

Drótos László – Somogyi Tamás

Adattármustra. Társadalomtudomány

A sorozat keretében független szakértők értékelik a magyar könyvtárak tartalomszolgáltatásait megadott szempontok szerint; első alkalommal a társadalomtudományok témájában. A bírálókat a használók számára fontos keresőfelülettel, keresőrendszerrel, tartalommal és metaadatokkal foglalkoznak. Az észrevételek egyrészt segítséget nyújtanak a rendszerek minél jobb és eredményesebb használatához, másrészt felhívják a fejlesztők és üzemeltetők figyelmét a szolgáltatás hiányosságaira, a jövőbeli minőségi javítások érdekében.

Erdélyi Magyar Adatbank

Honlap: <http://adatbank.ro>

E-mail: ugyelet@transindex.ro

Fenntartó: Jakabffy Elemér Alapítvány és Transindex

Értékelés dátuma: 2006. október 27.

Értékelő: Drótos László

Általános értékelés

Az Erdélyi Magyar Adatbank az egyik legjelentősebb határon túli magyar digitális gyűjtemény, amelyet a Jakabffy Elemér Alapítvány és a Transindex professzionális módon fejleszt 2004 eleje óta. Középtávon egy – Erdélyre összpontosító – társadalomtudományi referenciakönyvtár felépítése a cél, amely egyaránt szolgálja a kutatók és a szélesebb nagyközönség igényeit. Ennek megfelelően 2006 októberében már több tucat „tár”: névtárak, bibliográfiák, repertóriumok, katalógusok, térképtárak, szótárak, statisztikai adatok, kutatások nyilvántartásai, kataszterek stb., valamint több száz digitális könyv, tanulmány, cikk, irodalmi mű alkotta a gyűjteményt. A gyűjtőkört jelenleg már a szépirodalom, illetve általában a művészetek irányában is bővítik (és hangfelvételeket, videókat is felvesznek), a távlati tervek között pedig egy természettudományi alapkönyvtár kialakítása is szerepel. Most folyik a szolgáltatás portállá való átalakítása: a központi gyűjtemény mellett írók és tudósok „vendégoldalai”, magyarlakta települések honlapjai, és külső ugrópontok is helyet kapnak a szerveren.

A szolgáltatás tesztelése során megállapítható, hogy a tervezésnél megfogalmazott célok megvalósítása jó úton halad:

- a digitalizált anyagok fontosak, szakmai színvonaluk magas;
- az Erdéllyel kapcsolatos ismeretek egyre nagyobb körét fogják át;
- a statikus dokumentumok mellett sokféle kereshető tartalom is van;
- a gyűjtemény internetre optimalizált, és a felület használóbarát.

Az adatbank.ro vizsgálata három jelentős hiányosságot mutatott ki:

1. A tájékoztatók, súgók hiánya a kezdőlapról elérhető keresőknél, amelyek e nélkül „fekete dobozként” működnek, pedig a keresők egyre fontosabbak lesznek, mivel a gyűjtemény növekedésével a böngészés már nem elég hatékony.
2. A géppel olvasható, nyilvános metaadatok teljes hiánya, amelyekre a más gyűjteményekkel való közös kereshetőség miatt volna szükség.
3. Az angol és/vagy egyéb világnyelvű felület és dokumentumok hiánya, ami pedig nagyban segítené az erdélyi magyarság e fontos információforrásának világméretű használatát, és egyben a nyelvi elszigeteltségből való kiszabadulást.

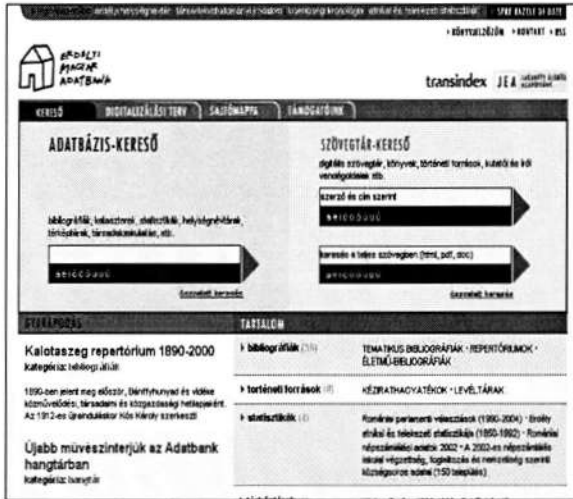
Összességében azonban az Erdélyi Magyar Adatbank példamutatóan gazdag és színvonalas szolgáltatás, amely minta lehet nemcsak más magyar kisebbségi adattáraknak, hanem általában az egy konkrét téma köré szerveződő digitális gyűjteményeknek is.

Részletes értékelés

Felület, használhatóság, segítség

- A kezdőlapon (1. ábra) rajta van minden fontos funkció: a háromféle gyorskereső, a részletes

tartalomjegyzék, az újdonságok listája és a négy legnépszerűbb gyűjteményegység, így könnyen áttekinthetjük az egyébként már most is zavarba ejtően bőséges kínálatot.



1. ábra A honlap felső része a keresőkkel és a tartalomjegyzékkel

- A projekt saját bemutatása viszont meglehetősen el van dugva. A digitalizálni tervezett kiadványok köre külön menüpontot kapott ugyan a mindössze négy részből álló főmenüben, de a projekt célját és jelenlegi állását ismertető szövegek a félrevezető nevű „Sajtómappa” rovat alá kerültek, néhány friss hírrel együtt. Valódi sajtóanyagokat (sajtóközleményeket, képeket, statisztikákat stb.) viszont nem találunk itt.
- Nem készítettek megfelelő tájékoztatót, illetve segítséget sem a laikus, sem a professzionális felhasználóknak, így például a gyors-, illetve az összetett keresők működését is csak találgatni lehet, pedig a hatékony keresés egy ekkora gyűjteménynél már nagyon fontos. A kevés tájékoztató szöveg is téves néhol: az *adatbank.transindex.ro/fkereso.php* lapon lévő teljes szövegű keresőnél az adatbázis-kereső (*abkereso.php*) magyarázata felejtődött ott.
- A navigáció viszont nagyon jó, az aloldalak egy részén is megjelenő rövid tartalomjegyzéknek, a kapcsolódó és ajánlott oldalak listáinak, valamint a funkciógazdag kezdőlapnak köszönhetően, ahová szinte mindenhol van visszatérés (bár például a *kuszalik2.adatbank.transindex.ro* oldalról visszamutató nyíl egy hibaüzenetre visz, a *kataszter.adatbank.transindex.ro* lapon pedig egyáltalán nem működik).
- Mivel a kezelőfelület nagyrészt karakteres, ezért elvben a hátrányos helyzetű felhasználók is jól

elbaldogulnak vele. A lassú kapcsolattal rendelkezőknek kétségtelenül gyorsan letöltődnek a kevés és kisméretű grafikus elemet tartalmazó oldalak. A gyengén látók számára hasznos az egyszerű, néhány erőteljes alapszint használó külalak, de ők például már hiába próbálnák a szövegméretet nagyobbra állítani a böngészőben, erre az oldal nem reagál. A vakok pedig a kevés, de fontos grafikus elem jelentését nem tudják kitalálni, mivel nincs hozzájuk szöveges leírás, így a keresést indító ikon helyett csak ennyit ír ki a lynx böngésző: *rozsaszingomb.gif*. A szolgáltatás nevét sem könnyű kideríteniük egyébként, mert az „Erdélyi Magyar Adatbank” felirat a nyitólapon csak a kis grafikus logón bogarasztható ki, a honlap „title” sora csupán a semmitmondó „Adatbank” szót tartalmazza. Hasznos viszont, hogy sok helyen van nyomtató nézet, vagyis navigációs oszlopok és menük nélküli „nyers” szöveg, ami a felolvasó programoknak is megfelelő.

- A magyar mellett egy erősen csökkentett funkcionalitású és tartalmú román kezdőlap van, angol vagy más nyelvű kezelőfelület nincs, csak egy egy mélyebben levő oldalon bukkan fel néha angol felirat vagy rövid kiadványismertető, pedig a gyűjtemény jellege igencsak alkalmas lenne világméretű használatra. A „nyelvszegénység”-et egyetlen jó ötlet enyhíti valamelyest: a kezdőlapon a gyorskeresőknél a magyar ékezetek egérrel is bevitethők (de itt a román karakterekre nem gondoltak, pedig valószínűbb, hogy – a szervert statisztika szerint – a romániaiaknál kétszer több magyarországi felhasználó inkább azok beírásával nem boldogul).
- A szolgáltatást fejlesztők és működtetők logója a kezdőlap tetején, megnevezése a lap alján fel van tüntetve, utóbbi helyen a két vezető személyes e-mail címe is megtalálható. A felhasználók visszajelzésére mindössze a „KONTAKT” gombra kattintva indítható levélküldés szolgál, hagyományos elérési címeket és telefonszámokat csak a fenntartók saját honlapjain böngészve lehet találni. Általában az egyes gyűjteményrészek aloldalain is megjelenik vagy a központi ügyelet, vagy az illetékes szerző/szerkesztő postafiókjának címe, de itt már egy „észrevételeim vannak” vagy „írok a szerzőnek” menüpontra kötvé. Ugyancsak állandó elem a „könyvjelzőm” menüpont (a „bookmark” szó Transindex-féle magyarítása), ami megkönnyíti a kezdő inter-nethasználóknak, hogy felvegyenek egy dokumentumot a kedvencek listájába. Ennél is jobb azonban, hogy egyes aloldalaknak saját, domén-szintű URL címük van (pl. *csikszereda.adatbank*).

transindex.ro), így könnyen lehet rájuk hivatkozni.

Kereső funkciók, találatok

- A kezdőlapra három gyorskereső van (2. ábra): egy az adattárjellegű forrásokhoz, kettő pedig a szöveges dokumentumok szerző+cím szerinti, illetve a *html*-, *doc*-, és *pdf*-állományokban való teljes szövegű kereséséhez (utóbbi két formátumnál csak nagyon korlátozottan, a tesztek szerint). A komolyabb feladatokhoz három „összetett kereső” áll rendelkezésre, az egyik szintén az adatbázisokhoz (ez csak annyiban különbözik az egyszerűtől, hogy kiválaszthatjuk az adattárak körét); a másikkal a szöveges dokumentumok egyéb adatai (pl. téma, műfaj, év, kiadó) között is kereshetünk; végül a harmadik a szerverten szintén helyett kapott írói és kutatói vendégoldalakon elérhető publikációk között keres. Ennyi keresési lehetőség az átlagfelhasználónak általában elegendő is, különösen hogy ott van még a dokumentumtípusok szerinti böngészés lehetősége. De professzionális célokra ezek az egyszerű AND kapcsolatot biztosító és kevés mezőtípust tartalmazó keresők nem alkalmasak, bár valószínűleg nem is lehetne egy igazán részletes keresőt írni egy ennyire heterogén gyűjteményre. Szerencsére az egyes nagyobb adat-, név-, térképtáraknak és katalógusoknak külön speciális keresőjük és böngészőfelületük van, ezekkel többnyire már jó találati listákat lehet elérni.

2. ábra A szöveges dokumentumok összetett keresője

- A sűgók hiányát a gyorskeresők beviteli mezőjébe írt, odakattintásra eltűnő tippek próbálják pótolni, de ezek egyrészt félrevezetők (a „szerző és cím szerint” mező valójában a megjelenési adatok közt is keres), másrészt nagyon zavaró, hogy ha mi írunk be egy keresőkérdést, majd javítás céljából megpróbáljuk egy hibás betű mellé tenni a kurzort az egérrel, akkor a mi eddig beírt kérdésünk is eltűnik a kattintás miatt, és kezdetjük előlről a gépelést. Keresőkérdéseink egyébként

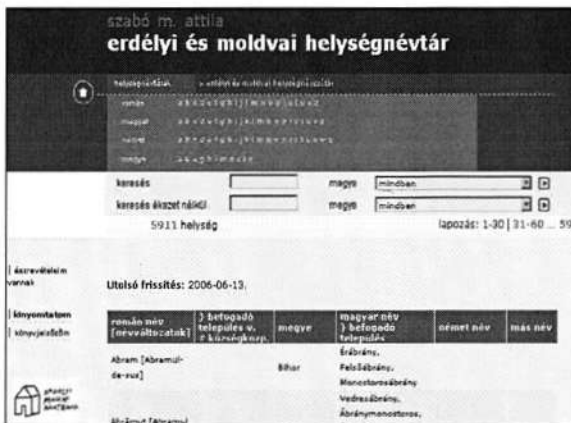
a találati lista megjelenésekor is eltűnnek, így bár hasznos ötlet, hogy az űrlap ilyenkor is látható a lista felett, valójában vissza kell lépnünk a böngészővel az előző oldalra, ha módosítani akarunk a keresésen, és nem akarunk mindent újra begépelni.

- A kezdőlapról elérhető keresők úgy tűnik, hogy csak egyetlen mezőn belül keresnek, annak is csupán egyszeri előfordulására illesztik a mintát. Így például egy többszerzős mű esetén nincs találat, ha egynél több nevet próbálunk beírni. Ugyanakkor a tárgyszó mezőnél sikerült néha két fogalmat összekombinálni vesszővel elválasztva, de ez sem következetes, mert például a „kisebbség, Románia” kérdésre van találat, a „kisebbség Románia” vagy a „Románia, kisebbség” kérdésre már nincs. A találatokat elemezve valószínűsíthető, hogy valamennyi tárgyszó egyetlen szöveges mezőben van, és a kereső erre illeszti a mintát egyszerű belső kereséssel. Erre a működési módra utal az is, hogy bár csonkolásra szolgáló speciális karakter nincs, a belső mintaillesztés miatt végül is a mindkét oldali automatikus csonkolás minden mezőben megvalósul (pl. a címben az „ormány” betűcsoportra keresve megtalálja az „önkormányzat” és „kormánykoalíciós” szavakat is). Automatikusan működik az ékezet nélküli keresés is, legalábbis nagyjából, mert például a „Bűvös erdő” nevű óvodát ékezet nélkül nem sikerült megtalálni az adattárakban. A kiadási évszámot is egyszerűen kezeli a rendszer: csak egyetlen évre lehet keresni, nincs intervallumra való szűkítési lehetőség. Nagyon hiányzik a gyűjteménybe való bekerülés dátuma szerinti keresés: mivel a kezdőlapra gyűjteményrészekenként csak egy-két újdonság van kiemelve, az RSS fájl pedig csak az utolsó 6 tétel címét tartalmazza, így nem lehet átnézni egy nagyobb időszak gyarapodásának ütemét és összetételét.
- Az egyedi szöveges dokumentumok összetett keresőjénél a témakör és a műfaj lenyíló menüből választható, a tárgyszónál pedig külön ablakban megjelenő index kérhető, ezek sokat segítenek abban, hogy ne kelljen találgatni; jó lenne ilyen a személy- és helynév mezőknél is.
- A válaszdíkok igen jók, még a teljes szövegű keresőnél is szinte azonnal jönnek a találatok. Nagyon hasznos, hogy az adattárjellegű gyűjteményrészekenél előbb egy összesített listát kapunk, ami csak azt mutatja, hogy melyik adatbázisban hányszor fordul elő a keresett betűcsoport, majd a kiválasztott tételre kattintva a kérdésünk továbbítódik az adott gyűjteményrész saját keresőjének, és onnan kapjuk azután a részletes és teljes találati listát. A találatok jól áttekinthe-

tők, minden fontos adatot tartalmaznak, az egyedi dokumentumoknál még a formátumot és néhol a fájlméretet is. Időnként azonban üres – talán időközben törölt – *html* oldalak is felbukkannak a listában (pl. a „mellet” szóra keresve). A teljes szövegű keresőnél – Google-logikával – csak 10 találat jelenik meg egyszerre, de szerencsére utána ezt akár 500-ra is növelhetjük (egy ekkora lista megjelenítése persze már jóval tovább tart). Kár, hogy itt nincs nyomtatási nézet, amivel a találatok „akadálymentesen”, a navigációs keretek nélkül kinyomtathatók, elmenthetők vagy felolvastathatók lennének.

Tartalom, minőség

- 2004 óta a fejlesztők tekintélyes és változatos gyűjteményt alakítottak ki, amely témájában természetesen a romániai magyarságra helyezi a hangsúlyt (3. ábra), azon belül is a társadalomtudományi, történelmi és közigazgatási adatokra és szakirodalomra, de az utóbbi időben erősödik a szépirodalmi gyűjteményrész is. Az eredetileg főleg adattárakból álló szolgáltatás ma már nagy számban tartalmaz digitális könyveket, térképeket, sőt újabb hangfelvételeket is, valamint a – portállá való átalakulás jeleként – külső hivatkozásokat hasonló témájú forrásokra. A 2006 szeptemberében frissített digitalizálási terv pedig újabb jelentős tartalombővítést ígér. A teljes szolgáltatás ingyenesen használható, és az egyedi dokumentumok le is tölthetők – bár többnyire csak fejezetenként, egybecsomagolt verziók nincsenek.



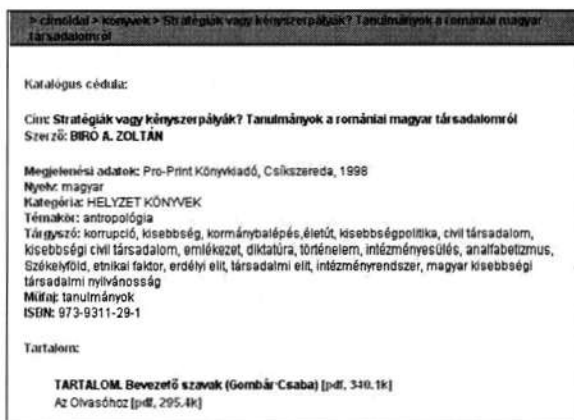
3. ábra Egy adattár kezdőoldala

- A formátumok között az interneten elterjedt *html* és *pdf* dominál, időnként Word dokumentum is előfordul; a többféle formában is elérhető szöveg azonban ritka. A térképek jó minőségű *jpg*, a hangok *mp3* állományok.

- A szövegek minősége jó, szembeötlő hibákat nem sikerült találni, a tipushibákra (pl. „visssza”, „viszsa”, „viszza”) csak egy-két találat van.
- A karakterkódolás a megvizsgált *html* oldalak fejlécében többnyire közép-európai (ISO 8859-2), de a fájlok belsejében már nem ilyen következetes a dolog: nyugat-európai „háztetős ő” betűk, „acute” kódolású magyar karakterek, és Unicode kódolású „ő” és „ű” betűk egyaránt előfordultak, és olyan írásjelek vagy nem magyar karakterek is, amelyek nincsenek a 8859-2-es szabványban, így helytelenül jelennek meg. (Valószínűleg kódhiba lehet az írói vendégoldalakon lapozásra szolgáló „következő szöveg” felirat hibás „ő” betűje mögött is.) Egyes adattárak pedig nyugat-európai, illetve utf-8 kódlapokkal készültek. A kódolások keveredése főleg a teljes szövegű keresésnél okoz gondot, a tesztek azt mutatják, hogy nem is lehet megtalálni egyes szavakat (pl. a „kellető” szóra nulla találatot ad a rendszer saját belső keresője, pedig többek között a bogdanlaszlo.adatbank.transindex.ro/belso.php?k=14&p=389#11 oldalon is előfordul – és a Google meg is találja).

Metaadatok, bibliográfiai információk

- A honlap az „Adatbank” szón kívül nem tartalmaz metaadatokat, és az egyes aloldalakon, a gyűjteményrészek nyitólapjain, sőt magukban a dokumentumokban sem sikerült ilyeneket találni. A számítógéppel olvasható metaadatok hiánya a legnagyobb probléma ezzel az amúgy professzionális módon fejlesztett, és elsődlegesen kutatási célokra készült gyűjteménnyel. Az összetett keresés űrlapján látható mezők ugyanakkor azt mutatják, hogy az alapvető adatok az egyedi szöveges dokumentumok egy részénél rendelkezésre állnak a háttérben működő adatbázisban (4. ábra), így megvan a remény rá, hogy ezek idővel bekerülhetnek majd a *Nemzeti Digitális Adattárba*, és így más gyűjteményekkel együtt – legalább részben, a fontosabb egyedi publikációk körére kiterjedve – közösen kereshetővé válik a határon túli magyarságnak ez a jelentős digitális információforrása.
- Ami a felhasználóknak szánt „szabad szemmel” kísérőadatokat illeti: itt nagyon vegyes a kép. Csak az egyedi, könyvjellegű dokumentumok egy részéhez van rendes „katalóguscédula”, de sok esetben – különösen a speciális dokumentumok gyűjteményrészeinél – kideríthetetlen a digitalizáláshoz használt nyomtatott forrás, vagy legalábbis nyomozást igényel, mert nem jelenik meg mindenhol következetesen egységes bibliográfiai leírás.



4. ábra Digitális könyv katalóguscédulája és tartalomjegyzéke

TÁRKI Adatbank

Honlap: <http://www.tarki.hu/adatbank-h/>

E-mail: tarki@tarki.hu

Fenntartó: TÁRKI Rt.

Értékelés dátuma: 2006. október 29.

Értékelő: Somogyi Tamás

Általános értékelés

A Társadalomkutatási Informatikai Egyesülés működtette TÁRKI Adatbankban (TDA) empirikus társadalomtudományi kutatások adatbázisai kereshetők. Jelenleg közel 650 adatgyűjtemény – többségében kérdőíves szociológiai kutatás – áll a hazai és külföldi társadalomtudományi kutatók, oktatók és hallgatók rendelkezésére. A 2001–2002-ben megvalósult *Empirikus Társadalomtudományi Adattár*ra épülve az *Informatikai és Hírközlési Minisztérium* támogatásával jött létre a TDA. A TDA honlapjáról elérhető még:

- a *Magyar Háztartás Panel* CD-ROM, amely az 1991 és 1997 közötti magyar háztartásokról készült empirikus tanulmányokat tartalmazza;
- a *TÁRKI Könyvtár* és *Andorka Rudolf* hagyatékából származó könyvtári gyűjtemény közös katalógusa;
- a *Női Adattár*, a női nemmel kapcsolatos társadalomtudományi kutatások nyitólapja;
- a *Kutatásjelentések*, a *Nemzeti Kulturális Alap-program* támogatásával létrejött kutatási összefoglalók, beszámolók és tanulmányok katalógusa;
- a *CHER* projekt (*Consortium of Household Panels for European Socio-economic Research*)

magyar oldala, egy 2002-ben és 2003-ban EU-támogatással megvalósult összehasonlító adatbázisprojekt, amely konzorciumának a TÁRKI is tagja volt.

Az Országos Tudományos Kutatási Alapprogramok (OTKA) és a TÁRKI kutatásainak eredményét őrző TDA-ban bárki elhelyezheti társadalomkutatási eredményeit egy letéti nyilatkozat kitöltésével. Ezzel egyidejűleg meg is határozhatja azt, hogy ki férhet hozzá az adatbázisához. A TDA négyféle hozzáférési szintet ismer:

- bárki számára hozzáférhető,
- magyarországi kutatóhelyek és közintézmények számára hozzáférhető,
- a TÁRKI Egyesülés tagjai számára hozzáférhető,
- csak az adattulajdonos engedélyével hozzáférhető.

A TDA-t a TÁRKI Rt. finanszírozza, kiegészítve a TÁRKI Egyesülés tagjainak tagdíjával és pályázati pénzzel. Az adathasználatért a TDA díjat számol fel, amelyet az adatok aktualitása és az adatbázis jellege alapján határoz meg. A külföldi és a hazai kutatók, valamint az egyesülés tagjai kedvezményben részesülnek.

A TDA legfőbb céljának, a tudományos nyilvánosságnak és az erre épülő tudományos másodelemzésnek a megvalósulását három kereső segíti. Keresni lehet az adatgyűjteményeket leíró metaadatok adatlapjában, a kérdőívekben és a publikációs listákban. Véleményem szerint a TDA céljainak megvalósítását jobban segítené az egyes keresők által elérhető tartalmak gondosabb bemutatása, valamint a honlap és a keresők több nyelven való megjelenése. A tudományos nyilvánosságot teljes mértékben csak gondozott metaadatok közzétételével, és az adatgyűjtemények szabad hozzáféréseinek megoldásával lehet elérni, amihez az első lépéseket minél hamarabb meg kellene tenni.

Részletes értékelés

Felület, használhatóság, segítség

- A TDA bemutatását és a keresők használatát lényegre törő, és általában érthető leírás segíti, bár sajnos néhány helyen értelmetlen mondatokkal is találkozhatunk.
- A honlapon több projekt is elérhető, ám ezek kapcsolata nem derül ki egyértelműen, mint ahogy az sem, hogy ezeknek mely tartalmi kereshetők a TDA különböző keresőfelületein (5. ábra). A TDA felülete kétnyelvű: magyar és angol. Sajnos a nyelvek közötti váltás lehetőségét

az adatbank oldalain nem találtam meg. A sokat ígérő „In English” menüpont a TÁRKI főoldalának angol verziójára juttatja el a felhasználót az adatbank angol felülete helyett. (Rutinos keresők azonban az adatbank URL-jében a „h” betűt „e”-re cserélve már tudják is olvasni a TDA angol felületét.) Egy ilyen komoly és értékes adatgyűjtemény honlapja az angolon kívül további nyelvek szerinti olvashatóságot is megkövetelne.



5. ábra A TDA kezdőoldala

- A honlap stílusa egyszerű, a háttérszín és a betűméret könnyű olvashatóságot tesz lehetővé, szinte felbontástól függetlenül. Kimondottan akadálymentes verzió nincs. Így tehát a honlap akadálymentes is, meg nem is.
- A honlapon lévő utolsó frissítés dátuma 2006. júliusi, a TDA adatfelhasználási díjait bemutató aktuális árlista pedig 2002. január 2-i.
- A TDA munkatársaival az impresszumban megadott elérhetőségeken keresztül lehet felvenni a kapcsolatot, ezenkívül az informatikai osztály vezetőjének egy kattintással lehet levelet írni.

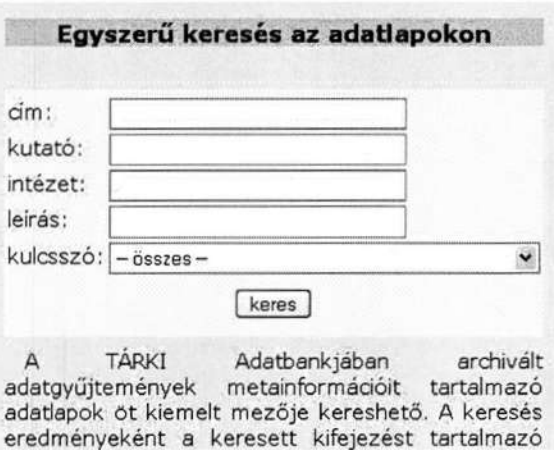
Kereső funkciók, találatok

Három keresőfelületet használhatunk: az adatgyűjteményeket leíró metaadatok adatlapjában, a kérdőíveken és a publikációs listákban. Ezenkívül külön keresőfelületet találunk a Kutatásjelentésekre.

Keresés az adatlapokon

- Az adatlapokon való kereső magyar és angol felülettel rendelkezik, külön „cím”, „kutató”, „intézmény” és „leírás” szabad mezőkkel (6. ábra). A „kulcsszó” mező tartalma kötött, egy előre megadott listából választhatunk. A kereső érzékenyen a kis- és nagybetűkre, jól működik a szó-tövező része.
- A nem üres mezők tartalmát „ÉS” kapcsolóval veszi figyelembe a keresőmotor. Ennél összetettebb keresést nem lehet végezni.
- A találati lista a kutatás címéből és azonosítójából áll, a címre kattintva jelenik meg több adat. A

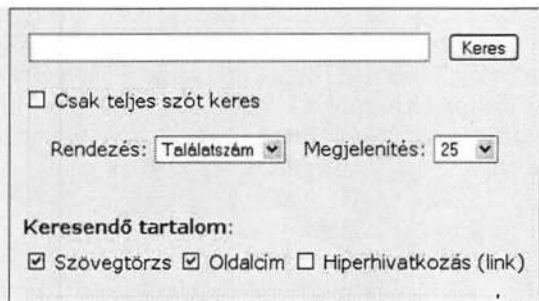
találati lista nem rendezhető, a lista mérete állandó: minden találat egy oldalon jelenik meg, a találati számtól függetlenül.



6. ábra Keresőfelület az adatgyűjtemények metaadataihoz

Keresés a kutatások kérdőíveiben

- Itt csak egyetlen mező áll rendelkezésünkre, a beírt kifejezéshez megadhatjuk, hogy csak egész szóra keressen-e a keresőmotor (7. ábra). Megválaszthatjuk a keresés helyét is: a szövegtörzsben, az oldalcímekben és a hiperhivatkozásokban.



7. ábra A tanulmányokhoz használt kérdőívek keresőfelülete

- A találati listát több szempont szerint rendezhetjük, illetve az egy oldalon megjelenő eredmények számát is maximalizálhatjuk.
- A keresést tanácsok és példák segítik. Több kereső megadásakor beállíthatjuk a logikai kapcsolatukat, valamint súlyozhatjuk is a keresőszavakat.
- A keresőfelületnek csak magyar nyelvű verzióját találtam. A sugó szerint a keresőmotor a nem magyar ékezetes betűket a kereséshez ékezetlen betűkre cseréli (pl. ä → a, ç → c, ô → o), de

tapasztalatom szerint magyar szavakat ékezetek használata nélkül nem lehet keresni.

- Az „életmód” szóra rengeteg találatot kapok, míg az „életmod”-ra egyet sem.

Keresés a kutatásokhoz kapcsolódó publikációs adatbázisban

- Itt „szerző” és „cím” szabad mező, valamint kötött „kulcsszó” mező áll rendelkezésünkre.
- A találati lista nem rendezhető, és nem méretezhető, minden találat egy oldalon jelenik meg. A szabadon elérhető tartalmak egy kattintással tölthetők le vagy olvashatók el, ugyanígy egy kattintással jelenik meg a publikáció adatai.
- Ezt a keresőfelületet is csak magyar nyelven találtam meg.

Keresés a kutatásjelentések között

- „Szerző” és „cím” szabad mezőt tölthetünk ki, a „nyelv” mező kötött (angol, magyar vagy bármely nyelv), ahogyan kötött a kulcsszó mező is (8. ábra).

8. ábra A kutatásjelentések keresőfelülete

- A keresőfelület alatt keresést segítő tanácsokat olvashatunk, valamint egy központi cím is szerepel, amelyre kattintva már írhatjuk is észrevételeinket. A keresőfelületen megtalálható „szerző” és „cím” mező mellett egy „támogató” mező használatához is kapunk tanácsot, amelyet azonban nem tudunk megfogadni, mert a mező hollétével kapcsolatban már nem találunk segítséget.
- A tanácsok alatt a készítőik feltüntették az oldal támogatóit, és a megvalósításához hozzájáruló nyertes pályázat nevét.
- Ez a felület is csak magyarul érhető el, annak ellenére, hogy a tartalmak egy része angol nyelvű, és az EU zászlaja is feltűnik a felületen.

- Leírás hiányában nem világos számomra, hogy pontosan miben is kereshetek itt, mik is pontosan a kutatási jelentések.

Tartalom, minőség

A vizsgálat során nem találtam a tartalmak minőségében kifogásolnivalót.

- A tartalmakkal kapcsolatban csak a leírásukat tudom hiányolni: nem derül ki számomra egyértelműen, hogy mik is pontosan a tartalmak, amelyek között kereshetek.
- Az ékezetes betűk helyesen jelennek meg, kezelésükkel nem láttam problémát.
- A vizsgálat során végzett keresés eredményeit pdf formátumban olvashattam, más formátumú dokumentumra nem bukkantam. A pdf dokumentumok olvasásához szükséges szoftver letöltéséhez van segítség a felületen.

Metaadatok, bibliográfiai információk

A találati listákban elérhető adatlapokon az elvárható metaadatok többsége szerepel, a vizsgálat során kitöltetlen adatlapot nem találtam.

- A TDA csatlakozott a Nemzeti Digitális Adattár keresőjéhez, a metaadatait minősített Dublin Core metaadatsémában szolgáltatja. Ez a séma a honlapon is megtalálható, igaz, csak a nehezen értelmezhető XSD fájlban.
- Az adatlapok nem használják ki teljes mértékben a séma nyújtotta lehetőségeket, reméljük, ez a jövőben változni fog, és az adatlapokon egyre több információ szerepel majd.
- A honlapon nem találtam meg a TDA metaadatainak begyűjtéséhez szükséges információkat. Remélem, ezt hamarosan pótolja a TDA, hiszen csak így valósulhat meg nagyobb hozzáférés ezekhez az értékes társadalomtudományi tartalmakhoz.

A Magyar Társadalomtudományok Digitális Archívuma

Honlap: http://www.fszek.hu/adatbazisok/sajat_adatbazisok/

E-mail: reiszl@fszek.hu

Fenntartó: Fővárosi Szabó Ervin Könyvtár

Értékelés dátuma: 2006. október 31.

Értékelő: Drótos László

Általános értékelés

A Magyar Társadalomtudományok Digitális Archívuma (MTDA) célja az 1848–1948 közötti időszak

társadalomtudományi, szociológiatörténeti irányzatnak és gondolkodóinak bemutatása, a korszak fontosabb magyar – és a FSZEK gyűjteményében megtalálható – műveinek digitalizálása, online kutathatóvá tétele, illetve hosszú távú megőrzése. Kiemelt figyelmet fordítanak a „Huszedik Század” és köre klasszikus vitáira, valamint a magyar szerzőknek a „Társadalomtudományi Könyvtár”-ban megjelent írásaira, és a korabeli nőmozgalmak dokumentumaira. A könyv jellegű kiadványok mellett folyóiratok és fényképek digitalizálása is folyik. A megcélzott felhasználói kör elsősorban a tudományos kutatás és az oktatás; a szolgáltatás ingyenes. A szöveges dokumentumok automatikus karakterfelismeréssel, az eredeti oldaltördelés megőrzésével pdf állományokba kerülnek, és így tölthetők le. A gyűjtemény egy részét fejlett keresővel rendelkező szöveges adatbázis-kezelőbe is betöltötték, ahol html formátumban is olvashatók ezek a könyvek.

A tesztelés során talált három komolyabb hiányosság a következő:

1. Az OCR során keletkezett hibákat nem javították rendesen, így oldalanként néhány ilyen felismerési hiba előfordul. Sajnos a képként szkennelt könyvlapok sem a pdf, sem a html verzióban nem nézhetők meg, így az értelemzavaró hibák nem tisztázhatók, illetve a táblázatokban vagy a szövegben előforduló adatok hi-telessége sem ellenőrizhető.
2. Nincsenek részletes metaadatok, és nincs tematikus keresés vagy böngészés, csak egy szerzők szerint rendezett lista az alapvető kiadási adatokkal.
3. A projekt meglehetősen „szétdarabolódott” – nincs egyértelmű nyitólap és lehatárolható saját felület, és nincsenek összhangban az egyes részek: a tájékoztatókban említett könyvek száma és gyarapodási üteme nem egyezik a valóban szolgáltatottal, a szöveges keresőbe a könyveknek kevesebb mint fele van betöltve, ugyanígy az NDA-ban is csak egy részük kereshető vissza, a könyvtár saját OPAC-jában pedig szinte egyáltalán nem található meg a digitalizált verziók.

Összefoglalva: a célkitűzés nagyon jó, a digitalizálható dokumentumok körének lehatárolása egy tudományos szempontból igen értékes archívumot ígér. Ehhez azonban egyrészt hiteles digitális másolatokat kellene készíteni, másrészt növelni kellene a projekt láthatóságát, és felszámolni az elmaradásokat a metaadatok, illetve a szöveges adatbázis terén.

Részletes értékelés

Felület, használhatóság, segítség

- Nem egyértelmű a gyűjtemény neve: „A Magyar Társadalomtudományok Digitális Archívuma”, „Magyar Társadalomtörténeti Digitális Archívum” és „Magyar Társadalomtudományi Digitális Archívum”, illetve „Digitális Szociológiatörténeti Archívum”, „Szociológiatörténeti Digitális Archívum” és „A magyar szociológiatörténet klasszikusai” megnevezés egyaránt előfordul. A kezdőlap URL címét sem sikerült egyértelműen eldönteni, mivel az egész szolgáltatás be van ágyazva a FSZEK honlapjába (9. ábra), és csak az almenükből érhetők el az egyes részek. A beágyazottságnak előnyei is vannak, mert így rendelkezésre állnak az amúgy szépen megtervezett honlap állandó funkciói: a navigációs keret, az adott rovatra le-szűkíthető saját kereső, a betűméret nagyítása, a nyomtatási nézet stb. A felület egyetlen kicsit kellemetlen jellemzője, hogy az indokoltnál többször nyit új böngészőablakot, de ezt idővel meg lehet szokni.



9. ábra A projekt híreit tartalmazó oldal

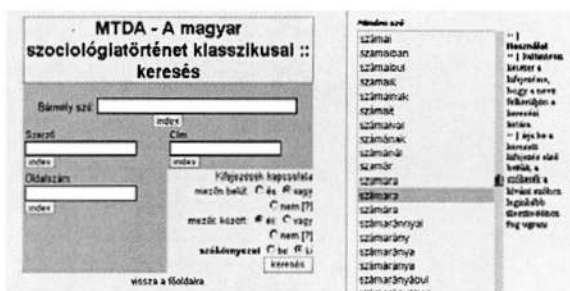
- A projekt céljáról, méretéről és állásáról, a gyűjtemény egyes részeiről rövid tájékoztató szövegek olvashatók. Sajnos egyiknél sincs dátum, így nem világos, hogy a bennük szereplő adatok mikorra érvényesek, például mikor került fel a 100. kötet, amelyről egy rövid hír tudósít. A gyűjtemény méretére vonatkozó adatok egyébként el-lentmondásosak – az ismertető 423 digitalizált könyvről beszél, de jelenleg csak 101 pdf állomány van az „MTDA E-könyvtár” nevű aloldalon. A honlap saját keresőjéhez (amellyel ezekben a fájlokban is keresni lehet) nincs semmilyen tájékoztató, és az adatbázisba töltött verziónál is minimális a használati utasítás az MTDA-úrlapon.
- Van viszont egy „Újdonságok” oldal, amely a gyűjtemény gyarapodásának figyelemmel kísé-résére lenne alkalmas. E cikk írásakor azonban csak egyetlen tétel árválkodott rajta, 2006. októ-ber 13-i dátummal.

- Bár a FSZEK honlapján van látássérülteknek szóló menüpont, de onnan a szolgáltatott adatbázisok nem érhetők el, így ez a digitális gyűjtemény sem. Persze a pdf könyvekkel a magyar felolvasóprogramok úgysem boldogulnának, de az adatbázisba töltött szövegek azért vakoknak is használhatók lennének; sajnos a NextPage kereső keretes felülete és kinyitogatható tartalomjegyzéke szintén nem kedvez a karakteres böngészőknek és a gépi felolvasásnak, az apró Garamond betűk pedig a gyengén látóknak okoznak nehézséget.
- A honlap angol felületén sem sikerült megtalálni ezt a társadalomtudományi archívumot. Ez részben érthető, mert a digitalizált források mind magyar nyelvűek, de éppen ritkaságuk miatt, valamint az archívum nyelvfüggetlen képi anyaga miatt legalább egy rövid idegen nyelvű ismertetőt megérne kitenni.
- Szintén a honlapba való beépülésnek köszönhetően a weblapok alján mindig ott szerepel a webmester levelezési címe, illetve egyes aloldalakon a gyűjtemény főszerkesztőjének elérhetősége is meg van adva, sőt kis szerencsével még a FSZEK adatbázisainak és digitális archívumának témájában indított fórumot is megtalálhatjuk (igaz, ebben az értékelés készítésekor még nem volt felhasználói visszajelzés). Az adatbázis-változatnál viszont nincs elérési cím megadva.

Kereső funkciók, találatok

- A pdf dokumentumokban a webszerver saját, teljes szövegű keresőjével találhatunk meg szavakat úgy, hogy a keresést az archívumot tartalmazó rovatra szűkítjük (10. ábra). (Hogy ez a rovat meddig terjed, milyen dokumentumok tartoznak bele és melyek nem, az sajnos nem derül ki egyértelműen.) A válaszdíó nagyon rövid, de a találati listában nincs szöveggörnyezet, és csak a találatokkal együtt jelenik meg az az „összetett” keresőmező, ahol ha mást nem is, de legalább annyit beállíthatunk, hogy a beírt szavak egymás mellett legyenek, vagy bárhol előfordulhatnak, illetve hogy bármelyik előfordulhat. Komolyabb keresésre ez nyilvánvalóan nem alkalmas.
- A gyűjtemény egy része (jelenleg 45 könyv) az Arcanum Adatbázis Kft. által üzemeltetett NextPage adatkezelő-rendszerbe is be lett töltve, így ezek szövege – többféle űrlappal is – kereshető a www.arcanum.hu/fszek/ oldalon. A találati listából kiválasztott dokumentum azután nemcsak html formátumban nézhető meg, hanem a szövegben elhelyezett oldalszámokra kattintva az eredeti tördelést megőrző pdf fájl is megjele-

nik – sajnos nem új ablakban, hanem az elég kis méretű keretben. A NextPage (korábbi nevén Folio) rendszer keresője a CD-ROM és online szakirodalom-kereséssel foglalkozó könyvtárak számára jól ismert. A laikusoknak azonban nem könnyű kiismerni a működését. Az egymazs gyorskereső és a „haladó”, valamint a „szakértői” szintű keresők mellett választhatunk egy „MTDA” nevű űrlapot is, amelyet kifejezetten erre az adatbázisra alakítottak ki. Itt szerző, cím és oldalszám mezők állnak rendelkezésre a teljes szövegben való keresés mellett. Mezőn belül és mezők között is lehetnek „AND”, „OR”, „NOT” műveletek, de a mezők nem ismételhetők, csokolni viszont – mindkét oldalról – lehet a * karakterrel. Mindegyik mezőnél van index is, amely a szavak kiválasztásában segít. Az indexablakra kattintva kezdetjük el a szó elejének begépelését, és ha türelmesek vagyunk, a lista néhány másodperc múlva a megadott helyre ugrik. Ha elrontottuk a gépelést, akkor kattintsunk a mező neve alatti „index”-gombra, vagy lapozzunk arébb a listában és kezdjük előlről. Ennél a „szókerék”-nek nevezett indexnél hosszú „ő” és „ű” betűket nem tudunk beírni, ezeket rövid vagy ékezet nélküli megfelelőikkel kereshetjük meg, de igazából nem érdemes ilyen szavakat választani az indexből, mivel rossz kódolással adódnak át az űrlapba, és így nem lesz találat.



10. ábra Az adatbázisba töltött dokumentumok részletes keresője

- A NextPage nagyon gyorsan adja a találatokat, és ezek azután kényelmesen végiglapozhatók, s a képernyő tetején levő ikonnal kinyitható az adott találat helyét mutató tartalomjegyzék is. Nincs azonban nyomtatási nézet sem a találati listához, sem a listából kiválasztott dokumentumhoz. (Érdemes ezért a Shift gombot lenyomva tartani, amikor valamelyik találatra kattintunk, hogy a dokumentum új ablakban nyíljon meg.) A gyorskereső kivételével kérhetünk szöveggörnyezetet is a találati listában, azonban a keresett szót vagy szavakat itt nem emeli ki rendszer.

Tartalom, minőség

- Az archívumnak jelenleg két gyűjteményrésze van: az egyik a magyar szociológiatörténet klasszikusainak digitalizált írásait tartalmazza, a másik az 1848–1919 között megjelent, a nők helyzetével és a feminizmussal foglalkozó kiadványok gyűjteménye. Utóbbiban a könyvek mellett *A Nő és a társadalom* c. lap egy évfolyama, valamint néhány – sajnos kis felbontású – szociofotó is megtalálható. A gyűjtemény magán-, illetve nonprofit célokra ingyenesen és szabadon használható.
- A dokumentumok szövegeit OCR programmal ismertették fel, majd az eredeti oldalelrendezést megőrizve, és a külalakot nagyjából utánóztatva pdf formátumba mentették őket (11. ábra). A szkennelt oldalképek nincsenek a fájlban, ami nagy kár, jó lett volna ezeket is megőrizni a pdf-ben, vagy legalább hozzákapcsolni őket az adatbázisból elérhető html változathoz. Arról nincs szó a projektet ismertető szövegben, hogy történelmi-e utólag javítás ezeken a felismertett szövegeken, és ha igen, akkor milyen alaposan. A megnézett dokumentumok mindenesetre olvasásra alkalmas minőségűek voltak, de oldalanként azért akad néhány tipikus OCR-hiba (pl. a *haboru_es_beke_kozott.pdf* 6. oldalán: „kéid&3”). Ezek miatt is, valamint az időnként előforduló – és szintén olvashatatlan – görög szavak miatt is szükséges lenne az oldalkép megtekinthetősége.



11. ábra Egy könyv pdf formátumban, és a találati lista

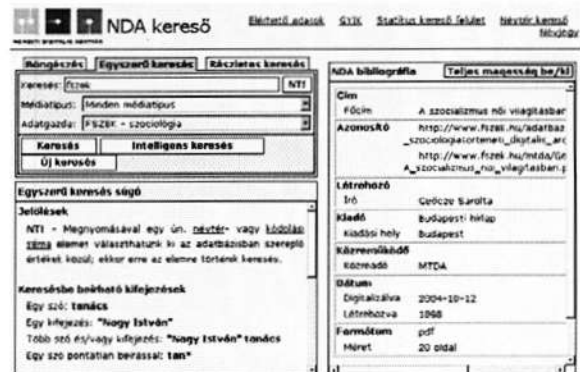
- A nyomtatás, másolás és megjegyzések hozzáférése nem volt letiltva a megvizsgált pdf-dokumentumokban, ami kifejezetten helyeseltető egy alapvetően tudományos kutatásra szánt digitális könyvgyűjteménynél.

Metaadatok, bibliográfiai információk

- A pdf fájlokban csak a szerző+cím van a „Title” mezőben, ezenkívül a digitális változat szerkesz-

tőjének neve és az „MTDA” rövidítés. A már letehető vagy a még szkennelés alatt levő, illetve digitalizálni tervezett könyvek listáját tartalmazó weblapon ezenkívül még a kiadási adatok olvashatók, valamint az egyes pdf dokumentumoknál a fájl mérete bájtban, és a szerverre való feltöltésének időpontja. Más metaadatot nem sikerült találni a honlapon sem géppel, sem emberi szemmel olvasható formában. A FSZEK saját OPAC-jában természetesen le vannak írva az eredeti nyomtatott könyvek, de a megvizsgáltak közül csak egynél volt csatoló a digitalizált verzióra, és a „számítógépfájl” típus alatt sem sikerült megtalálni az MTDA E-könyvtár darabjait.

- A Nemzeti Digitális Adattárba 2005. december elején betöltötték 45 tétel leírását: itt a szerző, cím, kiadás mellett a digitalizálás dátuma és az oldalszám van még megadva (12. ábra). Tárgyszavak vagy egyéb leíró adatok itt sincsenek. A gyűjtemény metaadatai az azóta eltelt közel egy év alatt nem frissültek, és nem bővültek az NDA-ban.



12. ábra Metaadatok az NDA keresőjében

Beérkezett: 2006. XI. 2-án.



Drótos László

az Országos Széchényi Könyvtár Magyar Elektronikus Könyvtár Osztályán főkönyvtáros. A Magyar Elektronikus Könyvtárért Egyesület elnökségi tagja. E-mail: mekdl@iif.hu



Somogyi Tamás

a Nemzeti Digitális Adattár informatikusa, közbeszerzési pályázatok bírálója. E-mail: somogyit@freemail.hu