

Hagyományos kiadványok átültetése a Webre: Hungarika-WWW

Nyáryné Grófcsik Erika - Kolonits Zoltán - Pap Gáspár - Kovács Ilona - Gál Júlia

Előzmények

1997 nyarának végén az Országos Széchényi Könyvtár Hungarika Dokumentációs Osztályának munkatársai megkerestek minket (SOFTING Kft.), hogy segítségükre tudnánk-e lenni a Hungarika-WWW, azaz egy olyan új Web-szolgáltatás létrehozásában, amely a hungarika-anyagokat (vagyis magyar vagy nem magyar nyelvű, magyar vonatkozású könyvtári dokumentumokat, valamint magyar szerzők munkáit) őrző külföldi könyvtárakról ad tájékoztatást. A hungarika-gyűjteményekről bőseges információ állt rendelkezésre, hiszen a Hungarika Dokumentációs Osztály munkatársai elkötelezett munkával évtizedek óta szisztematikusan gyűjtötték, rendezték az adatokat. Kutatómunkájuk eredményét a nyolcvanas évek végétől megjelenő „*Hungarika-anyagot őrző külföldi*

könyvtárak címjegyzéke” című sorozatban tették közzé. A sorozat egyes kötetei egy-egy ország hungarika-anyagot gyűjtő könyvtárait ismertetik. A kiadványok szerkesztése, ahogy az adatok tárolása is, eddig manuálisan történt. A Weben való megjelenés lehetősége egy UNESCO/IFLA projektnek köszönhető: a Hungarika Dokumentációs Osztály megbízást és szerény összegű támogatást kapott a világhálózaton hozzáférhető, egy adott kultúra patriotika-kutatását segítő címtár modelljének elkészítéséhez.

Követelmények, megoldási lehetőségek

Közhelyszámba megy, de változatlanul érvényes, hogy egy gépesítési feladat megoldását mindig a

követelmények elemzésével és rangsorolásával, a körülmények és a lehetőségek mérlegelésével kell kezdeni, és a megoldást is csak ezek szemszögéből szabad értékelni. A Hungarika-WWW tervezésekor megfontolásaink az alábbiak voltak:

- ☉ A rendszert úgy kell elkészíteni, hogy az a patriotika-gyűjtésre vonatkozó tájékoztatásnak a világhálón bárki számára hozzáférhető prototípusa lehessen. Ebből következően a funkcionalitás és a felhasználói felület teljes mértékű és míves kidolgozást, színvonalas dokumentálást kíván.
- ☉ A Web-hozzáférést nem egy működő számítógépes rendszerhez illesztve, annak új szolgáltatásaként kell megvalósítani, hanem egy manuális rendszerre építve, az elsődleges adatbevitellel járó problémákat, feladatokat is megoldva.
- ☉ A kidolgozáshoz rövid idő és szűkös erőforrás áll rendelkezésre.

Legfontosabb követelménynek – modell rendszerről lévén szó – az igényes, sok szempont szerint böngészhető felhasználói felület kialakítását tekintettük, amely kényelmes navigálási lehetőséget kínál. Ehhez a kiadványok jól bevált szerkezetét vetjük alapul, amely az egyes könyvtárak és azok hungarika-gyűjteményeinek leírásait földrajzi bontásban, valamint név-, intézménytípus- és tárgymutatók segítségével visszakereshetően közli. Figyelembe vettük, hogy az adattár kicsi: első lépésben néhány száz, teljes feltöltés után legfeljebb egy-kétezer könyvtár leírását tartalmazza majd. Úgy véltük, hogy a Web-lapok dinamikus generálásától eltekinthetünk, és a szolgáltatást standard böngésző programokkal kezelhető, időszakonként előállított és a szerverre felvitt HTML (HyperText Markup Language) formájú dokumentumok (fájlok) hierarchiájával statikusan is biztosíthatjuk.

A tájékoztatást jellegénél fogva két nyelven – a patriotika nyelvén (magyarul) és egy világnyelven (angolul) – kell nyújtani. Ez olyan alapvető követelménynek bizonyult, amelyet modell szinten is következetesen végig kellett vezetnünk. Legegyszerűbb megoldásnak az látszott, hogy a teljes lap-hierarchiát nyelvenként állítjuk elő. (Az adattár mérete ezt megengedte.)

A legnagyobb gondot az adatok bevitelére, egy olyan számítógépen hozzáférhető adattár létrehozására jelentette, amelyből a HTML-formájú fájlok programmal előállíthatók. A teljes körű adatrögzítés – idő és rögzítő kapacitás hiányában – nem jöhetett szóba, de nem állt rendelkezésre olyan adatbáziskezelő rendszer sem, amely – lehetőleg Windows felülettel – mind az adatok kényelmes editálását, mind azok elérését egy jól definiált programozói felületen (API) keresztül biztosította volna, netán a Web-lapok közvetlen előállítását is támogatta volna. Ilyen jellegű szoftver (pl. ORACLE) beszerzése és installálása mind az összeg nagysága, mind az ezzel járó egyéb ráfordítások és hosszútávú kihatások miatt ma már olyan stratégiai fontosságú döntést igényel, amelyet egy kisméretű rendszer nem indokol.

A megoldás módját végül az döntötte el, hogy az OSZK házi nyomdájában fellelhetőek voltak azok a szövegfájlok, amelyek a Word egy DOS alatt működő régi verziójával készültek a kiadványok manuálisan szerkesztett kézírataiból valamilyen DTP rendszerrel történő szedés céljából. Ezek a fájlok ugyan más-más kódban készültek, a borító adataitól kezdve az utószóig minden – részben felesleges – szöveget, értelmezhetetlen szedési parancsot tartalmaztak, de némi programozással WinCE (Windows Central European) kódra, az OSZK-ban általánosan használt Word 6-tal tovább editálható formára voltak konvertálhatók. Volt olyan kiadvány is, amelynek szedésre előkészített fájlja nem került elő. Ilyen esetben az optikai jelolvasással próbálkoztunk több-kevesebb sikerrel: a Recogni-

```

<COUNTRY ID=AT>
  <CITY>
    <NAME>Eisenstadt</NAME>
    <ALTNAM>Kismarton</ALTNAM>
    <LIB ID=2>
    <NAME>Burgenländische landesbibliothek</NAME>
    <ADDR>A-7001 Eisenstadt
    <L>Freiheitsplatz 1. Landhaus</ADDR>
    <TEL>02682/6002358</TEL>
    <FAX>61884</FAX>
    <BOSS>Dr. Johann Seedoch, Direktor</BOSS>
    <BOSS>W. Hofrat Dr. August Ernst, Leiter der Bibliothek</BOSS>
    <SINCE>1922</SINCE>
    <DESCR>A Burgenlandra vonatkozó anyagot gyűjtik kutatás és képzés céljából. Ezen belül hangsúlyt kap a történelem, a kultúrtörténet, az irodalomtörténet, az etnográfia, a politika és a tudomány, valamint a magyar és a horvát irodalom. Továbbá kiemelten gyűjtik a Liszt Ferencre vonatkozó anyagot is. A gyűjtemény nagysága:<VOL ID=36>96 800</VOL>kötet, <VOL ID=49>1300</VOL>kurrens periodika cím, <VOL ID=60>850</VOL>térkép, <VOL ID=44>7883</VOL>mikroforma, <VOL ID=11>217</VOL>disszertáció. Évi gyarapodása: 1500-2000 kötet. A magyar és magyar vonatkozású anyag az állomány része, külön nincs feltárva. Cédulakatalógus, olvasóterem van. Kölcsönöznek. Az 1801 előtti hungarikumok száma: <PVOL ID=1>75</PVOL>. Tájékoztatót nyújtanak: ORR Dr. Eva-Maria Folger. AR Franz Fazokas, ORR Mag. Norbert Frank. </DESCR>
    <REF>Ernst, August: 50 Jahre Burgenländische Landesbibliothek. = Burgenländische Heimatblätter. 1972. 34. 2. pp. 29-66.</REF>
    <KW ID=0107><KW ID=0058><KW ID=0288><KW ID=0319><KW ID=0048>
    <INSTTYPE ID=9>
    </LIB>
  </CITY>
</COUNTRY>

```

tát alkalmaztuk, amely a digitalizált kiadványszövegeket ugyancsak WinCE kódban, Word 6-tal beolvasható formában írta ki. A kiadványban még nem publikált újabb leírásokat és a kiegészítő adatokat közvetlenül WinWord-del rögzítettük.

A szövegfájlok szerkesztése

Miután eldöntöttük, hogy a feldolgozás alapját Word 6-tal editálható egyszerű szövegfájlok sorozata fogja képezni, ki kellett dolgoznunk egy olyan

jelölési módot, amellyel a szövegben előforduló információk egységei (a könyvtárakat leíró tételek), és a tételeket alkotó logikai adatelemek (a könyvtár neve, címe, a hungarika-gyűjtemény jellemzői stb.) könnyen kijelölhetők és programmal felismerhetővé tehetők.

A bevezetett jelölési mód teljes mértékben SGML (Standard Generalized Markup Language) kompatibilis. A szövegen belül egy adatelemet egy címke (angol kifejezéssel „tag”) jelöl, pontosabban címképár fog közre. A (nyitó)címke tulajdonképpen az adatelem relációjelek közé helyezett megnevezése

(<...>). A záró címke mindig a nyitó megismétlése azzal az eltéréssel, hogy a megnevezés előtt egy perjel áll (</...>). Mivel a HTML, vagyis a Web-dokumentumokat definiáló nyelv is az SGML egy implementációja, az SGML szintaxisi alapszabályait ismertnek tételezzük fel.

Az alkalmazható címkék teljes körű felsorolása helyett egy példával illusztráljuk használatukat:

A címkék tetszőlegesen egymásba ágyazhatók, ezek mindegyikének hatóköre a címképár záró eleméig tart. A fenti példa egy könyvtár valamennyi ismértét, így a helyét (város, ország) is tartalmazza. Ha ugyanott egy másik hungarika-anyagot gyűjtő könyvtár is van, a </LIB> zárócímke után következhet a másik könyvtár <LIB ID= > címkével kezdődő leírása, a korábban megadott város, illetve az ország arra a könyvtárra is érvényes lesz. Ha a szövegfájlban olyan szövegrészek is előfordulnak, amelyek nem jelöltek, más szóval egyetlen címképár értelmezési tartományába sem esnek bele, a szöveg feldolgozásánál figyelmen kívül maradnak. Hasonlóképpen hatástalanok a (jelölt) szövegben előforduló soremelések és lapdobások (CR, LF, FF) és egyet meghaladóan a térközök. A jelölt szövegben belüli formázást két speciális címke biztosítja: (új sor) és (paragrafus). A címkék transzparenssek, úgy például a fenti <DESCR> ... </DESCR> címképár közti szöveges leírást a benne előforduló címkék nélkül kell értelmezni. A szövegben elhelyezett <PVOL ID=>, </PVOL> címképár arra szolgál, hogy a közrefogott szám, mint a gyűjtemény egy jellemző méretadata, külön is elérhető, szövegekörnyezetéből kiragadva is felhasználható legyen.

Több olyan címkét vezetünk be, amelynek attribútuma (ID= vagy LANG=) van, és az attribútumnak az egyenlőségjel után megadott értéke a címkével minősített adatelem egyedi azonosítója, illetve azonosító kódja. Az ilyen címkék általában arra használatosak, hogy egy adatelemet ne kelljen minden esetben a tételbe beírni, hanem hivatkozás formá-

jában a kóddal is hozzá lehessen a tételhez rendelni. A kóddal azonosított adatelem önmaga is egy tétel, amelyet azonos szintaxissal, általában külön szövegfájlban írunk le.

A fenti példa egy sor <KW ID=> címkét tartalmaz. Ezek külön szövegfájlban tárolt tárgyszó tételekre hivatkoznak. A <KW ID=0048> hivatkozáshoz például a tárgyszó fájl alábbi tételei tartoznak:

```
<KW ID=0048 LANG=eng >periodicals</KW>
<KW ID=0048 LANG=eng SYN>serials</KW>
<KW ID=0048 LANG=hun >periodikumok</KW>
<KW ID=0048 LANG=hun SYN>folyóiratok</KW>
<KW ID=0048 LANG=hun SYN>időszaki kiadványok</KW>
```

Egy tárgyszó tétel – az azonosítón kívül – egyéb attribútumokat is tartalmaz: magát a tárgyszót, annak nyelvét, továbbá annak megjelölését, hogy az adott kifejezés a fogalom kitüntetett alakja vagy szinonímája.

A nyelv ugyancsak kódolt. Az ISO 639-2/B szabvány hárombetűs nyelvkódjait alkalmaztuk, amelyek kifejtése egy másik azonos szintaxisú szövegfájlba került. Például az „eng” kód az alábbi tételt azonosítja a nyelvfájlban:

```
<LANGUAGE ID=eng LANG=hun>angol</LANGUAGE>
<LANGUAGE ID=eng LANG=eng>English</LANGUAGE>
```

Hasonlóan történik az országok, intézménytípusok és dokumentumtípusok kódolása is. A módszer előnyeit az alábbiakban látjuk:

- ☞ Lehetővé teszi a kódolt információk nyelvfüggetlen rögzítését és nyelvfüggő megjelenítését. (A rendszer kétnyelvűsége, mint arra már utaltunk, alapkövetelmény volt.) A nyelvek száma

elvileg tetszőlegesen növelhető a már rögzített szövegek módosítása nélkül is.

- ☉ A kódok kifejtése egyetlen fájlban van. Ez könnyíti a karbantartást, az egységes használatot, ami az olyan indexelésre, illetve keresésre használt kifejezéseknél, mint a tárgyszavak vagy az intézménytípusok, rendkívül fontos.
- ☉ Csökkenti a rögzítendő adat mennyiségét.

Az adatszerkesztés megkönnyítésére Word-makrókat írtunk, amelyek működése a következő: kursorral kijelölhető a szövegrész, amelynek típusát (név, névvariáns, telefonszám stb.) egy menüből kell kiválasztani, ezután a megfelelő címkéket a szöveg elé és mögé a Word automatikusan befűzi. A címkék elhelyezése a szövegben lehet utószerkesztési folyamat, de történhet a rögzítéssel egy menetben is.

A szövegfájlok feldolgozása

Az adatfeldolgozás – a mai általános séma szerint – két lépésből áll:

- input adatok bevitele az adatbázisba,
- az adatbázisból kiválogatott adatokból különféle outputok előállítás.

Adatbáziskezelő rendszert, mint arra már utaltunk, a Hungarika-WWW kidolgozásakor nem használhattunk. Nem akartunk viszont lemondani a fenti séma lényegéről, nevezetesen arról, hogy az input és az output oldal egymástól viszonylag független, ami esetünkben azt jelenti, hogy a Web-lapok felépítése, megjelenítése változtatható a szövegfájlok módosítása nélkül is, és fordítva. Ezért egy olyan programot fejlesztettünk ki, amely ugyan egy menetben állítja elő a szövegfájlokból a Web-lapokat, koncepciójában azonban megfelel a sémának:

☉ A program input modulja beolvassa az összes szövegfájlt – ezek szám és sorrendje tetszőleges –, elemzi szintaxisukat, feltárja a szövegbe ágyazott logikai adatelemeket, és azokból egy relációs adatmodellt épít fel.

☉ A program output modulja a relációs modellből megszerkeszti a HTML-formájú fájlok kívánt hierarchiáját. A modell adatelemeihez függvényeken keresztül – nevezzük adatbázis filter könyvtárnak – fér hozzá, amelyek a relációs adatbázisoknál szokásos eléréseket és műveleteket biztosítják.

A felépített relációs adatmodell virtuális, hiszen csak a programfutás alatt él. Az adatbázis koncepció következetes végigvitele azonban lehetővé teszi azt, hogy – az input elemző és output szerkesztő programrészek lényegi módosítása nélkül – bármikor a modell mögé helyezhető legyen egy tényleges relációs adatbáziskezelő rendszer, ha ez az adatállomány méretének növekedése vagy bármely más ok miatt kívánatosá válik.

Az adatok tárolása, kiegészítése, módosítása jelenleg szövegfájl formában történik. A korábban rögzített adatok gyakori változása nem jellemző, az adatok karbantartása és új tételek beillesztése azonban szükséges. Ezt a műveletet támogatja a kihelyezett kérdőív. Az új tételek bevitele további szövegfájlok előállításával egyszerűen megoldható, a karbantartás pedig a meglévő szövegfájlok javításával, illetve kiegészítésével történik. A szintaktikai hibák és a hibás referenciák felismerésére és jelzésére ellenőrző programot készítettünk. Ennek alapján a javítások a szövegszerkesztővel elvégezhetők, majd a teljes Web-lap hierarchia újból előállítható. Egyelőre nem valószínű, hogy a rendszer túlnövi az adatbáziskezelő rendszer nélkül még adminisztrálható, kézben tartható méretet, így ez a megoldás elfogadható, sőt kifejezetten egyszerű és olcsó.

NetScape [http://www.soltmq.hu/~gazs/eng/de/index.html] File Edit View Go Bookmarks Options Directory Window Help

[home](#) | [countries](#) | [in Hungarian](#)

HUNGARICA COLLECTIONS - GERMANY

- [by City](#)
- [by Library](#)
- [by Type of Library](#)
- [by Subject](#)

Cities

- **Berlin**
 - [Universitätsbibliothek der Humboldt Universität zu Berlin](#)
 - [Universitätsbibliothek der Humboldt Universität, Zweigstelle Finnougristik](#)
- **Göttingen**
 - [Niedersächsische Staats- und Universitätsbibliothek, Referat für Uralistik und Osteuropäische Geschichte](#)
 - [Georg-August-Universität Göttingen, Finnisch-Ugrisches Seminar, Bibliothek](#)
- **Hamburg**
 - [Universität Hamburg, Finnisch-Ugrisches Seminar \(Zentrum für Hungarologie\), Bibliothek](#)
 - [Universität Hamburg, Historisches Seminar, Bibliothek](#)
- **Köln**
 - [Bibliothek des Seminars für Osteuropäische Geschichte, Universität zu Köln](#)

Document: Done

1. ábra

NetScape [http://www.soltmq.hu/~gazs/eng/fr/bib/23.html] File Edit View Go Bookmarks Options Directory Window Help

[home](#) | [countries](#) | [current country](#) | [in Hungarian](#)

HUNGARICA COLLECTIONS - FRANCE - Nanterre

BIBLIOTHÈQUE DE DOCUMENTATION INTERNATIONALE CONTEMPORAINE (BDIC)

F-92001 Nanterre, Cedex
 Centre Universitaire
 6, allée de l'Université
 Phone: 40-97-79-15 Fax: 40-97-79-40
 Internet address: <http://www.u-paris10fr/bdic/>

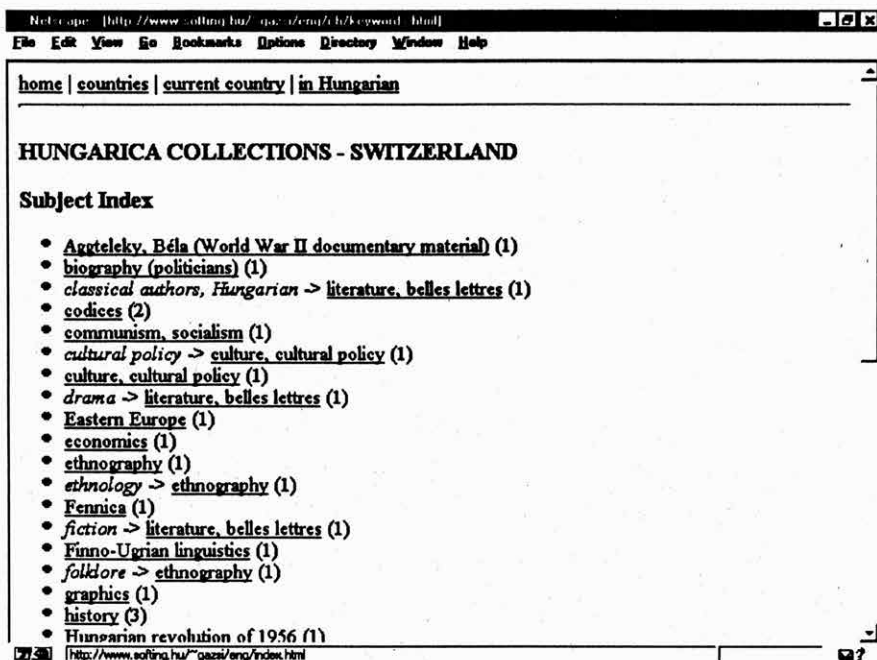
Head of the library and/or data supplier:
 Catherine Penin, Responsable de la Section Europe Centrale et Orientale de la BDIC
 Serolha Benzera, Spécialiste du hongrois

Foundation: 1914, collecting Hungarica since: 1917
 Library holdings: 1 000 000 vols, 90 000 item(s) periodicals
 Hungarica material/collection: 20 000 vols - Hungarian, French, German, Italian, Russian, multilingual

Subject profile:
[social sciences](#) - [social history](#) - [history](#) - [world wars](#) - [economics](#) - [economic history](#) - [agriculture](#) - [politics](#) - [political sciences](#) - [emigration affairs](#) - [international relations](#) - [women](#)

Document: Done

2. ábra



3. ábra

A Hungarika-WWW böngészése

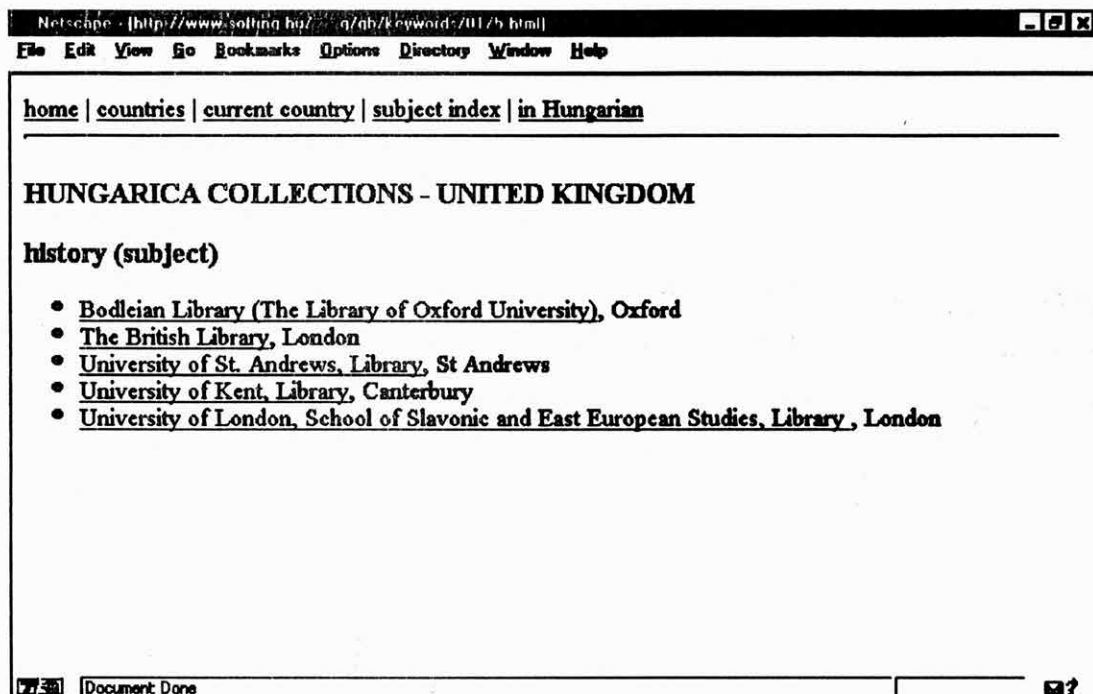
A Hungarika-WWW az OSZK honlapjáról érhető el. Címlapja, akárcsak az egész rendszer, kétnyelvű. Bármelyik nyelvet választjuk ki, a bejárható utak azonosak.

A keresés kiindulópontja a rendszer főlapja, amelyen egyrészt a keresés stratégiáját – országoként vagy témakörönként –, másrészt a böngészendő országot lehet kiválasztani, amennyiben az országokénti keresés mellett döntöttünk. Adott országon belül újból választhatunk: városoként vagy könyvtártípusoként kívánunk-e az ország hungarika-anyagot gyűjtő könyvtárai közt válogatni, esetleg a könyvtárnevek betűrendes mutatóját szeretnénk megnézni, netán témakörök szerint akarunk keresgélni. (1. ábra)

Az első három mutatótípus a könyvtárneveket sorolja fel más-más rendezettségben. Egy könyvtárnevet kiválasztva a megfelelő könyvtár, illetve hun-

garika-gyűjtemény részletes leírását kapjuk (2. ábra). A leírás tartalmazza a kapcsolatfelvétellel szükséges adatokat (név, cím, telefon- és faxszám, vezető és/vagy adatközlő munkatársak neve stb.), a hungarika-gyűjtés kezdetére, a gyűjtemény típusára, méretére, nyelvére utaló adatokat, és felsorolja azokat a tárgyköröket, amelyekben a gyűjtés folyik. A magyar nyelvű változat szöveges ismertetést is közöl a gyűjteményről, ami az angol változatban nem jelenik meg, vagy helyette egy rövid angol nyelvű összefoglalás olvasható. Fontos szerepe van az „Internet cím” elnevezésű adatnak, ha ismert: ez a szóban forgó könyvtár Web-címe (URL), amelyre rákattintva a böngészés a könyvtár honlapján, esetleg közvetlenül annak katalógusában folytatható.

Mind a teljes állományban, mind országoként lehetőség van tárgykör szerinti vizsgálódásra. A betűrendes tárgymutató (3. ábra) felsorolja a könyvtárak gyűjtésének jellemzésére használt vala-



4. ábra

mennyi tárgyszót zárójelben feltüntetve a témában gyűjtő könyvtárak számát. A tárgymutatóban dőlt betűvel megjelenő kifejezések szinonimák, amelyek a fogalom elsődleges megnevezésére mutatnak.

Egy tárgyszóra kattintva annak tárgykör-mutatójához jutunk (4. ábra), amely nem más, mint a tárgyszóval indexelt könyvtárak nevének felsorolása. A könyvtárnév – a többi mutatóhoz hasonlóan – a könyvtár leírásához vezet el. A tárgykör-mutatók nemcsak a betűrendes tárgymutatóból, hanem a tételek tárgyszó felsorolásaiból is elérhetők.

A kényelmes és gyors navigálást szolgálja a lapok tetején látható navigátorsáv. Ez visszalépést biztosít a laphierarchia magasabb szintjeire, minden oldalon megengedi a nyelvek közötti ide-oda váltást,

egy ország tárgymutatójából vagy annak valamilyen tárgykör-mutatójából pedig átlépést tesz lehetővé a teljes állományra vonatkozó tárgymutatóra, illetve tárgykör-mutatóba.

Zárszó

A Hungarika-WWW jelen állapotában egyedi szolgáltatás. A világháló használatának terjedésével azonban előbb-utóbb más olyan információs rendszereket is célszerű a Webre adaptálni, amelyek a manuális előállításához túl nagyok, egy nagyobb számítógépes apparátus gazdaságos igénybevételéhez azonban túl kicsik. Ilyen esetekben – akár régi, akár újonnan rögzített szövegfájlokra alapozva – az általunk kidolgozott módszer hasznosítható lehet.