

SOUBORNÁ DATABÁZE
KOOPERAČNÍHO SYSTÉMU ČLÁNKOVÉ BIBLIOGRAFIE
- OPTIMALIZACE INTEGRACE A SPRÁVY HETEROGENNÍCH DAT
Souhrnná zpráva za rok 2001

(Redakčně upraveno)

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



Praha, listopad 2001

OBSAH

A

Konstatační část

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

B

Analytická část

- B.1 Vlastní řešení**

- B.1.1 Linka automatické indexace
 - B.1.2 TOPIC a báze ANL FULL
 - B.1.3 Standardizace - bibliografická metadata ve formátu UNIMARC, DC, HTML, XHTML, XML, propojení záznamů s plnými texty
 - B.1.4 Skutečný stav propojení bibliografických záznamů s plnými texty
 - B.1.5 Souborná databáze ANL Kooperačního systému článkové bibliografie (KOSABI)
 - B.1.6 Česká národní bibliografie - řada Články v českých novinách, časopisech a sbornících na CD-ROM
 - B.1.7 Management Kooperačního systému článkové bibliografie (aplikace MNG KOSABI)
 - B.1.8 HW a SW podpora KOSABI a pro bázi ANL FULL
 - B.1.9 Smluvní ošetření projektu a Kooperačního systému článkové bibliografie
 - B.2 Přínos řešitele
 - B.3 Posun znalostí
- C

Návrhová část

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

Použití finančních prostředků

- D.1 Komentář
 - D.2 Tabulky (vynecháno)
- E

Resumé a klíčová slova

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

Přílohy

(vynecháno)

- F.1 Báze ANL v NK ČR
- F.2 Linka automatické indexace bibliografických záznamů z plných textů - bibliografická metadata pro bázi ANL (ALEPH) a plné texty s Dublin Core pro bázi ANL FULL (TOPIC)
- F.3 Server FULL.NKP.C
- F.4 Tituly v bázi ANL FULL, tabulka návštěví pro vyhledání v poli dotaz, tabulka formulářových polí, topiky, citace článku. ČNB. Aplikace pro správu a údržbu KOSABI. ProQuest. EBSCO.

A Konstatální část

A.1 Rešerše

Tištěné dokumenty:

1. KOSEK, J. XML pro každého : podrobný průvodce. 1. vyd. Praha : Grada 2000. 163 s.
2. OPPENHEIM, Charles and SMITHSON, Daniel. What is the hybrid library? *Journal of Information Science*. 1999, vol. 25, no. 2, s. 97-112.
3. Topic : systém pro inteligentní vyhledávání dokumentů. Praha : Tovek, 19?, 77 s.

Tištěné a elektronické dokumenty

4. ANDĚROVÁ, Ivana. 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. Též dostupný z: <http://full.nkp.cz/nkdb/index.html>.
5. ANDĚROVÁ, Ivana. Současný stav a perspektivy kooperačního systému článkové bibliografie. *Národní knihovna : knihovnická revue*. 1995, roč. 6, č. 1, s. 39-42. Též dostupný z: <http://full.nkp.cz/nkdb/index.html>.
6. 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. Též dostupný z: <http://full.nkp.cz/nkdb/index.html>.
7. BRATKOVÁ, Eva. Metadata jako nový nástroj pro komunikaci webovských informačních zdrojů. *Národní knihovna : knihovnická revue*. 1999, roč. 10, č. 4, s. 178-195. Též dostupný z: <http://full.nkp.cz/nkdb/index.html>.
8. ČERVENÝ, Vlastimil. Vyhledávání v databázích plných textů. *Národní knihovna : knihovnická revue*. 1999, roč. 10, č. 1, s. 6-12. Též dostupný z: <http://full.nkp.cz/nkdb/index.html>.
9. PSOHLAVEC, Stanislav: Z39.50 versus (?) XML. *Národní knihovna : knihovnická revue*. 2001, roč. 12, č. 1, s. 45-46. Též dostupný z: <http://full.nkp.cz/nkdb/index.html>.
10. Záznam pro soubornou databázi : UNIMARC. Fyzicky nesamostatné části dokumentů. Tištěné monografie a seriály. 1. vyd. Praha : Národní knihovna České republiky, 1999. 45 s. (Standardizace ; č. 19). Též dostupný z: nkpnew.2b.cz/pages/page.php3?page=fond_anal_unim_opr.htm.
11. Záznam pro soubornou databázi : Výměnný formát. Fyzicky nesamostatné části dokumentů. Tištěné monografie a seriály. 1. vyd. Praha : Národní knihovna České republiky, 1999. 39 s. (Standardizace ; č. 20). Též dostupný z: nkpnew.2b.cz/pages/page.php3?page=fond_ann_vf_opr.htm.

Elektronické dokumenty

12. 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: <http://ikaros.ff.cuni.cz/2000/c10/clanky.htm>.
13. ANDĚROVÁ, Ivana: propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů. Rok 1999, 2000, 2001 [online]. Dostupný z: nkpnew.2b.cz/pages/page.php3?page=oazp_propoj.htm.
14. ANDĚROVÁ, Ivana. Souborná databáze Kooperačního systému článkové bibliografie - optimalizace integrace a správy heterogenních dat. Souhrnná zpráva za rok ... [online]. Dostupný z: http://nkpnew.2b.cz/pages/page.php3?page=oazp_grant.htm.
15. Biblink [online]. Bath (Anglie) : UKOLN, last updated 12-Jul-2000 [cit. 2001-03-14]. Dostupný z: <http://hosted.ukoln.ac.uk/biblink/>.
16. 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 [cit. 2001-11-14]. Dostupný z: <http://www.inforum.cz/infomedia98/>.
17. CASLIN 2001. Popis a zpřístupnění dokumentů : nová výzva [online]. Beroun, 27.-31. května 2001 [cit. 14.11.2001]. Dostupný z: <http://www.caslin.cz:7777/caslin01/sbornik/index.htm>.
18. CELBOVÁ, Ludmila. Informace o projektu registrace domácích internetových zdrojů nově na serveru WebArchiv. *Ikaros* [online]. 2001, č. 5 [cit. 2001-05-01]. Dostupný z: <http://ikaros.ff.cuni.cz/2001/c05/webarchiv.htm>.
19. CELBOVÁ, Ludmila. Elektronické zdroje publikované v síti Internet jako součást České národní bibliografie. *Ikaros* [online]. 2000, č. 6 [cit. 2000-06-01]. Dostupný z: <http://ikaros.ff.cuni.cz/2000/c06/elzdroje.htm>.
20. Cobra+ : Computerised Bibliographic Record Actions [online]. Boston Spa (Velká Británie) : COBRA+, 1997 [cit. 2000-04-10]. Dostupný z: <http://portico.bl.uk/gabriel/en/projects/cobra.html>.
21. Dieper : digitised European periodicals [online]. Dostupný z: <http://gdz.sub.uni-goettingen.de/dieper/home.htm>.
22. DOI, the Digital Object Identifier System [online]. Kidlington (Oxford, Velká Británie) : International DOI Foundation, 1998, updated 4 April 2000 [cit. 2000-04-10]. Dostupný z: <http://www.doi.org/>.
23. Dublin Core Metadata Initiative [online]. Dublin (Ohio, USA) : OCLC, 2000 [cit. 2000-04-10]. Dostupný z: <http://purl.org/dc/>.
24. H.W. Wilson Company Selects Verity to Power the New WilsonWeb Site - the Premier Reference Resource for Librarians and Researchers [online]. Sunnyvale, Calif. and New York, NY, January 8, 2001 [cit. 2001-11-14]. Dostupný z: <http://www.verity.com/press/2001/20010108.html>.
25. ELAG. Integration Heterogeneous Resources [on line]. Prague, 6-8 June 2001 [cit. 2001-11-14]. Dostupný z: <http://www.stk.cz/elag2001/ELAG2001.html>.
26. HEIJTING, Inge. Interconnectivity and the Hybrid Library. *Ikaros* [online]. 1999, č. 10 [cit. 1999-11-01]. Dostupný z: <http://ikaros.ff.cuni.cz/ikaros/1999/c10/ebsco.htm>.
27. JONÁK, Zdeněk. Inteligence systémů zpracování textů. *Ikaros* [online]. 2000, č. 1 [cit. 2000-01-05]. Dostupný z: http://ikaros.ff.cuni.cz/ikaros/2000/c01/isko/z_jonak.htm.

28. JONÁK, Zdeněk. Krize mezilidské komunikace v období komunikační a informační exploze. *Ikaros* [online]. 1999, č. 5 [cit. 1999-05-01]. Dostupný z: <http://ikaros.ff.cuni.cz/ikaros/1999/c05/veda4.htm>.
29. JONÁK, Zdeněk. Pojem "informace" ve světě sdíleného pojetí skutečnosti. *Ikaros* [online], 2000, č. 2 [cit. 2000-02-01]. Dostupný z: <http://ikaros.ff.cuni.cz/ikaros/2000/c02/veda.htm>.
30. JONÁK, Zdeněk. Pokles důvěry ve vědu jako důsledek změny paradigmatu vědy : důsledky změny paradigmatu v informační vědě. Část 1. *Ikaros* [online]. 1999, č. 2 [cit. 1999-02-01]. Dostupný z: <http://ikaros.ff.cuni.cz/ikaros/1999/c02/veda.htm>.
31. JONÁK, Zdeněk. 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: <http://ikaros.ff.cuni.cz/ikaros/1999/c03/veda2.htm>.
32. JONÁK, Zdeněk. Vztah komunikační a obsahové struktury literárního díla. *Ikaros* [online], 1999, č. 6 [cit. 1999-06-01]. Dostupný z: <http://ikaros.ff.cuni.cz/ikaros/1999/c06/kom.htm>.
33. KOCH, Traugott and BORELL, Mattias. Dublin Core Metadata Template [online]. Lund (Švédsko) : Lund universitetsbibliotek, 1997, last update 1997-08-20 [2000-04-10]. Dostupný z: http://www.lub.lu.se/metadata/DC_creator.html.
34. Metadata [online]. Bath (Anglie) : UKOLN, last updated 16-Feb-2000 [cit. 2000-04-10]. Dostupný z: <http://www.ukoln.ac.uk/metadata/>.
35. NEDLIB : Networked European Deposit Library [online]. Hague (Nizozemí) : Koninklijke Bibliotheek, c1998, last updated 11-Mar-2001 [cit. 2001-04-28]. Dostupný z: <http://www.kb.nl/nedlib>.
36. Networked European Deposit Library [online]. Hague (Nizozemí) : Koninklijke Bibliotheek, last upd. 11-Mar-2001 [cit. 2001-03-14]. Dostupný z: <http://www.kb.nl/nedlib>.
37. Nordic Countries URN-generator : provided by the Nordic Libraries [online]. Lund (Švédsko) : Lund universitetsbibliotek, 1997 [cit. 2000-04-10]. Dostupný z: <http://www.lub.lu.se/cgi-bin/nmurn.pl>.
38. The Nordic Metadata projects [online]. Helsinki (Finsko) : Helsinki University, 1996, last updated 21 February 2000 [cit. 2001-04-28]. Dostupný z: <http://www.lib.helsinki.fi/meta>.
39. OLSON, Nancy B. Cataloging Internet Resources [online]. Dublin (Ohio, USA) : OCLC, 1997 [cit. 2000-04-10]. Dostupný z: <http://www.purl.org/oclc/cataloging-internet>.
40. PAPÍK, Richard. Trendy v rozvoji informačních služeb. *Ikaros* [online]. 1999, č. 8 [cit. 1999-09-01]. Dostupný z: http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_papik.htm.
41. Projects at the Royal Library in Stockholm, Sweden [online]. Stockholm : Royal Library, updated July 1, 1999 [cit. 2000-11-14]. Dostupný z: <http://www.kb.se/ENG/projekt.htm>.
42. Serial Item and Contribution Identifier [cit. 2000-11-14]. Dostupný z: <http://sunsite.berkeley.edu/SICI/version2.html>.
43. SICI Generator [cit. 2000-11-14]. Dostupný z: <http://www.ep.cs.nott.ac.uk/~sgp/sicisend.html>.
44. SVOBODA, Martin. Elektronické publikování. *Ikaros* [online], 1999, č. 3 [cit. 1999-03-01]. Dostupný z: <http://ikaros.ff.cuni.cz/ikaros/1999/c03/elpubl98/index.htm>.
45. The Nordic Metadata projects [online]. Helsinki (Finsko) : Helsinki University Library, 1996, last updated 21-Feb-2000 [cit. 2001-04-28]. Dostupný z: <http://www.lib.helsinki.fi/meta>.

46. TKACÍKOVÁ, Daniela. Když se řekne digitální knihovna ... *Ikaros* [online], 1999, č. 8 [cit. 1999-09-01]. Dostupný z: http://ikaros.ff.cuni.cz/ikaros/1999/c08/usti/usti_tkacikova.htm.
47. UHLÍŘ, Zdeněk. "Computing in Humanities", čili: Táhneme, anebo jsme vlečeni? *Ikaros* [online], 1999, č. 11 [cit. 1999-12-01]. Dostupný z: <http://ikaros.ff.cuni.cz/ikaros/1999/c11/computing.htm>.
48. Uniform Resource Names (urn) Charter [online]. Reston (VA, USA) : IETF, last modified 03-Jun-99 [cit. 2000-04-10]. Dostupný z: <http://www.ietf.org/html.charters/urn-charter.html>.
49. VOJTÁŠEK, Filip. Služby iDNES se rozrostly o regionálně členěný katalog odkazů. *Ikaros* [online]. 2001, č. 4 [cit. 2001-02-04]. Dostupný z: <http://ikaros.ff.cuni.cz/2001/c04/welcome.htm>.
50. ŽABIČKA, Petr. Dublin Core - metadata pro popis elektronických dokumentů. Předneseno na konferenci DATASEM 2000, konané 21. až 24. října 2000 v Brně. Dostupný z: <http://webarchiv.nkp.cz/datasem2000.pdf>.

Praktické výsledky grantu:

51. ANL FULL - Plnotextové vyhledávání v článcích z tisku. Topic system (experiment) [online]. Dostupný z: <http://full.nkp.cz>.
52. Báze ANL [online]. Dostupný z: http://sigma.nkp.cz/F/3TLEBXX6XQ7FSA6637D7F4YBYMMSN271ASJC5YTXBJBXGH66CY-31565?func=file&file_name=find-a&local_base=anl .
53. Metodika popisu článků ve formátu UNIMARC [online]. Dostupný z: http://nkpnew.2b.cz/pages/page.php3?page=oazp_popis1.htm .
54. Seriály (periodika) a analytický popis (články) v České republice, plné texty Propojení bibliografických záznamů s plnými texty [online]. Dostupný z: http://nkpnew.2b.cz/pages/page.php3?page=oazp_anal_popis.htm.
55. [Výsledky práce společnosti ANOPRESS, s.r. o. a informace o veřejné dražbě na adrese, online]. Dostupný z: <http://www.anopress.cz>.

A.2 Současný stav ve světě a v ČR

A.2.1 Obecně

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

Elektronické dokumenty jsou zpřístupňovány prostřednictvím nakladatelství, distributorských firem, informačních institucí či služeb a jejich produktů, dále pak prostřednictvím digitálních knihoven a služeb vznikajících na základě projektů, konzorcií a licencí. Přístup k plným textům je zajišťován přes různé formy bibliografií a soupisů, obsahů časopisů a plnotextových databází. Vyhledávání v plných textech zvyšuje komfort přístupu uživatelů k informacím. Elektronické dokumenty jsou zpřístupňovány v dohodnutých formátech, např. JPEG, GIF, PDF, TIFF, HTML. Služby knihoven jsou založeny na typu služby "document delivery".

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

Kromě vyhledávačů typu "search engines" (Alta Vista aj.) nebo předmětových katalogů Internetu (Yahoo! aj.) registrující informační zdroje v nestructurované podobě a u nichž relevance jejich zpětného vyhledávání je značně problematická, se přímo v prostoru Internetu a webu objevují registrační systémy, které přistupují ke zpracování těchto zdrojů přes strukturované záznamy. Tyto údaje mohou být obsažené ve zdrojích samotných (metadata).

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

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

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

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

Propojují se katalogy knihoven, záznamy s plnými texty dokumentů, "síťové dokumenty", je podporována spolupráce s muzei, archívy apod. Hovoří se o popisu dokumentu v hierarchii jako manifestace díla (čtyřúrovňový model manifestace díla **FRBR**) - vztahy mezi dílem, jeho vyjádřením, projevem a exemplářem.

Velká pozornost je věnována standardizaci popisu fondů. **Informační brány a portály** jsou na pořadu dne.

K identifikaci služeb, zdrojů a objektů na internetu slouží nestabilní **URL (Uniform Resource Locator)**, **PURL (Persistent URL)**, **URN (Uniform Resource Name)**, **DOI (Document Object Identifier)**.

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

A.2.2 Současný stav v zahraničí

Některé (digitální) knihovny, služby, projekty aj. instituce zabývající se zpřístupněním sekundárních informací o člancích a zpřístupněním plných textů

- **UNCoverWeb - A Current Awareness and Un Cover** je databáze registrující články z časopisů na základě obsahů přebíraných asi z 17000 titulů periodik. Databáze obsahuje stručné záznamy více než 7000000 článků, které vyšly v odborných a vědeckých časopisech od r. 1988. Hlavním cílem služby je poskytování plných textů článků - placená služba.
- Nejkompletnější přístup k elektronickým časopisům nabízí **OCLC FirstSearch Electronic Collection Online**. Interface umožňuje prohledávání periodik i čísel podle různých kritérií, přístup k citacím časopisů je zdarma, přístup k abstraktům a plným textům je možný jen u předplacených titulů, OCLC podporuje konzorciální přístup včetně přístupu kombinovaného s individuálním a "document delivery" (viz též dále).
- Ve **Švédsku** jsou články prezentované v systému **LIBRIS**. Záznam článku je možno zobrazit ve zkrácené podobě i ve struktuře MARC. Formulář obsahuje hypertextové odkazy na knihovny, v jejichž fondu se titul nachází.
- V **DBC (Dánském knihovnickém centru)** se zpracovávají články a recenze v rámci báze **BASIS** (ročně 30000 článků a 20000 recenzí z 9000 dánských periodik). Toto centrum buduje souborný katalog **DANBIB**, přes který lze zaslat objednávku elektronické kopie článku. Je propojený se švédským souborným katalogem LIBRIS a norský souborným katalogem **BIBSYS**.
- Univerzitní knihovna v **Helsinkách** (plní funkci Finské národní knihovny) provozuje centrální knihovnický systém VTLS sítě **Linnea**, v rámci které je zpřístupňována databáze článků **ARTO** z tisíce finských periodik. Ve Finsku existuje rozsáhlý projekt pro zpřístupňování elektronických zdrojů - **FinLib**.
- **NORDINFO** - projekt skandinávského souborného virtuálního katalogu zohledňuje aspekty specifické pro severské země - předmětová hesla, klasifikační systémy, národní systémy identifikátorů.
- **The Nordic Metadata project** - kooperační projekt severských zemí jako jeden z prvních řeší problematiku metadat Dublin Core v rámci spolupráce Norska, Dánska, Švédska, Finska a Islandu.

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

Bibliografické a plnotextové báze **dat americké firmy H.W. Wilson** pro společenskovední a humanitní obory (WilsonWeb a OmniFile) patří mezi nejlepší na světě. **Tyto báze jsou do určité míry vzorem pro budování kooperační databáze článkové bibliografie a mají s ní mnoho společných metodických prvků** - proto se o nich zmiňujeme podrobněji.

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

Excerptovaná periodika/seriály a rovněž tak články v nich obsažené jsou vybírány s ohledem na názory a podněty amerických, zejména veřejných knihoven. Každá báze má stanovenou svoji excerptní základnu s tím, že některá periodika se excerptují i pro více bází dat. Cílem firmy je vytvářet báze záznamů/plných textů z klíčových periodik v dané oblasti.

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

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

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

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

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

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

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

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

Směry a nástroje pro integraci heterogenních zdrojů byly nosným tématem semináře ELAG, který se konal v r. 2001 v Praze. Zmíníme se o několika **tématech diskutovaných na tomto semináři a o některých projektech zde referovaných. Tyto otázky jsou na pořadu dne i v ČR.**

Projekty

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

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)

Nástroje integrace heterogenních dat: XML, identifikace zdrojů, propojování, protokoly, digitální knihovny

DC (Dublin Core)

Formát metadat pro popis webovských informačních zdrojů - formát definovaný na základě mezinárodního konsenzu - obsahuje 15 prvků k identifikaci zdroje.

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 je formát textový, tzn., že dokument je možno vytvářet, přenášet a zpracovávat na libovolná data, jakýmkoli počítačovým systémem, jsou vyřešeny jazykové a kódové problémy. 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, textový formát apod.). Každý dokument má definovanou svoji strukturu prostřednictvím tzv. DTD (Document Type Definition). Tato DTD může být unikátní pro každý dokument, ale může být také společná pro celou řadu dokumentů. 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 je tzv. kádové styly (CSS)

Oblasti potenciálního využití XML jsou široké. Otevírá možnosti v klasickém i elektronickém publikování.

RDF (Resource Description Framework)

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

Propojování informačních zdrojů

FRBR Functional Requirements for Bibliographic Records - relace mezi Dílem, Vyjádřením díla, Provedením díla a Exemplářem díla (Work, Expression, Manifestation, Item). Propojování se zavádí komerčně: databáze statických odkazů, z abstraktových a indexových databází k plnému textu, z citace v plném textu na plný text, z OPAC k časopisu a jeho obsahu a odtud k plnému textu, většina odkazů je statických pro konkrétní případy předem budovaná. Dynamické linky jsou budovány následně, v době potřeby, jsou pravděpodobnostní.

Propojování pomocí URL, PURL, URN a DOI.

SFX (Special Effects) je databáze, která na základě jednotně stavěné identifikace dokumentu (Open URL) odvozené od metadat dokumentu a podle práv uživatele (context sensitive) nabízí další služby, jejichž dostupnost má uživatel zajištěnu. Práva uživatele musí SFX zaznamenat ve zvláštní databázi předem nastavené a podle toho bude nabízet služby dostupné obsluhovanému uživateli. Nabízené služby závisí též od obsahu OpenURL. V Open URL se může využít i identifikace dokumentů pomocí DOI, kdy SFX je schopen se spojit s databází CrossRef a převést DOI na OpenURL s bibliografickými údaji. Open URL vytváří ze svých metadat řada předních světových poskytovatelů on-line zdrojů.

Metalib a SFX. Metalib je multivyhledávač, který vyhledává v několika zdrojích najednou a to pomocí protokolu Z39.50, Aleph proprietary protocol, Http protokol přizpůsobený na vybrané cíle. Metalib obsahuje Knowledge base (seznam zdrojů, které může prohledávat), nástroje pro údržbu této knowledge base, a vyhledávací a presentační SW. Zdroje dat mohou doplňovat svá data tak, že z metadat formují tzv. OpenURL.

Z39.50 je soustava protokolů sloužících ke komunikaci mezi heterogenními knihovnickými systémy. Umožňuje vyhledávat, editovat, upload a download záznamů.

Jednotné informační brány řeší přístup k různým zdrojům z jednotného prostředí.

Jejich funkčnost je podmíněna homogenním nástrojem pro indexaci zejména věcnou.

Řešením je harmonizace řízených slovníků či tezaurů.

Konzorcia

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

Členy některých konzorcií jsou i české knihovny a instituce .

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

JSTOR - Journal Storage

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

Zabezpečení vědy a výzkumu v humanitních oborech základními informačními zdroji -

- celonárodní licence na on-line přístup do [fulltextové databáze ProQuest 5000](#) (plné texty 5000 humanitně orientovaných časopisů) a bibliografické databáze PCI Web (Periodicals Contents Index). Databáze: Arts, Law, Humanities, Women, Social Sciences Plus Text, Education Complete, Career and Technical Education, Medical Library, Health, Pharmaceutical News Index, Applied Science and Technology, Computing, Telecommunications, Religion, ABI/Inform Global, European Business, Asian Business, Accounting and Tax, Banking Information Source.

Zpřístupnění plnotextových databází odborných zahraničních periodik na základě programu

Open Society Institute EIFL-Direct - plošná multilicence.

Plné texty celkem 3.300 časopisů od r. 1990 a další informační 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 titulů 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 12 dalších databází:

ArticleFirst (bibliografické citace článků z 13.000 periodik), ContentFirst (seznamy obsahů periodik), NetFirst (bibliografické údaje o odborných zdrojích na Internetu, včetně abstraktů a klasifikace), PapersFirst (referáty ze světových konferencí, kongresů, 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.

A.2.3 Současný stav v ČR

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

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

Jejich zařazování do tematicky či jinak utříděných portálů v knihovnických iniciativách je zatím v plenkách. V předkládaném projektu je mj. tento portál navržen.

Relativně velký rozvoj na Internetu nastal v nabídce **českých novinových a časopiseckých elektronických zdrojů** - jsou vystaveny deníky, týdeníky a časopisy s různou hloubkou retrospektivy a úplnosti od volně přístupných přes registraci a služby placené. V některých elektronických zdrojích lze vyhledávat plnotextově. Vystavené zdroje jsou zachyceny s různou hloubkou retrospektivy (aktuální číslo, poslední čísla či roky, objevují se archivy volně dostupné či pouze registrovaným uživatelům. Některé zdroje obsahují citaci, abstrakt, objevují se current contents. Někdy lze vyhledávat podle základních formálních údajů, kombinovat dotazy pomocí logických operátorů, vyhledávat plnotextově (např. časopisy Vesmír, Euro, Rock a pop, Harmonie - vystavený volně archiv článků).

Některé zpravodajské servery na českém Internetu :

České noviny <http://ctk.ceskenoviny.cz/>,
iDNES <http://zpravy.idnes.cz/>,
Lidové noviny <http://www.lidovky.cz/>,
iHNed <http://www.ihned.cz/>,
Právo <http://www.pravo.cz/>.

Server iDNES nabízí služby v podobě regionálně členěného katalogu odkazů, částečně funguje jako předmětový katalog odkazů. Na rozdíl od Seznamu.cz, Atlasu nebo Centra.cz jsou odkazy v Klikni.cz uspořádány do 14 regionálních sekcí odpovídajících současnému územnímu uspořádání České republiky.

Některé vyhledávače na českém webu:

Seznam <http://www.seznam.cz/>,
Centrum <http://www.centrum.cz/>,
Redbox <http://www.redbox.cz/>,
Quick <http://www.quick.cz/>,

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.

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

Sagit (elektronické nakladatelství a právní texty) - <http://www.sagit.cz/>.

Tigis, s.r.o. (časopisy pro lékařskou odbornou veřejnost a edukační časopisy pro veřejnost) - <http://www.tigis.cz/>

Nakladatelství Muzikus (hudební tematika) - <http://www.muzikus.cz/>.

Portál <http://www.portal.cz/>.

Current Contents, abstrakty, citace, plné texty

AVČR - Časopisy vydávané Akademií věd České republiky, http://www.lib.cas.cz/knav/journals/Casopisy_AVCR.htm. 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. <http://www.cuni.cz/cuni/uz/nk/>.

Odborná knihovnická periodika

Národní knihovna. Knihovnická revue, <http://full.nkp.cz/nkkr/NKKR.html> (formát pdf a html pro rok 1999, 2000 a 2001 - webová prezentace periodika v rámci Projektu propojení analytických záznamů s plnými texty), ostatní čísla plnotextově přístupná na serveru FULL.NKP.CZ a www.anopress.cz.

Ikaros, elektronický časopis o informační společnosti, <http://ikaros.ff.cuni.cz>

Daidalos, informační server pro knihovníky (činnost ukončena), <http://daidalos.ff.cuni.cz/>

Bulletin SKIP, <http://skip.nkp.cz/Bulletin/Bulletin.htm>

U nás, <http://www.svkhk.cz/unas/>.

Některé databáze, digitální knihovny

V České republice existují některé oborové báze plnotextových informací, např. **ASPI** (Automatizovaný systém právních informací) zachycuje vývoj právní kultury, vztahující se k území současné České republiky i Slovenské republiky v rozsahu dvou století s výhledem na legislativu Evropských společenství.

V rámci **Parlamentní knihovny** se buduje systém, ve kterém jsou zpřístupněna v plné formě parlamentária. Digitální knihovna "**Český parlament**", <http://www.psp.cz/eknih/>.

Dokumenty Senátu, <http://www.senat.cz/dokumenty/index.htm>.

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 přes 1000 elektronických titulů předních světových vydavatelství. Elektronické vydavatelství spolupracuje na vydávání ČNB na CD-ROM.

Newton I.T.

Elektronická výstřižková služba, elektronický archiv novin a časopisů celostátních a regionálních, <http://www.newtonit.cz/>

Neumožňuje přímý přístup do celé databanky. Buduje následující archívy:

Deník Právo - <http://www.pravo.cz>

Deník Pražské slovo - <http://zn.newton.cz>

Časopis 100+1 zahraniční zajímavost - <http://stoplusjedna.newton.cz>

Elektronický archiv deníku Právo - <http://pravo.newtonit.cz>

Elektronický archiv deníku MF Dnes - <http://mfdnes.newtonit.cz>

Elektronický archiv týdeníku Týden - <http://tyden.newtonit.cz>

Elektronický archiv týdeníku The Prague Post- <http://praguepost.newtonit.cz>

Elektronický archiv týdeníku Respekt - <http://respekt.newton.cz>

Týdeník Profit - <http://www.profit.cz>.

Anopress, s.r.o.

On-line databanka novin a časopisů celostátních a regionálních, monitoring na zakázku, vědomostní databáze, <http://www.anopress.cz/>.

Anopress, s.r.o. umožňuje on-line přístup do databanky plných textů **TamTam**, na jejíž bázi poskytuje následné služby. Společnost zpřístupňuje informace zákazníkovi na dané téma. Anopress s.r.o. umožňuje přístup do databanky novin on-line na základě licenčních smluv a umožňuje nákup celých titulů periodik. Společnost Anopress je výhradním zpracovatelem elektronické podoby většiny českých regionálních titulů (51 titulů nakladatelství Bohemia). Pro zpřístupnění plných textů ve veřejných knihovnách bylo založeno **Konzorcium Anopress**. Společnost je výhradním zástupcem slovenské firmy SLOVAKIA ONLINE v ČR, která zpracovává elektronickou podobu slovenských tištěných médií. Kromě mediální části obsahuje databanka TAMTAM i část vědomostní, v níž jsou k dispozici pro fulltextové vyhledávání různé encyklopedie, příručky a další knihy referenčního charakteru.

Agentura od roku 1998 průběžně vytváří rozsáhlou databanku, která v současné době obsahuje texty článků a zpráv z devíti celostátních deníků a z téměř 80ti deníků regionálních a dále i textové záznamy zpravodajských, publicistických a diskusních pořadů rozhlasu a televize.

Společnost vyvinula vlastní software ISA, který umožňuje všechna data dále analyticky zpracovávat, exportovat je v několika formátech, čteně HTML, 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 cca 35 titulů, které odpovídají excerpční základně Kooperativního systému článkové bibliografie.

Dne 18.12. 2001 má být v dobrovolné dražbě informační agentura Anopress prodána.

Související projekty, metody a nástroje, související se zpřístupněním plných textů

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.

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.

Jednotná informační brána

Programový projekt Ministerstva kultury ČR "**Jednotná informační brána pro hybridní knihovny**" - řeší otázku jednotného uživatelského rozhraní pro hybridní knihovny. Takovou knihovnou je Metalib. Databáze ALEPH NK jsou jedním ze zdrojů této knihovny.

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

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

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

Propojování - Metalib a SFX

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

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

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

Bibliografické zpracování článků v ČR

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

Národní knihovna ČR zpracovává výběrově bibliografické záznamy článků ze všech druhů seriálů (noviny, časopisy, odborná periodika, sborníky) v rámci **Kooperačního systému článkové bibliografie (KOSABI)**, ve kterém spolupracují SVK, resp. krajské knihovny a MZK, specializované odborné knihovny (STK, ÚZPI, SPKK-ÚIV, ČSAV, nově ČGÚ). Na základě této spolupráce vzniká souborná databáze **ANL**. V systému **LANIUS** se zpracovávají bibliografické záznamy článků v knihovnách na úrovni okresů. V budoucnu je třeba sladit systém KOSABI a LANIUS tak, aby nedocházelo k duplicitnímu zpracování. Instituce spolupracující v rámci KOSABI přecházejí na nové SW vyšší generace, zatím probíhá ve většině popis článků v ISISu. V SVK Kladno se články popisují v systému RAPID, v MZK v Brně v ALEPH, v Olomouci též. V Ostravě a Českých Budějovicích se přechází na T-Series.

Kvalitní automatizovaná správa souborné databáze nutná. Souborná databáze KOSABI ANL obsahuje přes cca 670 000 záznamů, v NKČR se excerpuje se cca 210 titulů, 469 titulů ve spolupracujících institucích (278 specializované knihovny, 191 titulů v SVK a MZK). Přechod na zpracování v systému v ALEPH 500 v dubnu 2000 posunulo zpracování na úroveň mezinárodního formátu UNIMARC a pravidel popisu AACR2 s respektováním mezinárodních standardů věcného popisu - MDT-MRF pro oblast systematické indexace. V oblasti verbální věcné indexace se kombinují klíčová slova, věcné obecné kategorie a předmětová hesla.

Vyváženost vazby mezi jednotlivými vrstvami popisu je klíčovým momentem. V rámci kooperačního systému byla stanovena pravidla pro výběr titulů k popisu (na základě územní gesce - tituly regionální a celostátní provenience a dále pak na základě odborného zaměření). Dále byly stanoveny zásady výběru článků co do úplnosti i co do typů.

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 provenience (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 Anopress 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. Zároveň jsou rok 2001 a některá čísla r. 2000 vystavovány ve formátu html a pdf ve speciální webovské aplikaci pro toto periodikum.

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 řešený **SVK Kladno**.

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

V roce 2001 pokračuje spolupráce v rámci KOSABI, probíhá experiment linky automatické indexace a aplikace pro správu a údržbu KOSABI, aktualizuje se databáze ANL, ANL FULL. Báze ANL FULL je vybavena metadaty DC ve formátech html, XHTML, XML v kvalifikované i nekvalifikované formě. Pro zpřístupnění plnotextových informací je navržen nový layout serveru FULL.NKP.CZ. Pozornost se soustřeďuje na definici báze FULL.NKP.CZ - zejména na vyhledávání.

A. 3 Cíl, vstupní data

A. 3.1 Cíl

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

Cílem projektu v r. 2001 je aktualizace databáze ANL, ANL FULL a tedy i spolupráce v rámci KOSABI (po stránce faktické, metodické i organizační), tj. průběžné doplňování báze ANL FULL plnými texty a metadaty, dále pak aktualizace báze ANL bibliografickými záznamy s propojením na plné texty za současného vydávání báze ANL v rámci ČNB. Další cílem je experimentální zpracování záznamů o člancích z plných textů v rámci linky automatické indexace a zkvalitnění pracovního formuláře, experimentální odzkoušení aplikace pro správu a údržbu KOSABI, navržení aplikace pro evidenci zpracovávaných titulů v rámci KOSABI a navržení testu na kontrolu duplicity přijímaných záznamů. Dále pak vybavení báze ANL FULL metadaty DC ve formátech HTML, XHTML, XML v kvalifikované i nekvalifikované formě. Dalším cílem je nový moderní layout a struktura serveru FULL.NKP.CZ. Pozornost se soustřeďuje na definici báze FULL.NKP.CZ - zejména na vyhledávání (jednoduchý dotaz, formulářový dotaz, topiky, rejstříky). Dalším cílem je návrh a založení portálu na serveru FULL.NKP.CZ pro volně dostupné zdroje na českém Internetu z hlediska krajů a tematiky. Cílem je navržení strukturovaného portálu s popisem těchto zdrojů.

Cílem je tedy propojení plného textu se záznamem, které probíhá na základě jednak statických URL (volně přístupné texty na Internetu), jednak dynamických URL (generovaných v rámci linky automatické indexace) a realizace dříve připravených odkazů ze záznamů na plné texty zakoupené v minulých letech metodou off-line. Cílem je i odzkoušení propojení up and down mezi analytickými záznamy a zdroji v rámci báze NKC. Dalším cílem je doplňování plných textů on-line (linka automatické indexace) a off-line pro retrospektivní články.

Řešení pro rok 2001 spočívá v realizaci navržených metod v roce 2000.

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.**

A. 3.2 Vstupní data

Vstupními daty pro bázi ANL jsou jednak bibliografické záznamy zpracovávané v r. 2001 v rámci KOSABI, v rámci KOSABI bibliografické záznamy zpracované v NKČR jednak v ALEPHU a jednak v rámci linky automatické indexace a doplňované dynamickými URL za současného generování formátů UNIMARC, DC v HTML XHTML a XML. V roce 2001 se v lince automatické indexace zpracovává výběr z celostátních deníků a některých časopisů (Týden, Ekonom, Respekt, Reflex).

Vstupními daty pro bázi ANL FULL jsou tedy plné texty, které odpovídají profilu zpracování bibliografických záznamů vybavené metadaty DC v různých aplikacích. Plné texty se doplňují on-line v rámci linky automatické indexace. Zároveň se doplňují retrospektivně off-line (v roce 2001 byly doplněny zejména plné texty článků vydaných v roce 1997, 1998).

Bibliografické záznamy zpracované v ALEPHu jsou také propojované na plné texty statickými URL adresami - záznamy především titulů z oboru knihovnictví (U nás, Ikaros, aj.) a oborů dalších (Harmonie, Vesmír, Psychiatrie, Obecná psychologie, Veřejná správa, Učitelské noviny aj.).

Vstupními daty pro léta minulá jsou bibliografické záznamy a adekvátní plné texty (v roce 2001 je to výběr produkce bibliografických záznamů a plných textů zejména za léta 1997, 1998).

Summa summarum:

Vstupními daty pro plnotextovou bázi ANL FULL jsou plné texty zejména celostátních deníků a některých odborných časopisů za rok 1997-2001 a jim odpovídající bibliografické záznamy v tomtéž období pro bázi ANL v rámci KOSABI.

Vstupními daty pro automatizovanou linky 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.

B Analytická část

B.1 Vlastní řešení

Stručný popis produktů TamTam:

1. **Hlavní linka pro zpracování dokumentů**
 - a. **TamTam NK_special**
 - stažení vybraných článků z TamTamu v Anopressu do NK přes Internet
 - b. **TTDE - TamTam Data Extractor**
 - úprava hlavičkových údajů ve stažených člancích prostřednictvím 3 formulářů
 - výstup Aleph (i Unicode) a Full
 - hlavičky Uni, Uni-A, D.C., H1 (zobrazovací), H2 (indexovací)
 - odeslání zpracovaných dat do importní schránky Aleph a k fulltextovému indexování (Topic)
 - c. **MKINDEX - indexování, generování XML**
 - slouží k on-line indexování upravených článků do fulltextové db (Topic)
 - zároveň se on-line generují XML- a XML + (nové od 9/2001)

2. **Jednorázové aplikace/moduly**
 - a. **TTXMLGEN - TamTam XML generátor**
 - generuje XML- i XML+ z celé DB (nové od 9/2001)
 - (program TTXMLGEN.EXE na severu FULL.NKP.CZ)
 - b. **TTDT - TamTam Data Tester**
 - slouží k testování obsahu DB (nové od 8/2001)
 - vyhledá dokumenty, které obsahují nebo naopak neobsahují zadaný řetězec
 - 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
 - d. **TTDC - TamTam Data Converter**
 - konverze článků, stažených metodou Standard do fulltextové DB
 - poslední verze k 2.6.2001
 - pro použití se současnými hlavičkami je nutná jeho podstatná úprava
 - stahování me touto Standard by se nemělo už používat
 - e. **TTEXPORT - TamTam Export**
 - export dat z DB Anopress pro NK
 - poslední verze k 27.4.2000

3. **Internet**
 - 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
- d. **Zobrazení výsledků hledání - článku**
 - zobrazení textu, hlaviček (bez/s odkazy pro další hledání) - MKTXT.PHP
 - generování a zobrazení UNImarcu MKUNI.PHP
 - generování a zobrazení DC, XHTML- +, XML- + - MKXML.PHP
 - zobrazení klíčových slov - MKKEY.PHP
 - zobrazení XML- + - vygenerování při indexování dokumentu - MOFXML.PHP
- e. **Zobrazení plného textu z Alephu**
 - modul MKDOC.PHP
- f. **Statistika**
 - statistika zobrazených informací přes internet
 - modul STAT.ASP
- 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

Popis řešení

Pro optimalizaci integrace a správy heterogenních dat souborné databáze kooperačního vyvinula česká firma Anopress na podkladě analýzy a funkčního zadání návrh speciální technologie - linky automatického získávání plných textů, indexace bibliografických záznamů a plných textů, propojování záznamů na plné texty a jejich zpřístupnění. Řešení je progresivní a odpovídá nejnovějším trendům v této oblasti, je podpořeno kvalitním technickým a programovým vybavením. Jednotlivé moduly lze použít i samostatně. **V rámci experimentu v r. 2001 byla doladěna linka automatického zpracování** ve všech jejích fázích a uvedena od května t. r. do provozu. **V roce 2000 vyvinutý jednoduchý formulář byl rozšířen a doplněn kromě jiného zejména o pole věcného popisu a pole pro vazbu k propojení analytického záznamu se zdrojovým dokumentem v bázi NKC (vazba LKR).** Automaticky je generována URL adresa, SICI, "provizorní" URN, které sleduje filozofii a strukturu stanovenou pro tvorbu URN.

Automaticky lze generovat klíčová slova a převádět do formuláře. Na výstupu procesu zpracování byla zabudována **funkce kódování výsledného souboru záznamů v Ansel, UNICODu a UTF-8.** Poté je záznam odeslán do určeného adresáře na serveru FULL.NKP.CZ. Programovou aplikací vyvinutou v NK je **soubor naimportován do báze ANL (ALEPH) a zaindexován a zpřístupněn v systému TOPIC na serveru FULL.NKP.CZ v podobě plného textu vybaveného metadaty (UNIMARC a aplikace Dublin Core v podobě HTML, XHTML, XML v kvalifikované a nekvalifikované formě).** Vyhledání je možno z plného textu, z metadat v něm uložených (vyhledávání pomocí polí, formulářů, topiků a rejstříků). Zobrazení je možno pomocí tří uživatelských formátů. Vyřešen je také **tisk jak seznamu vyhledaných článků, tak tisk jednotlivých plných textů.** Export metadat a plných textů bude řešen v roce 2002. **V roce 2001 byla vyvinuta aplikace pro registraci, autentifikaci a přihlášení externích uživatelů prozatím na dobu 7 dnů.** Otázka zpřístupnění externím uživatelům závisí na vyřešení legislativně právních otázek event. ekonomických. **Externím uživatelům jsou běžně zpřístupňována metadata. Plné texty a metadata jsou zpřístupňována interním uživatelům NK.**

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

Předpokládá však do jisté úrovně strukturovaný vstupní text, ze kterého data mohou být extrahována.

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

Automatická indexace věcná je závislá na vytvoření tezaurů řízených heslářů s vazbou na klasifikační notaci v podobě MDT či UDC oproti kterým je porovnáván plný text a na základě tohoto porovnání přidělovány příslušné věcné termíny.

B. 1. 1 Linka automatické indexace

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). Řešení spočívá ve speciální aplikaci v praxi již používané technologie firmy na získávání a zpřístupňování plných textů pro NK - TAMTAM Profesional NK (TTPNK) - TamTam NK_special.

Pomocí této technologie, aplikace se **stahují plné texty** článků z Internetu z báze TAMTAM, založené plnotextovým pojmovým vyhledávacím systémem TOPIC. 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 levém části formuláře seznam názvů článků. Po kliknutí na článek se automaticky objeví údaje o článku, které jsou obsaženy v jeho hlavičce ve formuláři a příslušných hlavičkách. Po kontrole takto extrahovaných údajů z plného textu se doplňují další bibliografická metadata zejména věcné povahy. Poté je soubor zpracovaných článků odeslán. Tímto způsobem jsou zpracovávány všechny celostátní deníky a některé odborné časopisy.

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**.

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

- 1. Rozšířený formulář je nástroj vyvinutý pro editaci a doplnění extrahovaných dat. Data je možno upravovat** také v hlavičce ANIMARC-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 příslušný funkce, je možno automaticky generovat klíčová slova. Další **údaje věcného popisu** - předmětové kategorie doplněné MDT, hesla se doplňují. Automaticky se generuje URL, SICI, téměř celé pole LKR (k propojení se zdrojovým dokumentem - vazba up a naopak vazba down na analytický záznam - experimentálně zatím vazby zkoušeny u deníku Právo, Magazín Práva, Reflex, Respekt, Týden).
(pozn: v bázi ANL LKR také u titulů Slovo a slovesnost, Křesťanská revue, Dějiny a současnost).
Formulář má tři strany. První obsahuje jmenné a věcné údaje, druhá pouze věcné, třetí jmenné a věcné údaje.
- 2. UNIMARC-A hlavička** je hlavička s bibliografickými údaji **pro importní vstupní soubor záznamů pro ALEPH (řádkový UNIMARC)**, do které se automaticky generují tato data jako do Formuláře a úpravy zanesené do Formuláře, tato hlavička je také přístupná pro editaci samostatně. Jde o řádkový UNIMARC.
- 3. UNIMARC hlavička**, hlavička pro klasický UNIMARC s těmiž vlastnostmi jako hlavička UNIMARC-A - slouží k eventuelnímu importu pro systémy, které jsou založeny na UNIMARCU - řádkový UNIMARC. Do hlavičky jsou generované tytéž údaje jako do výše jmenované hlavičky.
- 4. DUBLIN hlavička** s těmiž vlastnostmi jako předchozí dvě hlavičky sloužící ke generování metadat **DUBLIN CORE** - vychází z poslední verze Dublin Set Elements, z 15 jeho prvků aplikuje 14. Pro jemnější vyhledávání jsou určeny tzv. prvky Anl Core.
- 5. Zobrazovací hlavička** slouží k zobrazení údajů v hlavičce plného textu.
- 6. Indexovací hlavička** slouží k indexaci uložených údajů v databázi ANL FULL.

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 nas taven výstup UNIMARC-A v kódu Ansel.**

Následuje odeslání záznamů pomocí voby "odeslat" na dolní liště, import do Alephu (program vyvinutý v NK), import na server FULL.NKP.CZ

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

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

Pro **propojení plných textů s bibliografickými záznamy v systému ALEPH** (doplnění URL adres do záznamů) byl vytvořen **skript MKDOC.PHP.** Propojení probíhá ne 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 a citaci dokumentu v rámci databáze ANL FULL.

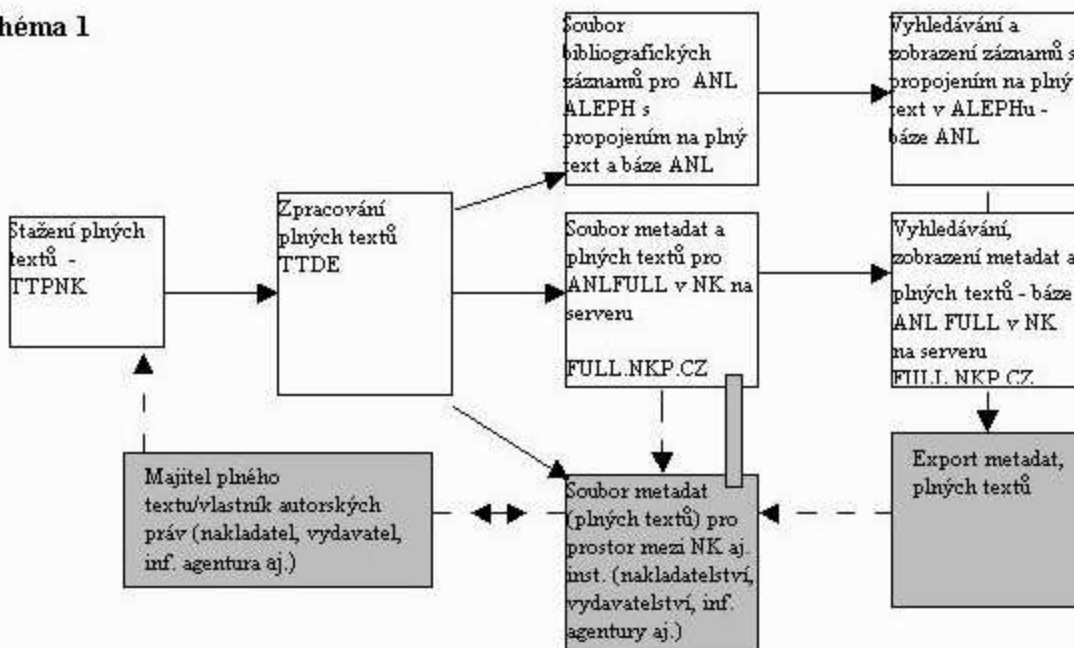
Zpracování článku v lince automatické indexace ilustruje následující sekvence obrázků:

[Obr. A](#), [Obr. B](#), [Obr. C](#), [Obr. D](#), [Obr. E](#), [Obr. F](#), [Obr. G](#), [Obr. H](#), [Obr. I](#)

Celý proces obecně ilustruje schéma na Obr. 1.

Obr. 1 Linka automatické indexace zabudovaná do předpokládaného možného vývoje

Schéma 1



Vysvětlivky: předpokládaný možný vývoj mezi autorem, nakladatelem, vydavatelem, informační agenturou, národní bibliografickou agenturou



B.1.2 [TOPIC a báze ANL FULL \(charakteristika serveru a databáze ANL FULL\)](#)

Společnost Anopress, s.r.o. 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 automatické indexace a zpřístupňovány v systému **TOPIC** v bázi ANL FULL s ohledem na nejnovější trendy v této oblasti (plnotextové pojmové vyhledávání) pomocí tzv. topiků. Pro přístup do celé databáze TAMTAM bylo založeno konzorcium Anopress v r. 2000.

Plné texty báze **ANL FULL** jsou zatím přístupné interním uživatelům NKČR. Externím uživatelům je běžně k dispozici pouze bibliografický popis (metadata), na dobu 7 dnů je možno získat i přístup k plným textům na základě zkušební registrace. 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 a v neposlední řadě na budoucím majiteli firmy Anopress.

Plný text je funkční v kódu v kódu CP 1250.

Inspirací a do jisté míry vzorem pro koncipování báze ANL FULL je databáze [ProQuest](#).

TOPIC (pojmově orientovaný vyhledávací systém, concept based retrieval) je systém třetí generace založený na následujících principech: rozklad pojmu na podpojmy, vážení jednotlivých podpojmů (větvi pojmového stromu), neostře vyhodnocování dotazů.

Dotaz v systému třetí generace reprezentuje pojem, resp. ideu vyhledávaného tématu. Jádrem dotazu je stromová hierarchická struktura, která rozkládá hledané téma na podtémata a přiřazuje jednotlivým částem váhy, které vyjadřují do jaké míry příslušné téma přispívá k celkovému určení tématu. Systém dále vypočítá míru relevance vyhledaných dokumentů. Oproti běžně používaným operátorům TOPIC používá logický operátor ACCRUE se specifickými vlastnostmi. Tento operátor sblízuje operátory and a or. Každý topik obsahuje tedy tři charakteristiky - strukturu, váhy a operátory

Nabízí se zde jistá formální analogie k hierarchickému selekčnímu jazyku systémové notace MDT.

Je však třeba zdůraznit, že topiky jsou tvořeny podle skutečnosti, MDT je víceméně taxativní systém jednotlivých oborů, nikoli témat. Proto je třeba k definici topiků přistupovat svébytně.

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

B.1.2.1 Charakteristika serveru [FULL.NKP.CZ](#)

Server FULL.NKP.CZ je určen primárně ke zpřístupňování plnotextových dokumentů, v současné fázi zejména těch, které jsou obsaženy v periodicky vydávaných dokumentech české provenience.

B.1.2.1.1 Základní členění stránek - základní a dílčí nabídky

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

Obecné (základní) informace

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

Další databáze (databáze NK - charakteristika báze ANL)

Periodika na WWW (volně dostupná periodika na www)

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

Vstup do db (ANL FULL)

Jak se zaregistrovat (komentář)

Registrace (formulář)

Přihlášení (formulář)

Vyhledání (ANL FULL)

Jak vyhledávat (metodika)

Formuláře

Rejstříky

B.1.2.1.2 Charakteristika databáze ANL FULL (obsah, zdroje)

Databáze ANL FULL - fulltextové vyhledávání ve vybraných člancích z novin a časopisů zatím s retrospektivou od r. 1997. Databáze je provozována na 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) a pomocí rejstříků

Báze obsahuje následující skupiny témat:

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

Výběr plných textů koresponduje s výběrem článků ke zpracování pro bázi ANL, která je vydávána též na CD-ROM jako řada Články v českých novinách, časopisech a sbornících (viz Další databáze).

Jsou zpřístupňovány faktograficky přínosné články, odborné články, články s dokumentární a uměleckou hodnotou, oficiální dokumenty (projevy, usnesení apod.), informace o kulturních, politických, vědeckých aj. akcích (konferencích, festivalech, jednáních, zasedáních, výstavách apod.), divadelních, rozhlasových a televizních inscenacích, závažné polemiky, diskuze, komentáře. Dále jsou zahrnuty reportáže, biografické články (životopisné, jubilejní, nekrology, rozhovory, profily osobností, vzpomínky), literární texty (básně, povídky apod., recenze publikací různě zaměřených, kompaktních desek, filmů apod.

Seznam titulů, jejichž články v plnotextové formě obsahuje báze ANL FULL - viz [Charakteristika databáze](#).

Databáze je v současné době on-line aktualizována v rámci linky automatické indexace za spolupráce se společností Anopress, s.r.o. Je také průběžně retrospektivně doplňována off-line.

B.1.2.1.3 Další databáze (ANL)

Souborná databáze báze **ANL** obsahuje bibliografické záznamy vybraných článků novin, časopisů a sborníků zpracovávaných v rámci Kooperativního systému článkové bibliografie (KOSABI). Databáze vychází jako řada České národní bibliografie. Vyhledávání je možné z údajů bibliografického popisu. Některé záznamy báze jsou propojeny na plné texty umístěné v bázi ANL FULL. Báze ANL obsahuje navíc linky na plné texty některých periodik a plná znění některých zákonů volně dostupných na na WWW (odborná knihovnická periodika - **Ikaros, U nás, Bulletin SKIP, Daidalos**), další odborová periodika (**Lesnická práce, Psychiatri e, Psychologie dnes, Vesmír, Harmonie, Collection of Czechoslovak Chemical Communications, Veřejná správa**). K článkům pojednávajícím o zákonech jsou připojována **plná znění zákonů**.

Výběr článků ke zpracování (viz výše).

Obsah:

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

- v [NK ČR](#)

- ve [spolupracujících institucích](#).

B.1.2.1.4 Periodika na WWW - Portál

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, bude postupně ladit.**

Dílčí nabídky:

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

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

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

- **Další zdroje (institucionální členění)**

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

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

B.1.2.1.5 Jak se zaregistrovat (komentář k registraci v databázi ANL FULL)

Bibliografické údaje o člancích, resp. metadata a plné texty jsou k dispozici čtenářům Národní knihovny. Pro externí uživatele jsou běžně k dispozici pouze bibliografické údaje o člancích. Tito uživatelé se musí pro získání plného textu zaregistrovat pomocí formuláře v nabídce Registrace. Po vyplnění povinných údajů bude těmto uživatelům zasláno e-mailem potvrzení registrace s aktivačním odkazem a heslem - po odeslání tohoto odkazu je registrace aktivována a prostřednictvím přiděleného hesla zajištěn bezplatný zkušební přístup na dobu sedmi

dnů. Rutinní zpřístupňování plných textů bude možné po vyřešení autorskoprávních, popř. ekonomických otázek souvisejících se zpřístupňováním plných textů.

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

B.1.2.1.6 Jak vyhledávat

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 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 možností 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 - definujeme jednotlivá témata.

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

Topiky jsou rozděleny v bázi do tří úrovní. Lze zatím vyhledávat podle některých nadefinovaných topiků (označeny šipkou vpravo).

Tematický dotaz je vytvořen v podobě nadefinovaného pojmového stromu v bázi. Skládá se z pojmů - slov, charakteristických pro danou oblast. Ty pak tvoří listy stromu. Témata lze samostatně zároveň používat v prostých i formulářových dotazech. Napojení listů/témat na další úroveň je zajištěno operátorem a přidělením váhy tomuto listu/tématu. Váhy ovlivňují skóre dotazu. TOPIC postupuje od listů až k vrcholu stromu a postupně vypočítává skóre. Topiky viz dále.

Pole

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

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

Vazby mezi polem dotaz, dalšími poli a tématy

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

Způsoby zobrazení seznamu výsledků (názvů vyhledaných článků) - viz dále.

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

Seznam výsledků (názvů vyhledaných článků):

- **jednoduchý** (standardní zobrazení)
- **se souhrnem** (standardní zobrazení plus souhrn)
- **seskupený** (standardní zobrazení plus články uspořádané do skupin podle shody v klíčových slovech **podle společného kontextu, clustering**)

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

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

Zobrazení údajů o článku viz dále.

- uživatelské formáty
- pracovní formáty

B.1.2.1.6.1 Vyhledávání

Vyhledání probíhá fulltextovou formou, tj. z plného znění jednotlivých dokumentů a z jejich citace, resp. metadat, resp. bibliografického popisu. Systém rozeznává pádové koncovky podstatných a přídavných jmen. Lze vyhledávat podle jednoduchého slova nebo fráze. U rozšířeného formuláře a u rozšířeného formuláře s tématy lze nastavit (ve spodní části formuláře) pro implicitní pole formuláře vyhledávání podle výskytu řetězce kdekoli v poli pomocí operátoru <contains>. Pokud je vyžadováno přesné zadání i porovnání celého řetězce použijeme =. Tyto operátory lze zapsat také přímo do pole dotaz, resp. hledat v textu. Operátor <contains> lze v tomto případě zapsat jednoduše jako #. V tomto případě nastavení operátorů nemá vliv při hledání z polí pomocí návěští a při použití různých konvencí (saved, *, atd.). Doporučujeme však v případě nejistoty nastavení <contains>.

B.1.2.1.6.2 Formuláře

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

Základní formulář

Nabídky:

Dotaz - hledat v textu (formulace dotazu)

Období pro vyhledání od-do

Typ seznamu výsledků (článků):

jednoduchý (pouze název s nabídkou zobrazovacích formátů)

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

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

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

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

a to sestupně nebo vzestupně.

Rozšířený formulář s tématy**Nabídky:**

Tento formulář má **stejné nabídky pro vyhledávání jako rozšířený formulář, umožňuje vyhledávat navíc podle témat, resp. topiků, resp. dotazů. Pro bázi ANL FULL jsou nadefinovány některé topiky, rozdělené do tří úrovní.** Topik slouží k tomu, že hledaný výraz vyplněný do textu dotazu a různě zpřesňovaný je navíc upřesněn topikem, tj. tématem. Tři úrovně topiků viz dále 7.3.3.
Např. hledáme-li v bázi výraz hvězdy, výsledek obsahuje několik tisíc dokumentů. Specifikujeme-li tento výraz pro vyhledávání v rámci astronomie, najdeme pouze několik set dokumentů.

B.1.2.1.6.3 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 téma t - 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)

B.1.2.1.6.3.1 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 osbaženy všechny hledané výrazy

or - pole obsahuje alespoň jeden hledaný výraz

not- hledaný výraz nesmí být v poli přítomen

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

Př. 1

hvězdy

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

Př. 2

hvězdy, asteroidy, planetky

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

Př.: 3

(hvězdy, asteriody) and komety

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

Př.: 4

(hvězdy and asteriody) not komety

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

Př.: 5

komety <near> kolize

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

Př.: 6

"meteorický roj" or "padající hvězdy" .

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

Př.: 7

meteo*

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

Př.: 8

*stvo

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

Př.: 9

??běr

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

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

Př.: 10

Data

dat <contains> 8.10.2001

dat# 8.10.2001

dat=8.10.2001

Vyhledá všechny dokumenty vydané v tomto dni.

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

dac=29.10.2001

dac <contains> 29.10.2001

dat#29.10.2001

Vyhledá všechny dokumenty zpracované v tomto dni.

Př.:11

zdr=Respekt and naz=Rafinovaný odraz skutečnosti

src=Respekt and ti=Rafinovaný odraz skutečnosti

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

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

Př. 12

zdr <contains>Respekt and naz <contains>odraz

src<contains>Respekt and ti<contains>odraz

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

Operátor <contains> lze nahradit #:

zdr#Respekt and naz#odraz

src#Respekt and ti#odraz

dc.source#Respekt and dc.title#odraz

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

[Tabulka návěští a prvky Dublin Core a Anl Core aplikované v bázi ANL FULL viz Charakteristika databáze.](#)

Je-li případě tečkové konvence s dc (Dublin Core) použita ještě tečková konvence s anl, používá se konvence s anl (Anl Core) - Dublin Core není takto jemně definován nebo TOPIC v současné verzi systému toto vyhledávání nepodporuje. Rovněž nepoužívejte návěští phnk, ale jeho alternativu.

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

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.

B.1.2.1.6.3.2 Formulářová pole s implicitními údaji v rozšířených formulářích

Tabulka formulářových polí a metodika vyhledávání v těchto polích viz [Charakteristika databáze - Jak vyhledávat - Formuláře.](#)

Jednotlivé údaje ze stejných polí nebo různých polí z rolovacího menu 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.

B.1.2.1.6.3.3 Vyhledávání pomocí topiků - Rozšířený formulář s tématy viz [Charakteristika databáze - Jak vyhledávat - Formuláře.](#)

Tématem se rozumí předem připravený složitý dotaz. Jedno nebo více témat je možno vybrat označením v seznamu v pravém pruhu od 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, druhá skupiny témat v těchto oblastech, třetí obsahuje detailní témata ve skupinách témat. Z první úrovně nelze zatím vyhledávat, z druhé a třetí ano - tam, kde je topik nadefinován (označeno šipkou vpravo). Tematická nabídka bude v budoucnu postupně rozšiřována a témata upřesňována.**

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 [Charakteristika databáze - Jak vyhledávat - Formuláře.](#)

B.1.2.1.6.3.4 Rejstříky viz [Charakteristika databáze - Jak vyhledávat - Rejstříky.](#)

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.

B.1.2.1.6.3.5 Některá doporučení, jak nejlépe vyhledávat

- Údaje obsažené v polích bibliografického popisu doporučujeme vyhledávat spíše pomocí implicitních polí v rozšířených formulářích než vypisováním návěstí 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.1.2.1.7 Výsledky vyhledávání, zobrazení, tisk (všechny formuláře) viz [Charakteristika databáze](#).

B.1.2.1.7.1 Seznam výsledků (seznam článků)

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

Druhy seznamu výsledků (seznam článků):

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

Seznam výsledků (názvů vyhledaných článků):

- **jednoduchý** (standardní zobrazení)
- **se souhrnem** (standardní zobrazení plus souhrn)
- **seskupený** (standardní zobrazení plus články uspořádané do skupin podle shody v klíčových slovech **podle společného kontextu, clustering**).

Třídění seznamu výsledků (názvů vyhledaných článků)

Skóre relevance, od které zobrazovat názvy článků)

Výsl./str. (počet článků na stranu)

Třídění: dle:

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

to sestupně nebo vzestupně.

B.1.2.1.7.2 Zobrazení údajů o článku

B.1.2.1.7.2.1 Formáty zobrazení

Uživatelské formáty

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



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



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



DC/HTML - Dublin Core Metadata

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

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

Pracovní formáty

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

B.1.2.1.7.2.2 Metadata

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

Kompletní seznam metadat v uživatelském formátu, resp. citace viz Příloha 4:

Př.1 :

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

Př. 2:

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

B.1.2.1.7.3 Tisk

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

Předpokládá se, že v budoucnu bude možné texty dotazu zaslat e-mailem, stahovat větší počet článků, které uživatel zaškrtně ve formátu html, textovém ASCII apod. a ukládat soubory na disketu. Bude řešen export jak citací, resp. bibliografických údajů, resp. metadat, tak plných textů.

Předkládá metodika vyhledávání je první verzí nápovědy pro vyhledávání v databázi ANL FULL. Bude se dále optimalizovat.

B.1.3 Standardizace - UNIMARC, DUBLIN CORE, HTML, XHTML, XML, propojení záznamů s plnými texty

Struktura bibliografických dat respektuje formát **UNIMARC** a knihovnická pravidla **AACR2** v oblasti jmenného popisu. V oblasti věcného popisu se používá aktualizovaná verze **MDT-MRF**. Verbální věcný popis obsahuje **předmětové kategorie**, které zasazují dokument do širších souvislostí v rámci databáze z hlediska obecných témat, jež by se měla sblížovat s tématy systému TOPIC. Předmětové kategorie do jisté míry konvenují metodě konspektu aplikované pro popis a mapování fondů. Dále se používají **klíčová slova**, která jsou dále částečně řízená a **předmětová hesla**. Automaticky je generován **souhrn článku**, automaticky lze také generovat klíčová slova, která však mají v současném stádiu "počítačovou formu". V budoucnu lze předpokládat i automatické generování termínů předmětové indexace. Tyto forma však vyžaduje intenzivní vývoj řízených slovníků/heslářů či tezaurů a jejich integraci do "subject gateways".

V budoucnu se předpokládá intenzivnější využívání vznikajících souborů autorit jmenných i věcných .

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

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

Formáty zobrazení jsou popsány dříve.

B.1.4 Skutečný stav propojení bibliografických záznamů s plnými texty

Bylo zakoupeno cca 24426 plných textů článků publikovaných v zejména roce 1997, 1998 (rok 1999, 2000 doplněn převážně v minulých letech). V prvním pololetí r. 2001 byly připraveny k propojení a následně propojeny plné texty získané v rámci Konzorcia Anopress. **Od května jsou propojovány plné texty se záznamy v rámci linky automatické indexace.** Zhruba do konce října bylo takto zpracováno a propojeno s plnými texty **8869 záznamů** (včetně práce v lince v rámci souvisejícího grantu Propojení analytických záznamů s plnými texty). V r. 2001 bylo **realizováno připojení off-line záznamů k plným textům, které byly k tomuto propojení připraveny v r. 1999 a 2000** (i v rámci projektu Propojení analytických záznamů s plnými texty). Programově bylo propojeno takto připravených textů **cca 51 000**.

Výše uvedenými metodami byly propojovány zejména celostátní deníky, dále periodika **Týden, Respekt, Reflex, Ekonom, Profit, Bankovníctví**.

V roce dále pokračovalo **ruční propojování záznamů zpracovávaných v ALEPHu s volně přístupnými texty na WWW**. Do konce října bylo propojeno **cca 1047** záznamů (periodika: Veřejná správa, Lesnická práce, Obecná psychologie, Psychiatrie, Lesnictví, Harmonie, Vesmír, Učiteléské noviny).

Souhrn za r. 200-2001: na serveru FULL.NKP.CZ jsou přístupné plné texty článků vydané v letech 1997-2001 a získané v rámci tohoto grantu a grantu Propojení analytických záznamů v letech 1999-2001, říjen v počtu cca 82 160. Zhruba stejný počet je propojen se záznamy v ANL

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

Celkem bylo za r. 2000-2001, říjen propojeno **přes 2000 záznamů s volně dostupnými texty** na Internetu.

V oddělení analytického zpracování bylo zpracováno v **rámci linky automatické indexace v roce 2001, květen-říjen 8869** záznamů umístěných zároveň v bázi ANL.

B.1.5 Souborná databáze ANL Kooperčního systému článkové bibliografie (KOSABI)

V roce 2001 pokračovalo oddělení ve zpracování záznamů pro bázi ANL(ALEPH). Kooperující instituce **# SVK, resp. krajské knihovny a jiné instituce** - přispívají pravidelně do souborné databáze. Záznamy nejsou zatím dodávány SVK České Budějovice vzhledem k implementaci nové verze Tinlibu (též konverzi T-Series do UNIMARCu). Ostrava přechází také na systém T-Series, do srpna byly záznamy dodávány v ISISU, v současné době se ladí dodávání ze systému T-Series. MZK v Brně odevzdává záznamy v delších intervalech, letos přechází na ALEPH 500. Na ALEPH 500 přešla Olomouc. Kladno zpracovává záznamy v systému Rapid, ostatní spolupracující instituce setrvávají v systému ISIS. Kromě záznamů zasílaných do souborné databáze ANL spolupracující knihovny budují své lokální databáze. Byla vyzkoušena kooperace s **Českým geologickým ústavem** a zkonvertováno cca 2000 záznamů článků publikovaných v r. 1998-1999 do báze ANL.

Záznamy respektují metodický materiál **Záznam pro soubornou databázi : UNIMARC a Záznam pro soubornou databázi : Výměnný formát**. V roce 2001 byla **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 jednotlivým prvkem souborné databáze. Oddělení analytického zpracování přistoupilo k **intenzivnějšímu využívání souborů autorit v NK a zároveň v říjnu přistoupilo k přípravě dat pro soubory jmenných a věcných autorit**.

Hlavní řešitelka projektu se zúčastnila **zasedání SDRUK, sekce pro bibliografii**, kde byla diskutována problematika funkce KOSABI v době přechodu některých knihoven na krajskou úroveň a

vzájemných vazeb v KOSABI. Bylo konstatováno, že krajské knihovny v nových krajích nejsou na spolupráci připraveny, resp. personálně vybaveny.

Pro rok 2001 bude zachována kompetence stávajících SVK, resp. krajských knihoven v rozsahu regionů před územní reformou.

V květnu proběhlo setkání s **pracovníky SKAT** a diskutována možná spolupráce v budoucnu z hlediska přebírání záznamů a excerptce titulů v systému Lanios, resp. Clavius.

V červnu 2001 se konal **seminář Pracovní skupiny pro analytické zpracování**, kde byla mimo jiné diskutována problematika přechodu od klíčových slov k předmětovým heslům a vazby mezi předmětovými termíny.

V r. 2001 se hlavní řešitelka zúčastnila semináře **ELAG** s pracovním názvem **Integrating of Heterogeneous Resources**. Samotný název toho prestižního semináře napovídá o důležitosti otázek souvisejících s řešením otázek týkajících se integrace heterogenních dat do kooperujících systémů za účelem rychlejšího a kvalitnějšího zpřístupnění informací, které spolu souvisejí. Autorka zprávy se zúčastnila **workshopu Automatic vs. manual indexing**. Na tomto workshopu byly diskutovány výhody a nevýhody ruční a automatické indexace a nutnost koexistence obou metod.

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

V roce 2001 byla pravidelně poskytována data KOSABI pro měsíční aktualizace a **CD-ROM ČNB - řada Články v českých novinách, časopisech a sbornících**.

B.1.7 Management Kooperačního systému článkové bibliografie (aplikace MNG KOSABI) (viz Příloha F4)

Na základě zadání pro řízení a správu kooperačního systému a za využití již vyvinutých řešení v rámci Souborného katalogu CASLIN probíhají práce na vývoji aplikace pro KOSABI na serveru ANL(systém **LINUX a ORACLE**) **Ve stádiu experimentu je aplikace pro příjem a automatizované zpracování dat (přijímání analytických záznamů, integrace stávajících programů pro konverzi analytických záznamů, globální úpravy analytických záznamů, vývoj programů na formálně logické kontroly kooperujících knihoven - test na UNIMARC pro analytické záznamy)**. V r. 2001 byla vyvinuta aplikace pro evidenci zpracovávaných titulů a test na duplicitu záznamů (klíčů).

B.1.7.1 Aplikace pro příjem a zpracování dat, export do ALEPhu

Aplikace pro správu a údržbu KOSABI - přehled dílčích funkcí

[Obr. L](#), [Obr. M](#), [Obr. N](#), [Obr. O](#)

1. Příjem dat pro systém

- způsoby příjmu dat: ftp, upload, disketa
- formát ty: Aleph, Aleph500, UNIMARC řádkový, UNIMARC ISO 2709, Výměnný formát
- kódování: GIZMO, PC Latin 2, Kamenicky, ISO 8859-2, UTF-8, Ansel, ISO 646

Experimentálně vyzkoušen upload ve všech zmíněných formátech a kódování, nutné průběžné ladění.

2. Zpracování přijímaných dat

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

Všechny funkce experimentálně odzkoušeny, nutné průběžné ladění .

3. Export dat z baze SKA 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 Ansel a formátu Aleph500
- uložení výstupního souboru do příslušného adresáře pro import do Aleph.
- statistika zpracování

Export do ALEPHu bude řešen

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.1.7.2 Aplikace pro evidenci zpracovávaných titulů (nedefinována pole odpovídají SK CASLIN)

V r. 2001 byl nedefinován formát pro bázi titulů. K tomuto účelu byl modifikován formát SHORT SK CASLIN.

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

Sigla vlastníka: [910a].[910a]

Region: [952a].[952a]

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

K této bázi budou nedefinovány přístupové soubory pro možnost hledání .

B.1.7.3 Test na duplicitu klíčů

V r. 2001 byly stanoveny klíče, podle nichž je prováděna kontrola na duplicitu přijímaných dat.

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

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

B.1.8 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), do konce roku 2001 má být proveden upgrade ORACLu.

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 jednodprocesorový 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 roce 2001 bylo zakoupeno PC DURON .V příštím roce je třeba nainstalovat WINDOWS NT 2000.**

B.1.9 Smluvní ošetření projektu

Pro automatickou indexaci, správu (údržbu) plnotextové databáze a dodávku plných textů byly uzavřeny dvě smlouvy: s ing. I. Maternem a Anopressem, s.r.o dne 15.6.2000. Smlouva na vývoj aplikace pro management kooperačního systému s ing. Koktanem byla uzavřena dne 26.9.2000.Smlouva s Anopressem **zůstává v platnosti pro rok 2001. Pro další roky zůstává v**

platnosti smlouva s hlavním programátorem projektu - Ing. Maternem.

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.

V současné době se nachází Anopress, s.r.o. v konkurzu. Jednou z podmínek stanovených pro konkurzní řízení je respektování smluvních závazků Anopressu. Jak bude projekt pokračovat

závisí na budoucím majiteli firmy, jeho schopnostech, možnostech a strategii. Jde zejména o získávání plných textů pro linku automatické Indexace.

B.2 Přínos řešitele

Přínos projektu v r. 2001 spočívá v realizaci metod navržených v r. 2000 ve stadiu úspěšného experimentu, který měl až poloprovozní charakter v případě budování báze ANL FULL.

Aplikace pro správu a údržbu KOSABI je ve stadiu experimentu a dalšího ladění.

Přínos projektu v r. 2001 spočívá v praktické integraci elektronických zdrojů do služeb knihovny pomocí technologicky nejvypělejších nástrojů pro zpřístupnění těchto zdrojů při zachování tradičně zpřístupňovaných sekundární informací formou bibliografických záznamů. Jde o integraci heterogenních dat do Kooperačního systému článkové bibliografie, v němž dochází k propojení tradičních knihovnických postupů a fondů s určitými prvky digitální knihovny.

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

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

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

Další přínos projektu spočívá v rychlém **zpřístupnění analytických záznamů provázaných s plnými texty zdrojových dokumentů v rámci Kooperačního systému článkové bibliografie v postupném propojování záznamů báze ANL se záznamy o zdrojových dokumentech v bázi NKC. Báze ANL je jedním ze zdrojů Metalib.**

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

Pro toto zpřístupnění plných textů byl navržen **moderní layout stránek serveru FULL.NKP.CZ. Velká pozornost byla věnována definici báze ANL.FULL, zejména pak systému vyhledávání.**

Na server byl **zabudován portál** pro zpřístupnění volně dostupných elektronických periodik na Internetu, která jsou členěna na principu krajů, tematiky a oblastí, institucionálním.

V roce 2001 byla **posílena kooperace v rámci KOSABI** po stránce metodické i organizační. Do kooperace byl přibrán Český geologický ústav, byla realizována jednání se systémem LANIUS ohledně budoucí možné spolupráce. Dále byla posouzena konverze T-Series UNIMARC vzhledem k systému zavedenému v bázi ANL a navrženy některé korekce.

Možnost vzniku **duplicit či multiplacit** při zpracování je minimalizováno delimitací periodik mezi spolupracujícími institucemi. Tím má projekt další i ekonomický význam.

Další přínos spočívá ve zvýšení **uživatelského komfortu** - v nalezení příslušného článku z novin či časopisu (v budoucnu snad i statě se sborníku) v elektronické formě Navigace k primárním dokumentům patří k základním trendům v oblasti knihovnictví a informatiky. Byla vyvinuta aplikace k zajištění registrovaného přístupu externích uživatelů k plným textům. Tito uživatelé se mohou zkušebně zaregistrovat na dobu 7 dnů.

V případě vhodných podmínek přichází v úvahu možnost aplikace vypracovaných metod na některé spolupracující subjekty v Kooperačním systému článkové bibliografie v budoucnu a po jisté úpravě aplikace linky automatické indexace v komunikaci mezi autorem, nakladatelem, informační institucí, bibliografickou agenturou.

Další přínos spočívá v **nákupu plných textů v rámci Konzorcia Anopresss**, což má nemalý ekonomický efekt.

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

Řešení experimentálně realizovaná praxi v roce 2001 jsou příspěvkem do programu univerzální bibliografické kontroly (UBC) a všeobecné dostupnosti publikací (UAP) IFLA a jsou také součástí

praktické realizace čl. 61 kulturní politiky o kooperaci knihoven a především čl. 63 o podpoře automatizace knihovnických systémů a propojení do globálních informačních struktur. Přispívají tak k optimalizaci veřejných informačních služeb (VIS).

B.3 Posun znalostí

K významnému posunu znalostí došlo především v těchto oblastech:

- Poměrně dobrá **orientace v nových trendech zpracování** a zpřístupňování informací.
- Praktické **realizace těchto trendů na konkrétní fungující systém**, která umožní jeho přetrvání v budoucnu (metadata, pojmové vyhledávání, propojování informací, vazba na klasický fond, fond elektronických dokumentů a registrace v ČNB, budování digitální knihovny).
- **Fungující linka automatické indexace.**
- **Propojení bibliografických záznamů s elektronickými zdroji na základě dynamických i statických URL adres, propojení analytických záznamů na zdrojový dokument a naopak, kombinace pojmového vyhledávání s metadaty jako předpoklad pro implementaci kooperačního systému do konceptu metaknihovny, informačních bran a "subject gateways".**
- **Vytvoření předpokladů pro distribuované vyhledávání** informací za současného využití tradičního slučování informací do souborné databáze založené na architektuře statických bází.
- **Zpřístupnění plných textů v rámci moderně koncipované báze ANL FULL** na základě vyspělých technologií pro zpřístupňování elektronických dokumentů, **praktická implementace Dublin Core v rámci HTML, XHTML, XML.**
- Praktická **realizace pojmového vyhledávání pomocí topiků** a budování tzv. znalostní 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 údajů analytického popisu.**
- **Byl položen základ pro přehledně strukturovaný portál volně dostupných zdrojů na Internetu s popisem těchto zdrojů.**
- **Možnost využití metody vypracované pro linku automatické indexace pro zpracování informací a toku informací mezi autorem, nakladatelem, informační agenturou, bibliografickou agenturou/knihovnou.**
- **Posílení vazby v rámci KOSABI** v situaci reformy státní správy a samosprávy z hlediska metodického i organizačního - navázání spolupráce s nově vznikajícími krajskými knihovnami.
- **Nová forma poskytování informačních služeb za spolupráce knihovnické a moderní informační instituce. Bez kvalitního know how těchto společností není možný vývoj v oblasti knihoven.**

C Návrhová část

C.1 Výsledky řešení

- **Nová struktura serveru FULL.NKP.CZ, layout www stránek.**
- **Praktická realizace nástrojů pro optimalizaci integrace a správy heterogenních dat v rámci Kooperačního systému článkové bibliografie. Vznikl tak nástroj pro automatickou přípravu dat pro bibliografickou databázi založenou na UNIMARCu a plnotextovou databázi založenou na pojmovém vyhledávání systému TOPIC a metadatech zpřístupňovaných v Dublin Core a jeho aplikacích v HTML, XHTML, XML v kvalifikované a nekvalifikované formě .**
- **Realizace aplikace pro získávání dat TAMTAM SPECIAL NK (TTSNK).**
- **Realizace aplikace pro automatické generování bibliografických údajů do analytických záznamů a metadat DUBLIN CORE a jejich umístění do plných textů i mimo ně - TAMTAM DATA EXTRACTOR (TTDE).**
- **Aplikace pro indexaci dat do plnotextové databáze ANL FULL.**
- **Aplikace pro vyhledávání a zpřístupnění plných textů formou uživatelských formátů - citací s hypertextovými odkazy i bez nich, citací se souhrnem, v DC, ale i pracovních formátů ve formě HTML, XHTML, XML. Návrh a realizace Anl Core pro české analytické záznamy.**
- **Definice topiků.**
- **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, vazba up and down analytický záznam a zdrojový dokument.**
- **Průběžné doplňování báze ANL bibliografickými záznamy a báze ANL FULL plnými texty (on-line), propojení dříve získaných plných textů se záznamy (off-line).**
- **Zakoupení plných textů pro plnotextovou databázi, průběžné doplňování databáze ANL FULL, propojování bibliografických záznamů s plnými texty na základě dynamických a statických URL adres. Průběžné propojování záznamů s plnými texty na základě statických URL adres.**
- **Převedení periodika Národní knihovna z báze Anopress do báze ANL FULL v HTML tvaru.**
- **Návrh strukturovaného portálu pro volně přístupné zdroje na WWW po regionální, tematické a institucionální úrovni.**
- **Registrace a autentifikace uživatelů.**
- **Aplikace pro management Kooperačního systému článkové bibliografie (MNG KOSABI).**
- **Báze zpracovávaných titulů na serveru ANL.**
- **Aplikace pro kontrolu duplicitních záznamů.**
- **Standardizace týkající se popisu článků v UNIMARCu a AACR2R.**
- **Zabezpečení HW a SW podpory kooperačního systému.**
- **Zabezpečení KOSABI po stránce metodické i organizační.**

C.2 Závěr

Výsledkem řešení projektu v r. 2001 nové pojetí serveru FULL.NKP.CZ a způsobu vyhledávání v plnotextové databázi ANL FULL (vyhledávání pomocí metadat, rejstříků, topiků). Báze vzniká v rámci linky automatické indexace plných textů za současného generování formátů pro zpřístupnění elektronických dokumentů a bibliografických záznamů pro bázi ANL. Aplikace umožňuje přípravu importního souboru bibliografických záznamů (s automaticky generovanou URL adresou) pro ALEPH a jiné systémy, založené na UNIMARCu a doplnění plných textů o metadata Dublin Core v HTML,

XHTML, XML.

Formuláře pro vyhledávání a následné zobrazení výsledků vyhledávání umožňují uživateli získat relevantní informace na základě kombinace vyhledávání pomocí metadat a pojmového vyhledávání za současně hodnocení relevance a třídění vyhledaných dat. Vyhledávat lze i podle rejstříků.

Propojení záznamů s plnými texty je realizováno on-line i off-line, dynamickými i statickými URL.

Pro uživatele bylo nadefinováno několik formátů.

Byl vytvořen portál pro zpřístupnění volně dostupných textů periodik vystavených na Internetu.

Takto připravená metadata mohou sloužit k tvorbě tzv. open URL adres v rámci metaknihovny.

Praktická implementace předmětových kategorií umožňuje zabudovat analytické záznamy do konceptu metody konspektu a informačních bran.

Experimentální realizace programové aplikace na provoz, správu a údržbu databáze článkové bibliografie, tj. aplikace pro management KOSABI usnadní správu systému v budoucnu .

Po úspěšném experimentálním odzkoušení systému automatické indexace v r. 2001 je možno zahájit poloprovoz systému v r. 2002.

C.3 Návrhy opatření

- **Zajištění financování projektu v roce 2002 a dále zajištění souvisejícího projektu Propojení analytických záznamů s plnými texty - optimalizace zpřístupnění plných textů, který tento projekt saturuje koncepčně a doplňuje finančně. Koordinace s ostatními stávajícími i budoucími projekty v NK i jinde.**
- **Posílení vazeb stávajícího KOSABI na existující kooperační systémy na nižší úrovni z hlediska správního, zejména systém LANIUS.**
- Personální zajištění projektu z hlediska počtu pracovníků oddělení analytického zpracování v NK.
- **Existence souborů autorit v oblasti jmenného a věcného popisu** jsou nutným předpokladem kvalitního zpracování a vyhledávání informací. Přesunutí kapacity do příslušných oddělení tuto situaci pomohou řešit.
- **Přísnější výběr článků k indexaci deníků s cílem vyloučení subjektivního faktoru při excerpci deníků.**
- **Funkčnost propojovacích vazeb v systému ALEPH** a možnosti expanze a spolehlivosti systému v tomto ohledu. Moderní informační systémy jsou založeny na propojování sekundárních informací s primárními jak klasickými tak elektronickými, ale také na vzájemném propojování sekundárních informací o různých typech dokumentů. Nejde pouze o propojení záznamů s plnými texty, ale také o provázání seriálů a jednotlivých čísel na analytický rozpis článků obsažených v seriálu v rámci báze NKC či Souborného katalogu CASLIN, dále pak připojení článků - recenzí k recenzovaným dokumentům v rámci těchto bází. Navigační systémy na úplné obsahy čísel seriálů mimo záběr abstraktových databází jsou v zahraničí zcela běžné.
- **Ošetření autorskoprávních aspektů, event. ekonomických aspektů zpřístupňování plných textů externím a vzdáleným uživatelům**
- **Pokračování Konzorcia Anopress?**
- **Kladný postoj, schopnosti, možnosti a vhodná strategie nového majitele společnosti Anopress ke spolupráci s knihovnou (knihovnami).**
- **Posílení serveru Anopress a posílení jeho internetové linky**

D Použití finančních prostředků

D.1 Komentář

Komentář ke konečnému použití finančních prostředků je upraven oproti původní zprávě, protože ta se podávala k datu 15.11.2001. Čerpání finančních prostředků je v konečném vyúčtování projektu zohledněno v tabulkách a doloženo fakturami.

Výše státní dotace: 1 056 000.-

Neinvestiční prostředky: 1 015 000.-

Služby 924 000.-

Mzdové prostředky 91 000.-

Investiční prostředky: 41 000.-

Finanční prostředky získané z jiných zdrojů (vklad do projektu):

Národní knihovna ČR - odd. analytického zpracování **200 000.-**

Anopress, s.r.o. **120 000.-**

Použití finančních prostředků je zohledněno v následujících tabulkách. První z nich zachycuje využití neinvestičních prostředků. V rámci neinvestičních prostředků jsou odděleny placené služby, materiál, mzdy. Druhá zachycuje použití investic.

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

Plánované čerpání finančních neinvestic do konce roku 2001 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 vkládá zatím do projektu cca 30 000 za přístup NK do databanky TAMTAM za měsíc listopad a prosinec 2001.

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 a pracovníků oddělení poměrně velký.

E Resumé a klíčová slova

E.1 Resumé a klíčová slova v češtině

Resumé:

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

Výsledkem řešení projektu v r. 2001 je experimentální aplikace pro získávání a automatickou indexaci bibliografických záznamů z plných textů a následné vytvoření importního souboru záznamů pro databázi bibliografických záznamů ANL a databázi plných textů ANL FULL v NK s implementací metadat DC v plných textech včetně automaticky generovaných formátů HTML, XHTML, XML v kvalifikované i nekvalifikované formě.

V r. 2001 byla navržena nová struktura a nový design serveru FULL.NKP.CZ. Velká pozornost byly věnována definici báze ANL FULL a definici vyhledávání v bázi (prostý dotaz, formulářový dotaz, topiky, rejstříky). Indexace plnotextové databáze v systému TOPIC umožňuje pojmové vyhledávání informací. Byl instalován formulář pro registraci uživatelů. Plné texty článků byly v rámci Konzorcia Anopress průběžně stahovány a připraveny k dynamickému propojení s bibliografickými záznamy v 1. pololetí tohoto roku. Od května jsou zpracovávány záznamy z deníků a některých časopisů v rámci linky automatické indexace. Plnotextová databáze byla dále průběžně doplňována články z deníků časopisů vydanými v r. 1997 a 1998. Průběžně byly staticky propojovány záznamy a plné texty z volně dostupných stabilních titulů na Internetu. Byl založen portál pro zpřístupnění volně dostupných textů na Internetu ve struktuře tematické a regionální s popisem zdrojů.

V roce 2001 byla průběžně aktualizována souborná databáze kooperačního systému ANL a vydáván CD-ROM s Českou národní bibliografií.

Další výsledkem řešení v r. 2000 je experimentální aplikace pro správu a údržbu KOSABI.

Klíčová slova:

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

15. listopadu 2001

PhDr. Vojtěch Balík, ředitel NK

PhDr. Ivana Anděrová, hlavní řešitelka