

MÁTRIKSZ - Magyar Átfogó Tárgyi Információkereső-nyelvi Szótár

Új fejezet a tartalmi visszakeresésben

Bánki Zsolt

1. A MÁTRIKSZ projekt története

1.1. Van-e esély?

A Könyvtári Figyelő 2001. 1. számában *Ungváry Rudolf* az OSZK teaurusz/KÖZTAURUSZ-ról írott tanulmánya elején elmondta a véleményét az olyan nagyszabású tervekről, melyekkel „kívülről” (minisztériumok, vagy központi fejlesztő intézetek) állnak elő¹. Ma azonban már – lupus in fabula – ő is részese egy hasonló körülmények között született kezdeményezésnek, egy országos tárgyi információkereső szolgáltatás kialakításáról, mely közös hajóba ültetett olyan könyvtárakat, melyek között korábban szakmai véleménykülönbségek voltak.

Hogyan indult tehát az országos tárgyi információkereső szolgáltatás kialakításának terve?

2001. május 3-án az Országos Széchényi Könyvtárban a Nemzeti Kulturális Örökség Minisztériuma (NKÖM) könyvtári osztályának képviselője megbeszélést folytatott a nemzeti könyvtár és a MOKKA (Magyar Országos Közös Katalógus) illetékeseivel a MOKKA előtt álló feladatokról, s felmerültek a tárgyszavak kezelésével kapcsolatos szakmai kérdések is. *Monok István* az OSZK főigazgatója javasolta egy szakmai bizottság felállítását, amely olyan intézmények képviselőiből szerveződik, ahol már eddig is mélyrehatóan foglalkoztak a természetes nyelven alapuló osztályozással, és a gyakorlatban is alkalmazzák az ilyen technikákat. A bizottság felállí-

tásánál lényeges szempont volt, hogy ne kapcsolódjon közvetlenül a MOKKA-hoz, és a tagokat felkérés útján válasszák ki.

1.2. A Tárgyszó Bizottság tevékenysége

A megalakulásakor egyszerűen csak „Tárgyszó Bizottságnak” titulált testületbe a következő szakértőket kérték fel: *Aszalós Károly* (Szegedi Tudományegyetem Könyvtára), *Bakonyi Géza* (Szegedi Tudományegyetem Könyvtára), *Bánki Zsolt* (OSZK), *Czinkné Bükkösi Valéria* (Kaposvári Városi és Megyei Könyvtár), *Feimer Ágnes* (Könyvtári Intézet), *Koltay Klára* (Debreceni Egyetem Egyetemi és Nemzeti Könyvtár), *Ungváry Rudolf* (OSZK).

A bizottság feladata az volt, hogy megvizsgálja egy közös tárgyi visszakereső rendszer létrehozásának lehetőségét, s hogy megalkotható-e olyan szolgáltatás, amely központi szolgáltatásként lekérdezi a különböző könyvtári bibliográfiai, ill. egyéb nem könyvtári jellegű adatbázisokat is.

Megállapításairól a NKÖM számára 2001. június 15-ére előterjesztést kellett benyújtani, amely javaslatot is tartalmazott a tartalmi visszakeresés eszközeire vonatkozóan.

Az előterjesztés elkészítése után a bizottság tagjai úgy döntöttek, hogy a közösen kialakított koncepciót megvalósítják, ezért a szükséges feltételek megteremtése érdekében konzorciumot alakítanak és az Országos Széchényi Könyvtár égisze alatt pályázatot nyújtanak be a Széchényi Terv keretében a Miniszterelnöki Hivatal Informatikai Kormánybizottsága (MEH-IKB) által meghirdetett „*Az információs társadalommal kapcsolatos hazai kutatások készítésének, adatbázisok képzésének, karbantartásának, publikálásának, támogatása*” című pályázatra. A konzorciumnak azonos döntési joggal tagja lett az összes résztvevő intézmény:

- az Országos Széchényi Könyvtár,
- a Debreceni Egyetem Egyetemi és Nemzeti Könyvtár,

- a Szegedi Egyetem Egyetemi Könyvtár
- a Kaposvári Megyei és Városi könyvtár és
- a Könyvtári Intézet.

A konzorciumban a hazai könyvtárügy „nagy könyvtárai” mellett képviselve vannak a közművelődési könyvtárak és az elméleti háttérrel nyújtó Könyvtári Intézet is, így tevékenységüket az egész szakmát érintő, képviselő vállalkozásnak tekinthetjük.

A konzorcium a pályázati úton elnyert támogatás felhasználásával megkezdte a munkáját, melynek eredményeképpen a szolgáltatás várhatóan 2002 áprilisában elindul.

2. A MÁTRIKSZ – Magyar Átfogó Tárgyi InformációKereső-nyelvi Szótár fogalma

A „születendő gyermeknek” olyan nevet akartunk adni, mely könnyen megjegyezhető és kifejtve érthetővé teszi tartalmát. Így született meg a MÁTRIKSZ mozaikszó.

2.1. Mit takar ez a koncepció?

Átfogó tárgyi információkereső-nyelvi szótáron általánosságban az osztályozáshoz/indexeléshez és a kereséshez használt természetes nyelven alapuló nyelvek strukturált állományát értjük. Ilyenek a tárgyszójegyzékek, teauruszok. A tárgyszójegyzékekbe fölvetett tárgyszavak lehetnek hagyományos, összetett (fő-, al-, mellék-, formai, földrajzi stb. tárgyszóelemeket tartalmazó) tárgyszavak és egyedi, egymással egyenrangú tárgyszavak. A teauruszok lexikai egységei lehetnek deszkriptorok és nemdeszkriptorok. Mindezeket a kifejezéseket összefoglalóan tárgyi kifejezéseknek nevezzük (topical term). Az információkereső-nyelvi szótárak tárgyi kifejezései a bibliográfiai/katalogizálási

rendszerekben a tartalmi feltárás besorolási adatainak szerepét látják el.

A Magyar Átfogó Tárgyi Információkereső-nyelvi Szótáron (MÁTRIKSZ) meghatározott rész szótárak együttműködő rendszerét értjük, melyek között az egyiket kitüntetettnek tekintjük, de használat szempontjából a résztvevő szótárak egymással egyenrangúak. Jelen esetben a kitüntetett szótár az OSZK teaurusz/KÖZTAURUSZ lett, melynek szerepéről a 3. fejezetben részletesebben szólok.

Az együttműködés során megőrződik a résztvevő rendszerek önállósága, miközben kialakítjuk közöttük a közös kereshetőséget. Ennek értelmében a MÁTRIKSZ nem egyetlen szótár, hanem a résztvevő részrendszerek közös szótára. Alkalmas keresőfelület segítségével a felhasználó mindegyik résztvevő szótár állományáról tájékozódhat és használhatja tárgyi kifejezéseit mind osztályozásra/indexelésre, mind keresőkérdéseinek megfogalmazására, a résztvevő intézmények bibliográfiai adatbázisainak lekérdezésére. A MÁTRIKSZ önálló tárgyszó adatbázist alkot, mely független a résztvevő intézmények saját bibliográfiai adatbázisaitól, és lekérdezi azokat.

2.2. A közös kereshetőség módja

A résztvevő szótárak egy közös on-line mutatóban jelennek meg. Az egyes szótárak tárgyi kifejezéseit forrásuk szerint a mutatóban minősítik, belőlük kiindulva tekinthetők meg egyrészt a résztvevő adatbázisok bibliográfiai rekordjai, másrészt a tárgyi kifejezések kapcsolatai (tárgyszó-, ill. teauruszcikkek), valamint az adott kifejezések besorolási rekordjai. A kapcsolódó kifejezésekből kiindulva vagy újra az on-line mutatóhoz, vagy a kapcsolódó kifejezésekhez tartozó találatokhoz lehet továbbjutni.

Az egyes osztályozási rendszerek szótárai és a kapcsolódó bibliográfiai adatbázisok állománya egységes környezetben jelenik meg.

A résztvevő rendszerek tulajdonosai elhatározták, hogy tárgyi kifejezéseiket megfeleltetik az OSZK teaurusz/KÖZTAURUSZ kifejezéseivel. Megállapodásra jutottak az egyes részrendszerek hosszú távú szerkesztésének elveiben. Tárgyi kifejezések megfeleltetésére több, szabadon választható lehetőség kínálkozott:

- ☉ az azonos tartalmú, de eltérő alakú kifejezések megegyezésének biztosítása (alaki azonosság létrehozása) azzal, hogy átvesszik a kitüntetett szótár megfelelő kifejezését;
- ☉ az eredeti, az OSZK teaurusz/KÖZTAURUSZ szóhasználatától eltérő saját kifejezést megtartva a különböző alakú, de tartalmilag azonosítható kifejezések között a tárgyi kifejezés MARC besorolási rekordjának 750-es (ekvivalencia) mezőjébe beépítik az OSZK teaurusz/KÖZTAURUSZ megfelelő kifejezését. Ennek a folyamatnak az inverze lejátszódik a kitüntetett szótárban is, így a megfeleltetés kölcsönös. Ez az eljárás lehetővé teszi, hogy a kitüntetett szótár kifejezésére való keresés találatot eredményezhessen az összes könyvtár bibliográfiai adatbázisában;
- ☉ azokban az esetekben, ahol a részrendszerekben többlet jelentkezik az OSZK teauruszhoz/KÖZTAURUSZ-hoz képest, ott ezeket a kifejezéseket az OSZK teaurusz/KÖZTAURUSZ-ba beépítik annak szerkesztői.

Az így egységesített kifejezések már MÁTRIKSZ-szócikknek tekinthetők és az összes olyan intézmény bibliográfiai adatbázisában sikerrel folytathatunk keresést, ahonnan a megfeleltetett tárgyszavak származtak. A MÁTRIKSZ szótárában a megfeleltetés folyamatos munka. A projekt jelenlegi első szakaszában a matematika és a szociológia ismeretterületeinek szakkifejezéseivel végezzük el a megfeleltetést. Mivel a munka első fázisában még nem létezik közös adatbázis, ezt a lépést még mindenki a saját adatbázisában végzi el, egymásnak elektroni-

kus úton juttatva el az állományokat. Az első fázis befejeztével létrejön a közös szótári rész. A továbbiakban helyi szinten végrehajtandó megfeleltetések on-line töltődnek fel a közös MÁTRIKSZ szótárba. Az áprilisban induló MÁTRIKSZ mutatójában státuszát tekintve három típusú tárgyi kifejezés lesz található:

- ☞ Különböző rendszerekből származó alakilag és tartalmilag megfeleltetett kifejezések – ez MÁTRIKSZ tárgyi kifejezés, segítségével az összes elérhető bibliográfiai adatbázisban lehetséges találat;
- ☞ Különböző rendszerek alakilag különböző kifejezései, melyek a tárgyi kifejezés MARC besorolási rekordjának 750-es ekvivalencia mezőjében vannak megfeleltetve a kitüntetett szótár lexikai egységével – ez MÁTRIKSZ tárgyi kifejezés, segítségével az összes elérhető bibliográfiai adatbázisban lehetséges találat;
- ☞ Különböző rendszerek alakilag különböző kifejezései, melyek nincsenek megfeleltetve egymás-

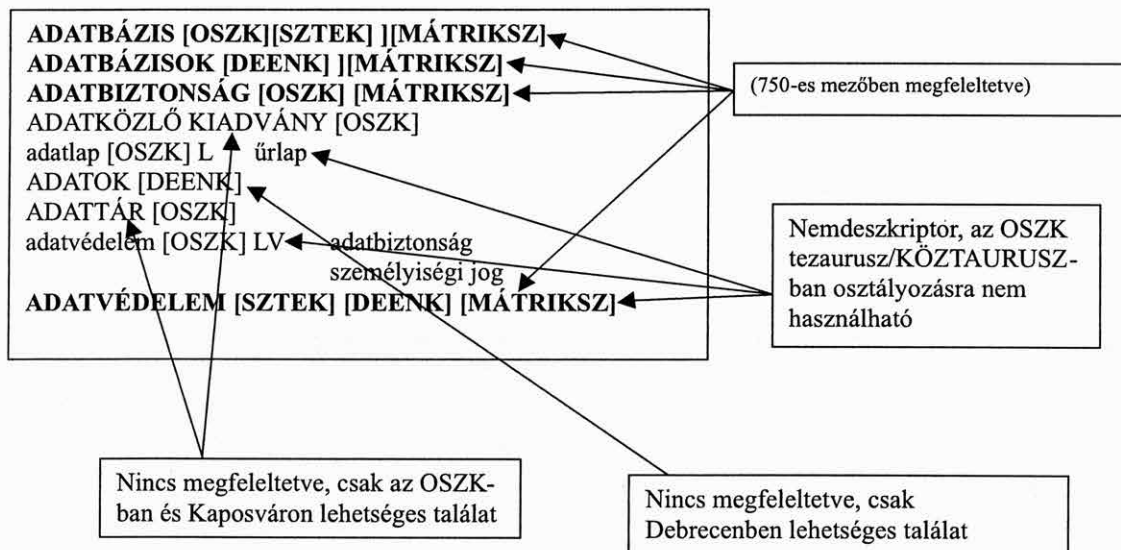
nak – csak a tárgyi kifejezés forrásául szolgáló intézmény bibliográfiai adatbázisában lehet találat.

3. A kitüntetett szótár és a résztvevők

Mint már említettem, a kitüntetett szótár szerepét az OSZK teaurusz/KÖZTAURUSZ tölti be. Az OSZK teaurusz és a közművelődési könyvtárak számára készült KÖZTAURUSZ általános, átfogó teaurusz. A két teauruszt egyszerre szerkesztették, jelenleg mindenben megegyeznek.²

3.1. A kitüntetett szótár szerepe

A kitüntetett szótárt ajánljuk alkalmazásra azon könyvtárak, illetve más felhasználók számára, akik nem rendelkeznek saját természetes nyelvű információkereső szótárral (tárgyszórendszerrel, teaurusssal). Az OSZK teaurusz/KÖZTAURUSZ kiér-



1. ábra

Példa az on-line mutató megjelenítésére

lelt, megfelelően strukturált állományú, folyamatosan karbantartják, és egységes szóhasználatot, terminológiát, ezáltal átjárhatóságot biztosít a különböző rendszerek között, valamint rendelkezik MARC adatsere-formátummal.

Az OSZK teaurusz/KÖZTAURUSZ azon könyvtárak, felhasználók körének, akik rendelkeznek természetes nyelvű tárgyi információkereső rendszerrel, vagy résztvevői a MÁTRIKSZ-nak, konzultációs segédeszközként szolgál meglévő tárgyi kifejezések (tárgyszócikkek, teauruszcikkek) kiegészítése céljából, illetve új tárgyi kifejezések felvételekor. A hosszú távú konvergencia alapvető feltétele a kitüntetett szótárhoz való igazodás a meglévő rekordok kiegészítésével (ekvivalencia mező beépítésével), valamint azzal, hogy új kifejezés felvételekor lehetőleg az OSZK teaurusz/KÖZTAURUSZ kifejezését használják vagy feleltessék meg annak, a fent leírt módon.

3.2. A MÁTRIKSZ-ban induláskor résztvevő rendszerek

- ☒ az OSZK teaurusz/KÖZTAURUSZ;
- ☒ a Debreceni Egyetem Könyvtárának átfogó tárgyszórendszere – a Library of Congress Subject Headings (LCSH) magyar adaptációja;
- ☒ a Szegedi Tudományegyetem Könyvtárának átfogó tárgyszórendszere – hagyományos tárgyszórendszer;
- ☒ az ETO mutató;

A résztvevő információkereső-nyelvi szótárak köre a későbbiekben még szabadon bővíthető.

4. Kapcsolat az ETO-val

A MÁTRIKSZ projekt támogatja az ETO középkiadásához készülő tárgymutató munkálatait, amelyek készítésekor a kitüntetett szótár szerepét betöl-

tő OSZK teaurusz/KÖZTAURUSZ szókészletét is figyelembe veszik, és amely lehetővé teszi, hogy az ETO mutató egybeépüljön az OSZK teaurusz/KÖZTAURUSZ releváns kifejezéseivel. A mutató egyes tételeiből képzett ETO besorolási rekordok részét képezik a leendő Átfogó Tárgyi Információkereső-nyelvi Szótárnak. A MÁTRIKSZ keresőfelülete opcionálisan felkínálja a tárgyszó rekordba beépített ETO jelzettel vagy jelzetekkel való keresést. A felhasználó eldöntheti, hogy csak a tárgyszóval, vagy a fogalomnak megfelelő ETO jelzettel akar-e keresni. Így a MÁTRIKSZ-on keresztül természetes nyelvű kifejezések segítségével lekérdezhetők az ETO-val osztályozott dokumentumok bibliográfiai rekordjai. E megoldás jelentőségét nem kell külön hangsúlyoznom akkor, ha figyelembe vesszük, hogy megítélésem szerint az ETO-val, illetve a természetes nyelvű kifejezéssel osztályozott bibliográfiai rekordok aránya kb. 70-30%.

5. A tárgyi kifejezések adatsere formátuma

Az Átfogó Tárgyi Információkereső-nyelvi Szótár kifejezéseinek szócikkei (tárgyszócikkek, teauruszcikkek) HUNMARC/USMARC besorolási rekord adatsere formátumban exportálhatók és importálhatók.

6. Felhasználói szoftver és keresőfelület

6.1. Elvárások a szoftverrel szemben

A két legnagyobb közös katalogizálási vállalkozás, a MOKKA és a VOCAL is saját, közös adatbázist létesített és tart fenn. Ezeknek a rendszereknek a lekérdezése nem okoz problémát, miután a bibliográ-

fiai rekordok közös adatbázisban vannak. A MÁTRIKSZ-nak meg kell oldania a különböző szoftverekben kezelt, fizikailag is több helyen lévő adatbázisok egységes, biztonságos lekérdezését. A MÁTRIKSZ szoftvere a Z39.50 kapun keresztül kommunikál az egyes bibliográfiai adatbázisokkal, úgy, hogy az azokból nyert találatokat a saját felületén egységes struktúrában jeleníti meg. A felhasználó egységes felületen szabadon választhat, hogy a résztvevő adatbázisok mindegyikét, vagy csak a kiválasztottakat akarja lekérdezni, párhuzamosan megjelenítve az egyes bibliográfiai adatbázisokból kapott találatokat.

A rendszer lényegéhez tartozik, hogy távolról elérhető központi szolgáltatást kíván nyújtani abban a tekintetben is, hogy a könyvtárak és más felhasználók számára szabadon letölthető tárgyszó rekordokat biztosít. A tezauszkezelés szoftveres megoldásának megvitatásakor kiindulópontnak tekintettük az Ungváry Rudolf által a Tudományos és Műszaki Tájékoztatás 2001. 3. számában e tárgyban publikáltakat.³ A MÁTRIKSZ szoftvere az első lépcsőben még nem fogja megvalósítani a komplex tezauszkezelés követelményét, de megfelelő keresési lehetőségeket biztosít.

6.2. Karbantartás, fejlesztés

A MÁTRIKSZ karbantartására, fejlesztésére közös karbantartó/kommunikációs mechanizmus és felület fog szolgálni, mely a résztvevő szakemberek és a külső felhasználók számára egyaránt biztosítja a rendszerhez való visszacsatolás lehetőségét.

7. A szoftver és a műszaki háttér

A szoftver szállítója a debis IT Services Dataware kft. Kiindulópontnak tekintjük a VOCAL tárgyszó-adatbázisát kezelő Corvina rendszert. A jelenleg ismert állapota a fent leírtak értelmében jelentős át-

alakuláson fog keresztülmenni a szolgáltatás elindulásáig. Ez azonban a MÁTRIKSZ projektnek csupán az első szakasza, így a szoftverfejlesztésben is tervezünk olyan lépéseket, melyek a későbbi periódusokban történnek majd meg. Ebbe a körbe elsősorban a szoftver komplex tezauszkezelő szolgáltatásainak kialakítása tartozik, a szintaktikai és tranzitív ellenőrzésekkel.

7.1 A MÁTRIKSZ adatbázis technikai alapjai

A tárgyszó adatbázis SUN ultra 2 (2x300Mhz 1GB) szerveren fut, SUN Solaris (Unix) operációs rendszer vezérli, Ingres adatbázis kezelőt használva. A szervert a Debreceni Egyetem Egyetemi és Nemzeti Könyvtára a Vocal adatbázis számára tartja fenn, és helyet biztosít a MÁTRIKSZ adatbázis számára is.

A szerver 100 Mbites UTP kábellel, switchen keresztül kapcsolódik a Debreceni Egyetem számítógépes hálózatára, amely 155 Mbittel éri el a HBONE magot. Ez a hálózati kapcsolat jelenleg jelentős tartalékokkal rendelkezik.

A felhasználók web interfészen keresztül kérdezhetik le az adatbázist. Az adatok feltöltésére a rendszer saját internet protokollt használ, ami bármilyen internetbe kötött számítógépen megvalósítható. (Unix, PC). A feltöltés automatikus.

7.2 A MÁTRIKSZ induló becsült rekordszáma

- ☉ OSZK tezausz/KÖZTAURUSZ – cca. 28.000 tárgyi kifejezés,
- ☉ Debreceni Egyetem Egyetemi és Nemzeti Könyvtár – cca. 25.000 tárgyi kifejezés,
- ☉ Szegedi Tudományegyetem Egyetemi Könyvtár – cca. 60.000 tárgyi kifejezés,
- ☉ a Magyar Átfogó Tárgyi Információkereső-nyelvi Szótár közös adatbázisa – cca. 113.000 tárgyi kifejezés.

8. További terveink

A szolgáltatás indulásának pillanatában – áprilisban – terveink megvalósításának csupán az első szakaszán jutottunk túl. Ezt a félkész állapotot a MÁTRIKSZ indexe híven fogja tükrözni, hiszen csak a két fent említett ismeretterületen készült el a megfeleltetés, a tárgyszórekordok jelentősebb része abban az állapotban fog megjelenni a indexben ahogy az egyes forrásokból érkezett.

A továbblépés útja tehát adott: újabb források megteremtésével, már on-line módon folytatni kell a megfeleltetést, egységesítést a MÁTRIKSZ és az egyes részrendszerek között.

A fejlesztés másik területe a szoftver szolgáltatásainak bővítése. A keresőszolgáltatás szélesítését, a karbantartás automatizálását, és a lekérdezhető adatbázisok körének bővítését szintén a következő

szakaszokban kívánjuk megvalósítani.

Reményeink szerint ebben a munkában partnerre találunk a MÁTRIKSZ-ot megismerő könyvtári környezetben, és felfedezik az elképzelésben rejlő nagyszerű lehetőséget, és így a MÁTRIKSZ előbb-utóbb valóban országos jelentőségű szolgáltatássá nővi ki magát.

A cikk megjelenésekor a MÁTRIX már on-line elérhető a következő címen:

<http://www.matrix.hu>

JEGYZETEK

- 1 UNGVÁRY Rudolf: Az OSZK teaurusza és a KÖZTAURUSZ
In: Könyvtári Figyelő. 2001. 47. évf. 1. sz. 11-41. p.
- 2 i. m.
- 3 UNGVÁRY Rudolf: Teaurusz a felhasználói felületen. Az optimális megjelenítés problémái. In: Tudományos és Műszaki Tájékoztatás. 2001. 48. évf. 3. sz. 99-109. p.

A Könyvtártudományi Szakkönyvtárban készülő MANCI (Magyar és Nemzetközi Cikk) adatbázis web felületen elérhető változata két új szolgáltatással bővült:

1. Válogatás a találatokból, találatok letöltése

A Keresési eredmények képernyőn ezután lehetőség van arra, hogy a bibliográfiai tételek felett megjelenő négyzetbe kattintva válogassunk a találatokból. Ezután a letöltés (kijelöltek) opció kiválasztásával a Letöltés képernyőhöz jutunk, ahol a kívánt formátum és karakterkészlet megadása után letölthetjük a kiválasztott találatokat. Itt a letöltés gombra kattintva a rendszer válasza az internet-böngésző típusától és beállításaitól függ: pl. a Netscape felajánlja a választást a fájl megnyitása és lemezre mentése között. Az Explorer automatikusan a hozzákapcsolt szövegszerkesztőben (ha van ilyen) jeleníti meg a találatokat. Ugyancsak a Keresési eredmények képernyőn kezdeményezhetjük az összes találat letöltését, illetve az összes kijelölés letöltését is.

2. Referátumok megjelenítése

1991-től a Könyvtári Figyelő Külföldi folyóirat-figyelő rovata számára kiválogatott jelentősebb külföldi cikkek magyar nyelvű referátumai, illetve annotációi elektronikus formában is rendelkezésre állnak. Ezeket a bibliográfiai tételek végén levő Ref. vagy Autoref. adatelemre kattintva lehet megjeleníteni.

A fenti fejlesztések kivitelezője és tervezője Kiefner Ferenc.

Az adatbázis elérhető a Könyvtári Intézet honlapján keresztül a www.ki.oszk.hu címen, az OSZK honlapjáról a www.oszk.hu címen, illetve közvetlenül a <http://w3.oszk.hu/manci.htm> címen is.

Észrevételeiket, javaslataikat Novák István szerkesztőnek lehet megírni a novak@oszk.hu címen.

(Rácz Ágnes tudósítása. Katalist, 2001. dec. 21.)
