky dotazu | Předchozí dotazy | Schránka

Základní vyhledávání | Z různých polí | Z více bází | Vyhledávání CCL

**ANL - Články v českých novinách, časopisech a sbornících**  
Základní vyhledávání [Informace o bázi](#)

Kombinace slov z různých polí (spojte pomocí AND, NOT nebo OR)

Autor	<input type="text" value="Slomek"/>	Slova z názvu	<input type="text" value="nezavděčil"/>
Rok/datum vydání	<input type="text"/>	Předmět (klíčová slova)	<input type="text"/>
Zdrojový dokument	<input type="text"/>	Blížkost slov?	<input type="radio"/> Ne <input type="radio"/> Ano

Příklad záznamu propojeného s plným textem, který vznikl v lince TTDE,  
propojení na plný text v bázi ANL FULL

Standardní zobrazení záznamu - propojení na plný text v 856, předmětové kategorie  
s MDT

## ANL - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Katalogizační záznam](#) -- [Citace](#) -- [UNIMARC](#) -- [S návěštími poli](#)

Záznam 1 z 1

◀ Předchozí záznam

Další záznam ▶

**Název** ● [Básník, jenž se národu nezavděčil / Jaromír Slomek](#)  
**Autor** ● [Slomek, Jaromír](#)  
**Rozsah** : 1 il  
**Zdroj. dokument** Lidové noviny. -- [Praha]. -- ISSN 0862-5921. -- Roč. 15, č. 67 (20020320), s. 6  
**Plný text** ● <http://full.nkp.cz/nkdb/view/mkdoc.asp?p1=LN20020320000027>

**Poznámka** ↕ Rubrika: Domov  
**Forma** ● [portréty](#)  
**Předmět. kat.** ● [literatura -- česká literatura -- 82](#)  
**Předmět. hesla** ● [Machar Josef Svatopluk, 1864-1942](#)  
**Předmět. hesla** ● [básníci -- Česko -- stol. 19.-20.](#)  
**Klíč. slova** ● [čeští básníci](#)  
**MDT** ● [821.162.3-051](#)  
● [\(437.3\)](#)

**Jazyk** cze

**Anotace** Před šedesáti lety, jen pár týdnů předtím, než si stateční parašutisté dokázali počíhat na Reinharda Heydricha, zemřel v Praze **básník** Josef Svatopluk Machar

**Vazba nahoru** [vazba na zdrojový dokument](#)  
**C.zázn. v bázi** 000704530

### JIB CASLIN a báze ANL, ANL FULL

Báze ANL je připojena pro vyhledávání, báze ANL FULL zatím jako odkaz

The screenshot displays the JIB search interface. At the top, it says "JIB Jednotná informační brána" and "REGISTROVANÝ UŽIVATEL | ODHLÁSIT SE | ENGLISH | NASTAVENÍ | NÁPOVĚDA". Below this is a navigation bar with "ÚVODNÍ STRÁNKA", "VYHLEDÁVÁNÍ", "VÝSLEDEK DOTAZU", "NALEZENÉ ZÁZNAMY", and "HISTORIE". The main content area is titled "Host" and contains several search-related sections: "VÝPIS ZDROJŮ", "KATEGORIE ZDROJŮ", "VÝBĚR ZDROJŮ", and "Vyhledávání". The "Vyhledávání" section has two search boxes, both set to "Všechna pole" and "Slovo nebo fráze:". Below the search boxes are "Hledat" and "Smazat" buttons. The "VÝBĚR ZDROJŮ" section has input fields for "Jméno zdroje", "Klíčová slova", and "Jakékoli slovo". The "Vyhledávání" section also has tabs for "Všechny zdroje", "Prohledatelné zdroje", and "Odkazové zdroje". The search results are displayed in a table with columns "Typ" and "Jméno zdroje". The results are: "Fulltext NK - ANL FULL (plné texty)" and "Bibliography NK - článková bibl. (ANL)".

Tentýž záznam v JIB CASLIN v ANL s propojením na plný text v ANL FULL

ANL - Úplné zobrazení

Zvolte formát: Standardní

Záznam 1 z 2

Název:  Básník...

Autor:  Slomek...

Rozsah: 1 il

Zdroj. dokument: Lidové noviny

Plný text:  <http://ful...>

Poznámka: Rubrika: Domov

Forma:  portréty

Předmět. kat.:  literatura

Název: Básník, jenž se národu nezavděčil

Hlavní autor: Jaromír Slomek

Zdroj: Lidové noviny

ISSN: 0862-5921

Roč.: 15, č. 67 (20.3.2002), s. 6

Rubrika: Domov

Předmět. ktg.: literatura: česká literatura

MDT: 821.162.3-051, (437.3)

Osoba jako předmět: Machar, Josef Svatopluk, 1864-1942

Téma jako předmět: básníci-Česko-stol. 19.-20.

Typ dokumentu: portréty

Geograf. kód: e-11--

Příběh J. S. MACHARA, "vídeňského Čecha" a kritika domácích poměrů

ČESKÉ LEGENDY 20. STOLETÍ

Před šedesáti lety, jen pár týdnů předtím, než si stateční parašutisté dokázali počíhat na Reinharda Heyd Josefa Svatopluka Machara. Toho roku dovršil 78 let, i když výroč...

**PRO ZOBRAZENÍ PLNÝCH TEXTŮ JE TŘEBA SE PŘIHÁSIT NEBO ZAREGISTROVAT NA SERVERU IIR I MKD C7**

Příklad článku - recenze zpracovaného v lince TTDE a zpřístupněného v v ANL - standardní zobrazení a UNIMARC, resp. v JIB s propojením na plný text



## ANL - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Katalogizační záznam](#) -- [Citace](#) -- [UNIMARC](#) -- [S návěstími polí](#)

Záznam 6 z 494

[◀ Předchozí záznam](#)

[▶ Další záznam](#)

- Název** ● [Angličan se dívá na Evropu / Tomáš Zahradníček](#)  
**Autor** ● [Zahradníček, Tomáš](#)  
**Rozsah** : 1 il  
**Zdroj. dokument** ● [Lidové noviny](#) -- [Praha]. -- ISSN 0862-5921. -- Roč. 16, č. 185 (20030809), s. 15  
**Plný text** ● <http://full.nkp.cz/nkdb/view/mkdoc.asp?p1=LN2003080900007300>
- Poznámka** Rubrika: Orientace  
**Forma** ● **recenze**  
**Předmět. kat.** ● [politika -- 32](#)  
**Předmět. hesla** ● [Ash Timothy Garton, 1955-](#)  
**Předmět. hesla** ● [Dějiny přítomnosti](#)  
**Předmět. hesla** ● [politologie -- Evropa -- stol. 20. , léta 90.](#)  
**MDT** ● [321](#)  
● [\(4\)](#)  
● [\(070.447\)](#)
- Jazyk** cze
- Anotace** Mezi západními odborníky na střední a východní Evropu si vydobyl prominentní postavení britský historik a novinář Timothy Garton Ash
- Vazba nahoru** [vazba na zdrojový dokument](#)  
**C.zázn. v bázi** 000793083



## ANL - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Katalogizační záznam](#) -- [Citace](#) -- [UNIMARC](#) -- [S návěštími polí](#)

Záznam 1 z 1

◀ Předchozí záznam

Další záznam ▶

**LDR** -----naa 22----- 450  
**FMT** RD  
**001** an2003abdf0001918  
**014** |a 0862-5921(20030809)16:185[15:LN2003080900007300]3.0.CO;2-1 |2 sici |9  
LN2003080900007300  
**100** |a 20030910j20030809u y0czey0103 ba  
**1010** |a cze  
**102** |a CZ  
**105** |a a r 000yy  
**2001** |a Angličan se dívá na Evropu |f Tomáš Zahradníček  
**215** |c 1 il  
**304** |a Rubrika: Orientace  
**330** |a Mezi západními odborníky na střední a východní Evropu si vydobyl prominentní postavení britský  
historik a novinář Timothy Garton Ash  
**461 0** |1 011 |a 0862-5921 |1 2001 |a Lidové noviny |v Roč. 16, č. 185 (20030809), s. 15 |1 210 |a  
[Praha]  
**5451** |a Orientace  
**600 1** |a Ash |b Timothy Garton |f 1955-  
**605** |a Dějiny přítomnosti |n kniha |k 2003  
**6060** |a politologie |y Evropa |z stol. 20., léta 90.  
**608** |a recenze  
**615** |a politika |n 32

660 |a e-----  
 675 |a 321  
 675 |a (4)  
 675 |a (070.447)  
 700 1 |a Zahradníček |b Tomáš  
 801 1 |a CZ |b ANOPRESS |c 20030910  
 801 2 |a CZ |b ABA001 |c 20030910 |g AACR2  
 8564 |u <http://full.nkp.cz/nkdb/view/mkdoc.asp?p1=LN2003080900007300>  
 910 |a ABA001 |t rd  
 9564 |a full.nkp.cz |d /nkdb/view/mkdoc.asp?p1= |f LN2003080900007300 |z CP 1250 |u <http://full.nkp.cz/nkdb/view/mkdoc.asp?p1=LN2003080900007300> |9 HTM  
 ZAZ0 |g Lidové noviny |h ISSN 0862-5921 |s Roč. 16 |z č. 185 |d 20030809 |l s. 15 |m [Praha]  
 STZ |a sa20030910 |b jale  
 VAH 16  
 LKR |a UP || NKC01 |b |n vazba na zdrojový dokument |m Angličan se dívá na Evropu ...  
 CAT |c 20030910 || ANL01 |h 2030  
 SYS 000793083



**Z** [Tisk článek] [vst] [dajmeta dočíst] [dočíst] [dočíst] [dočíst] [dočíst]

**N** **Název:** Angličan se dívá na Evropu  
**A** **Hlavní autor:** Tomáš Zahradníček  
**R** **Zdroj:** Lidové noviny  
**R** **ISSN:** 0862-5921  
**Z** **Roč.:** 16, č. 185 (9.8.2003), s. 15  
**P** **Rubrika:** Orientace  
**P** **Předmět. ktg.:** politika  
**MDI:** 321, (4), (070.447)

**P** **Osoba jako předmět:** Ash, Timothy Garton, 1955-  
**P** **Dílo jako předmět:** Dějiny přítomnosti (knihy). 2003  
**P** **Téma jako předmět:** politologie-Evropa-stol. 20., léta 90.  
**P** **Typ dokumentu:** recenze  
**General text:** e-----



CAT |c 20030910 |l ANLD01 |h 2030  
 SYS 000793083

**Z** (Tisk článku) vni dořmeta dořbřtni: dořbřtnit dořbřni: dořbřnit

**N** **Název:** Angličan se dívá na Evropu  
**A** **Hlavní autor:** Tomáš Zahradníček  
**R** **Zdroj:** Lidové noviny  
**P** **ISSN:** 0962-5921  
**Z** **Roč.:** 16, č. 185 (9.8.2003), s. 15  
**P** **Rubrika:** Orientace  
**P** **Předmět. křg.:** politika  
**P** **MDT:** 321, (4), (070.447)

**P** **Osoba jako předmět:** Ash, Timothy Garton, 1955-  
**P** **Dělo jako předmět:** Dějiny přítomnosti (kniha), 2003  
**P** **Tema jako předmět:** politologie-Evropa-stol. 20., léta 90.  
**P** **Typ dokumentu:** recenze  
**P** **Geograf. kód:** e-----

**M** Reportáže a úvahy Timothyho Gartona Ashe z devadesátých let

Mezi západními odborníky na střední a východní Evropu si vydobyl prominentní postavení britský historik a novinář Timothy Garton Ash. Výbor z jeho článků, reportáží a esejů, tištěných v 90. letech v amerických a anglických časopisech a novinách...

**J**

**A**

**PRO ZOBRAZENÍ PLNÝCH TEXTŮ JE TŘEBA SE  
 PŘIHLÁSIT NEBO ZAREGISTROVAT  
 NA SERVERU [FULL.NKP.CZ](http://FULL.NKP.CZ)**

*Příklady článků zpracovaných v ALEPHu a zpřístupnění v ANL, resp. JIB*

Příklady článků zpracovaných v ALEPHu a zpřístupnění v ANL, resp. JIB

## ANL - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Katalogizační záznam](#) -- [Citace](#) -- [UNIMARC](#) -- [S návěštími polí](#)

Záznam 16 z 42

[◀ Předchozí záznam](#)

[▶ Další záznam](#)

**Název** ● [Korovnicovití : Adelgidae / Petr Kapitola](#)  
**Autor** ● [Kapitola, Petr](#)  
**Rozsah** : 11 il.  
**Zdroj. dokument** Lesnická práce. -- ISSN 0322-9254. -- Roč. 82, č. 5 (2003)  
Lesní ochranná služba : [příloha], S. [I]-IV

**Poznámka** 9 obrázků, 2 tab.  
**Předmět. kat.** ● [zoologie](#) -- [hmyz](#) ● [lesnictví](#) -- [ochrana lesa](#)  
**Předmět. hesla** ● [korovnicovití](#) -- [biologie](#) -- [rozšíření](#) ● [jehličnaté dřeviny](#) -- [hmyzí škůdci](#)  
**Klíč. slova** ● [Adelgidae](#) ● [diagnostika](#) ● [obrana](#)  
**MDT** ● [595.2/7](#)  
● [591.5](#)  
● [630.4](#)

**Jazyk** cze

**C.zázn. v bázi** 000780265

## ANL - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Katalogizační záznam](#) -- [Citace](#) -- [UNIMARC](#) -- [S návěštími polí](#)

Záznam 6 z 484

[◀ Předchozí záznam](#)

[▶ Další záznam](#)

**Název** ● [Absurdní svět Fridricha\[!\] Dürrenmatta / Karel Franczyk](#)  
**Autor** ● [Franczyk, Karel](#)  
**Zdroj. dokument** **Tvar.** -- ISSN 0862-657X. -- Roč. 14, č. 2 (20030123), s. 22

**Forma** ● [recenze](#)  
**Předmět. kat.** ● [literatura](#) -- [švýcarská literatura](#)  
**Předmět. hesla** ● [německy psaná próza](#) -- [Svýcarsko](#) -- [stol. 20., polovina 2.](#)  
**Klíč. slova** ● [švýcarská próza](#)  
**MDT** ● [821.112\(494\)-31](#)  
● [\(070.447\)](#)

**Jazyk** cze

**Recenze na** ● [Rek hledá Rekyni / Friedrich Dürrenmatt. -- Praha : Havran. -- 2002](#)

**C.zázn. v bázi** 000759459



Záznam 2 z 42

◀ Předchozí záznam

Další záznam ▶

- Název** ● [Daten von früher Abten und Lehren des Klosters Kah thong in Derge \(Khams\) / Helmut Eimer](#)
- Autor** ● [Eimer, Helmut](#)
- Zdroj. dokument** Archiv orientální. -- ISSN 0044-8699. -- Roč. 71, č. 3 (2003), s. 318-322
- Předmět. kat.** ● [náboženství -- buddhismus](#) ● [rukopisy](#) ● [jazykověda -- tibetské jazyky](#)
- Předmět. hesla** ● [Kah thog rdo rje gdan \(tibetský klášter\)](#)
- Předmět. hesla** ● [tibetanistika](#) ● [tibetský buddhismus](#) ● [buddhistické kláštery -- Tibet \(Čína\)](#) ● [tibetské rukopisy](#) ● [tibetština](#)
- MDT** ● [80\(=584\)+908\(515\)](#)  
● [243.4](#)  
● [242-523.6](#)  
● [091\(515\)](#)  
● [811.584.6](#)  
● [\(515\)](#)
- Jazyk** ger
- C.zázn. v bázi** 000792771

## ANL - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Katalogizační záznam](#) -- [Citace](#) -- [UNIMARC](#) -- [S návěštými poli](#)

Záznam 2 z 117

◀ Předchozí záznam

Další záznam ▶

- Název** ● [100. výročí objevu chromatografie : život a dílo M.S. Cvěta / Eva Smolková-Keulemansová](#)
- Autor** ● [Smolková-Keulemansová, Eva](#)
- Rozsah** : 3 il.
- Zdroj. dokument** Chemické listy. -- ISSN 0009-2770. -- Roč. 97, č. 3 (2003), s. 134-139
- Plný text** ● [http://chemicke-listy.vscht.cz/full/2003\\_03\\_01.pdf](http://chemicke-listy.vscht.cz/full/2003_03_01.pdf)
- Poznámka** anglické resumé  
3 obrázky
- Předmět. kat.** ● [chemie -- analytická chemie](#)
- Předmět. hesla** ● [Cvět Michail Semjonovič, 1872-1919](#)
- Předmět. hesla** ● [fyzici -- Rusko -- stol. 19.-20.](#) ● [chemici -- Rusko -- stol. 19.-20.](#) ● [chromatografie -- historie -- stol. 20.](#)
- MDT** ● [543.5](#)  
● [54-051](#)  
● [\(47+57\)](#)
- Jazyk** cze
- C.zázn. v bázi** 000779645

## ANL - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Katalogizační záznam](#) -- [Citace](#) -- [UNIMARC](#) -- [S návěštími polí](#)

Záznam 15 z 117

[◀ Předchozí záznam](#)

[Další záznam ▶](#)

**Název** ● [Dětský autismus / Michal Hrdlička](#)  
**Autor** ● [Hrdlička, Michal](#)  
**Rozsah** 2 il.  
**Zdroj. dokument** [Psychiatrie](#). -- ISSN 1211-7579. -- Roč. 7, č. 2 (2003), s. 117-122

**Poznámka** Anglické resumé  
2 tab.

**Předmět. kat.** ● [lékařství -- psychiatrie](#)  
**Předmět. hesla** ● [dětský autismus -- historie](#) -- [příznaky](#) ● [dětský autismus -- epidemiologie](#) -- [etiologie](#) ● [dětský autismus -- terapie](#)

**MDT** ● [616.89](#)  
**Jazyk** cze

**C.zázn. v bázi** 000787826

[Příklad záznamu z kooperující instituce v KOSABI v bázi ANL, resp. JIB](#)

## ANL - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Katalogizační záznam](#) -- [Citace](#) -- [UNIMARC](#) -- [S návěštími polí](#)

Záznam 1 z 42

[◀ Předchozí záznam](#)

[Další záznam ▶](#)

**Název** ● [Bohuslav Havránek \(30.1.1893-2.3.1978\) / M. Cornejová](#)  
**Autor** ● [Cornejová, M.](#)  
**Zdroj. dokument** [Universitas : revue Masarykovy univerzity v Brně](#). -- ISSN 1211-3384. -- Roč. 2002, č. 1 (2003), s. 20-21

**Forma** ● [jubilejní články](#)  
**Předmět. kat.** ● [jazykověda](#)  
**Předmět. hesla** ● [Havránek Bohuslav, 1893-1978](#)  
**Předmět. hesla** ● [spisovná čeština -- gramatika](#) ● [staroslovenština](#)  
**MDT** ● [929\(Havránek.B.\)](#)  
● [811.162.3](#)  
● [81`1](#)

**Jazyk** cze

**C.zázn. v bázi** 000772930

## ANL - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Katalogizační záznam](#) -- [Citace](#) -- [UNIMARC](#) -- [S návěstími polí](#)

Záznam 10 z 470

[◀ Předchozí záznam](#) [▶ Další záznam](#)

**Název** [Azbuky volí studenti dobrovolně / Hana Hulíková](#)  
**Autor** [Hulíková, Hana](#)  
**Zdroj. dokument** Olomoucký den. -- [Ceské Budějovice]. -- ISSN 1212-4699, C. 202 (**20030829**), s. 9  
**Plný text** <http://kat.vkol.cz/ALEPH/-/ext-find?base=SVKD1%26find=WLC=III.302.107>

**Předmět. kat.** [školství -- střední školy](#)  
**Předmět. hesla** [ruština](#) [studium jazyků](#) -- **Olomouc (Cesko)** -- r. 2003  
**Předmět. hesla** [Olomouc \(Cesko\)](#)  
**MDT** [373.5/6](#)  
[811.161](#)

**Jazyk** cze

**C.zázn. v bázi** 000798335

Báze ANL pravidelně zpřístupňovaná jako řada ČNB Články v českých novinách, časopisech a sbornících (měsíční diskety, Aktualizace na CD-ROM)

The screenshot shows a web application interface. On the left, there is a table with columns 'Rel.' and 'Autoři článku'. The table lists various authors and their associated works. The row for 'Jungmannová, Lenka (recenzent)' is highlighted. On the right, there is a detailed view of the selected article, including fields for 'Název článku', 'Název zdroj. dokumentu', 'Datum zdroj. dokumentu', 'Lokace ve zdroj. dok.', 'Autoři recenz. dokumentu', 'Název recenz. dokumentu', 'MDT', 'Předmětové kategorie', 'Předmětová hesla', 'Klíčová slova', 'Formální označení', and 'Jazyk dokumentu'.

Rel.	Autoři článku
	Jiřinská, Veronika
	Dvořáková, Alena
	Krejčí, Marek
	Staněk, Jiří
	Hendych, Jan   Laňovička, Jiří
	Bradna, Jan
	Chrobák, Jakub
	Jesený, Vít
	Kovářík, Miroslav
	Kubelík, Martin   Musilek, Ladislav   Pavelka, (Informace o č)
	Trávníček, Mojmír
	Příša, Vladimír
	Novotný, Vladimír
	Štípl, Zdeněk
	Haman, Aleš
	Schauer, Hubert Gordon, 1862-1892
	Váňaník, Jiří
	Horák, Ondřej
	Jareš, Michal (sestavovatel)
	Slavík, Jiří
	Poslední, Petr
	Gísa, Václav
	Blažek, Bohuslav
	Rejchrt, Miloš, 1946- (dotazovaný)   Vančura, (Byla to snad
	Šimon, Petr
	Bishop, Elizabeth, 1911-1979   Shanfeldová, Yveta   Děť
	Lowell, Robert   Shanfeldová, Yveta   překlada   Elizabeth E
	Sedláč, David
	Kolář, Viktor
	Just, Vladimír, 1946-
	Jungmannová, Lenka (recenzent)
	Tvarůžková, Lucie

**Autoři článku**  
Jungmannová, Lenka (recenzent)

**Název článku**  
Vše i nic

**Název zdroj. dokumentu**  
Literární noviny

**Datum zdroj. dokumentu**  
20010509

**Lokace ve zdroj. dok.**  
Roč. 12, č. 19 (20010509), s. 12

**Autoři recenz. dokumentu**  
Ravenhill, Mark  
Sloupová, Jitka

**Název recenz. dokumentu**  
Faust

**MDT**  
792.2  
821.111-2  
(070.447)

**Předmětové kategorie**  
divadlo - české divadlo  
literatura - anglická literatura

**Předmětová hesla**  
Faust je mrtvý  
Činohry - divadelní inscenace - Brno (Česko) - r. 2001  
anglické drama - Velká Británie - stol. 20., léta 90.

**Klíčová slova**  
anglická dramata  
divadelní hry

**Formální označení**  
recenze

**Jazyk dokumentu**  
cze

## Příloha F 2

**Linka zpracování bibliografických záznamů z plných textů - TamTam Data Extractor (TTDE) - bibliografická metadata pro bázi ANL (ALEPH) a plné texty s Dublin Core pro bázi ANL FULL (TOPIC)**

### a) Funkční schéma linky TTDE



**Stažení článků z Internetu (databáze TAMTAM) pomocí TAMTAM Special NK (TTSNK) - možno stahovat více článků najednou na základě tematu, názvu článku i zdrojového dokumentu aj. Pomocí TAMTAM Standard (TTS) je možno stahovat vždy jeden článek. Dále je popsána verze TTSNK**

**Připojení do databáze TAMTAM - speciální přístup pro NK**





✖
Varování!

[OK]

Dnes je středa  
17. 4. 2002  
Svítek má Rudolt

---

🔍 VYHLEDÁVÁNÍ
📰 MÉDIA
📰 ZPRÁVODAJSTVÍ
📄 PLATNÉ PŘEDPISY
📄 VĚDOMOSTNÍ DB
📄 INFORMACE
📄 KONTAKT
📄 ZDROJE
📄 LOKAZY

**PRIHLÁŠENÍ**

**PRIHLÁŠEN UŽIVATEL:**  
Jméno: nkp1a  
Verze: speciální [NĚK]

Odhlásit se

[Úprava registrace]

**VYBRANÉ SLUŽBY**

- 📄 PŘEHLED PRODUKTŮ
- 📄 Mediální db 1, 2, 3
- 📄 Rejstříky (nedání)
- 📄 Archivy zdrojů
- 📄 3dP - předpisy
- 📄 vědomostní báze
- 📄 GLUCH-monitoring

**RYCHLÉ HLEDÁNÍ**

🔍 média 🔍 tamtam.cz

[Vyhledávací formulář]

**POČASI**

Čechy	
Střední	14°
Jihní	12°
Západní	13°
Severní	13°

## Server TAMTAM.CZ je tu pro Vás

Dodáme Vám v elektronické podobě plné znění článků z novin a časopisů i zpravodajské relace rozhlasových a televizních stanic. Navíc Vám umožníme v těchto informacích photoaktivně vyhledávat prostřednictvím Internetu.



- **INFORMACE PRO VÁS:** Aktuální informace o změnách a novinkách na serveru jsou **ZDE**
- **ZKOUŠEBNÍ PŘÍSTUP:** týdenní zkušební přístup zdarma. Stačí se **zaregistrovat** a vyzkoušet
- **PŘEHLED PRODUKTŮ:** kompletní přehled našich produktů a služeb pro Vás naleznete **ZDE**
- **VYHLEDÁVACÍ SLUŽBY:** umožňují fulltextové prohledávání dat v reálném čase (online) prostřednictvím Internetu. Uživatel pracuje s vybranou databází sám dle svého uvážení (nedání databáze, zákony, vědomostní databáze). Archivy informačních zdrojů **NOVE** (na data lze nahlížet dle stran a rubrik, resp. relací). Vyhledávání dle rejstříků.
- **MONITOROVACÍ SLUŽBY:** monitorování předem připravených témat. **Privátní monitoring** (personalizovaný online monitoring témat vypracovaných přesně podle požadavků uživatele), **GLUCH-monitoring NOVE** (personalizovaný online monitoring veřejně dostupných témat, možnost dalšího upřesnění dotazu), **ISA** (informační servis Anopress)
- **OSTATNÍ SLUŽBY:** výběry dat, rychlá a automatická dodávka informací (Internet, E-mail, FTP). Analytické výstupy, referáže, projektová řešení, poradenství, školení
- **INFORMAČNÍ ZDROJE:** monitorujeme většinu tuzemských informačních zdrojů. Kompletní seznam **ZDE**. Dodáváme výběry ze slovenských informačních zdrojů

**UDÁLOST DNE**

**Cibrian usadil Ferrerovou-Waldnerovou**



**Telefon:** 0274007156  
**Adresa:** Hokučická 3  
Praha 4, 148 00  
Podrobné kontakty  
Kde nás najdete  
Napíšte nám

Novinky pro Vás  
Sestavte si web  
Mapa serveru

Událost dne  
Hlavní události  
Aktuality Anopress  
Zpravodajský servis  
Informace z televizí  
Předpověď počasí  
Dopravní servis 1, 2



Kronika 20 století a Kronika Českých zemí  
v elektronické podobě  
získaly prestižní ocenění  
CHIP Tip. Zde je recenze a  
internetová podoba a navíc i  
Kronika listova.

Dotaz

## Dotaz

The screenshot shows the ANO PRESS search interface. At the top, there is a warning box: "Varování: Chybí Vám potřebné informace! Neprodleně použijte Tamtám!". The main navigation bar includes: "VYHLEDÁVÁNÍ - MÉDIA | ZPRAVODAJSTVÍ | PLATNÉ PŘEDPISY | VĚDOMOSTNÍ DB | INFORMACE | KONTAKT | ZDROJE | ODKAZY".

The left sidebar contains sections: "PŘIHLÁŠENÍ" (with user info: jméno: nřpta, verze: speciální [NK]), "VYBRANÉ SLUŽBY" (including "PŘEHLED PRODUKTŮ" and "RYCHLÉ HLEDÁNÍ"), and "NASTAVENÍ, TISK".

The main content area is titled "MEDIÁLNÍ DATABÁZE - ROZŠÍŘENÝ FORMULÁŘ (pro experty) [3]". It features a search query "Lidové noviny" in a text box, which is circled in red. Below the query are filters for "Období pro hledání v archivních databázích" (set to 20 Březen 2002) and "Databáze" (set to "Aktuálních 14 dnů").

The search criteria are defined in a table:

Rozšířené	Vazba	Dotaz (jeden výraz)	Hledat v poli
<input type="checkbox"/>	Přesně	AND (a)	Autor, redaktor
<input type="checkbox"/>	Přesně	AND (a)	Nadpis článku
<input type="checkbox"/>	Přesně	AND (a)	Rubrika

Additional options include "Volby" (Min. skóre: vše, Čístr: 10, Třídění: Skóre, Sestup), "Seznam" (Jednoduchý, Se souhrnem, Seskupený, Nedělený), and "Omezení" (Omezení dle oblastí: Tisk celostátní, Casopisy, TV a rozhlas, Střední Čechy; Omezení dle zdrojů: Blek, Malé noviny, Hospodářské noviny, Lidové noviny).

Buttons at the bottom include "HLEDAT ČLÁNKY", "Uložit dotaz", and "Načíst, zrušit".

The right sidebar contains "HLEDAT ČLÁNKY", "ANO PRESS" logo, "Podrobné kontakty", "ÚVODNÍ STRANA", "Účinnost dne", and "Společnost Anopress".

Označení a stažení článku

## Označení a stažení článku

The screenshot shows a web browser window with the URL [http://www.anopress.cz/T1\\_CZ/media/dotaz.asp?im=3&lepicim=](http://www.anopress.cz/T1_CZ/media/dotaz.asp?im=3&lepicim=). The main content area is divided into two panes. The left pane, titled "Celkem nalezeno 103 dokumentů (Tisk)", displays a list of 11 articles. Each article entry includes a checkbox, a number (71-80), a date (20.03.2002), a source (Lidové noviny), a page number (str. 7 or 5), and a file size in bytes. The articles are: 71. Bin Ládín je podle bratra v pořádku (963 B); 72. Brřítké Společenství vyloučilo Zimbabwe (1117 B); 73. Afghánci popřeli údaje USA o mrtvých (3545 B); 74. Ženu, která pobodala otce, uchránila před trestem smrti (1593 B); 75. Běsník, jenž se národu nezvdělil (5210 B); 76. Pířva prokázala, že dítě bylo jeden den staré (872 B); 77. Petřinská rozhledna má nový uzhled (1062 B); 78. Na výřet do Vítkovic jako do Pířy? (3625 B); 79. Vřr na horách v nárazech dosáhl síly vichřice, hrozí i povodně (1630 B); 80. Razię ve zřinské řirně? Možná omýř policie (3439 B). Below the list are buttons for "Zobrazř označené" and "Stáhnout označené", and a checkbox for "Označř všechny řlánky v seznamu" with a "Zobrazř dotaz" link. The right pane shows a confirmation message: "Stáženř souboru s texty do NK" and "Soubor NKP6442.ZIP je přřpraven ke staženř" with a file size of 3763 B. A "Stáhnout soubor" button is present, with an orange arrow pointing to it. Below this is a link for "Statistika v aktuálním měsřci (4/2002)" and a "Zobrazř statistiku - detailní logový soubor" link.

Extrahovaná data z báze TamTam do formuláře ke zpracování (má tři strany)

Rozřřřený formulář pro editaci - 1. strana - jmenné a věcné údaje.

Katalogizátor doplňuje předevřím věcné údaje - předm., kategorie, klíčová slova, MDT (na horní liřtě nabídka formulář 1,2,3, dále text, Unim-A, Unim, Dublin, hlavička 1, hlavička 2; na dolní liřtě zpracovat záznam, zruřit záznam, odeslat data, vyprázdnit, novinky, nepřenášet do hlaviček, nenačřtat data)

Přenesení bibliografických dat z plného textu, generování do UNIMARCU (pro bibliografický záznam) a metadat Dublin Core (pro plný text)

(Bibliografické údaje se konvertují z plných textů do formuláře pro editaci a zároveň do všech ostatních hlaviček (UNIMARC - tj. klasický, UNIMARC-A tj. UNIMARC pro Aleph), Dublin Core (DC), indexovací hlavička, zobrazovací hlavička), podle potřeby se formulář edituje, opravy se přenášejř zároveň do UNIMARCU, UNIMARCU-A a Dublinu. Dále lze editovat samostatně UNIMARC-A, UNIMARC. Automaticky se generují údaje jmenného popisu, URL, SICI, částečně LKR (po zařkrtnutř v nabídce Různě lze stahovat i automaticky generovaná klíčová slova.

Těmata a skupiny témat ANL FULL se budou "sbliřžovat" s předmětovými kategoriemi a hesly v ANL).



Doplnění údajů katalogizátorem - 1. strana formuláře

TamTam Data Extractor [30.10.2001 - 2.2.6]

Form.1 Form.2 Form.3 Text Unim. Unim. Dublín Hlav.1 Hlav.2 Běžné

1. Datum uložení 20020417 [100a:601c] ID záznamu an2002ab00000792 [001] Periodičita [100a]

2. Datum vydání 20020320 [100a:461v:ZAZd] 105 a [0] z [4] y [11] y [12]

3. Název článku Básník, jenž se národu nezavděčil [200e]

3a. Číslo části [2001h] 3b. Název části [2001i]

4. Podnázev [200e]

5. Hlavní autoři Jaromír Slomek [200.700-701.710-711]

6. Další autoři [200.702.712]

7. Rubrika Domov [304a.545a]

8. Souhrn Před šedesáti lety, jen pár týdnů předtím, než si stateční parašutisté dokázali počíhat na Reinharda Heydricha, zemřel v Praze básník Josef Svatopluk Machar [330e]

9. Název zdroje Lidové noviny [461a:ZAZg] Sa  Nov.  Čas.  Ost. 9d. LKR [LKRb]

9b. Číslo části zdr [461h] 9c. Název části zdr [461i]

10. Název přílohy [461a:ZAZg] 11. ISSN 0862-5921 [461a:ZAZh]

12. Místo vydání Praha [461v:ZAZm] 13. Vydavatel Lidové noviny

14. Ročník 15 [461v:ZAZn] 15. Číslo 67 [461v:ZAZh] 16. Strana 6 [461v:ZAZl]

18. Křesová slova čeští básníci [610a]

17. Předmět. klč. literatura: česká literatura[82] [615a-615e]

19. MDT 821.162.3-051. [437.3] [675a] 20. VAH 16

21. Geograf. kód e-xx

22. URL odkaz http://full.nkp.cz [a] /nkdb/mkdoc.php?o= [6] LN2002032000002700 [056i]

23. Dokument CP 1250 HTM [z] [9] 24. Typ dokumentu portréty

25. STZ sa20020417 26. ID Annot. - NBN LN2002032000002700

27. SICI 0862-5921[20020320]15:67[6:LN2002032000002700]3.0.CO;2-Y 28. IL 1 il. [215c]

Počet 4 ZPRACOVAT ZÁZNAM Zrušit záznam Odeslat data Vyprázdnit Novinky ?? Nepřenášet do hlaviček Nenačítat články

Doplnění údajů věcného popisu - 2. strana formuláře

TamTam Data Extractor [30.10.2001 - 2.2.6]

Form.1 Form.2 Form.3 Text Unim. A Unim. Dublin Hlav.1 Hlav.2 Bůzné

**A) Téma jako předmět [6060]**

Téma [a]	Věcné zpřesnění [x]	Zeměpisné zpřesnění [y]	Chronologické zpřesnění [z]
1 básníci		Česko	stol. 19.-20.
2			
3			

**B) Geografické jméno jako předmět [607]**

Téma [a]	Věcné zpřesnění [x]	Zeměpisné zpřesnění [y]	Chronologické zpřesnění [z]
1			
2			
3			

**C) Osoba jako předmět [600 1/0]**

Příjmení [a]	Křestní jm. [b]	Doplněk [c]	Rímské čís. [d]	Data [f]	Věc. zpř. [x]	Zem. zpř. [y]	Chron. zpř. [z]
1 Machar	Josef			1864-194			
2							
3							

**D) Dílo jako předmět [605]**

Název [a]	Číslo částí [h]	Název částí [i]	Druh [n]	Datum vydání [k]
1				
2				
3				

**G) Korporace jako předmět [60102/01]**

Instituce, Místo [a]	Zpřesnění [b]	Doplněk [c]
1		
2		
3		

Počet: 4 ZPRACOVAT ZÁZNAM Zrušit záznam Odeslat data Vyprázdnit Novinky ??  Nepřenášet do hlaviček  Nenačítat články

Vygenerovaná data ve hlavičce UNIMARC-A a klasický UNIMARC

Form.1	Form.2	Form.3	Text	Unim.-A	Unim.	Dublin	Hlav.1	Hlav.2	Bůznt
FMT	L	RD							
LDR	L	-----naa^^22-----^^^450^							
001	L	an2002abdf0000792							
100	L	\$\$a20020417j20020320u^^y0czey0103^^^ba							
1010	L	\$\$acze							
102	L	\$\$aCZ							
105	L	\$\$aa^^^z^^^000yy							
014	L	\$\$a0862-5921(20020320)15:67[6:LN2002032000002700]3.0.CO;2- Y\$\$2sici\$\$9LN2002032000002700							
2001	L	\$\$aBásnik, jenž se národu nezavděčil\$\$fJaromir Slomek							
215	L	\$\$c1 il.							
304	L	\$\$aRubrika: Domov							
330	L	\$\$aPřed šedesáti lety, jen pár týdnů předtím, než si stateční parašutisté dokázali počíhat na Reinharda Heydricha, zemřel v Praze básník Josef Svatopluk Machar							
461	0	L \$\$1011^^\$\$a0862-5921\$\$12001^\$\$aLidové noviny\$\$vRoč. 15, č. 67 (20020320), s. 6\$\$1210^^\$\$a[Praha]							
5451	L	\$\$aDomov							
600	1	L \$\$aMachar\$\$bJosef\$\$f1864-1942							
6060	L	\$\$abásníci\$\$yČesko\$\$zstol. 19.-20.							
608	L	\$\$aportréty							
6100	L	\$\$ačeští básníci							
615	L	\$\$aliteratura\$\$xčeská literatura\$\$n82							
660	L	\$\$ae-xr---							
675	L	\$\$a821.162.3-051							
675	L	\$\$a(437.3)							
700	1	L \$\$aSlomek\$\$bJaromir							
801	1	L \$\$aCZ\$\$bANOPRESS\$\$c20020417							
801	2	L \$\$aCZ\$\$bABA001\$\$c20020417\$\$gARCR2							
8564	L	\$\$afull.nkp.cz\$\$d/nkdb/mkdoc.php?p1=\$\$fLN2002032000002700\$\$zCP 1250\$\$uhttp://full.nkp.cz/nkdb/mkdoc.php?p1=LN2002032000002700\$\$9HTM							

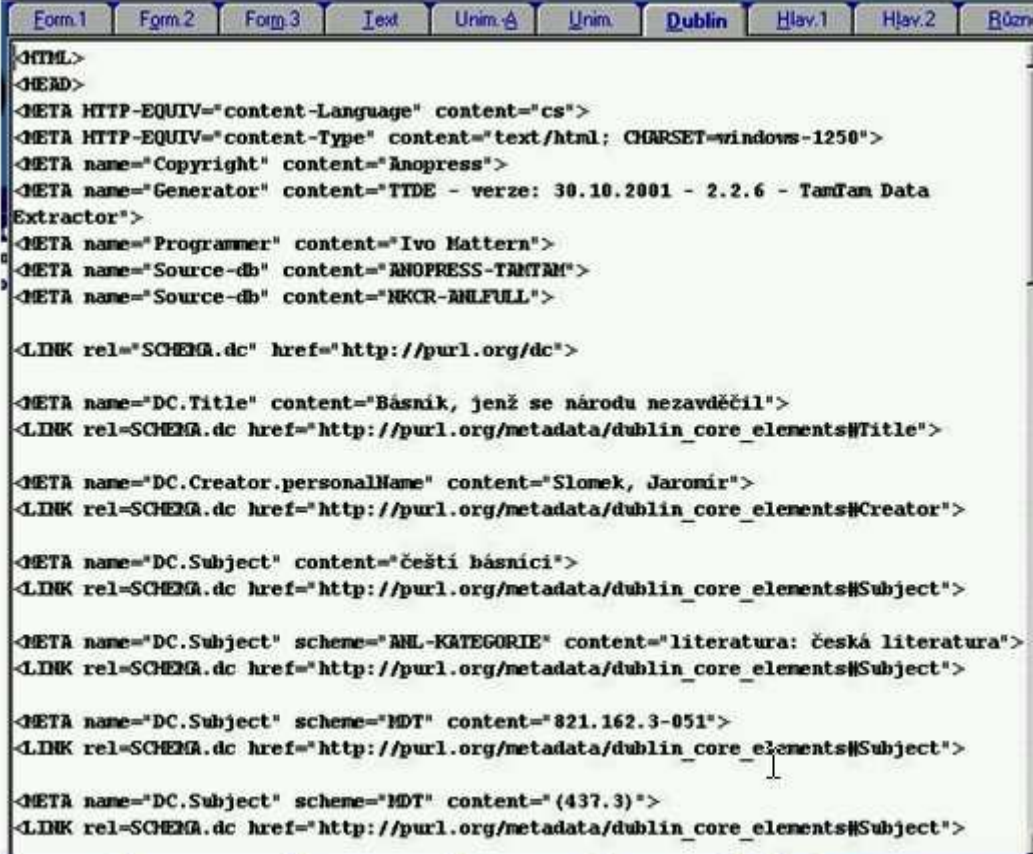
Form.1	Form.2	Form.3	Text	Unim.-A	Unim.	Dublin	Hlav.1	Hlav.2	Bůznt
LAB	L	-----naa##22-----###450#							
001	L	an2002abdf0000792							
100##	L	\$\$a20020417j20020320u##y0czey0103####ba							
1010##	L	\$\$acze							
102##	L	\$\$aCZ							
105##	L	\$\$aa####z####000yy							
014##	L	\$\$a0862-5921(20020320)15:67[6:LN2002032000002700]3.0.CO;2- Y\$2sici-i\$9LN2002032000002700							
2001##	L	\$\$aBásnik, jenž se národu nezavděčil\$\$fJaromir Slomek							
215##	L	\$\$c1 il.							
304##	L	\$\$aRubrika: Domov							
330##	L	\$\$aPřed šedesáti lety, jen pár týdnů předtím, než si stateční parašutisté dokázali počíhat na Reinharda Heydricha, zemřel v Praze básník Josef Svatopluk Machar							
461##	0	L \$\$1011##\$\$a0862-5921\$\$12001##\$\$aLidové noviny\$\$vRoč. 15, č. 67 (20020320), s. 6\$1210##\$\$a[Praha]							
5451##	L	\$\$aDomov							
600	1	L \$\$aMachar\$\$bJosef\$\$f1864-1942							
6060##	L	\$\$abásníci\$\$yČesko\$\$zstol. 19.-20.							
608##	L	\$\$aportréty							
6100##	L	\$\$ačeští básníci							
615##	L	\$\$aliteratura\$\$xčeská literatura\$\$n82							
660##	L	\$\$ae-xr---							
675##	L	\$\$a821.162.3-051							
675##	L	\$\$a(437.3)							
700##	1	L \$\$aSlomek\$\$bJaromir							
801##	1	L \$\$aCZ\$\$bANOPRESS\$\$c20020417							
801##	2	L \$\$aCZ\$\$bABA001\$\$c20020417\$\$gARCR2							
8564##	L	\$\$afull.nkp.cz\$\$d/nkdb/mkdoc.php?p1=\$\$fLN2002032000002700\$\$zCP 1250\$\$uhttp://full.nkp.cz/nkdb/mkdoc.php?p1=LN2002032000002700\$\$9HTM							
910##	L	\$\$aABA001\$\$trd							



Vygenerovaná data v hlavičce Dublin Core

Hlavička Dublin Core (pro plný text)

(dále v aplikacích XHTML a XML v "pracovních formátech")



```
<HTML>
<HEAD>
<META HTTP-EQUIV="content-Language" content="cs">
<META HTTP-EQUIV="content-Type" content="text/html; CHARSET=windows-1250">
<META name="Copyright" content="Anopress">
<META name="Generator" content="TTDE - verze: 30.10.2001 - 2.2.6 - TamTam Data
Extractor">
<META name="Programmer" content="Ivo Hattern">
<META name="Source-db" content="ANOPRESS-TAMTAM">
<META name="Source-db" content="NKCR-ANLFULL">

<LINK rel="SCHEMA.dc" href="http://purl.org/dc">

<META name="DC.Title" content="Básnik, jenž se národu nezavděčil">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Title">

<META name="DC.Creator.personalName" content="Slonek, Jaronír">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Creator">

<META name="DC.Subject" content="čeští básníci">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

<META name="DC.Subject" scheme="ANL-KATEGORIE" content="literatura: česká literatura">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

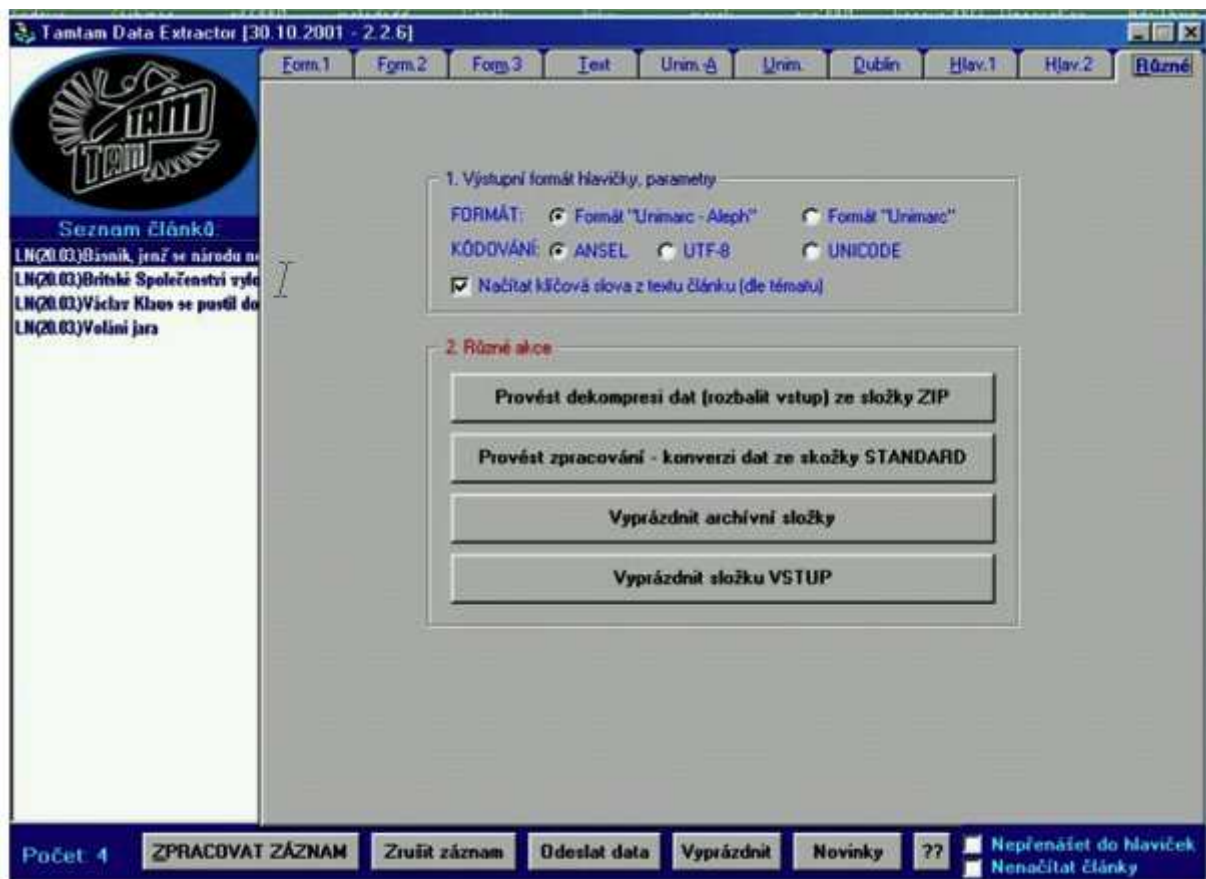
<META name="DC.Subject" scheme="MDT" content="821.162.3-051">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

<META name="DC.Subject" scheme="MDT" content="(437.3)">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">
```

Odeslání dat

Po odrážce "Různé" je možné nastavit tvar výstupní hlavičky, spustit ruční vstup dat, vyprázdnit archivní složky

Nabídka "různé" pro navolení výstupního formátu bibliografických údajů, resp. metadat a kódování pro bázi ANL (ALEPH - Ansel, UTF 8, UNICODE)



Záznam (metadata FULL) importovaný z TTDE do ANL, resp. JIB (v pozadí), přes automaticky generovanou URL v TTDE je propojený na plný text v ANL





Zpracovaný článek V TTDE v bázi ANL FULL (TOPIC) v Seznamu výsledků - tři zobrazovací formáty

The screenshot displays the ANL FULL search interface. At the top, there is a navigation menu with three columns: 'Základní informace' (Basic information), 'Vstup do databáze' (Database access), and 'Vyhledávání' (Searching). Below this is a search form titled 'ZÁKLADNÍ FORMULÁŘ' (Basic Form) with a search box containing 'Machar' and a 'HLEDAT' (Search) button. The search parameters are set to 'od: 1 Leden 2002' and 'do: 26 Duben 2002'. The results section shows 'Celkem nalezeno 4 článků' (Total 4 articles found) and lists four search results with their respective dates, page numbers, and titles.

Číslo	Relevance	Datum	Titulek	Stránky	Velikost
1	0.96	20.03.2002	Lidové noviny	str. 6	[13325 B]
2	0.77	08.04.2002	Lidové noviny	str. 6	[13000 B]
3	0.77	08.02.2002	Haló noviny	str. 1, 5	[15565 B]
4	0.77	09.01.2002	Haló noviny	str. 10	[13571 B]

Copyright © | Design © NKP, Anopress - Ivana Anděrová, Ivo Mattern, 2000-2002

Zpracovaný plný text v bázi ANL FULL - plný text s metadaty

The screenshot displays the ANL FULL search interface. At the top, there are navigation links: 'Základní informace', 'Vstup do databáze', and 'Vyhledávání'. Below these are sub-links for 'Charakteristika DB', 'Jak se zaregistrovat', 'Jak vyhledávat', 'Další databáze NK', 'Registrační formulář', 'Formuláře', and 'Periodika na WWW', 'Přihlášení', and 'Rejstříky'. The main header reads 'ANL FULL - plnotextové vyhledávání v článcích z tisku'. The search form includes a search box with the query 'zahr<contains>Lidové and naz<contains>báseň', a search button, and filters for 'Období pro vyhledávání' (from September 2001 to April 2002) and 'Typ seznamu' (set to 'Sešupený'). The results section shows 'Celkem nalezeno 7 článků'. The first result is 'Báseň, jenž se národu nezavděčil' by Jaromír Slomek, published in 'Lidové noviny' on 19.12.2001. The right sidebar provides detailed metadata for the selected article, including author, source, ISSN, issue, rubric, and subject terms. A 'HOME' button is visible in the bottom right corner.

Období pro vyhledávání:	od:	do:
	18 Zář 2001	18 Duben 2002

**TEXT ČLÁNKU S ODKAZY**

**Název:** Báseň, jenž se národu nezavděčil  
**Hlavní autor:** Jaromír Slomek  
**Zdroj:** Lidové noviny  
**ISSN:** 0862-5921  
**Roč.:** 15, č. 67 (20.3.2002), s. 6  
**Rubrika:** Domov  
**Předmět. ktg.:** literatura: česká literatura  
**MDT:** 821.162.3-051. (437.3)

**Osoba jako předmět:** Machar, Josef Svatopluk, 1864-1942  
**Téma jako předmět:** básníci-Česko-stol. 19.-20.  
**Typ dokumentu:** portréty  
**Geograf. kód:** @-@---

**Příběh J. S. MACHARA, "vídeňského Čecha" a kritika domácích poměrů**

**ČESKÉ LEGENDY 20. STOLETÍ**

Před šedesáti lety, jen pár týdnů předtím, než si stateční parašutisté dokázali počítat na Reinharda Heydricha, zemřel v Praze básník Josef Svatopluk Machar. Toho roku dovršil 78 let, i když v době svého paronání mohl v životě stvořit jen...

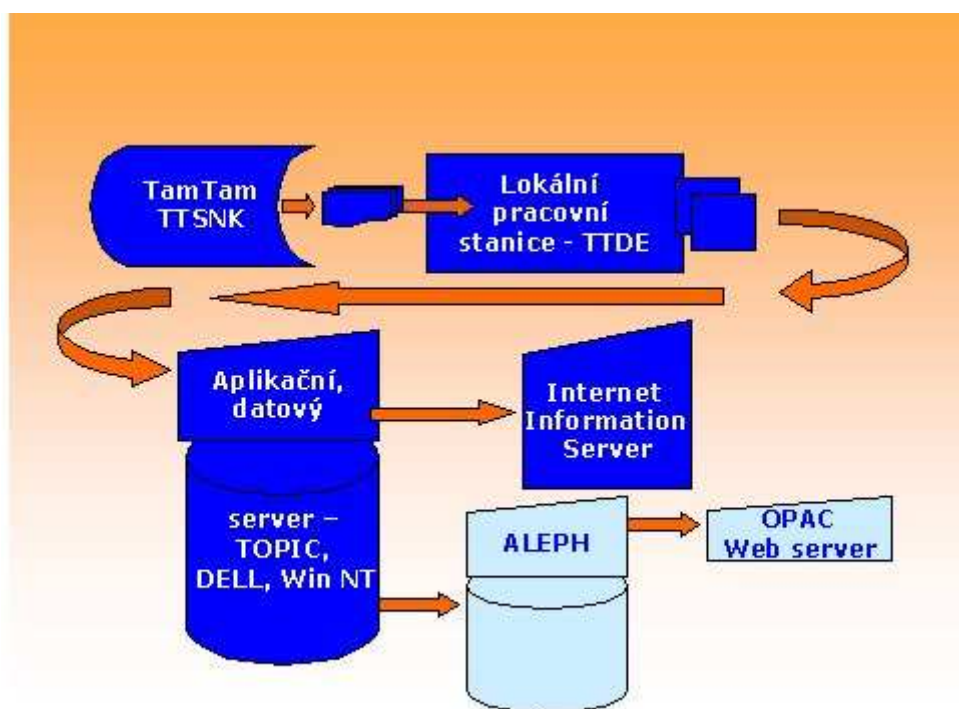
Copyright © Design © NKP, Anxpress - Ivana Anděrová, Ivo Matern, 2000-2002

Linka TTDE je použitelná po úpravě vstupním filtrem i pro spolupráci mezi autorem, nakladatelem/vydavatelem a bibliografickou agenturou



### Příloha F 3

Architektura systému získávání, zpracování a zpřístupnění plných textů v systému TOPIC (báze ANL FULL a portál Periodika na WWW) – detailní ukázky formulářů, formátů, výstupů. Administrace (opravy a statistiky)



Server full.nkp.cz - homepage



	Základní informace	Vstup do databáze	Vyhledávání
	Charakteristika DB	Jak se zaregistrovat	Jak vyhledávat
	Další databáze NK	Registrační formulář	Formuláře
	Periodika na WWW	Přihlášení	Rejstříky

**ANL FULL**


**ANL FULL - Články v českých novinách a časopisech (plně texty)**  
 Plnotextové vyhledávání ve vybraných článcích z českých novin a časopisů. Interní uživatelé NK ČR: k dispozici plně texty článků a metadata. Externí uživatelé NK ČR: běžně k dispozici pouze metadata, zkušební registrace pro přístup k plným textům na 7 dnů. Portál volných www periodik.

**ANL FULL - Articles in Czech newspapers and magazines (full texts)**  
 Full text retrieval in selected articles from the Czech newspapers and journals. Internal users of NK ČR: full texts and metadata available. External users of NK ČR: only metadata available, 7 days' test registration for full text access. Portal of free www periodicals.

**Management Kooperačního systému článkové bibliografie (MNG KOSABI)**

[Plně texty v českých novinách a časopisech \(ANL FULL, ANL, Anopress, WWW\) - přehled](#)  
[Souborná databáze Kooperačního systému ... \(programový projekt\)](#)  
[Propojení analytických záznamů ... \(výzkumný záměr\)](#)  
[Co je nového ...](#)

## Registrace

	Základní informace	Vstup do databáze	Vyhledávání
	Charakteristika DB	Jak se zaregistrovat	Jak vyhledávat
	Další databáze NK	Registrační formulář	Formuláře ▾
	Periodika na WWW	Přihlášení	Rejstříky

**ANL FULL - plnotextové vyhledávání v článcích z tisku**

**REGISTRACE NOVÉHO UŽIVATELE**

Zde se můžete zaregistrovat jako nový uživatel vyhledávací služby. Povinné položky jsou označeny hvězdičkou\*. Přístup do databázi aktivujete kliknutím na odkaz, který Vám dojde e-mailem jako potvrzení Vaší registrace - získáte tím plný přístup **na dobu 7 dní z tohoto počítače** (IP adresa: 10.1.2.163)

Typ Vašeho připojení k Internetu*	<input type="radio"/> Přímá linka	<input checked="" type="radio"/> Připojení modemem	(důležitý údaj - vyberte správnou volbu)
Přístupové jméno (login)*	<input type="text"/>		(písmena bez diakritiky, 3-20 znaků)
Přístupové heslo*	<input type="text"/>		(písmena bez diakritiky, 3-20 znaků)
Ověření přístupového hesla*	<input type="text"/>		(shoda s předchozím údajem)
E-mailová adresa pro aktivaci*	<input type="text"/>		(e-mail slouží k aktivaci, 3-50 znaků)
Název (firma)	<input type="text"/>		(v případě registrace firmy, 2-50 znaků)
Vaše jméno, příjmení, titul*	<input type="text"/>		(2-100 znaků)
Ulice	<input type="text"/>		(2-50 znaků)
Město, místo*	<input type="text"/>		(2-50 znaků)
PSČ*	<input type="text"/>		(5 čísel bez mezery)
Telefon, fax*	<input type="text"/>		(rychlý kontakt, 3-50 znaků)

**Odeslat registraci**

Copyright © | Design © NKP, Anopress - Ivana Anděrková, Jvo Mittern, 2000-2002 NOM

## Přihlášení

Základní informace    Vstup do databáze    Vyhledávání  
 Charakteristika DB    Jak se zaregistrovat    Jak vyhledávat  
 Další databáze NK    Registrační formulář    Formuláře ≡  
 ≡ Periodika na WWW    Přihlášení    Rejstříky

ANL FULL - plnotextové vyhledávání v člancích z tisku

**PŘIHLÁŠENÍ**  
 Jméno   
 Heslo   
 Přihlášení

Copyright © IJ Design © NKP - Access - Ivana Andříková, Ivo Matěrn, 2000, 2002

## Periodika na WWW - portál

Adresa <http://full.nkp.cz/> Přej

Základní informace    Vstup do databáze    Vyhledávání  
 Charakteristika DB    Jak se zaregistrovat    Jak vyhledávat  
 Další databáze NK    Registrační formulář    Formuláře  
 Periodika na WWW    Přihlášení    Rejstříky

**PERIODIKA NA WWW:**

[I. Regionální periodika](#)  
[II. Oborová periodika](#)  
[III. Další zdroje](#)  
[IV. Další informace](#)  
[V. Popis nabídky](#)

## Nabídka Periodika na WWW – Regionální periodika s detailem

**VYBERTE REGION (KRAJ)**

<a href="#">Praha</a>	<a href="#">Středočeský</a>	<a href="#">Jihočeský</a>
<a href="#">Plzeňský</a>	<a href="#">Karlovarský</a>	<a href="#">Ústecký</a>
<a href="#">Liberecký</a>	<a href="#">Královéhradecký</a>	<a href="#">Pardubický</a>
<a href="#">Moravskoslezský</a>	<a href="#">Zlínský</a>	<a href="#">Olomoucký</a>
<a href="#">Vysočina</a>	<a href="#">Jihomoravský</a>	

[\[Tisk detailu\]](#)

Název	Mariánskolázeňské listy
WWW	<a href="http://www.ml-listy.cz">http://www.ml-listy.cz</a>
Popis	týdeník o dění v Mariánských Lázních
Vydavatel	Art Gallery Nataly
Telefon	01 65/5636
Fax	01 65/3528
E-mail	
Adresa	Ruská 216/48, 353 01 Mariánské Lázně
ISSN	

- [Kv36001.cz \(pozn.: nepřístupné\)](#)  
Popis: týdeník sledující dění v Karlových Varech, zpravodajství, politika, kultura, lidé  
Vydavatel: Ivana Kallinová  
www: <http://www.kv36001.cz>
- [Mariánskolázeňské listy](#)  
Popis: týdeník o dění v Mariánských Lázních  
Vydavatel: Art Gallery Nataly  
www: <http://www.ml-listy.cz>

CELKEM: 2

Číslo 45      10.11.2002      Ročník 11

# MARIÁNSKOLÁZEŇSKÉ LISTY

NEZÁVISLÝ TÝDENÍK ZÁVISLÝ NA DOBRÉ VŮLI OBČANŮ MĚSTA

## MARIÁNSKOLÁZEŇSKÉ LISTY

**ANKETNÍ OTÁZKA**

**Jste pro výstavbu nové radnice s obchodním centrem na místě „centrálního parkoviště“?**

**Ano** (345)     

**Ne** (175)     

Od 20.5. 2002 hlasovalo celkem 520 čtenářů.  
Hlasuj kliknutím na příslušnou odpověď (Hlasovat lze pouze 1x!)

**Vítejte na internetových stránkách Mariánskolázeňských Listů**

*Počínaje 1. prosincem 2001 jsme zprovozili novou službu v rámci tohoto serveru a tou je*

Jste návštěvník číslo: 6592

Přetiskování článků pouze se svolením [majitele](#) ©

Design: © 2001 [Kazon](#)

**Nabídka Periodika na WWW – Oborová periodika**





**Báze ANL FULL – formuláře pro vyhledávání**



**Základní formulář a seznam vyhledaných článků, detail vyhledaného plného textu s metadaty**



	Základní informace	Vstup do databáze	Vyhledávání
	Charakteristika DB	Jak se zaregistrovat	Jak vyhledávat
	Další databáze NK	Registrační formulář	Formuláře
	Periodika na WWW	Přihlášení	Rejstříky

**ANL FULL**

**ZÁKLADNÍ FORMULÁŘ**

Dotaz:  **HLEDAT**

Období pro vyhledávání: od: 10 Listopad 2000 do: 10 Listopad 2002

Typ seznamu:  Jednoduchý  Se souhrnem

**Celkem nalezeno 10 článků** [Tisk seznamu]

1	0.87	20.03.2002	Lidové noviny	str. 6	[13325 B]
   <a href="#">Básník, jenž se národu nezavděčil</a>					
2	0.80	28.06.2002	Pátek magazín LN	str. 12-20	[23394 B]
   <a href="#">Mami, jak jsi tak mohla žít?</a>					
3	0.77	22.12.2001	Mladá fronta Dnes	str. B/2: Civil...	[17152 B]
   <a href="#">Zpěvujte, lidičky, tv naše písničky</a>					
4	0.77	21.07.2001	Magazín Práva	str. 28-29	[18908 B]
   <a href="#">Machar, Josef Svatopluk, 1864-1942</a>					

**Název:** Básník, jenž se národu nezavděčil

**Hlavní autor:** Jaromír Slomek

**Zdroj:** Lidové noviny

**ISSN:** 0862-5921

**Roč.:** 15, č. 67 (20.3.2002), s. 6

**Rubrika:** Domov

**Předmět. ktg.:** literatura: česká literatura

**MDT:** 821.162.3-051, (437.3)

**Osoba jako předmět:** Machar, Josef Svatopluk, 1864-1942

**Téma jako předmět:** básníci-Česko-stol. 19.-20.

**Typ dokumentu:** portréty

**Geograf. kód:** e-xr---

Příběh J. S. MACHARA, "vídeňského Čecha" a kritika domácích poměrů

Seznam výsledků – více názvů článků se souhrnem

	Základní informace	Vstup do databáze	Vyhledávání
	Charakteristika DB	Jak se zaregistrovat	Jak vyhledávat
	Další databáze NK	Registrační formulář	Formuláře
	Periodika na WWW	Přihlášení	Rejstříky

**ANL FULL**

**ZÁKLADNÍ FORMULÁŘ**

Dotaz:  **HLEDAT**

Období pro vyhledávání: od: 10 Listopad 2000 do: 10 Listopad 2002

Typ seznamu:  Jednoduchý  Se souhrnem

**Celkem nalezeno 10 článků** [Tisk seznamu]

1	0.87	20.03.2002	Lidové noviny	str. 6	[13325 B]
   <a href="#">Básník, jenž se národu nezavděčil</a>					
<p>Před šedesáti lety, jen pár týdnů předtím, než si stateční parašutisté dokázali počítat na Reinharda Heydricha, zemřel v Praze básník Josef Svatopluk Machar</p>					
2	0.80	28.06.2002	Pátek magazín LN	str. 12-20	[23394 B]
   <a href="#">Mami, jak jsi tak mohla žít?</a>					
<p>Jsou dnes ženy v Česku - po macharovsku řečeno - stejnými otrokyněmi otroka, za které básník považoval ještě generaci jejich babiček?</p>					

## Vyhledání, označení, stažení a export plného textu s metadaty v různých formátech

**Základní informace**  
 Charakteristika DB  
 Další databáze NK  
 Periodika na WWW

**Vstup do databáze**  
 Jak se zaregistrovat  
 Registrační formulář  
 Odhlášení

**Vyhledávání**  
 Jak vyhledávat  
 Formuláře  
 Rejstříky

**ZÁKLADNÍ FORMULÁŘ**  
 Dotaz:  **HLEDAT**

Období pro vyhledávání: od: 10 Listopad 2000 do: 10 Listopad 2002

Typ seznamu:  Jednoduchý  Se souhrnem

9 0.77 09.05.2002 Reflex str. 80-83 [24726 B]  
 Politik v bílém plášti

10 0.77 06.04.2002 Lidové noviny str. 06 [13000 B]  
 Poslední dny lidstva

Anarchista, jenž pěšky prošel celou Evropou

Zobrazit označené Stáhnout označené

Označit všechny články v seznamu

**Zvolte tvar výstupního souboru:**

**Přípravit texty**

HTML tvar se zvýrazněním klíčových slov v textu:  
 červené  tučné  kurzívou  podtrhnout

HTML tvar bez zvýraznění klíčových slov v textu

Textový tvar se "[ ]" zvýrazněním "[ ]" klíčových slov v textu

Textový tvar bez zvýraznění klíčových slov v textu

RTF formát se zvýrazněním slov (jako u HTML tvaru)

RTF formát bez zvýraznění klíčových slov v textu

XML formát (eXtensible Markup Language)

Soubor Úpravy Zobrazit Oblíbené Nástroje nápověda

Zpět Hledat Oblíbené Média

Adresa  Přejít

**Základní informace**  
 Charakteristika DB  
 Další databáze NK  
 Periodika na WWW

**Vstup do databáze**  
 Jak se zaregistrovat  
 Registrační formulář  
 Odhlášení

**Vyhledávání**  
 Jak vyhledávat  
 Formuláře  
 Rejstříky

**ZÁKLADNÍ FORMULÁŘ**  
 Dotaz:  **HLEDAT**

Období pro vyhledávání: od: 10 Listopad 2000 do: 10 Listopad 2002

Typ seznamu:  Jednoduchý  Se souhrnem

9 0.77 09.05.2002 Reflex str. 80-83 [24726 B]  
 Politik v bílém plášti

10 0.77 06.04.2002 Lidové noviny str. 06 [13000 B]  
 Poslední dny lidstva

Anarchista, jenž pěšky prošel celou Evropou

Zobrazit označené Stáhnout označené

Označit všechny články v seznamu

**Stáhnout soubor**

Stáhnete soubor:  
 and31820.zip z full.nkp.cz

Chcete soubor otevřít nebo jej uložit v počítači?

Otevřít Uložit Storno Další informace


Před otevřením souboru tohoto typu se vždy dotázat

**Stáhnout soubor**

and31820.ZIP je připraven ke stažení s komprimovanými články: 15958 B

## Rozšířený formulář a formulace dotazu



	Základní informace	Vstup do databáze	Vyhledávání
	Charakteristika DB	Jak se zaregistrovat	Jak vyhledávat
	Další databáze NK	Registrační formulář	Formuláře
	Periodika na WWW	Odhlášení	Rejstříky
<b>ANL FULL</b>			
Období od: 10 Listopad 2000 do: 10 Listopad 2002			datum vydání
Název, autor:	AND básník		název článku
Číselné údaje:	AND		název článku
Předmět:	AND Machar		ročník/svazek
Další výběry:	AND --- Všechny typy článků ---		ročník/svazek
Úsob vyhledávání řetězců:	Postačí výskyt zadaných řetězců kdekoliv v poli: <CONTAINS>		osoby
			předmět, ktg.
			<typ článku>
			<název zdroje>
			(neplatí pro čísla)
Seznam: Jednoduchý	Skóre: 0	Výsíl./str: 10	Třídění: Skóre Sestupně
			<a href="#">Search Tips</a> , <a href="#">Jak vyhledávat</a> <a href="#">Stahování a export</a>

**Uživatelský formát: údaje o článku s odkazy (citace, event. souhrn)**

Celkem nalezeno 1 článků [Tisk seznamu]		[Tisk článku] ani dořmeta dořhítmí: dořhítmí: dořhítmí: dořhítmí:	
1	0.87	20.03.2002	Lidové noviny str. 6 [13325 B]
<input type="checkbox"/> <a href="#">Básník, jenž se národu nezavděčil</a>			
<input type="checkbox"/> <a href="#">Údaje o článku s odkazy pro další hledání</a>		<input type="checkbox"/> <a href="#">Stáhnout označené</a>	
<input type="checkbox"/> <a href="#">Označit všechny články v seznamu</a>			
		<b>ÚDAJE O ČLÁNKU S ODKAZY</b>	
<b>Název:</b>	<a href="#">Básník, jenž se národu nezavděčil</a>		
<b>Hlavní autor:</b>	<a href="#">Jaromír Slomek</a>		
<b>Zdroj:</b>	<a href="#">Lidové noviny</a>		
<b>ISSN:</b>	<a href="#">0862-5921</a>		
<b>Roč.:</b>	15, č. 67 (20.3.2002), s. 6		
<b>Rubrika:</b>	<a href="#">Domov</a>		
<b>Předmět. ktg.:</b>	<a href="#">literatura: česká literatura</a>		
<b>MDT:</b>	<a href="#">821.162.3-051, (437.3)</a>		
<b>Osoba jako předmět:</b>	<a href="#">Machar, Josef Svatopluk, 18641942</a>		
<b>Téma jako předmět:</b>	<a href="#">básníci-Česko-stol. 19.-20.</a>		
<b>Typ dokumentu:</b>	<a href="#">portréty</a>		
<b>Geograf. kód:</b>	<a href="#">e-xf---</a>		
...			

## Uživatelský formát: údaje o článku s plným textem

The screenshot shows a web interface with a search results list on the left and a detailed view of an article on the right. The search results list includes the following information:

- Celkem nalezeno 1 článků [Tisk seznamu]
- 1 0.87 20.03.2002 Lidové noviny str. 6 [13325 B]
- [Básník, jenž se národu nezavděčil](#)

A tooltip is visible over the article title, containing the text: "Text článku se zvýrazněním kl. slov a s odkazy pro další hledání".

The detailed view on the right is titled "TEXT ČLÁNKU S ODKAZY" and contains the following metadata:

- Název:** [Básník, jenž se národu nezavděčil](#)
- Hlavní autor:** [Jaromír Slomek](#)
- Zdroj:** [Lidové noviny](#)
- ISSN:** [0862-5921](#)
- Roč.:** 15, č. 67 (20.3.2002), s. 6
- Rubrika:** [Domov](#)
- Předmět. ktg.:** [literatura česká literatura](#)
- MDT:** [821.162.3-051 \(437.3\)](#)

Below the metadata, there are additional fields:

- Osoba jako předmět:** [Machar, Josef Svatopluk, 18641942](#)
- Téma jako předmět:** [básníci-Česko-stol. 19.-20.](#)
- Typ dokumentu:** [portréty](#)
- Geograf. kód:** [8-XI---](#)

The main text of the article begins with: "Příběh J. S. MACHARA, 'vídeňského Čecha' a kritika domácích poměrů". Below this, the text is categorized as "ČESKÉ LEGENDY 20. STOLETÍ" and starts with the sentence: "Před šedesáti lety, jen pár týdnů předtím, než si stateční parašutisté dokázali počítat na Reinharda Heydricha,".

## Uživatelský formát: Dublin Core

The screenshot shows the same web interface as above, but with the "Dublin Core Metadata" tab selected. The metadata is displayed in a structured format:

- Title:** [Básník, jenž se národu nezavděčil](#)
- Creator.personalName:** Slomek, Jaromir
- Subject:** čeští básníci
- Subject, scheme=ANL-KATEGORIE:** literatura: česká literatura
- Subject, scheme=MDT:** 821.162.3-051
- Subject, scheme=MDT:** (437.3)
- Subject, scheme=PRNK:** Machar, Josef Svatopluk, 1864-1942
- Subject, scheme=PRNK:** básníci-Česko-stol. 19.-20.
- Coverage.spatial:** Česko
- Coverage.temporal:** stol. 19.-20.
- Description.abstract:** Před šedesáti lety, jen pár týdnů předtím,

## Pracovní formát: UNIMARC

The screenshot shows a web interface with two main panels. The left panel displays article metadata: 'Celkem nalezeno 1 článků', '1 0.87 20.03.2002 Lidové noviny str. 6 [13325 B]', and the title 'Básník, jenž se národu nezavděčil'. Below this are buttons for 'Zobrazit označené' and 'Stáhnout označené', and a checkbox 'Označit všechny články v seznamu'. The right panel, titled 'UNIMARC', displays a list of MARC fields in a text-based format. The fields include:

- LAB -----naa##22-----###450#
- 001 an2002abdf0000792
- 100##a20020326j20020320u##y0czey0103###ba
- 1010#caeze
- 102##caCZ
- 105##caaa###z###000yy
- 014##ca0862-5921(20020320)15:67<6:LN2002032000
- 2001#caBásník, jenž se národu nezavděčilfJar
- 215##c1 1l
- 304##caRubrika: Domov
- 330##caPřed šedesáti lety, jen pár týdnů předt
- 461#l1011##ca0862-5921#l2001#caLidové noviny
- 5451#caDomov
- 600 1#aMachar#bJosef Svatopluk#f1864-1942
- 6060#cabásníci#yČesko#zsol. 19.-20.
- 608##caportréty
- 6100#cačeští básníci
- 615##caliteratura#xčeská literatura#n82
- 660##cae-Xr---
- 675##ca821.162.3-051
- 675##ca(437.3)
- 700#l#aSlomek#bJaromír

## Pracovní formát: DC

The screenshot shows a web interface similar to the UNIMARC example. The left panel displays the same article metadata. The right panel, titled 'DC/HTML/META', displays Dublin Core metadata in HTML format. The metadata includes:

- <META name="DC.Title" content="Básník, jenž se národu nezavděčil">
- <LINK rel="SCHEMA.dc" href="http://purl.org/metc/1.11">
- <META name="DC.Creator.personalName" content="Jaromír Slomek">
- <LINK rel="SCHEMA.dc" href="http://purl.org/metc/1.11">
- <META name="DC.Subject" content="čeští básníci">
- <LINK rel="SCHEMA.dc" href="http://purl.org/metc/1.11">
- <META name="DC.Subject" scheme="ANL-KATEGORIE" content="ANL-82">
- <LINK rel="SCHEMA.dc" href="http://purl.org/metc/1.11">
- <META name="DC.Subject" scheme="HDT" content="HDT-1567">
- <LINK rel="SCHEMA.dc" href="http://purl.org/metc/1.11">
- <META name="DC.Subject" scheme="HDT" content="HDT-1567">
- <LINK rel="SCHEMA.dc" href="http://purl.org/metc/1.11">
- <META name="DC.Subject" scheme="PHNK" content="PHNK-437.3">
- <LINK rel="SCHEMA.dc" href="http://purl.org/metc/1.11">
- <META name="DC.Subject" scheme="PHNK" content="PHNK-437.3">
- <LINK rel="SCHEMA.dc" href="http://purl.org/metc/1.11">



## Pracovní formát: XHTML kvalifikovaný

The screenshot shows a search result interface. The top bar indicates 'Celkem nalezeno 1 článků' and '[Tisk seznamu]'. Below it, search details include '1 0.87 20.03.2002 Lidové noviny str. 6 [13325 B]'. The article title is 'Básník, jenž se národu nezavděčil'. There are buttons for 'Zobrazit označené' and 'Stáhnout označené', and a checkbox for 'Označit všechny články v seznamu'. The right pane, titled '[Tisk detailu]', displays the DC/XHTML metadata for the article.

```
DC/XHTML - KVALIFIKOVANÝ

<link rel="schema.DC" href="http://purl.org/dc/
<meta name="DC.Title" content="Básník, jenž se
<meta name="DC.Creator.personalName" content='
<meta name="DC.Subject" content="česti básníci
<meta name="DC.Subject" scheme="ANL-KATEGORIE'
<meta name="DC.Subject" scheme="MDT" content='
<meta name="DC.Subject" scheme="MDT" content='
<meta name="DC.Subject" scheme="PHNK" content=
<meta name="DC.Subject" scheme="PHNK" content=
<meta name="DC.Coverage.spatial" content="Česk
<meta name="DC.Coverage.temporal" content="st
<meta name="DC.Description.abstract" content='
<meta name="DC.Publisher" content="Lidové nov
<meta name="DC.Date.issued" scheme="W3C-DTF" c
<meta name="DC.Type" scheme="DCMI" content="te
<meta name="DC.Type" content="article"/>
<meta name="DC.Type" content="článek"/>
<meta name="DC.Type" content="portréty"/>
<meta name="DC.Format" scheme="IMT" content="t
<meta name="DC.Identifier" content="http://ful
<meta name="DC.Identifier" scheme="SICI" conte
<meta name="DC.Identifier" scheme="URN" conter
<meta name="DC.Source" content="Lidové noviny
```

## Pracovní formát: XML kvalifikovaný

This screenshot shows the same search result interface as above, but the right pane displays the XML metadata for the article. The XML is an RDF document with namespaces for rdf, dc, and dcq.

```
<?xml version="1.0" encoding="UTF-8" ?
>
- <rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/
  -rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements
  xmlns:dcq="http://purl.org/dc/docume
  -qualifiers-20000711.htm">
- <rdf:Description rdf:about="Básník,
  jenž se národu nezavděčil">
  <dc:Title>Básník, jenž se
  národu
  nezavděčil</dc:Title>
- <dc:Creator>
  - <rdf:Description>
    <dcq:CreatorType>personalName
    <rdf:value>Slomek,
    Jaromír</rdf:value>
  </rdf:Description>
  </dc:Creator>
- <dc:Subject>
- <rdf:Bag>
  <rdf:li>Čeští
```

## Nabídka rejstříků

The screenshot shows the NK database website interface. At the top, there is a navigation menu with three columns: 'Základní informace', 'Vstup do databáze', and 'Vyhledávání'. Below this, there is a 'VÝBĚR REJSTŘÍKU' (Index Selection) section with a list of index types: 'Auroři', 'Název zdroje', 'Zdroj, ročník, číslo', 'Datum vydání', 'Předmět, kategorie', 'Předmět', 'Místo', 'Hesla', 'Osoby', 'Korporace', 'Akce', and 'Klíčová slova'. The 'Předmět' (Subject) option is highlighted.

## Detail předmětového rejstříku – Předmětové kategorie (termíny NT a BT) – nutná redakce rejstříku za účelem vytvoření tezauru

The screenshot shows the NK database website interface, specifically the 'PŘEDMĚT. KTG.' (Subject Categories) detail view. The page title is '[Výběr rejstříku] [Tisk] PŘEDMĚT. KTG.'. Below the title, there is a search filter for the period: 'Období pro vyhledávání: od 2001 do 2002'. The main content area displays a list of subject categories under the letter 'M': 'maďarská kinematografie NT', 'maďarská literatura NT', 'maďarská literatura NT', 'maďarské dějiny NT', 'maďarské divadlo NT', 'maďarské hrané filmy NT', 'malířství BT', 'malířství NT', 'malířství BT', and 'malířství BT'. The 'malířství BT' category is highlighted.

## Předmětový rejstřík - Osoby - se seznamem vyhledaných článků

	Základní informace	Vstup do databáze	Vyhledávání
	Charakteristika DB	Jak se zaregistrovat	Jak vyhledávat
	Další databáze NK	Registrační formulář	Formuláře
	Periodika na WWW	Přihlášení	Rejstříky

<p><b>Celkem nalezeno 5 článků</b> <a href="#">[Tisk seznamu]</a></p>					
1	0.77	21.12.2001	Mladá fronta Dnes	str. C/8: Sport	[14587 B]
<p><a href="#">Zvonokosům věvodí Jitka Molavcová - slídí, donáší a intrikuje</a></p>					
2	0.77	17.12.2001	Právo	str. 15	[14821 B]
<p><a href="#">Karlínské Zvonokosy roby tragédie staré panny</a></p>					
3	0.77	10.12.2001	Pražské Slovo	str. 28	[12753 B]
<p><a href="#">Šťavnaté Zvonokosy</a></p>					
4	0.77	10.12.2001	Lidové noviny	str. 30	[14447 B]
<p><a href="#">Socialistický muzikál znovu na jevišti</a></p>					
5	0.77	06.12.2001	Pražské Slovo	str. 26	[12694 B]
<p><a href="#">Zvonokosy se vraceli do Karlína</a></p>					

Vyhledávání pomocí topiků – navolena tematická oblast (přírodní a matematické obory), skupina témat (matematika, fyzika), detailní téma (astronomie, astrofyzika); stejný dotaz bez navoleného topiků

	Základní informace	Vstup do databáze	Vyhledávání
	Charakteristika DB	Jak se zaregistrovat	Jak vyhledávat
	Další databáze NK	Registrační formulář	Formuláře
	Periodika na WWW	Přihlášení	Rejstříky

<p><a href="#">Návrat na homepage FULL.NKP.CZ</a></p>		<p><b>ROZŠÍŘENÝ FORMULÁŘ S TÉMATY</b></p>		<p><b>SEZNAM TÉMAT</b></p>
Hledat v textu:	hvězdy	<p><b>VYHLEDAT</b></p>		<p>Tematické oblasti:</p> <p>Přir. a matematické ob</p>
Období od:	10 Listopad 2000 do: 10 Listopad 2002	datum vydání		<p>Skupiny témat:</p> <p>Matematika, fyzika</p>
Název, autor:	AND	název článku		<p>Detailní témata:</p> <p>...Bez detailních témat...</p> <p>» Astronomie, astrofyz</p>
Číselné údaje:	AND	název článku		<p>» Fyzika</p>
Předmět:	AND	ročník/svazek		<p>» Kybernetika, robotika</p>
Další výběry:	AND	ročník/svazek		<p>» Matematika</p>
	AND	předmět, ktg.		<p>» Statistika</p>
	AND	předmět, ktg.		<p>» Umělá inteligence</p>
	AND	<typ článku>		<p>Vazba mezi tématy</p> <p><input checked="" type="radio"/> A <input type="radio"/> Nebo</p> <p>Vazba dotaz - témat</p>
	AND	<název zdroje>		



Celkem nalezeno 60 článků [Tisk seznamu]				
1 2 3 4 5 6 Následující				
1	0.93	19.10.2002	Mladá fronta Dnes	str. C/9: Sport [10819 B]
<a href="#">Astronomové objevili dvě nové exoplanety</a>				
2	0.92	25.05.2002	Mladá fronta Dnes	str. C/8: Věda [12123 B]
<a href="#">Svědkové vesmírných katastrof</a>				
3	0.92	13.07.2002	Mladá fronta Dnes	str. C/9: Sport [11897 B]
<a href="#">Nejtěžší hvězda ve vesmíru</a>				
4	0.91	11.04.2002	Reflex	str. 54-56 [21626 B]
<a href="#">Slunce by zářilo modře</a>				
5	0.91	06.04.2002	Mladá fronta Dnes	str. C/7: Sport [11124 B]
<a href="#">Kolem Věgy asi obíhá planeta</a>				
6	0.89	19.02.2002	Hospodářské noviny	str. 01: Věda a ... [15033 B]
<a href="#">Hvězdárna snů na odvrácené straně Měsíce</a>				
7	0.89	20.10.2001	Magazín Práva	str. 36-37 [20895 B]
<a href="#">Pražská léta dánského hvězdáře</a>				
8	0.89	11.05.2001	Mladá fronta Dnes	str. 14 [6896 B]
<a href="#">Vesmír, hvězda spojila planetu</a>				
9	0.87	18.06.2002	Týden	str. 40-45 [27851 B]
<a href="#">Češi jdou do nebe</a>				
100.85		30.11.2000	Reflex	str. 84 [22265 B]
<a href="#">MYSLET VESMÍR</a>				
1 2 3 4 5 6 Následující				

[Tisk článku] [uni](#) [do/meta](#) [do/xhtml](#) [do/xhtml](#) [do/xml](#) [do/xml](#)

**Název:** Nejtěžší hvězda ve vesmíru  
**Podnázev:** Hubbleův kosmický dalekohled umožnil as  
**Hlavní autor:** Petr Sobotka  
**Zdroj:** Mladá fronta Dnes Sport  
**ISSN:** 1210-1168  
**Roč.:** 13, č. 161 (13.7.2002), s. C/9  
**Rubrika:** Věda  
**Předmět. ktg.:** astronomie  
**MDT:** 524.3

**Téma jako předmět:** hvězdy-výzkum-stol. 21., počátek

PETR SOBOTKA  
 Naše Slunce je ve vesmíru jen nepatrným drobečkem. Američtí astronomové našli zatím nejtěžší známou hvězdu. Má hmotnost téměř jako šedesát Sluncí. Pomocí Hubbleova kosmického dalekohledu vědci objevili na jižní obloze ve hvězdokupě R136 v galaxii zvané Velké Magellanovo mračno tři nové zářivé ...

**PRO ZOBRAZENÍ PLNÝCH TEXTŮ JE TŘEBA SE PŘIHLÁSIT NEBO ZAREGISTROVAT NA SERVERU [FULL.NKP.CZ](http://FULL.NKP.CZ)**

Celkem nalezeno 2214 článků [Tisk seznamu]				
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 Následující				
1	1.00	13.07.2002	Mladá fronta Dnes	str. C/9: Sport [11897 B]
<a href="#">Nejtěžší hvězda ve vesmíru</a>				
2	0.97	12.04.2002	Haló noviny	str. 2-3: Obrys ... [26172 B]
<a href="#">Vaše čokoládová hvězda</a>				
3	0.94	19.10.2002	Mladá fronta Dnes	str. C/9: Sport [10819 B]
<a href="#">Astronomové objevili dvě nové exoplanety</a>				
4	0.93	06.04.2002	Mladá fronta Dnes	str. C/7: Sport [11124 B]
<a href="#">Kolem Věgy asi obíhá planeta</a>				
5	0.92	01.09.2002	Kapitál	str. 14-16 [29220 B]
<a href="#">Alchymie vztahů</a>				
6	0.92	25.05.2002	Mladá fronta Dnes	str. C/8: Věda [12123 B]
<a href="#">Svědkové vesmírných katastrof</a>				
7	0.92	02.01.2001	Týden	str. 26 [23512 B]
<a href="#">Chrtiči televizní srandy</a>				
8	0.91	11.04.2002	Reflex	str. 54-56 [21626 B]
<a href="#">Slunce by zářilo modře</a>				
9	0.91	25.04.2002	Právo	str. 1, 2: Salon [19573 B]
<a href="#">Poetická revoluce</a>				
100.91		05.03.2001	Respekt	str. 18 [10259 B]
<a href="#">Koruna z dvanácti hvězd nad hlavou</a>				

[Tisk článku] [uni](#) [do/meta](#) [do/xhtml](#) [do/xhtml](#) [do/xml](#) [do/xml](#)

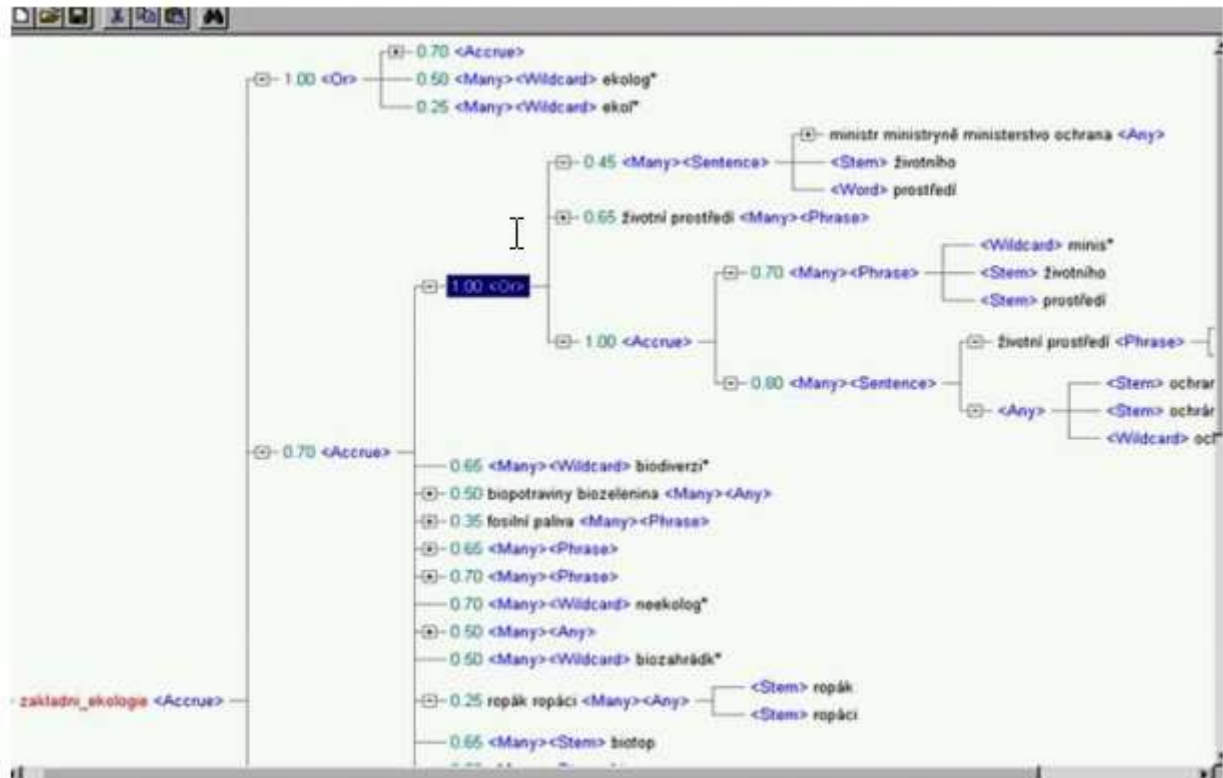
**Název:** Slunce by zářilo modře  
**Hlavní autor:** Jan Palouš  
**Další autor:** Jan Zvelebil  
**Zdroj:** Reflex  
**ISSN:** 0862-6634  
**Roč.:** 13, č. 15 (11.4.2002), s. 54-56  
**Rubrika:** Rozhovor  
**Předmět. ktg.:** astronomie  
**MDT:** 52-051, (047.53), (437.3)

**Osoba jako předmět:** Palouš, Jan, 1949-  
**Téma jako předmět:** astronomové-Česko-r. 2002  
**Typ dokumentu:** rozhovory  
**Geograf. kód:** e-xr--

Astronomie je snad jediná z přírodních věd, která si na předměty svého bádání nemůže "posvítit" zblízka. Zároveň jako jediná věda vůbec zkoumá vesmír jako celek, od zemské atmosféry až po vzdálené kvasary. Předpovídá budoucnost Slunce, hledí miliardy let do minulosti. Co si myslet o obrazu světa, který nabízí? Profesor J...

**PRO ZOBRAZENÍ PLNÝCH TEXTŮ JE TŘEBA SE PŘIHLÁSIT NEBO ZAREGISTROVAT NA SERVERU [FULL.NKP.CZ](http://FULL.NKP.CZ)**

## Ukázka topiku ekologie



## Administrace

### NKP - Administrace

[Úprava textu článku](#) **NOVÉ**

[Statistika přístupů](#)

[Údržba rejstříků podle počátečních písmen](#)

[Údržba rejstříků podle data zápisu](#)

[Údržba regionálních periodik](#)

[Údržba oborových periodik](#)

[Informační zdroje](#)

[Skupiny témat](#)

[Uživatelé \(přidávání, rušení, změny\)](#)

[Změny hesel uživatelů](#)

[Mailů \(komu se posílá registrace\)](#)





## Údržba regionálních a oborových periodik.

Pomocí této části aplikace je možno aktualizovat regionální a oborová periodika v nabídce Portál

**Údržba**

[Přidat](#) [Zpět k nabídce](#)

10 01 - Čas videa  
10 02 - Harmonie  
10 02 - Hudební rozhledy  
10 02 - Opus Musicum  
10 03 - Divadelní noviny  
10 03 - Divadelní noviny  
10 04 - HOST  
10 04 - Literární noviny  
10 04 - Tramvaj NAČERNO  
10 04 - Tvar  
20 01 - ARCHINET  
20 01 - Architekt  
20 01 - Fórum architektury a stavitelství  
20 02 - ASTROPIS  
20 02 - JihoČAS  
20 03 - Československý časopis pro fyziku  
20 03 - CHEMagazin  
20 03 - Chemické Listy

Oblast: Odborné - Přírodní vědy  
Název: Chemické Listy  
WWW: http://chemicke-listy.vscht.cz/  
Popis: měsíčník, abstrakty publikovaných článků  
Vydavatel: Chemické listy  
Telefon: 0222 22 17 78  
Fax: 0222 22 01 84  
E-mail: jiratova@csvts.cz  
Adresa: Novotného lávka 5, 116 68 Praha 1  
ISSN: 1213-7103

[Změnit](#) [Zrušit detail](#)

## Údržba seznamu zdrojů

Pomocí této aplikace je možno udržovat seznam titulů, které jsou obsaženy v bázi ANL FULL

**Údržba seznamu zdrojů**

[Přidat nový zdroj](#) [Zpět k nabídce](#)

Řazení: [Podle zkratky zdroje](#) | [Podle názvu zdroje](#)

CELOSTÁTNÍ TITULY  
[1] [BL Blesk](#)  
[1] [HA Haló noviny](#)  
[1] [HN Hospodářské noviny](#)  
[1] [LN Lidové noviny](#)  
[1] [MF Mladá fronta Dnes](#)  
[1] [PR Právo](#)  
[1] [S1 Slovo](#)  
[1] [Z1 ZN Zemské noviny](#)

ČASOPISY A MAGAZÍNY  
[1] [BA Bankovníctví](#)  
[1] [EK Ekonom](#)  
[1] [L Magazín LN](#)  
[1] [NK Národní knihovna](#)  
[1] [P Magazín Práva](#)

**Nový zdroj**

Historický zdroj:   
Zkratka zdroje:   
Název zdroje:   
ISSN zdroje:   
Vydavatel:   
Místo vydání:   
Oblast: Celostátní tituly  
Relace:   
Poznámka:   
Zobrazovat:   
[Přidat](#)

## Údržba témat

Pomocí této aplikace lze udržovat témata (topiky)



The screenshot displays the 'Údržba témat' (Topic Maintenance) application interface. It is divided into two main sections: 'Údržba témat' (Topic Maintenance) and 'Detail'.

**Údržba témat**

Buttons: [Přidat skupinu témat](#) | [Zpět k nabídce](#)

- [001 - Ekonomika, ekonomie \[SEZNAM TÉMAT\]](#)
- [002 - Finance \[SEZNAM TÉMAT\]](#)
- [003 - > Obchod \[SEZNAM TÉMAT\]](#)
- [004 - > Práce \[SEZNAM TÉMAT\]](#)
- [024 - > Vpočetní technika \[SEZNAM TÉMAT\]](#)
- [025 - Doprava, spoje \[SEZNAM TÉMAT\]](#)
- [026 - Energetika \[SEZNAM TÉMAT\]](#)
- [027 - > Průmysl \[SEZNAM TÉMAT\]](#)
- [028 - Zemědělství \[SEZNAM TÉMAT\]](#)
- [029 - > Metrologie, normy, standardy \[SEZNAM TÉMAT\]](#)
- [030 - > Kultura \[SEZNAM TÉMAT\]](#)

**Detail**

Typ (číslo):

Název skupiny:

[Zrušit detail](#)

Aplikace umožňuje podobně udržovat i seznamy uživatelů, hesel, mailů.

## Příloha F 4

Tituly v bázi ANL FULL, ANL, volně na WWW, Dublin Core a ANL Core - tabulka návěští pro vyhledávání v poli dotaz, tabulka formulářových polí, topiky, citace článku

a) Tituly v bázi ANL FULL, ANL, Anopress, volně přístupné na www

### Plné texty v českých novinách a časopisech – báze ANL FULL, ANL

NOVINY A ČASOPISY ZPŘÍSTUPŇOVANÉ V NKČR A V ANOPRESSU (REGISTRACE, KONSORCIUM)	2003	2002	2001	2000	1999	1998	1990 - 1996	1997	Suma
Titul, počet článků v <b>ANL FULL</b> vydaných v 1990-31.10.2002 (zpracovaných v 1999-2003, 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 <b>TamTam (Anopress)</b>									
V bázi <b>ANL</b> jsou přístupné tytéž tituly/články jako v ANL FULL, a to přes bibliografický záznam a link do ANL FULL									
<b>Bankovníctví &gt;&gt;&gt;</b>	101	117	33	72				20	343
<b>Berounský deník &gt;&gt;&gt;</b>					96	154		189	439
<b>České Slovo &gt;&gt;&gt;</b>			310						310
<b>Deník Jablonecka &gt;&gt;&gt;</b>					199	94		444	737
<b>Euro &gt;&gt;&gt;</b>	30	163							193
<b>Ekonom &gt;&gt;&gt;</b>	451	475	1002	1071	46	14		385	3444
<b>Haló noviny &gt;&gt;&gt;</b>	1216	1293	1074	997	1340	1327		780	8027
<b>Hanácký a středomoravský den &gt;&gt;&gt;</b>				182					182
<b>Hospodářské noviny &gt;&gt;&gt;</b>	1660	1906	1652	2227	3094	2631		2782	15954
<b>Hradecké noviny &gt;&gt;&gt;</b>				23	148	260		389	820
<b>Chebský deník &gt;&gt;&gt;</b>					32	148		156	336
<b>Kapitál &gt;&gt;&gt;</b>	100	77	3	97	178				430
<b>Kladenský deník &gt;&gt;&gt;</b>					256	19			275
<b>Liberecký den &gt;&gt;&gt;</b>					1			73	74
<b>Lidové noviny &gt;&gt;&gt;</b>	1662	2910	2265	1841	3575	4789		4110	21153
<b>Magazín LN &gt;&gt;&gt;</b>			22	56					78
<b>Magazín Práva &gt;&gt;&gt;</b>	171	168	165	47					521
<b>Mladá fronta Dnes &gt;&gt;&gt;</b>	2361	2718	3140	2401	2489	2998		4104	20754





<b>Odpady</b>								9
<b>Psychiatrie &gt;&gt;</b>				9	15			24
<b>Psychologie dnes &gt;&gt;</b>	28	120	119	76	29			372
<b>Revue církevního práva</b>								63
<b>Souvislosti</b>								29
<b>Učitel'ské noviny &gt;&gt;</b>								
<b>U nás &gt;&gt;</b>		69	16	67	24			107
<b>Veřejná správa &gt;&gt;</b>	168	318	397	324	249	35		1491
<b>Vesmír &gt;&gt;</b>	85	145	101	73	86	69		559
<b>Věstník (příl. Obec a finance)</b>								128
<b>Země světa</b>								111
<b>Celkem</b>	486	1777	984	1084	785	289	22	6408

10.11.2003

*b) Dublin Core, ANL CORE s vyhledáváním v bázi ANL FULL, metadata*

<b>Dublin Core prvky s kvalifikátorem</b>	<b>Použité Schéma (kvalifikátor hodnoty)</b>	<b>ANL CORE Prvky*</b>	<b>Hodnota</b>	<b>Prefixy pro vyhledávání (Čes.)***</b>	<b>Prefixy pro vyhledávání (Ang.)***</b>
DC.Title			Název	Naz	ti
DC.Title		anl.tipart	Název části/Název článku na pokračování	Nan	tip
DC.Subtitle*			podnázev	Pdn	sbt
DC.Creator.personal Name			Autor (hlavní, primární)	Auh	aup
DC.Creator.corporateName			Autor (hlavní, primární)	Auh	aup
DC.Contributor			Přispěvatel (vedlejší, sekundární, další autor)	Auv	aus
DC.Publisher			Vydavatel	Vyd	pub
DC.Source			Zdroj (název, roč., č. datum, strany)	Zdr	Src
		anl.sopart	Zdroj (název části)	Zdn	sop
		anl.cap	Rubrika	Rbr	cap

DC.Source	ISSN	anl.issn	Zdroj (ISSN)	Issn	issn
		anl.vol	Svazek/ročník	Roc	vol
		anl.nr	Číslo	Cis	nr
DC.Date.issued	W3C-DTF		Datum (vydání)		
DC.Date.modified	W3C-DTF		Datum (zpracování)	Dac	dac
DC.Type	DCMI		Typ (popisovanéh o dokumentu)		
DC.Type			Typ/druh (popisovanéh o dokumentu)		
DC.Format			Format (popisovanéh o dokumentu)		
DC.Language	RFC1766		Jazyk		
DC.Rights			Práva		
DC.Identifier	SICI		Identifikátor		
DC.Identifier	URN		Identifikátor		
DC.Identifier			Identifikátor (URL)		
		anl.id	Identifikátor (identifikační číslo článku)	Id	id
DC.Coverage.spatial	GAC		Pokrytí (prostorové)	Zek	gac
DC.Coverage.spatial			Pokrytí (prostorové)		
DC.Coverage.temporal			Pokrytí (časové)		
DC.Subject			Předmět (klíčová slova)	Kls	key
		anl.key	Předmět (klíčová slova)	Kls	key
DC.Subject	PHNK		Předmět, téma jako předmět, heslo (PHNK)	Phnk	phnk
		anl.sub	Předmět, téma jako předmět, heslo (PHNK)	Sub	sub

DC.Subject	PHNK		Předmět, geografické jméno jako předmět	Phnk	phnk
		anl.geo		Geo	geo
DC.Subject	ANL- KATEGOR IE	anl.topic	Předmět (předmětové kategorie ANL)	Top	top
DC.Subject					
DC.Subject	PHNK		Předmět (osoba jako předmět)	Phnk	Phnk
		anl.pers on	Předmět (osoba jako předmět)	Oso	Per
DC.Subject	PHNK		Předmět (korporace jako předmět)	Phnk	Phnk
		anl.corp oration	Předmět (korporace jako předmět)	Kor	cor
DC.Subject	PHNK		Předmět (akce jako předmět)	Phnk	phnk
		anl.act	Předmět (akce jako předmět)	Akc	act
DC.Relation	zatím nedefinován				

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

*Je-li případně tečkové 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

**c) Tabulka formulářových polí a metodika vyhledávání v těchto polích rozšířený formulář pro vyhledávání a rozšířený formulář pro vyhledávání s tématy) viz hlavní text**

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>.**

Poznámka 1: př. 1 je vždy pro operátor <contains>, př. 2 pro =.

Poznámka 2: doporučuje se nastavení <contains>, přesné znění je lépe vyhledávat z rejstříku.

Poznámka 3. pozor: Příjmení, Křestní jméno - vyhledají se všechna jména obsahující některé ze slov, tvar s čárkou lépe hledat v rejstříku. Příklady uvedené pro výběr "všichni autoři" platí i pro hlavní a vedlejší autory.

Poznámka 4: doporučujeme navolit zdroj spíše mocí nabídky Název zdroje ve spodní části obrazovky.

Poznámka 5: nezobrazuje se v uživatelských formátech.

Poznámka 6: určeno spíše k pracovním účelům.

Poznámka 7: nezobrazuje se v uživatelských formátech, je možno je zobrazit v pracovním formátu při zobrazení plného textu - nabídka key).

Poznámka 8: tento typ vyhledávání se bude ještě vyvíjet.

Poznámka 9: jednotlivé nabídky: všechny články, komentáře, rozhovory, jubilejní články, recenze, .sloupky, novely, povídky, básně .

Poznámka 10: jednotlivé nabídky: názvy jednotlivých zdrojových dokumentů.

<b>Způsoby vyhledávání v bázi ANL FULL</b>  - rozšířený formulář	<u><b>Implicitní pole</b></u>		
<i>Název údaje v češtině</i>	<i>Typ údaje</i>	<b>Příklady pro kombinace polí</b>  <b>Poznámka 2</b>	<b>Příklady pro vyhledání údajů v rámci jednoho pole</b>  <b>Poznámka 1</b>  <b>Poznámka 2</b>
<b>Období od do</b>	<i>Datum vydání</i>		Je třeba navolit
	<i>Datum zpracování</i>		Je třeba navolit
<b>Název, autor</b>	<i>Všechny názvy</i>	Př:  Rafinovaný  and  Odraz	1.cesta  2.Cesta mezi hlavou a rukou  2. Rafinovaný odraz skutečnosti
	<i>Všichni autoři</i>	Př.:  Kosatík  and  Pavel  Př.:  Kosatík  and  Hvížd'ala  <b>Poznámka 2</b>  <b>Poznámka 3</b>  Př.:  Vladimír Dlouhý  and  Petr Šimůnek	Př.:  1.Kosatík  2.Pavel Kosatík  <b>Poznámka 2</b>  <b>Poznámka 3</b>

	<i>Hlavní autor</i>	Př:  Kosatík  and  Pavel	1.Kosatík  2. Pavel Kosatík
	<i>Další autor</i>	Př:  Hvížd'ala  and  Karel	1.Hvížd'ala  2. Karel Hvížd'ala
	<i>Název článku</i>	Př:  Čti  and  Skákej	1.skákej  2. Čti a skákej: Nebe, peklo, ráj
	<i>Podnázev článku</i>	Př.:  Publicista  and  Spisovatel	1.publicista  2. Když některé věci nenapíšu, nikdy se je nedozvím, říká publicista a spisovatel Pavel Kosatík
	<i>Název zdroje</i>	Př.:  Fronta  and  Dnes  Př. fronta  and  Slovo  <b>Poznámka 4</b>	1.fronta  2.Mladá fronta Dnes  <b>Poznámka 4</b>
	<i>Název přílohy zdroje</i>	Př.  Čtení  and  Víkend	1.Ekonomika  2. Čtení na víkend

	<i>Název rubriky</i>	Př. Kultura and Pohledy	1.kultura 2.Kultura-pohledy
	<i>Název vydavatele</i>	Př.: Vltava and Press <b>Poznámka 5</b>	1.Vltava 2.Vltava-Labe-Press <b>Poznámka 5</b>
<b>Číselné údaje</b>	<i>Ročník/svazek</i>	Př.: 12	12
	<i>Číslo vydání</i>	204	204
	<i>Strana</i>	C/5	C/5
	<i>Identifikační číslo</i>	Př.: HN20010424000077 and PR20010907000027 <b>Poznámka 5, 6</b>	HN20010424000077 <b>Poznámka 5, 6</b>
	<i>ISSN</i>	1210-1168 and 1212-4273	1210-1168
<b>Předmět</b> <b>Poznámka 2</b>	<i>Předmětové kategorie</i>	Př.: Literatura and Česká <b>Poznámka 2</b>	1.literatura 2.literatura: česká literatura <b>Poznámka 2, 3</b>



	<i>Klíčová slova</i>	Př:  Američtí  and  Spisovatelé  <b>Poznámka 7)</b>	1.američtí  2.američtí spisovatelé  <b>Poznámka 7</b>
	<i>Předmětová hesla</i>	Př.:  Spisovatelé  and  Československo  Př.:  Spisovatelé  and  r. 2001	1.spisovatelé  2. spisovatelé- Československo-r. 2001  <b>Poznámka 2</b>
	<i>Osoby</i>	Př.:  Jan  and  Masaryk  Př.:  Masaryk  and  Kohout  <b>Poznámka 3</b>	1.Masaryk  <b>Poznámka 3</b>
	<i>Akce</i>	Př.: Portréty duší  Př.: portréty  and  Duší	1. portréty  2. Portréty duší

	<i>Korporace</i>	Př.: Divadlo and Řeznické <b>Poznámka 2</b>	1.Divadlo 2.Divadlo v Řeznické <b>Poznámka 2</b>
	<i>Dílo/dokument</i>	Př.: Stíny and Hudsonem <b>Poznámka 2</b>	1.Stíny 2.Stíny nad Hudsonem <b>Poznámka 2</b>
	<i>Geografické vymezení</i>	e-xr--- and e-fr---  (pozn.:kód označující Českou republiku a Francii) <b>Poznámka 8</b>	e-xr---  (pozn.:kód označující Českou republiku) <b>Poznámka 8</b>
	<i>Chronologické vymezení</i>		
<b>Další výběry</b>	<i>Typ článku</i> <b>Poznámka 9</b>	Je třeba navolit	Je třeba navolit
	Název zdroje	Je třeba navolit	Je třeba navolit

**c)Přehled dosud nadefinovaných 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			
Geografie	Ostatní země světa	Filipíny		
		Malajsie		
		Srí Lanka		
		USA		

	<b>Evropa</b>	<b>Albánie</b>	
		<b>Andora</b>	
		<b>Arménie</b>	
		<b>Azerbajdžán</b>	
		<b>Belgie</b>	
		<b>Bělorusko</b>	
		<b>další země (celkem 49)</b>	
	<b>→→→Geografie</b>		
<b>Hospodářství, výroba</b>	<b>Doprava, spoje →→→</b>	<b>→→→&gt;Doprava</b> <b>→→→&gt;Letecká doprava</b> <b>→→→&gt;MHD</b> <b>Pošta, filatelie</b> <b>Silniční doprava</b> <b>Taxislužba</b> <b>Telekomunikace</b> <b>Vodní doprava</b> <b>Železniční doprava</b>	
	<b>Energetika→→→</b>	<b>→→→&gt;Energetika</b> <b>→→→&gt;Hornictví, těžební průmysl</b> <b>→→→&gt;Jaderná energetika</b> <b>→→→&gt;Plynárenství</b>	
	<b>&gt;Meteorologie, normy, standardy</b>		
	<b>&gt;Průmysl →→→</b>	<b>→→→&gt;Automobilový</b> <b>→→→&gt;Dřevozpracující</b> <b>→→→&gt;Elektrotechnika</b> <b>→→→&gt;Farmaceutický</b> <b>→→→&gt;Hutnictví</b> <b>→→→&gt;Chemický</b> <b>→→→&gt;Papírenství</b> <b>→→→&gt;Petrochemie</b> <b>→→→&gt;Polygrafie, tiskařský</b>	

		<ul style="list-style-type: none"> <li>→→→&gt;Stavebnictví</li> <li>→→→&gt;Strojírenství</li> <li>→→→&gt;Textilní</li> <li>→→→&gt;Zbrojní</li> </ul>	
	>Výpočetní technika		
	Zemědělství→→→	<ul style="list-style-type: none"> <li>→→→&gt;Lesnictví, myslivectví</li> <li>→→→&gt;Mlékárenství</li> <li>→→→&gt;Pivovarnictví</li> <li>→→→&gt;Potravinářství</li> <li>→→→&gt;Zemědělství</li> </ul>	
<b>Kultura, umění</b>	<b>Kultura</b>		
	Literatura, písemnictví →→→	<ul style="list-style-type: none"> <li>→→→&gt;Písmo</li> <li>→→→&gt;Literatura</li> <li>→→→&gt;Tisk</li> <li>→→→&gt;Vzácné tisky</li> </ul>	
	Umění →→→	<ul style="list-style-type: none"> <li>→→→&gt;Architektura</li> <li>→→→&gt;Divadlo</li> <li>→→→&gt;Film, kinematografie</li> <li>→→→&gt;Fotografie</li> <li>→→→&gt;Hudba</li> <li>→→→&gt;Urbanismus</li> <li>→→→&gt;Výtvarné umění</li> </ul>	
<b>Přírodní a matematické obory</b>	<b>Matematika, fyzika →→→</b>	<ul style="list-style-type: none"> <li>→→→&gt;Astronomie, astrofyzika</li> <li>→→→&gt;Fyzika</li> <li>→→→&gt;Kybernetika, robotika</li> <li>→→→&gt;Matematika</li> </ul>	

		→→→>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 vědy →→→	→→→>Archeologie →→→>Demografie →→→>Etika →→→>Etnografie →→→>Filozofie →→→>Historie →→→>Jazykověda, lingvistika, filologie →→→>Psychologie →→→>Sociologie, společnost →→→>Společ. vědy obecně	
	Knihovnictví a informační věda →→→	→→→>Autority	Tyto topiky jsou provizorní. Pro

		<ul style="list-style-type: none"> <li>→→→&gt;Bibliografie</li> <li>→→→&gt;Databáze</li> <li>→→→&gt;Dokumentace</li> <li>→→→&gt;Knihovnictví</li> <li>→→→&gt;Knihovny</li> <li>→→→&gt;Kybernetika</li> <li>→→→&gt;Informační služby</li> <li>→→→&gt;Jmenný popis</li> <li>→→→&gt;Písmo</li> <li>→→→&gt;Příručky</li> <li>→→→&gt;Seriály</li> <li>→→→&gt;Standardizace</li> <li>→→→&gt;Věcný popis</li> <li>→→→&gt;Zpracování dokumentů</li> </ul>	<p>periodikum Národní knihovna se počítá s tvorbou tezauru/říze ného hesláře, který bude podkladem topiku pro periodikum Národní knihovna</p>
	Společnost→→→	<ul style="list-style-type: none"> <li>→→→&gt;Feminismus</li> <li>→→→&gt;Náboženství</li> <li>→→→&gt;Politika</li> <li>→→→&gt;Právo</li> <li>→→→&gt;Sociální zabezpečení</li> <li>→→→&gt;Školství</li> <li>→→→&gt;Věda a technika</li> <li>→→→&gt;Veřejná správa</li> <li>→→→&gt;Vojenství, vojenská technika</li> </ul>	
Sport, volný čas	>Společenské a lidové zábavy		
	Sport →→→	→→→>Atletika	



		→→→>Basketbal →→→>Cyklistika →→→>Fotbal →→→>Gymnastika →→→>Lední hokej →→→>Motorismus →→→>Tenis →→→>Volejbal	
	>Volný čas		
Zdravotnictví, lékařství (bude lépe specifikováno a rozpracováno)	Zdravotnictví	Červený kříž Léčitelství Terapie Toxikologie Úrazy Vitamíny Zdravotnická zařízení Zdravotnictví	Bude rozpracováno
	Farmacie, farmakologie	totéž	
	Lékařské obory	Cestovní medicína	
		Gynekologie, porodnictví	
		Kardiologie	
		Neurologie	
		Oční	
		Ortopedie	
		Stomatologie	
		Urologie	
		Ušní, nosní, krční	
	Nemoci a onemocnění	AIDS	
		Alergie	
		Cukrovka	
		Encefalitida	
		Epilepsie	
		Hepatitida	
		Chřipka	
		Mozková mrtvice	
		Pohlavní choroby	

		<b>Rakovina</b>	
		<b>Salmonelóza</b>	
		<b>Vzteklina</b>	

**e) Seznam metadat v uživatelském formátu, resp. citace**

Název:

Podnázev:

Hlavní autor:

Další autor:

Zdroj (název, ročník, číslo, datum):

Zdroj-příl.:

Nakladatel/vydavatel

Místo vydání

ISSN:

Rubrika:

Předmět. ktg.:

MDT:

Osoba jako předmět:

Korporace jako předmět:

Akce jako předmět:

Dílo jako předmět:

Téma jako předmět:

Geograf. jm. jako př.:

Typ dokumentu:

Geograf. kód:

Časový kód:

SICI

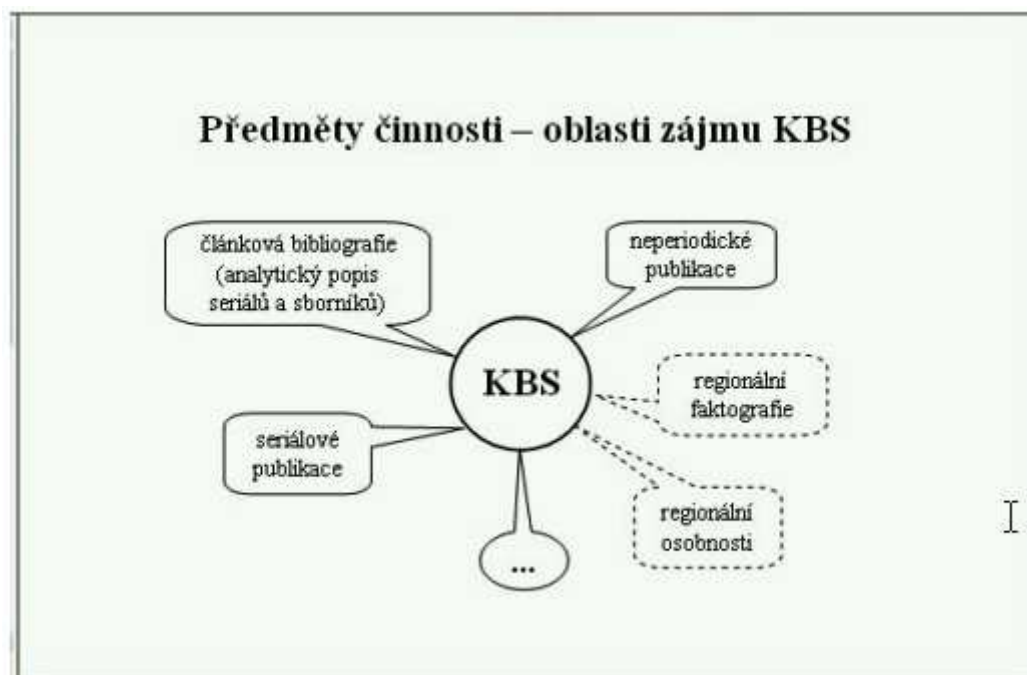
URN

## Příloha F 5

### Krajský bibliografický systém. Aplikace pro správu a údržbu KOSABI (architektura systému, jednotlivé aplikace pro příjem a zpracování dat, báze titulů)

#### a) Krajský bibliografický systém

#### Model Krajského bibliografického systému jako předpoklad funkce KOSABI



## Legislativní rámec fungování KBS

**Knihovní zákon (257/2001) – §11 Krajská knihovna, odst. 2:**  
(...)

- a. spolupracuje s Národní knihovnou při zpracování národní bibliografie a při zpracování souborného katalogu,
- b. zpracovává a zpřístupňuje regionální informační databáze a zabezpečuje koordinaci krajského bibliografického systému,

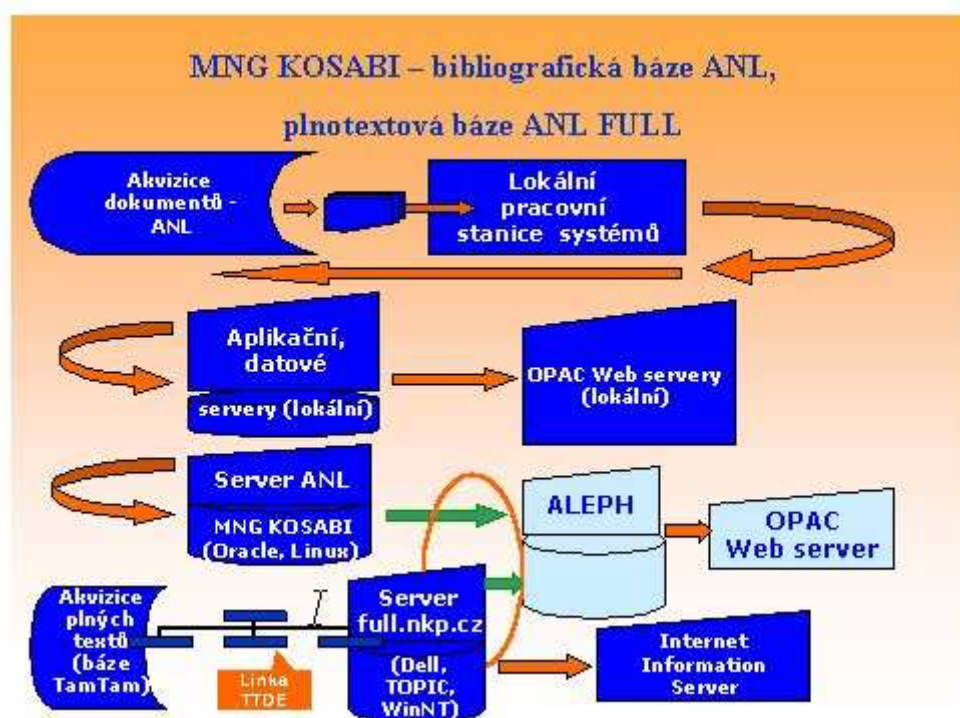
**Zákon o neperiodických publikacích (37/1995)**  
§3, odst. 1, pís. d  
(příjemci reg. povinného výtisku)

**Tiskový zákon (46/2000)**  
§9, odst. 1, pís. f  
(příjemci povinného výtisku)

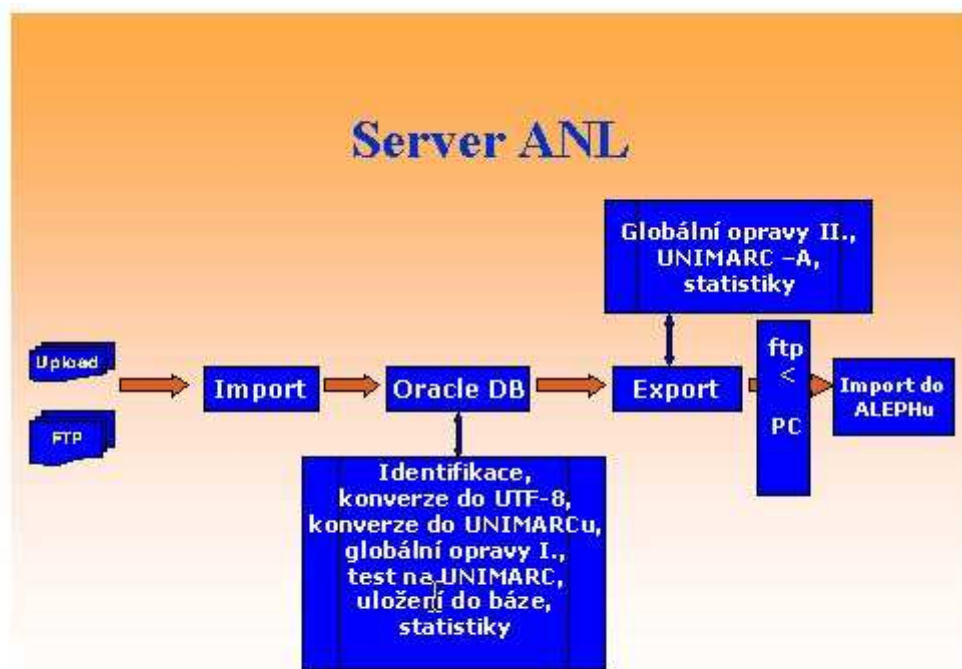
**Zákon 320/2002 o změně a zrušení některých zákonů v souvislosti s ukončením činnosti okresních úřadů**  
čl. 22 a 23 (účinnost od 1. 1. 2003)

### b) Management KOSABI

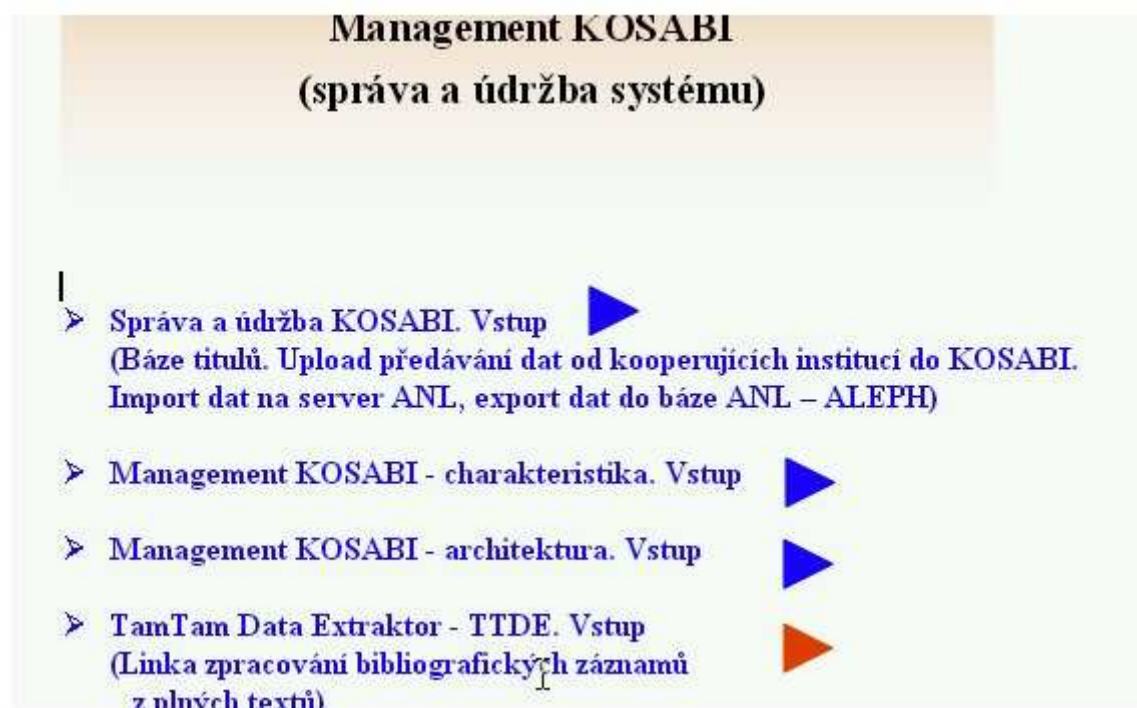
#### Management KOSABI - architektura



#### Server ANL



Jednotné rozhraní na serveru full.nkp.cz pro přístup do aplikace pro MNG KOSABI na serveru ANL



Vstup do aplikace pro MNG KOSABI

## KOSABI - souborná databáze ANL

[Báze Titulů](#) Volně přístupná

[Upload předávání dat do kooperačního systému](#) Registrace

[Import - statistiky](#) Registrace

[Klíče](#) Registrace

[Kosabi Admin](#) Registrace

### Báze titulů

## KOSABI - souborná databáze ANL

Báze Titulů  
Třídění: kategorie titulů

[Třídít dle názvu](#)

[Třídít dle zpracovávající instituce](#)

[Pouze tituly s plným textem](#)

- [2](#) Acta technica CSAV. ABA001. A
- [3](#) Acta universitatis Carolinae. Biologica. ABA001. A
- [7](#) Archeologické rozhledy. ABA001. A
- [9](#) Archiv orientální. ABA001. A
- [14](#) Biologia plantarum. ABA001. A
- [15](#) Biologické listy. ABA001. A
- [19](#) Collection of Czechoslovak Chemical Communication. ABA001. Báze ANL. A
- [20](#) Czechoslovak Journal of Physics. ABA001. A
- [22](#) Časopis pro moderní filologii. ABA001. A
- [23](#) Časopis pro právní vědu a praxi. ABA001. A
- [25](#) Česká literatura. ABA001. A

### Zobrazení konkrétního titulu – formát SHORT

## KOSABI - souborná databáze ANL

Báze Titulů  
Třídění: kategorie titulů

[Třídít dle názvu](#)

[Třídít dle zpracovávající instituce](#)

[Pouze tituly s plným textem](#)

- [2](#) Acta technica CSAV. ABA001. A
- [3](#) Acta universitatis Carolinae. Biologica. ABA001. A
- [7](#) Archeologické rozhledy. ABA001. A
- [9](#) Archiv orientální. ABA001. A
- [14](#) Biologia plantarum. ABA001. A
- [15](#) Biologické listy. ABA001. A
- [19](#) Collection of Czechoslovak Chemical Communication. ABA001. Báze ANL. A
- [20](#) Czechoslovak Journal of Physics. ABA001. A
- [22](#) Časopis pro moderní filologii. ABA001. A
- [23](#) Časopis pro právní vědu a praxi. ABA001. A
- [25](#) Česká literatura. ABA001. A

**Zadat heslo k síti** [?] [X]

 Zadejte své uživatelské jméno a heslo.

Server: anl.nkp.cz

Sféra: BAZE\_SKK

Uživatelské jméno:

Heslo:

Uložit toto heslo do seznamu hesel

OK Storno

**Upload záznamů**



**KOSABI - souborná databáze ANL**

Upload předávání dat do kooperačního systému

*Uživatel: aba001\_user*  
*Stgla: ABA001*

Data format:

Data kod:

Data:

**KOSABI - souborná databáze ANL**

Upload předávání dat do kooperačního systému

*Uživatel: aba001\_user*  
*Stgla: ABA001*

Data format:

Data kod:

Data:

- ISO 646 nebo ISO 5426 [UM]
- veškerá diakritika pomocí GIZMO [Gi]
- PC Latin2 + GIZMO [LG]
- kód Kamenických + GIZMO [KG]
- UNICODE UTF8 [UC]
- ISO 8859-2 + GIZMO [SG]
- kód Kamenických [KA]
- ANSEL [AN]

**Import - statistiky**

**Přehled dávek k dalšímu zpracování a statistiky**

## KOSABI - souborná databáze ANL

Statistiky batch

<u>2002-09</u>	15
<u>2002-10</u>	47
<u>2002-11</u>	2
<u>2003-01</u>	5
<u>2003-02</u>	22
<u>2003-05</u>	4
<u>2003-06</u>	30

## KOSABI - souborná databáze ANL

Správa dat  
Přehled o dávkách: 2002-10

<b>EXP. 16</b>	2002-10-08 23:41:52	ABA001 08 aba001an.dtt_20021008_234151	256 ./ul.c (2307) lcb
<b>EXP. 17</b>	2002-10-08 23:41:52	ABA001 08 aba001an.dtt_h105_20021008_234151	0 null
<b>EXP. 44</b>	2002-10-08 08:00:02	ABA012 25 aba012sg.vfi_20021025_080001	256 ./suf.c (887) REC_SEP_KC
<b>EXP. 45</b>	2002-10-08 08:00:02	ABA013 25 aba013kg.vfi.iso_20021025_080001	256 ./suf.c (887) REC_SEP_KC
<b>EXP. 46</b>	2002-10-08 08:00:02	CBA001 25 cba001sg.uls.mrc_20021025_080000	0 null
<b>EXP. 47</b>	2002-10-08 08:00:47	HKA001 25 hka001sg.vfi.iso_20021025_080000	256 ./suf.c (887) REC_SEP_KC
<b>EXP. 48</b>	2002-10-08 08:00:47	KLG001 25 klg001kg.rum.iso_20021025_080000	256 ./suf.c (270) 1.zaznam ne lbl ev. fmt
<b>EXP. 49</b>	2002-10-08 08:00:47	KLG001 25 klg001kg.uls.iso_20021025_080000	256 ./suf.c (887) REC_SEP_KC
<b>EXP. 50</b>	2002-10-08 08:00:47	LIA001 25 lia001sg.uls.iso_20021025_080000	0 null
<b>EXP. 51</b>	2002-10-08 08:01:52	OLA001 25 ola001uc.dtt.ola_20021025_080001	256 ./ul.c (2307) lcb->d_pmax

Statistiky dávek

<b>KOSABI - souborná databáze ANL</b>										
Statistiky batch: 52										
ownerdoku	rctxt	vaha	inp	get	move	konv	uni	outadd	outnew	
OSA001	neznamy	0	null	18	136	1350	0	0	0	1
<u>4666</u>										

<b>KOSABI - souborná databáze ANL</b>										
Statistiky batch: 57										
ownerdoku	rctxt	vaha	inp	get	move	konv	uni	outadd	outnew	
ABA001	neznamy	0	null	18	779	7640	0	0	0	15
5348										
5349										
5350										
5351										
5352										
5353										
5354										
5355										
5356										
5357										
5358										
5359										
5360										
5361										
5362										

### Statistika chyb v dávce

<b>KOSABI - souborná databáze ANL</b>	
Správa dat	
Statistiky batch: 54	
RecOwner: ABA001 RecDoku: 0	
<b>Vaha:</b>	<b>18</b>
<b>Vstup:</b>	<b>681</b>
<b>Chyba vazebnich poli:</b>	<b>681</b>
<b>Chyba CRERR:</b>	<b>0</b>
<b>Davka prijata do baze</b>	
<b>Rec: 1 O projekt Mosty se zacina zajimat stale vice mest v zemi</b>	
<b>ERR vazebni TAG: 001 SERIAL_GENERAL*981350*2001*001*276</b>	
<b>Rec: 2 Uroven vystavnich trhu je rok od roku horsi</b>	
<b>ERR vazebni TAG: 001 SERIAL_GENERAL*980624*2002*002*160</b>	

## Export dat do ALEPHu nebo k dalším opravám

**KOSABI - souborná databáze ANL**  
EXPORT

BATCH 6

Kritické chyby ne

Unimarc chyby ano

Adresar Export-ANL

Export souboru

Export-Aleph

Export ANL

## Klíče

### Test na duplicitu klíčů

<b>CASLIN - KOSABI - souborná databáze ANL</b>	
TEST KEY PRIM / rec_id / batch_id	
<a href="#">zásady provozního aduřzení školství2001</a>	36
<a href="#">aktuální seznam avostiveterinářství2002</a>	30
<a href="#">právní předpisy veterinářství2002</a>	20
<a href="#">celostátní olo50 učebnice matematik a fyzik a informatik a2001.02</a>	18
<a href="#">čas s autem d450our agandlectví kosmonautika200209</a>	18
<a href="#">dítě housle ? dnes ? hudební výchova2001</a>	18
<a href="#">naše vojenská "čtyřtádesítka" letectví kosmonautika200209</a>	18
<a href="#">nedokončené projekty z letectví kosmonautika200209</a>	18
<a href="#">počítač hudební výchovy hudební výchova2001</a>	18
<a href="#">zobcová flétna škol hudební výchova2001</a>	18
<a href="#">náhled na historii našeho centra v Jablonci nad Sázavou2002</a>	16
<a href="#">představujeme Jablonci nad Sázavou2002</a>	16
<a href="#">představujeme jeho české kandidáty v eskobudejovicích listy 10.05.2002</a>	16
<a href="#">představujeme jeho české kandidáty v eskobudejovicích listy 24.05.2002</a>	16
<a href="#">experimentální a odhalení identifikace technická diagnostika astronomů</a>	14

## KOSABI - souborná databáze ANL

Správa dat  
Zaznam RecId: 4666

001\_\_ 20021004 13502200  
005\_\_ 20020925000000  
100\_\_ a20020925d2001 ||||0czey0103 ||  
1010\_ aslo  
102\_\_ aCZ  
2001\_ aPoznatky k vėskytu obojšiveInAkov (Amphibia) v altlvu potoka ParAšfJana Bartov  
463\_1 1010\_\_ a80-7042-811-212001\_ aBiologie, ekologied= Biologia, ekologiaf[odpovRdné redaktor Pavel Drozd, Aleç  
Dolné]vs. 84-871205\_\_ aVyd. 1. 1210\_\_ aOstravacOstravsk univerzita2001  
607\_\_ aParAš (potok, Slovensko)9607  
6100\_ abiomonitorov nA  
6100\_ aobojšiveInAcl  
6100\_ apotoky  
675\_\_ a591.99n  
675\_\_ a597.6/.99n  
700\_1 aBartov bJana4070  
801\_0 aCZbOSA001c20020925gAACR291  
910\_\_ aOSA001trm

Vaha: 18  
Test UNIMARC: Ano  
Test CRERR: Ne  
Klic: zivotlibicletechpolabivlastivednyzpravodaj

967\_\_ aLibice nad Cidlinou (Česko)bNymburkccz-cc  
910\_\_ aKLG001bRR.V/382trs

Vaha: 18  
Test UNIMARC: Ano  
Test CRERR: Ne  
Klic: zivotlibicletechpolabivlastivednyzpravodaj

Administrace – test URL

## KOSABI - souborná databáze ANL

Admin Kosabi

[Test URL adres](#)

[Statistika testu](#)

[Start testu](#)

[Přehled aktualizace URL adres](#)

*Volba typu adres k testu*

## KOSABI - souborná databáze ANL

Start testu

Adresy vyma full nka.cz

Pouze ERROR Adresy

Odeslat

Vymaz

Testy přehled

<b>Id</b>	<b>test start</b>	<b>test end</b>	<b>pocet testovanych adres</b>	<b>Error</b>
42	2003-11-06 14:40:04	2003-11-06 14:58:57	1248	346
41	2003-11-05 19:20:05	2003-11-05 19:36:35	4050	1
34	2003-06-23 11:30:04	2003-06-23 11:42:01	575	339



## KOSABI - souborná databáze ANL

Start testu

Adresy vyjma full.nkp.cz

Pouze ERROR Adresy

Odeslat

Vymaz

Testy přehled

<b>Id</b>	<b>test start</b>	<b>test end</b>	<b>pocet testovanych adres</b>	<b>Error</b>
42	2003-11-06 14:40:04	2003-11-06 14:58:57	1248	346
41	2003-11-05 19:20:05	2003-11-05 19:36:35	4050	1

Výsledek testu – statistiky neplatných a platných URL

## KOSABI - souborná databáze ANL

Navrat kody

<b>Kod</b>	<b>text</b>	<b>pocet URL adres</b>
<u>RC: -8</u>	Error FATAL	50
<u>RC: -4</u>	Error FATAL	59
<u>RC: 0</u>	neexistovano	6352
<u>RC: 200</u>	OK	107753
<u>RC: 400</u>	<u>Error viz RFC 2616</u>	74
<u>RC: 403</u>	<u>Error viz RFC 2616</u>	5
<u>RC: 404</u>	<u>Error viz RFC 2616</u>	171
<u>RC: 500</u>	<u>Error viz RFC 2616</u>	7

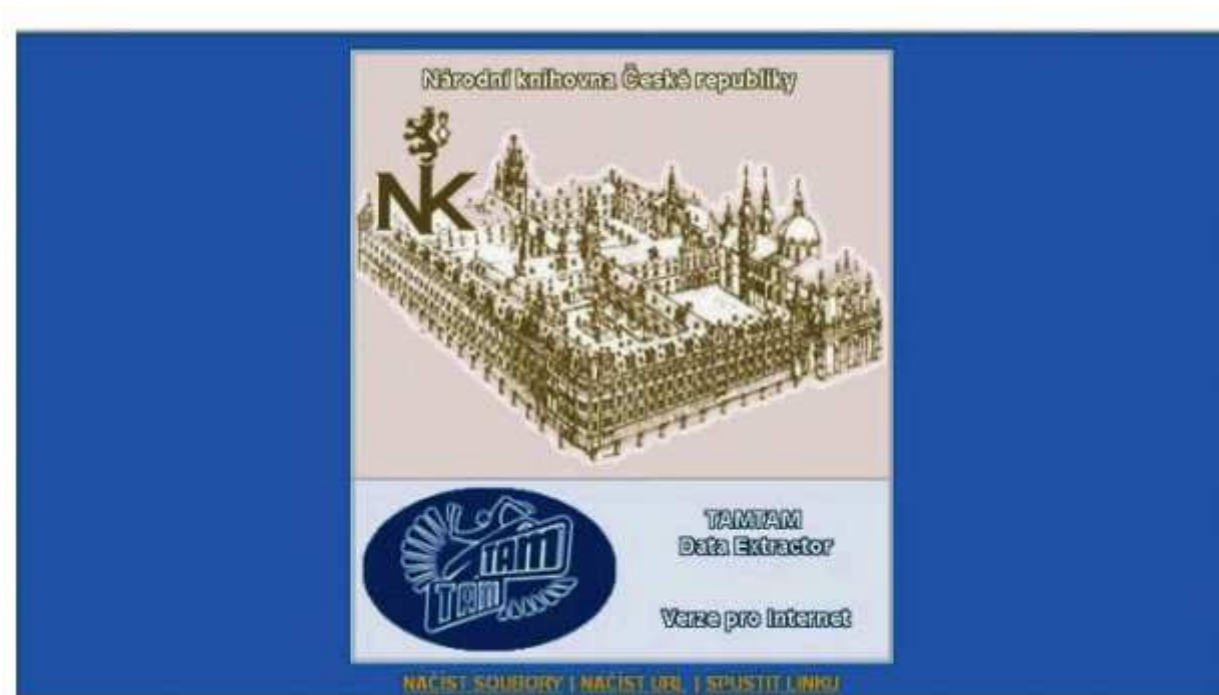
Rueme kody



## Příloha F 6

**Linka zpracování bibliografických záznamů z plných textů – WWW TamTam Data Extractor ( WWW TTDE) - bibliografická metadata pro bázi ANL (ALEPH), plné texty s Dublin Core pro bázi ANL FULL (TOPIC) nebo jen metadata pro obě báze s propojením na plný text. Základní funkce TTDE pro lokální stanice jsou zachovány. WWW TTDE rozšířena o vstup v HTML z libovolné URL a aplikaci Opravy**

Vstup



**Formulář pro načtení dat z TamTam Anopress, ANL FULL, URL – soubory ukládány na disk, odeslání ke zpracování v lince**

TTDE [3.0.014 - 21.10.2003] - Microsoft Internet Explorer

**VYBERTE SOUBORY PRO ZPRACOVÁNÍ - pouze ZIP, TXT nebo HTML soubory**

1.  Procházet...

2.  Procházet...

3.  Procházet...

4.  Procházet...

5.  Procházet...

6.  Procházet...

7.  Procházet...

8.  Procházet...

9.  Procházet...

10.  Procházet...

**TYP VSTUPU:** Linka TTDE - Anopress, NK

**POSLAT SOUBORY NA SERVER FULL.NKP.CZ A SPUSTIT TTDE**

2.  Procházet...

3.  Procházet...

4.  Procházet...

5.  Procházet...

6.  Procházet...

7.  Procházet...

8.  Procházet...

9.  Procházet...

10.  Procházet...

**TYP VSTUPU:**

- Linka TTDE - Anopress, NK
- Linka TTDE - Anopress, NK
- Anopress - mediální databáze
- Full.nkp.cz - databáze
- Obecné URL

**POSLAT SOUBORY NA SERVER FULL.NKP.CZ A SPUSTIT TTDE**

**ZADEJTE URL PRO ZPRACOVÁNÍ - pouze ve tvaru "http://..."**

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

**STÁHNOUT ZADANÁ URL A SPUSTIT TTDE**

Formulář pro editaci má 4 strany, je rozšířen o údaje zejména s ohledem na kontrolu údajů proti rejstříkům ANL FULL, v současné době probíhají práce na komponentě pro kontrolu údajů proti autoritám

	Form.1	Form.2	Form.3	Form.4	Text	Unm.-A	Unm.	Marč.21	Dubín	Hlav.1	Hlav.2	Opr
1. Datum uložení	20031022	[100a,001c]	<input checked="" type="checkbox"/> záznamu	an2003abd-	[001]	Periodičita						
2. Datum vydání	20030701	[100a,401v_zazd]				105 Y	[0] z	[4] Y	[11] Y			
3. Název článku	Humorista Kishon nerad vypráví vtípy											
3a. Číslo části		[2001h]	3b. Název části									
4. Podnázev												
7. Rubrika	Styl											
8. Souhrn	Ne to, že je jeden z nejčtenějších humoristů na světě, ale že vyhrál mistrovství sv											
9. Název zdroje >	Právo	[401a_zazg]	9a. <input checked="" type="radio"/> Nov <input type="radio"/> Čas <input type="radio"/> Ost	9d. LKR								
9b. Číslo části zdr.		[401h]	9c. Název části zdr.									
10. Název přílohy		[401a_zazg]	11. <input checked="" type="radio"/> isbn <input type="radio"/> isbn	1211-2119								
12. Místo vydání	Praha	[401v_zazn]	12a. Datum vydání									
13. Vydavatel >	Borgis	[401c]	13a. Adresa vyd.									
14. Ročník	13	[401v_zazg]	15. Číslo	152	[401v_zazh]	16. Strana	2					
17. Klíčová slova												
18. Předmět. ktg. >												
18b. MDT_konспект		[615a]	18c. DDC_konспект									
19. MDT > /DDC		[675a]	20. VAH	16								
21. Geograf. kód >		[600a]	21b. Časový kód >									
22a. URL interní	http://full.nkp.cz/~nkdb/view/mkdoc.asp?p1=PR20030701000215											

Form.1	Form.2	Form.3	Form.4	Text	Unim.-A	Unim.	Marc. 21	Dublin	Hlav.1	Hlav.2	Opravy...	
1. Datum uložení	20031022	[100a,801c]	D. záznamu	an2003abd-	[001]	Periodičta	J	[100a]				
2. Datum vydání	20030701	[100a,461v,2a2d]			105	Y	[0]	Z	[4]	Y	[11]	Y
3. Název článku	Humorista Kishon nerad vypráví vtipy										[200a]	
3a. Číslo části		[2001h]	3b. Název části								[200i]	
4. Podnázev											[200e]	
7. Rubrika	Styl										[304a,54b]	
8. Souhrn	Ne to, že je jeden z nejčtenějších humoristů na světě, ale že vyhrál mistrovství sv										[100k]	
9. Název zdroje >	Právo	[461a,232g]	9a.	<input checked="" type="radio"/> Nov	<input type="radio"/> Čes	<input type="radio"/> Ost	9d. LKR				[461i]	
9b. Číslo části zdr.		[461b]	9c. Název části zdr.								[461j]	
10. Název přílohy		[461a,232g]	11.	<input checked="" type="radio"/> issn	<input type="radio"/> isbn	1211-2119					[461a,232]	
12. Místo vydání	Praha	[461v,232m]	12a. Datum vydání								[461d]	
13. Vydavatel >	Borgis	[461c]	13a. Adresa vyd.								[461b]	
14. Ročník	13	[461v,232j]	15. Číslo	152	[461v,232h]	16. Strana	2				[461v,232]	
17. Klíčová slova											[610a]	
16. Předmět. idg. >											[615axm3]	
18b. MDT_konспект		[615an]	18c. DDC_konспект								[615an]	
19. MDT > / DDC		[675an]	20. VAH	16							[vsh]	
21. Geograf. kód >		[660a]	21b. Časový kód >								[661a]	
22a. URL_interní	http://full.nkp.cz		mkdb/view/mkdoc.asp?p1=	PR20030701000215							[660]	
22b. URL_externí	http://www.anopress.cz/frana_nk/mkdocnk.asp?p1=		PR20030701000215								[660]	
23. Dokument	CP 1250	HTM	Media	24. Typ dokum. >							[660]	
25. Form.řez.řvo	text/html	cze	Text	26. Ľ							[215c]	
27. STZ	se20031022	[ic]	28. URN:NDN	URN:NDN:cz-nkPR200307							[614]	

19

ZPRACOVAT ZÁZNAM

Zrušit záznam

Odeslat data

Vyprázdnit

Novinky



TTDE [3.0.014 - 21.10.2003] - Microsoft Internet Explorer

Form.1 Form.2 Form.3 Form.4 Text Úřim.-A Úřim. Marc 21 Dubín Hlav.1 Hlav.2 Opravy...

Hlavní autoři > **Zuzana Fialová** Další autoři >

Odkazuje

Odkazovaný

**5.1 Hlavní autoři [700, 701]**

Příjmení [a]	Jméno [b]	Doplněk [c]	PJM	Římské č. [d]	Data [f]	[g]	[p]	Čís. aut. [3]	[4]
>									
>									

**6.1 Další autoři [702]**

Příjmení [a]	Jméno [b]	Doplněk [c]	PJM	Římské č. [d]	Data [f]	[g]	[p]	Čís. aut. [3]	[4]
>									
>									

**6.2 Další autoři [702]**

Příjmení [a]	Jméno [b]	Doplněk [c]	PJM	Římské č. [d]	Data [f]	[g]	[p]	Čís. aut. [3]	[4]
>									
>									

**KA. Autor osoba [autor, dílo - 604, 700, 500]**

Příjmení [a]	Jméno [b]	Doplněk [c]	PJM	Římské č. [d]	Data [f]	[g]	[p]	Čís. aut. [3]	[4]
>									
>									

**Název [a]** **Číslo části [h]** **Název části [i]** **Druh [n]** **Číslo autority [3]** **Datum vydání [k]**

**KB. Autor korporace [autor, dílo - 604, 710, 500]**

Instituce, Místo [a]	Zpřesnění [b]	Doplněk [c]	Číslo autority [3]	[4]
>				
>				

**Název [a]** **Číslo části [h]** **Název části [i]** **Druh [n]** **Číslo autority [3]** **Datum vydání [k]**

Počet: 19

ZPRACOVAT ZÁZNAM Zrušit záznam Odeslat data Vyprázdnit Novinky

TTDE [3.0.014 - 21.10.2003] - Microsoft Internet Explorer

Form.1 Form.2 Form.3 Form.4 Text Úřim.-A Úřim. Marc 21 Dubín Hlav.1 Hlav.2 Opravy...

**Téma jako předmět [6060 - A]**

Téma [a]	Číslo autority [3]	Věcné zpřesnění [x]	Zeměpis. zpřesnění [y]	Chronol. zpřesnění [z]
>				
>				
>				

**Geografické jméno jako předmět [607 - B]**

Téma [a]	Číslo autority [3]	Věcné zpřesnění [x]	Zeměpis. zpřesnění [y]	Chronol. zpřesnění [z]
>				
>				
>				

**Osoba jako předmět [600 1/0 - C]**

Příjmení [a]	Jméno [b]	[c]	PJM	[d]	Data [f]	[g]	[p]	[3]	[x]	[y]	[z]
>											
>											

**Dílo jako předmět [605 - D]**

Název [a]	Číslo části [h]	Název části [i]	Druh [n]	Číslo autority [3]	Datum vydání [k]
>					
>					
>					

**Korporace jako předmět [60102/01 - G]**


Instituce, Místo [a]	Zpřesnění [b]	Doplněk [c]	Číslo autority [3]
>			
>			
>			

Počet: 19

ZPRACOVAT ZÁZNAM Zrušit záznam Odeslat data Vyprázdnit Novinky

TTDE [3.0.014 - 21.10.2003] - Microsoft Internet Explorer

[Fom.1](#)
[Fom.2](#)
[Fom.3](#)
[Fom.4](#)
[Text](#)
[Unim. A](#)
[Unim.](#)
[Marc.21](#)
[Dubn.](#)
[Hlav.1](#)
[Hlav.2](#)
[Opravy ...](#)



**Seznam článků**

[PP\(13.12\) Labant ve Sportě, Mareš za](#)  
[NO\(13.12\) Labant definitivně patří týmu](#)  
[PR\(01.07\) Josef Rakoncaj vymění Ma](#)  
[VQ\(01.07\) Historie Křehlého hradu v](#)  
[NQ\(01.07\) Příběhy ze života](#)  
[NQ\(01.07\) Život tenorů: rány i výhrů](#)  
[NQ\(01.07\) Obdiváme se hůř než mnoh](#)  
[NQ\(01.07\) Bída dělá z dítě vky](#)  
[JQ\(21.09\) Neřekám, že jsem svůj záp](#)  
[IA\(01.07\) KSČM se připravuje na sv](#)  
[IA\(01.07\) KSČMcccc se připravuje na](#)  
[TY\(03.02\) Ivan Havel: Kdybych byl st](#)  
[WR\(29.08\) Bilance pilotního projektu II](#)  
[WR\(29.08\) Z39.50 - protokol z minul](#)  
[WR\(29.08\) Jsem vytvářel optimism](#)  
[WR\(29.08\) K dobrotě KnowEurope m](#)  
[WR\(29.08\) Meze bezmezných inform](#)  
[WR\(29.08\) Mha přede mnou, mha za](#)  
[WR\(29.08\) Kde raději nepoužívat fi](#)

Téma jako předmět [6000 - A]					
Téma [a]	Číslo autority [3]	Věcné zpřesnění [x]	Zeměpis. zpřesnění [y]	Chronol. zpřesnění [z]	
1 ▶					
2 ▶					
3 ▶					
4 ▶					

Geografické jméno jako předmět [607 - B]					
Téma [a]	Číslo autority [3]	Věcné zpřesnění [x]	Zeměpis. zpřesnění [y]	Chronol. zpřesnění [z]	
1 ▶					
2 ▶					
3 ▶					

Osoba jako předmět [600 1A0 - C]											
Příjmení [a]	Jméno [b]	[c]	PJM	[d]	Data [f]	[g]	[p]	[3]	[x]	[y]	[z]
1 ▶											
2 ▶											
3 ▶											

Dílo jako předmět [605 - D]					
Název [a]	Číslo části [n]	Název části [j]	Druh [n]	Číslo autority [3]	Datum vydání [k]
1 ▶					
2 ▶					
3 ▶					


Korporace jako předmět [60102/01 - O]			
Instituce, Místo [a]	Zpřesnění [b]	Doplněk [c]	Číslo autority [3]
1 ▶			
2 ▶			
3 ▶			

Počet: 18

[Všechny články](#)
[Tisknout seznam](#)
[Přidat do seznamu](#)
[Upravit seznam](#)

TTDE [3.0.014 - 21. 10. 2003] - Microsoft Internet Explorer

Form.1 Form.2 Form.3 Form.4 Text Unim.-A Unim. Marc.21 Dubln Hlav.1 Hlav.2 Opravy...



**Seznam článků**

PP(13.12) Labant ve Spartě, Mareš za  
 PO(13.12) Labant definitivně patří týmu  
 PR(01.07) Josef Rakoncaj vymění Ha  
 -R(01.07) Hnutíto křehko nemá vř  
 HQ(01.07) Příběhy ze života  
 HQ(01.07) Život seniorů: rány i výtu  
 HQ(01.07) Chováte se hůř než mto  
 HQ(01.07) Bída dělá z dětí vlky  
 LN(21.09) Neřádné, že jsem svůj záp  
 HA(01.07) KSČM se připravuje na sv  
 HA(01.07) KSČM:cox se připravuje na  
 TY(03.02) Ivan Havel: Kdybych byl st  
 HR(26.06) Bilance pětiho projektu k  
 HR(26.06) Z39.50 - protokol z minulé  
 HR(29.06) Jsem vytrvalý optimista  
 HR(29.06) K databázi KnowEurope m  
 HR(29.06) Meze bezmezných inform  
 HR(29.06) Mha přede mnou, mha za  
 HR(29.06) Kate raději nepoužíval flash

**Akce jako předmět [60102/01/12 - H]**

Inst., Mís., Ak.[a]	Zpřesnění [b]	Kvalifikátor [c]	Číslo akce [d]	Místo konání [e]	Datum konání [f]	Číslo autority [3]

**Hlavní korporace [71002/01 - E]**

Instituce, Místo [a]	Zpřesnění [b]	Doplněk [c]	Číslo autority [3]	[4]

**Další korporace [71202/01 - I]**

Instituce, Místo [a]	Zpřesnění [b]	Doplněk [c]	Číslo autority [3]	[4]

**Hlavní akce [71002/01/12 - F]**

In., Mís., Ak [a]	Zpřesnění [b]	Kvalifikátor [c]	Číslo akce [d]	Místo konání [e]	Dat. konání [f]	Číslo aut [3]	[4]

**Další akce [71202/01/12 - J]**

In., Mís., Ak [a]	Zpřesnění [b]	Kvalifikátor [c]	Číslo akce [d]	Místo konání [e]	Dat. konání [f]	Číslo aut [3]	[4]

Počet: 19


ZPRACOVAT ZÁZNAM Zrušit záznam Odstranit data Vyprázdnit Novinky

Hlavička text



TTDE [3.0.014 - 21.10.2003] - Microsoft Internet Explorer

Form1 Form2 Form3 Form4 Text Unim-A Unim Marc 21 Dubln Hlav.1 Hlav.2 Opravy...



**Seznam článků**

PP(13.12) Labant ve Sportě, Mareš za  
 WO(13.12) Labant definitivně patří týmu  
 PR(01.07) Josef Rakoncaj vyměnil H  
 PR(01.07) Humorista Kishon nerad vy  
 WO(01.07) Příběhy ze života  
 WO(01.07) Život seniorů: rány i výhled  
 WO(01.07) Chováme se hůř než mnoh  
 WO(01.07) Bída dělá z dětí vlky  
 LN(21.09) Neřikám, že jsem svůj záp  
 HA(01.07) KSČM se připravuje na sv  
 HA(01.07) KSČM:co se připravuje na  
 TY(03.02) Ivan Havel: Kdybych byl st  
 WW(28.08) Bilance pilotního projektu II  
 WW(28.08) Z39.50 - protokol z minul  
 WW(29.08) Jsem vytvářel optimism  
 WW(29.08) K databázi KnowEurope m  
 WW(29.08) Meze bezmezných inform  
 WW(29.08) Mha přede mnou, mha za  
 WW(29.08) Kde raději nepoužívat flash

Jeden z nejčtenějších satiriků na světě Ephraim Kishon si stěžuje, že lidé od něj často chtějí slyšet vtipy. Nebo mu zase neopak sami chtějí nějaký vyprávět. On to ale neschází. Je to jeho profese. Copak byste chtěli, aby gynekolog také pořád vyprávěl o svém zaměstnání? podivuje se s úsměvem na rtech spisovatel a dramatik, který se původně vyučil šperkařem. Jak sám o sobě říká, v souvislosti s jeho osobou se objevují hlavně dvě fakta. Ne to, že je jeden z nejčtenějších humoristů na světě, ale že vyhrál mistrovství světa v kulečnicku a večerel s Charlem Chaplinem. Chaplin mě chytil za ruku. Prý plakal, když slyšel, co se stalo za druhé světové války Židům. Zeptal jsem se, jestli i on je Žid. Neodpověděl. Ale jeho žena ukázala, že napůl, vzpomíná S velkým konikem se setkali, když Charlie Chaplin dostal Oscara za životní zásluhy. Devětasedmdesátiletý Kishon se narodil v Maďarsku, ale v roce 1949 odešel do Izraele. V současnosti žije střídavě v Izraeli a ve švýcarském Appenzelu. Nedávno se po třetí oženil. Své knížky píše hebrejsky, devatenáct jich bylo přeloženo do češtiny.

\* Vzpomněl byste si na své začátky v Izraeli?

Všechno záleží na tom, umět hebrejsky. Jde o to, podrobně studovat písmo. Pro dospělého člověka je to téměř nemožné. A já jsem to zvládl jen proto, že jsem nevěděl, že je to téměř nemožné.

\* Prý jste tak pořádkumilovný, že máte své byty v Izraeli a ve Švýcarsku úplně stejně zařízené?

Některé náležitosti musí být všude stejné. Například na stole musí být šachový počítač. Protože psaní je někdy tak nudné, že musí mě ještě jinou činnost. Nerozděluj si ale deset minut na psaní a deset minut na šachy. Je to stejná pauza, jako když někdo kouří. Tak já udělám tah a počkám, až počítač udělá protáh. Šachy jsem hrál i s Gariem Kasparovem.

\* I ostatní zařízení je podobné?

Ve Švýcarsku mám bazén, v Izraeli kulečnickový stůl. Strašně dlouho hraju totiž kulečnic. Pětkrát jsem se účastnil i mistrovství světa v kulečnicku v Berlíně. Jednou jsem byl druhý, jednou třetí a pak také poslední a před třemi lety jsem byl ve své kategorii i první.

\* Vánujete se ještě někdy šperkařství?

V Izraeli mám takovou malou dílnu. Ještě v Maďarsku jsem navrhoval a vyráběl různé řady, ceny do soutěží a drobné šperky. Ale abych svoji manželku a děti přesvědčil, že to není jen taková legenda, že jsem byl kdysi šperkařem, tak mám nějaké své kusy v zlatu a občas jim je ukážu.


\* Napsal jste knížku Můj nejmenší pokrm je jídlo. Jaký máte vztah k vaření?

Počet 19

TPD&CVAT ZÁZNAM Zpráva seznam Otestat data Vymařdat Stránky

TTDE [3.0.014 - 21.10.2003] - Microsoft Internet Explorer

Form1 Form2 Form3 Form4 Text Unim-A Unim Marc 21 Dubln Hlav.1 Hlav.2 Opravy...



**Seznam článků**

PP(13.12) Labant ve Sportě, Mareš za  
 WO(13.12) Labant definitivně patří týmu  
 PR(01.07) Josef Rakoncaj vyměnil H  
 PR(01.07) Humorista Kishon nerad vy  
 WO(01.07) Příběhy ze života  
 WO(01.07) Život seniorů: rány i výhled  
 WO(01.07) Chováme se hůř než mnoh  
 WO(01.07) Bída dělá z dětí vlky  
 LN(21.09) Neřikám, že jsem svůj záp  
 MF(03.03) Václav Klaus: Jsem ve fon  
 HA(01.07) KSČM se připravuje na sv  
 HA(01.07) KSČM:co se připravuje na  
 TY(03.02) Ivan Havel: Kdybych byl st  
 WW(28.08) Bilance pilotního projektu II  
 WW(28.08) Z39.50 - protokol z minul  
 WW(29.08) Jsem vytvářel optimism  
 WW(29.08) K databázi KnowEurope m  
 WW(29.08) Meze bezmezných inform  
 WW(29.08) Mha přede mnou, mha za  
 WW(29.08) Kde raději nepoužívat flash

ka href="http://www.ikaros.cz/Clanek.asp?ID=200309008 target="\_blank">Odkaz na zdrojový dokument</a>

ŘÍJEN 2003  
 příští číslo vyjde 3.11.2003

Náš partner

Stáhněte si celé číslo

Copyright © 1998-2003 ikaros ISSN 1212-5075 Žádný příspěvek nebo jeho část nesmí být reprodukován bez souhlasu autora bez oznámení redakci.

```
<html>
<head>
<meta name="DC.Title" content="Kulturní dědictví, digitalizace a sémantický web" />
<meta name="DC.Creator.personalName" content="Uhlíř, Zdeněk" />
<meta name="DC.Creator.address" content="zdenek.uhlik@okp.cz" />
<meta name="DC.Subject" content="digitalizace" />
<meta name="DC.Subject" content="selekční jazyky" />
<meta name="DC.Subject" content="World Wide Web" />
<meta name="DC.Subject" content="metadata" />
<meta name="DC.Publisher" content="Ikaros, o. z." />
<meta name="DC.Publisher.address" content="redakce@ikaros.cz" />
<meta name="DC.Date.issued" scheme="W3C-DTF" content="2003-9-29" />
<meta name="DC.Type" scheme="DCHIType" content="Text" />
<meta name="DC.Format" scheme="IHT" content="text/asp" />
<meta name="DC.Format.medium" content="computerfile" />
<meta name="DC.Identifier" content="http://www.ikaros.cz/Clanek.asp?ID=200309008" />
<meta name="DC.Identifier" scheme="URN" content="URN:NBN:cz-sk20011447" />
```

## Generovaný UNIMARC-A pro ALEPH

TTDE [3.0.014 - 21.10.2003] - Microsoft Internet Explorer

Form1 Form2 Form3 Form4 Text Unim-A Unim. Marc 21 Dubň. Hlav.1 Hlav.2 Opravy...

**Seznam článků**

- PP(13.12) Labant ve Sportě, Mareš za
- NO(13.12) Labant delinivně patří týmu
- PR(01.07) Josef Rakoncaj vyměnil Hb
- 7F(01.07) Hranostola Kishon rovná se
- NX(01.07) Příběhy ze života
- NX(01.07) Život tenorů: rány i výhry
- NX(01.07) Chováte se hůř než mnou
- NX(01.07) Bika dělá z děs vky
- LN(21.09) Neříkám, že jsem svůj záp
- HA(01.07) KSČM se připravuje na svč
- HA(01.07) KSČM:xxx se připravuje ná
- TY(03.02) Ivan Havel: Kdybych byl st
- NR(28.08) Bilance pilotního projektu M
- NR(28.08) Z39.50 - protokol z měsí
- NR(29.08) Jsem vytrvalý optimista
- NR(29.08) K databázi KnowEurope m
- NR(29.08) Meze bezpečných inform
- NR(29.08) Mha přede mnou, mha za
- NR(29.08) Kde raději nepoužívat flash

FMT L ED  
 LDR L -----naa^^Z^-----^^450^  
 001 L an2003abd-  
 100 L \$\$a20031022j20030701u^^y0cney0103^^ba  
 1010 L \$facre  
 102 L \$faCE  
 105 L \$fay^^z^^000yy  
 014 L \$\$a1211-2119(20030701)13:152(2:URN:NBN:cs-nkPR20030701000215;3.0.CO;2-8  
 \$\$2sici\$\$9URN:NBN:cs-nkPR20030701000215  
 2001 L \$\$aHumorista Kishon nerad vypráví vtipy\$\$Zuzana Fialová  
 304 L \$\$aBubrika: Styl  
 330 L \$\$aNe to, že je jeden z nejčtenějších humoristů na světě, ale že vyhrál  
 mistrovství světa v kulečnicku a večeřel s Charliem Chaplinem  
 461 0 L \$\$1011^^\$\$a1211-2119\$\$1001^^\$aPrávo\$\$vRoč. 13, č. 152 (20030701), s. 2  
 \$\$1210^^\$aPraha\$\$cBorgis  
 5451 L \$\$aStyl  
 700 1 L \$\$aFialová\$\$bZuzana  
 801 1 L \$\$aCE\$\$bANOPRESS\$\$c20031022  
 801 2 L \$\$aCZ\$\$bABA001\$\$c20031022\$\$gAACB2  
 8564 L \$\$uhttp://full.nkp.cz/nkdb/view/nkdoc.asp?pi=PR20030701000215\$\$9HTM  
 8564 L \$\$uhttp://www.anopress.cz/brana\_nk/nkdocnk.asp?pi=PR20030701000215\$\$9HTM  
 9564 L \$\$afull.nkp.cz\$\$d/nkdb/view/nkdoc.asp?pi=PR20030701000215\$\$zCP 1250\$\$9HTM  
 910 L \$\$aABA001\$\$tRd  
 ZAZ0 L \$\$gPrávo\$\$hISSN 1211-2119\$\$sRoč. 13\$\$zč. 152\$\$d20030701\$\$1s. 2\$\$a(Praha)  
 ST2 L \$\$a20031022  
 LXR L \$\$aUP\$\$1NKC01\$\$nvazba na zdrojový  
 dokument\$\$uhttp://www.anopress.cz/brana\_nk/nkdocnk.asp?pi=PR20030701000215  
 VAI L 16


Negerovat Unimarc-A z formátu - zachovat současný stav

## Generovaný UNIMARC klasický



TTDE [3.0.014 - 21.10.2003] - Microsoft Internet Explorer

Form1 Form2 Form3 Form4 Text Urim\_A Urim Marc 21 Dubln Hlav.1 Hlav.2 Opravy ...



**Seznam článků**

PP(13.12) Labant ve Špatě, Mareš za  
 HQ(13.12) Labant definitivně patří týmu  
 PR(01.07) Josef Rakonajc vyměnil He  
 PR(01.07) Historista Kishon nerad vy  
 HQ(01.07) Příběhy ze života  
 HQ(01.07) Život seniorů: rány i výhry  
 HQ(01.07) Chováme se hůř než mmo  
 HQ(01.07) Bída dělá z dělá vky  
 JN(21.09) Neřekám, že jsem svůj záp  
 HA(01.07) KSČM se připravuje na tví  
 HA(01.07) KSČMcox se připravuje na  
 TY(03.02) Ivan Havel: Kdybych byl st  
 HQ(28.06) Bilance pilního projektu k  
 HQ(28.06) Z39.50 - protokol z minulé  
 HQ(29.06) Jsem vytrvalý optimista  
 HQ(29.06) K databázi KnowEurope m  
 HQ(29.06) Meze bezmezných inform  
 HQ(29.06) Mňo přede mnou, mňo za  
 HQ(29.06) Kde raději nepoužíval flash

LAB -----naa##22-----###450#  
 001 an2003abd-  
 100##fa20031022j20030701u##y0csey0103###ba  
 1010#facce  
 102##faCZ  
 105##fay###c###000yy  
 014##fa1211-2119(20030701)13:1S2[2:URN:NBN:cs-nkPR20030701000215]3.0.CO;E-8  
 #2xcic#9URN:NBN:cs-nkPR20030701000215  
 2001##fa#historista Kishon nerad vypráví vtipy#Zuzana Fialová  
 304##fa#ubrika: Styl  
 330##fa#Ne to, že je jeden z nejčtenějších humoristů na světě, ale že vyhrál  
 mistrovství světa v kulečnicku a večerel s Charliem Chaplinem  
 461#1#1011##fa1211-2119#12001##faPrávo#v#oč. 13, č. 1S2 (20030701), s. 2  
 #1210##fa#raha#f#Bergis  
 5451##fa#Styl  
 700#1#faFialová#bZuzana  
 801#1#faCZ#b#AN09P88S#c20031022  
 801#2#faCZ#b#AN01#c20031022#fgAACR2  
 8564#fu#http://full.nkp.cz/nkdb/view/nkdoc.asp?pl=PR20030701000215#9HTM  
 8564#fu#http://www.anopress.cz/hrana\_nk/nkdocnk.asp?pl=PR20030701000215#9HTM  
 9564#fa#full.nkp.cz#fd/nkdb/view/nkdoc.asp?pl=f#PR20030701000215#faCP 1250#9HTM  
 910##fa#AR#001#c#d  
 VAN 16

Regenerovat Urimarc z formátů - zachovat současný stav

Dublin

TTDE [3.0.014 - 21.10.2003] Microsoft Internet Explorer

Form1 Form2 Form3 Form4 Test Unim-A Unim Marc 21 Dubln Hlav.1 Hlav.2 Opravy...

Seznam článků

- PP(13.12) Labat ve Sportě, Mareš za
- MO(13.12) Labat definitivně patří týmu
- PR(01.07) Josef Rakoncaj vyměří Hs
- FP(01.07) Humorista Kishon nerad vy
- MX(01.07) Příběhy ze života
- MX(01.07) Život seniorů: rány i výhrů
- MX(01.07) Chováme se hůř než mnoh
- MX(01.07) Bída dělá z dětí věky
- LN(21.09) Neřekám, že jsem svůj záp
- KA(01.07) KSČM se připravuje na sv
- KA(01.07) KSČM:xxx se připravuje na
- TY(03.02) Ivan Havel: Kdybych byl st
- HW(26.08) Bilance plebního projektu: M
- HW(26.08) Z39.50 - protokol z minuléh
- HW(29.08) Jsem vytvářej opátnista
- HW(29.08) K databázi KnowEurope m
- HW(29.08) Meze bezmeznych inform
- HW(29.08) Mha přede mnou, rňho za
- HW(29.08) Kde raději nepoužívat flash

```

<LINK rel="SCHEMA.dc" href="http://purl.org/dc">
<META name="DC.Title" content="Humorista Kishon nerad vypráví vtipy">
<LINK rel="schema.DC" href="http://purl.org/metadata/dublin_core_elements#Title">
<META name="DC.Creator.personalName" content="Fialová, Zuzana">
<LINK rel="schema.DC" href="http://purl.org/metadata/dublin_core_elements#Creator">
<META name="DC.Description.abstract" content="Na to, že je jeden z nejčtenějších humoristů na světě, ale že vyhrál mistrovství světa v kulečnicku a večerel s Charliem Chaplinem">
<LINK rel="schema.DC" href="http://purl.org/metadata/dublin_core_elements#Description">
<META name="DC.Publisher" content="Borgis">
<LINK rel="schema.DC" href="http://purl.org/metadata/dublin_core_elements#Publisher">
<META name="DC.Date.issued" scheme="W3C-DTF" content="2003-07-01">
<LINK rel="schema.DC" href="http://purl.org/metadata/dublin_core_elements#Date.issued">
<META name="DC.Type" scheme="DCHIType" content="Text">
<LINK rel="schema.DC" href="http://purl.org/metadata/dublin_core_elements#Type">
<META name="DC.Type" scheme="DCHIType" content="Media">
<LINK rel="schema.DC" href="http://purl.org/metadata/dublin_core_elements#Type">
<META name="DC.Type" content="article">
<LINK rel="schema.DC" href="http://purl.org/metadata/dublin_core_elements#Type">
<META name="DC.Type" content="článek">
<LINK rel="schema.DC" href="http://purl.org/metadata/dublin_core_elements#Type">

```

Regenerovat Dublin z formuláře - zachovat současný stav

## Opravy rejstříků a metadat



TTDE [3.0.014 - 21. 10. 2003] - Microsoft Internet Explorer

Form.1 Form.2 Form.3 Form.4 Text Unim.A Unim. Marc 21 Dubín Hlav.1 Hlav.2 Opravy...



**Seznam článků**

PP(13.12) Labant ve Sportě, Mareš za  
 HO(13.12) Labant definitivně patří týmu  
 PR(01.07) Josef Rakoncaj vymění Ha  
 [R](01.07) Hnanová Křehou nerad vy  
 HO(01.07) Příběhy ze života  
 HO(01.07) Život seniorů: rány i výřeš  
 HO(01.07) Chováme se hůř než mnoh  
 HO(01.07) Bída dělá z dělí vky  
 JN(21.09) Neříkám, že jsem svůj záp  
 HA(01.07) KSČM se připravuje na sv  
 HA(01.07) KSČM:xxx se připravuje na  
 TY(03.02) Ivan Havel: Kdybych byl st  
 HO(26.08) Bilance pletního projektu H  
 HO(26.08) Z39.50 - protokol z minul  
 HO(29.08) Jsem vytrvalý optimista  
 HO(29.08) K databázi KnowEurope m  
 HO(29.08) Meze bezmezných inform  
 HO(29.08) Mha přede mnou, mha za  
 HO(29.08) Kde raději nepoužívat fash

**Dotaz do rejstříku:**

Počáteční řetězec:

Výběr rejstříku:

**Výpis z rejstříku:**

*(Přidat záznam do rejstříku)*

[Klaus Jr., Václav](#)  
[Klaus, Václav](#)  
[Klausmann, Alexandra](#)  
[Klausová, Kamila](#)  
[Klausová, Lívia](#)  
[Klausová, Marie](#)

**Detail záznamu:**

Heslo rejstříku:

**Celkem nalezeno 211 článků**

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 ;  
 Následující

0.94 28.02.2002 Madá fronta Dnes str. B/5 Ekor  
 Do voleb svého kandidáta neřekneme

0.91 22.01.2003 Madá fronta Dnes str. B/6 Ekor  
 Jenom Havel a můj otec jsou státníci

0.90 03.03.2003 Madá fronta Dnes str. A/1, A/6-A/7  
 Václav Klaus: Jsem ve formě

0.89 24.07.2003 Reflex str. 62-64  
 Klaus 1968

0.89 05.01.2002 Magazin Práva str. 4-7  
 Václav Klaus dědeček, chlapuď, čtenář

0.88 29.10.2002 Týden str. 24-25  
 Měkké přistávání Václava Klause

0.88 14.12.2002 Lidové noviny str. 11  
 Nikdo z kandidátů není dokonatý

0.88 11.12.2002 Hospodářské noviny str. 03  
 Havel byl stranický prezident. Já bych nebyl

0.87 16.01.2002 Madá fronta Dnes str. B/8 Ekor  
 Železného bych klidně přilal znovu

0.87 07.11.2001 Madá fronta Dnes str. B/8 Ekor  
 Někteří lidé zůstávají až do senility

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 ;  
 Následující

TTDE [3.0.014 - 21. 10. 2003] - Microsoft Internet Explorer

Form.1 Form.2 Form.3 Form.4 Text Unim.A Unim. Marc 21 Dubín Hlav.1 Hlav.2 Opravy...



**Seznam článků**

PP(13.12) Labant ve Sportě, Mareš za  
 HO(13.12) Labant definitivně patří týmu  
 PR(01.07) Josef Rakoncaj vymění Ha  
 [R](01.07) Hnanová Křehou nerad vy  
 HO(01.07) Příběhy ze života  
 HO(01.07) Život seniorů: rány i výřeš  
 HO(01.07) Chováme se hůř než mnoh  
 HO(01.07) Bída dělá z dělí vky  
 JN(21.09) Neříkám, že jsem svůj záp  
 HA(01.07) KSČM se připravuje na sv  
 HA(01.07) KSČM:xxx se připravuje na  
 TY(03.02) Ivan Havel: Kdybych byl st  
 HO(26.08) Bilance pletního projektu H  
 HO(26.08) Z39.50 - protokol z minul  
 HO(29.08) Jsem vytrvalý optimista  
 HO(29.08) K databázi KnowEurope m  
 HO(29.08) Meze bezmezných inform  
 HO(29.08) Mha přede mnou, mha za  
 HO(29.08) Kde raději nepoužívat fash

**Dotaz do rejstříku:**

Počáteční řetězec:

Výběr rejstříku:

**Výpis z rejstříku:**

*(Přidat záznam do rejstříku)*

[Klaus Jr., Václav](#)  
[Klaus, Václav](#)  
[Klausmann, Alexandra](#)  
[Klausová, Kamila](#)  
[Klausová, Lívia](#)  
[Klausová, Marie](#)

**Detail záznamu:**

Heslo rejstříku:

**Celkem nalezeno 211 článků**

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 ;  
 Následující

0.94 28.02.2002 Madá fronta Dnes str. B/5 Ekor  
 Do voleb svého kandidáta neřekneme

0.91 22.01.2003 Madá fronta Dnes str. B/6 Ekor  
 Jenom Havel a můj otec jsou státníci

0.90 03.03.2003 Madá fronta Dnes str. A/1, A/6-A/7  
 Václav Klaus: Jsem ve formě

0.89 24.07.2003 Reflex str. 62-64  
 Klaus 1968

0.89 05.01.2002 Magazin Práva str. 4-7  
 Václav Klaus dědeček, chlapuď, čtenář

0.88 29.10.2002 Týden str. 24-25  
 Měkké přistávání Václava Klause

0.88 14.12.2002 Lidové noviny str. 11  
 Nikdo z kandidátů není dokonatý

0.88 11.12.2002 Hospodářské noviny str. 03  
 Havel byl stranický prezident. Já bych nebyl

0.87 16.01.2002 Madá fronta Dnes str. B/8 Ekor  
 Železného bych klidně přilal znovu


0.87 07.11.2001 Madá fronta Dnes str. B/8 Ekor  
 Někteří lidé zůstávají až do senility

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 ;  
 Následující

## Odeslání dat pro ANL a ANL FULL

TTDI [3.0.014 - 21.10.2003] - Microsoft Internet Explorer

Form.1 Form.2 Form.3 Form.4 Text Unim.-A Unim. Marc.21 Dublin Hlav.1 Hlav.2 Opravy...



**Seznam článků:**

- PP(13.12) Labant ve Sportu - Marec 24
- WO(13.12) Labant dělničně patří tým
- PR(01.07) Josef Rakoncaj vyměnil Hř
- PR(01.07) Hanořista Kishon nerál vy
- WX(01.07) Příběhy ze života
- WX(01.07) Život seniorů - rány i výhny
- WX(01.07) Chováme se hůř než mnoh
- WX(01.07) Bída dělá z děje vlky
- LN(21.09) Neříkám, že jsem svůj záp.
- MF(03.03) Václav Klaus: Jsem ve fon
- HA(01.07) KSČM se připravuje na zvč
- HA(01.07) KSČMcoox se připravuje na
- TY(03.02) Ivan Havel: Kdybych byl st
- WR(26.08) Bilance pilotního projektu W
- WR(26.08) Z39.50 - protiskol z minulét
- WR(29.08) Jsem vytrvalý optimista
- WR(29.08) K databázi KnowEurope m
- WR(29.08) Měze bezmeznych informo
- WR(29.08) Mha přede mnou, mha za
- WR(29.08) Kde raději nepoužívat Bash

### 1. Formát a kódování

FORMÁT:  Formát 'Unimarc - Aleph'  Formát 'Unimarc'

KÓDOVÁNÍ:  ANSEL  UTF-8  UNICODE

### 2. Různé volby

Načítat klíčová slova z textu článku (dle tématu)

### 3. Další akce

[Obnovit zrušené články](#)

[Vyprázdnit vstupní archívy](#)

[Vyprázdnit výstupní archívy](#)

### 4. Načíst článek z URL adresy

[Načíst data](#)

I