

Adattármustra. Levéltári iratok

A sorozat egyes számaiban független szakemberek értékelik a magyar nyelvű internet könyvtári szempontból fontos tartalomszolgáltatásait megadott szempontok szerint; ez alkalommal a levéltári jellegű információforrásokat. A bírálatok a használók számára fontos kezelőfelülettel, keresőrendszerrel, tartalommal és metaadatokkal foglalkoznak. Az értékelések egyrészt segítséget nyújtanak a rendszerek minél eredményesebb használatához, másrészt felhívják a fejlesztők és karbantartók figyelmét a szolgáltatás hiányosságaira, a jövőbeli minőségi javítások érdekében.

Országgyűlési dokumentumok: naplók és irományok 1861–1990

Honlap: <http://mpgy.ogyk.hu>
E-mail: mpgy@ogyk.hu
Fenntartó: Országgyűlési Könyvtár
Értékelés dátuma: 2009. május 18.
Értékelő: Drótos László

Általános értékelés

Az Országgyűlési Könyvtár által létrehozott digitális dokumentumtár ünnepélyes megnyitója 2008 júniusában volt. Ekkor az 1861–1918 közötti parlamenti naplók és irományok váltak elérhetővé, vagyis több mint 300 ezer szkennelt oldal; ez a szám azóta már közel a duplájára nőtt és a záróév pedig 1990-ig tolódott ki. A sajtóhírekben megjelent tervek szerint a jövőben további évekkel és típusokkal – például jegyzőkönyvek, házszabályok, delegációk iratai, zárszámadások, költségvetések – bővül majd ez a szolgáltatás. A dokumentumok oldalanként különálló, kétrétegű PDF állományokban vannak: felül egy egybites (fekete-fehér) és mögötte a felismertett (javítatlan) szöveg. Keresni a teljes szövegben vagy az egyes kötetekhez készült mutatókban lehet. A szolgáltatásnak elég szokatlan, sokablakos kezelőfelülete van, ami az alkalmi felhasználóknak nehezen átlátható, megtanulható. Az egyes aloldalakon egy-két mondatos „tipp”-ek nyújtanak némi segítséget, de részletes útmutatót sehol nem találni. Jó tulajdonsága viszont a rendszernek, hogy a korábbi kereséseket megőrzi, így ezek később is előhívhatók és lefutathatók – igaz csak az adott gépről, mert ezek nem a szerveren tárolódnak.

A tesztelés során talált három komolyabb hiányosság a következő:

1. Nincsen rendes ismertető sem a digitalizált tartalomról, sem a felhasználói felületről, pedig a kívülállók (pl. újságírók, politikusok, jogászok) számára egyik sem magától értetődő, így könnyen lehet, hogy az „elveszettség érzése” miatt gyorsan otthagyják ezt a rengeteg munkával létrehozott, értékes szolgáltatást.
2. A speciális ablakkezelés nemcsak a rendszer átláthatóságát nehezíti, hanem az átlagfelhasználó számára elérhetetlenné válnak a böngészőjének megszokott funkciói: nem tud könyvjelzőt elhelyezni, találati listákat elmenteni vagy kinyomtatni, egyes ablakokat átméretezni vagy görgetni, betűméretet állítani, és azt sem látja, hogy éppen tölt-e valamit a kliense vagy csak rosszul kattintott és azért nem történik semmi. És természetesen egy ilyen felület a közintézményektől elvárt akadálymentes hozzáférésnek sem felel meg.
3. Különösen az irományok mutatói között böngészve elég gyakori, hogy megjelenik ugyan az iromány száma, de nem kattintható. Tartalmi ismertető vagy a projekt előrehaladását mutató adatok hiányában nem lehet tudni, hogy ennek az-e az oka, hogy az adott irat még nem került digitalizálásra (esetleg nincs is meg az OGYK-ban), vagy adatbeviteli hiba történt, ezért nem került rá a számra a hivatkozás. Fontos volna arra is felhívni a figyelmet, hogy nem minden időszakhoz készültek mutatók, így aki ezek alapján próbál keresni, az a teljes dokumentumhalmaznak csak egy részét találhatja így meg.

Összefoglalva: Nem könnyű feladat egy ekkora és elég kevésbé strukturált szöveghalmazt jól kezelhető és böngészhető adatbázissá alakítani

anélkül, hogy jelentős élömrunkát kelljen befektetni a metaadatolásba, illetve a szintén nyomtatott kötetekből digitalizálandó mutatók egységesítésébe és hibáinak javításába. Mindenesetre már az is nagy eredmény, hogy egyáltalán online hozzáférhetővé vált a magyar politika- és jogtörténetnek ez a jelentős szelete, és remélhetőleg még tovább is bővül majd a tartalom. Az erőforrások függvényében érdemes volna megfontolni egy lényegesen leegyszerűsített alternatív felület kialakítását is, vagy legalább egy helyi Google kereső beépítését a kezdőlapra azoknak, akik nem akarják/tudják megtanulni a jelenlegi komplex – és szinte egyáltalán nem dokumentált – felületet. Ezt azután nem volna nehéz angolra is lefordítani a külföldiek számára, mert bár a tartalom nagy része magyar, de vannak benne például nemzetközi szerződések is. A használhatóságot tovább javítaná, ha lenne külön személynév-index és egyszerűen lehetne év(ek)re szűkíteni a kereséseket.

Részletes értékelés

Felület, használhatóság, segítség

- Az Országgyűlési Könyvtár honlapján (www.ogyk.hu) balra fent egy kiemelt felirat hívja fel a figyelmet erre az adatbázisra, de aki a főmenüben próbálja megtalálni, az a „Könyvtári adatbázisok” alatt hiába keresgél, ehelyett az „e-Könyvtár” menüpontot kell választania. Valójában a szolgáltatás jellege miatt mindkét aloldalra érdemes volna bekötni. De legalább a nevét frissíteni az „e-Könyvtár” alatt, mert ott még mindig a régi címmel: „Képviselőházi naplók és irományok, főrendiházi naplók 1861–1918” szerepel.
- Az mpgy.ogyk.hu kezdőlapján (1. ábra) egy rövid üdvözlő és tájékoztató szöveg fogad, de ebben csak annyi plusz információ van a főcímmel képest, hogy mintegy 600 ezer oldalt tartalmaz az adatbázis. Nagyon hiányzik itt egy olyan gomb, amelyre kattintva részletes tájékoztatást kaphatna az érdeklődő, hogy milyen állományrész lett itt digitalizálva, melyek azok a naplók és irományok, mely részekhez vannak mutatók, mit jelentenek a rövidítések és kódok, mik az esetleges hiányok és a további tervek, hogyan vannak a dokumentumok PDF-ben tárolva, kik készítették a szolgáltatást, mikor és milyen ütemben bővül stb. De legalább a szolgáltatás megnyitásakor készült sajtó- és előadásanyagokra lehetne egy hivatkozás, például a www.ogyk.hu/legfr/hirek/20080611_mpgy.html oldalra, ahol több hasznos ismertető is össze van gyűjtve. Súlyos, FAQ, hírek, sajtóválogatás, impresszum, jogi nyilatkozat,

technikai követelmények, statisztika, linkgyűjtemény, üzenetküldés és egyéb, a nagyobb tartalomszolgáltatásoknál megszokott menüpontok helyett csak a kétfajta keresőhöz vezető gombok és egy – a háttérképbe eléggé belevesző – „Kapcsolat” felirat fogadja a felhasználót. Utóbbi egy e-mail címet és a könyvtár személyes meglátogatását ajánlja, de még innen sincs egy ugrópont a könyvtár honlapjára, vagyis az mpgy.ogyk.hu olyan szigetként létezik a kibertérben, ahonnan nincs kivezető út.



1. ábra Az adatbázis nyitólapja

- Az oldal külalakja a témához illő, archaizált felület, tiszta és következetesen végigvitt megjelenéssel; bár a design a korábbi keresések ablakában azért megtörik, mert ott már a böngésző saját nyomógombjai jelennek meg. Nagy felbontásban a círádás háttérkép szemfárasztó lehet, különösen azért, mert az oldalak általában csak 790 pont szélesek (ami azt eredményezi, hogy éppen nem férnek el egy 800x600-as képernyőn a görgetősáv és a margó miatt), így sok látszik a tapétából és kevés hely marad az érdemi tartalomnak. A kereső és böngésző ablakok mérete sem lett eltávolítva, mert ezeket meg a leggyakoribb 1024x768-as felbontásnál függőlegesen kell kicsit tologatni. Az indexablakoknál a Firefox nem teszi ki a fel-le görgetősávot, így kisebb felbontásban ezek alja egyáltalán nem látszik: vagyis lehetetlen az indexek lapozása, a korábbi keresések ablakánál pedig vízszintesen van ugyanilyen probléma. Más kompatibilitási hibát nem találtam, az adatbázis mindkét elterjedt böngészővel használható (ellentétben az OGYK saját honlapjával, amelyhez külön Explorerablaknyitó modult kell letölteni a Firefox használóknak).

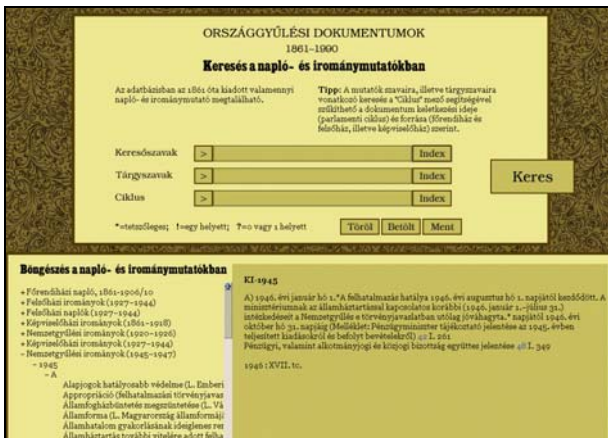
- Az oldalak többnyire keretből állanak, amelyek nem méretezhetők át és vizuálisan is összefolyanak, ha nincs éppen köztük a görgetősáv. Így a mutatókat böngésző felhasználó ideje nagy részét a keskeny frame-ben szorongó, sokszor nagyon hosszú sorok vízszintes tologatásával tölti – hiába van akármilyen nagyfelbontású monitorja.
- A kattintható hivatkozások színe zöld és kék (hogy melyik szín mit jelöl, arról nincs tájékoztatás), a találatokban a keresőszavak pirossal vannak kiemelve, úgyhogy ezek jól láthatók a zöldesbarna háttéren. Nem jó viszont, hogy a hivatkozások megtekintés után nem váltanak színt, így nem lehet követni, hogy mit néztünk már meg; és az sem jó, hogy a találati listákban a sorszámok nincsenek átszínezve, emiatt nehéz rájönni, hogy azokra kell kattintani és nem a találatok szövegére.
- Szinte minden kattintás új ablaknyitást eredményez (akkor is, ha csak új fület/lapot állítottunk be erre az esetre a böngészőnkben), ezért néhány perces keresgélés után zavarbaejtően sok ablak sorakozik a képernyőnk alján, a tálcán. Bár ennek a javascriptes megoldásnak megvan a maga logikája és szerepe, továbbá megnyitható egy külön ablakkezelő ablak is, de az alkalmi látogató nem látja át ezt a rendszert és gyorsan elvész az ablakok sorozatában. Ráadásul az indexablakok maguktól bezáródnak, ha kikerülnek a fókuszból. Az átlagfelhasználó kellemetlenségeit tovább növeli, hogy ezek az új ablakok cím-, menü-, ikon- és státuszsor nélküliek, ezért a megszokott funkciókat sehol sem találja (hacsak nem ismeri a jobb gombos menüket és a billentyűparancsokat), vagyis nem tud nyomtatni, menteni, visszalépni, URL címet kimásolni stb. Ezeket a veszteségeket a felület csak részben pótolja, mert visszalépésre, bezárásra és zárolásra (az adott ablak felülírás elleni védelmére) vannak ugyan néhol saját gombok, de a többi funkcióra nem. Mindig nagy az ára annak, ha lemondunk a böngészők beépített képességeiről és a felhasználókat újfajta navigáció és funkciók megtanulására kényszerítjük, de ilyen esetben legalább egy jól megírt súgóval segítsünk nekik. Sajnos a felület használatához semmilyen segítség nincs, még lebegő súgók sincsenek a gombokhoz. Az egyetlen fellelhető Súgó gomb a korábbi keresések ablakában van, de az egy üres képernyőt nyit meg.
- Maguk a dokumentumok 1 bites képként vannak eltárolva, minden oldal külön, egy-egy – védelem nélküli – PDF fájlban. A fekete-fehér oldalképek nem szépek (fax-minőség), de legalább gyorsan

letöltődnek és általában az olvashatóság határa fölött vannak, viszont a dokumentumok eredeti külalakjának visszaadására és professzionális célokra természetesen nem alkalmasak. (Érdeemes lenne néhány dokumentumot vagy legalább oldalt mintaként színesben, jobb minőségben is feltenni az adatbázis honlapjára, hogy az emberek lássák, hogyan is néznek ki ezek a valóságban.) Az egyoldalas megoldás csak a letöltési idő szempontjából jó, de a felhasználó így állandó lapozásra és rövid várakozásra kényszerül, és ha úgy van beállítva a Reader a gépén, hogy a böngészőbe ágyazva jelenjen meg, akkor egy nem teljesképernyős keretben kell görgetnie a dokumentumot, ráadásul a számára kényelmes nagyítás a következő lapozásnál visszaáll az eredetire, úgyhogy a képernyőn való hosszabb olvasásra kevesen vállalkoznak ilyen körülmények között. Ha túllapozunk egy kötet végén, akkor vagy egy „File doesn't begin with %PDF” vagy egy „nincs kép” üzenet jelenik meg. Mivel minden oldal külön töltődik le, a kötetben való navigálás, illetve egyáltalán egy megtalált iromány terjedelmének felmérése is körülményes – egy hosszabb dokumentum kinyomtatásáról nem is beszélve. Alternatív megoldás lehetett volna (legalább opcionálisan) a kötetek egészének vagy nagyobb egységeinek PDF-be mentése és az egyes oldalakra való hivatkozás *URL/fájlnev.pdf#page=oldalaszám* formában való megoldása, vagy pedig egy online PDF megjelenítő beépítése a felületbe.

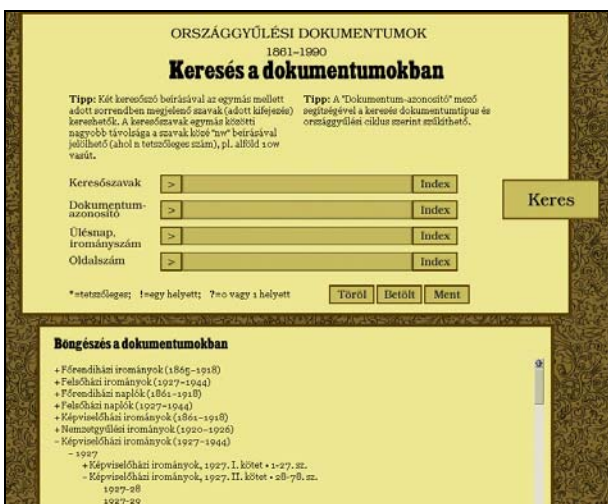
Kereső funkciók, találatok

- Kétfajta keresési és egyben böngészési lehetőség van, ezek külön ablakokban nyílnak meg és már az *nfo.arcanum.hu* szerverről jönnek. A mutatók űrlapján *keresőszavak* és *tárgyszavak* szerinti keresés és ciklus szerinti szűkítés, valamint ciklusok és azon belül betűrendben való böngészés választható (2. ábra). Hogy mi a különbség, illetve kapcsolat a keresőszavak és tárgyszavak, valamint a böngészhető mutatók között, azt a laikus nehezen deríti ki: de úgy tűnik, hogy valójában a *mutatócímek* maguk a *tárgyszavak*, és ezek szavankénti indexe jelenti a *keresőszavakat*. Lehetett volna szerencsésebb elnevezéseket választani ezen az űrlapon, mert a *tárgyszó* és a *keresőszó* alatt mást szoktunk érteni (előbbi egy teauruszból vett deskriptor, utóbbi pedig a felhasználó által begépelte keresőkérdés egy szava). A teljes szövegű űrlapon is egy *keresőszavak* sor, valamint – elsősorban szűkítés céljából – dokumentumazonosító, ülésnap/irományszám és oldalszám mezők vannak (3. ábra). Itt nagyon

hiányzik az év szerinti szűkítés lehetősége, amely az azonosítóból nem szűrhető ki, tekintve hogy az csak a ciklusok kezdő évszámát tartalmazza. Ezen a lapon a keresés mellett kötetenként és azon belül oldalszámoként tudunk böngészni.



2. ábra A mutatók kereső- és böngészőlapja



3. ábra A dokumentumok teljes szövegű kereső- és böngészőlapja

- A böngészés a sorok előtti kis + és - jelekkel történik, ezekkel – a nem kézenfekvő és az egérrel nehezen eltalálható – jelekkel nyithatjuk és zárhatjuk az egyes ágakat (nagyobb ágak kinyitására néha várni kell pár másodpercet). Nem szerencsés, hogy a kétféle űrlapon a legfelső szintű ágak sorrendje nem azonos, így nehezebben tűnik fel az is, hogy nincs minden dokumentumtípushoz és időszakhoz mutató, vagyis hogy a mutatók űrlapjával csak egy részhalmazt érünk el. Az abc-sorrendű böngészésnél a számok a

betűk között vannak (pl. a 100.800/1945. F. M. sz. *rendelet* az SZ betűnél).

- A keresőmezőkhöz indexablakok (4. ábra) tartoznak, ezekhez pedig szókezdet és szórésztel segédmezők, melyek nagyon hasznosak, tekintve hogy esetenként hatalmas indexekről van szó, amelyekben nem könnyű navigálni. Különösen, hogy a lapozógombok viszont nem sokat segítenek: az első lapozás után ugyan egyszerre két „Következő 100” feliratú gomb is megjelenik (majd később néha mindkettő eltűnik valami programhiba miatt), de csak egyik működik, a másik az index elejére visz, és visszafelé lapozni vagy nagyobb lépésekben ugrálni egyáltalán nem lehet. A dokumentumazonosítók indexében az 1-es szám után jönnek a 10-es sorszámúak, a 2-es után a 20-asok stb., vagyis nem numerikusan, hanem abc-ben vannak rendezve. A keresőszavak indexe persze tele van hibával, mert hogy javítatlan OCR-es szövegből készült; elég például a *hallj* kezdetű szavakat megnézni az indexben, hogy lássuk, hányféleképpen fordul elő a jegyzőkönyvekben gyakori *Halljuk!* felkiáltás.



4. ábra A mutatók indexablaka

- A kereső működéséhez mindössze annyi segítség van, hogy a beírt szavak kifejezésnek számítsanak, de írhatunk közéjük egy w operátort (amivel a szavak távolsága adható meg), továbbá használhatók a *, ! és ? jelek (tetszőleges számú, pontosan 1, illetve 0 vagy 1 karakter helyettesítésére). Valójában a kereső ennél sokkal többet tud, ismeri például a Boole-műveleteket és zárójellezni is lehet: például az *(orosz or szovjet) 4n hadifog** keresőkérdés is működik (itt a 4n max. négyzónyi távolságot jelent,etszőleges sorrendben). Egy ekkora teljes szövegű adatbá-

zsnál nagyon fontos lenne a kereső szintaxisának pontos leírása, mert csak annak ismeretében lehet bonyolultabb kereséseket hatékonyan elvégezni.

- A kereső jó tulajdonsága, hogy automatikusan megőrzi a korábbi kérdéseket (egy-egy cookie-ban tárolva a felhasználó gépén) és ezek egy kisablakban átnevezhetők, törölhetők vagy újra átnevezhetők az űrlapba. Vannak azért furcsaságok ebben a kisablakban is, mert nem mindig frissül magától; és jó volna, ha időrendben vagy betűrendben lehetne rendezni a keresőkérdéseket.
- A találatok rendszerint gyorsan megjelennek, de ezek sorrendje nem világos és nem állítható be (5. ábra). Számuk a lista tetején látható, elég szerencsétlenül kiírva: *Találatok: 37 találat*. Viszont az indexablakokhoz képest egy sokkal ügyesebb, kétszintű lapozómenü van hozzájuk, amivel egy nagyobb halmaz is könnyen bejárható. Kevésbé jó, hogy hiába kattintunk ösztönösen egy tételre a listában, valójában a szöveggel azonos kinézetű sorszám van az a hivatkozás, ami a dokumentumhoz visz. És azt sem látjuk, hogy mely találatokat néztünk már meg, mert az ugrópontok nem váltanak színt. A keresett szavak viszont jól láthatóan ki vannak emelve a listában (6. ábra), sőt még magukon a PDF képalapokon is (itt sajnos olvashatatlanul, és ezen csak a Windows beállításai között, az „Asztal/Kijelölt elemek” színének átállításával tudunk segíteni).

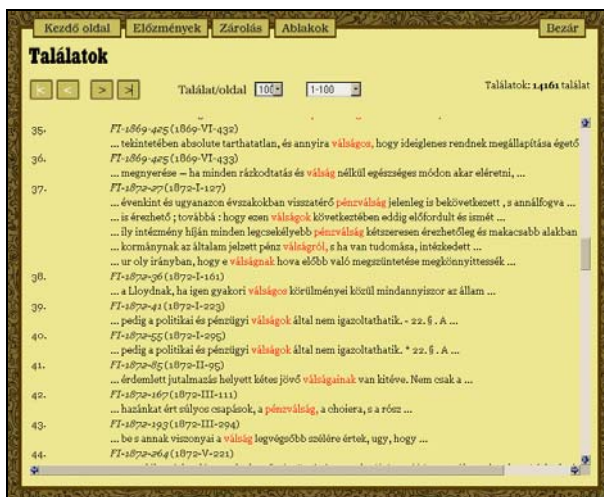


5. ábra A mutatók egyik találati listája és az ablakkezelő ablak

Tartalom, minőség

- Az adatbázis az 1861 és 1990 közötti időszakból tartalmaz digitalizált dokumentumokat: főrendházi, felsőházi, képviselőházi, nemzetgyűlési, or-

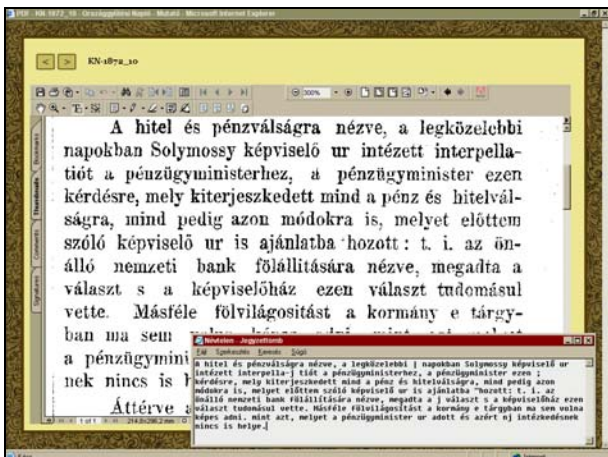
szággyűlési naplókat és irományokat, a lehető leghitelesebb forrásból: az Országgyűlési Könyvtár állományából. Hogy ez a lefedett időszak és dokumentumhalmaz mennyire teljes, vagyis hogy vannak-e hiányok az eredeti gyűjteményben vagy a digitalizált anyagban, az sajnos nem derül ki, mint ahogy az sem nyilvánvaló a külső felhasználóknak, hogy nem minden kötethez készültek mutatók. A nyitólapon levő tájékoztató csak annyit közöl, hogy mintegy 600 ezer oldalt digitalizáltak, ez nagyjából helyes is, mert a dokumentumkereső űrlapon egy * karaktert beírva a tesztelés időpontjában 587 412 találat volt (és ebben még a mutatókötetek feltehetően nincsenek is benne). Viszont az a tény, hogy az irományok mutatóiban böngészve elég sok tételnél nincs kattintható hivatkozás a teljes szövegre, vagy azt jelzi, hogy nem teljes a (digitális) állomány, vagy hibák történtek a linkelésnél.



6. ábra A teljes szövegű kereső egyik találati listája

- A teljes szöveg és a mutatók is OCR-es, vagyis géppel felismertetett szöveget tartalmaznak, ezért soronként 1-2 hibával számolni kell, erre figyelmeztetnek is a feliratok, de csak az indexablakokban, a keresőűrlapokon nem. Hogy mennyire hibás egy dokumentum szövege, azt vágólapon át való kimásolással (7. ábra), vagy a legtöbb Readerben a Reflow ikon (Ctrl/4 billentyű) megnyomásával ellenőrizhetjük. Ezek a hibák természetesen befolyásolják a keresések pontosságát. Hogy a keresés szempontjából fontosabb mutatóköteteket valamennyire javították-e a karakterfelismerés után, azt szintén nem tudni, de a helyesíráellenőrrrel gyorsan lehet találni ezekben a tárgyszóindexekben is hibákat (pl.

felhatalmazás, státusrendezése, öntözésrol, Zálólevelek, Arpád, Barankovics stb.).



7. ábra Egy oldalkép részlete PDF-ben és mellette a kimásolt OCR-es szöveg

Metaadatok, bibliográfiai információk

- Az mpgy.ogyk.hu oldal fejlécében csak egy <title> sor van metaadatként, az is csupán ennyit tartalmaz: *Országgyűlési dokumentumok*. Az adatbázis teljes neve csak képként van a kezdőlapra, ezért is lehet az, hogy a Google nem az első találatok között adja a „naplók és irományok” kifejezésre keresve.
- A PDF dokumentumok szintén csak egy Title adatot hordoznak a metaadat blokkban (Ctrl/D gomb az Adobe Readerben), de abban korrektül benne van az eredeti forrás, például *Nemzetgyűlési irományok, 1920. XI. kötet • 320–362., XLIX. sz.*, ami nagyon fontos, tekintve, hogy mindegyik PDF csak egyetlen oldalképet tartalmaz, amelyen nem látszik, hogy melyik kötet melyik részéből származik. A forrás a fájlnevből sem derül ki (az mindig *a.pdf*), és ha a mutatókban keresünk, akkor a webfelületen is csak a kötetkódok látszanak (pl. *KI-1920_11*) mind a találati listában, mind pedig a dokumentummegjelenítő ablakban, nem pedig azok részletes feloldása. Ráadásul a találati listában még ezek a kódok sem minden tételnél jelennek meg, ami végképp megítélhetlenné teszi, hogy egy adott találat milyen típusú dokumentumból és melyik időszakból származik.
- A közel 60 ezer PDF fájlból a Google csak néhány százat talál meg, a többi a „mélyweb” részét képezi. Ezen olyan statikus HTML link-listák létrehozásával lehetne segíteni, melyeket bejárva a keresőgép robotja könnyen és biztosan rá talál valamennyi PDF-re.

Budapest Főváros Levéltára Központi Kereső, Társadalomtörténeti Adattár

Honlap:

http://lear3.bparchiv.hu/kozponti/simplese_arch.html

E-mail: bfl@bparchiv.hu

Fenntartó: Budapest Főváros Levéltára

Értékelés dátuma: 2009. május 9–17.

Értékelő: Körös Kata

Általános értékelés

Budapest Főváros Levéltára (BFL) kiemelkedő helyet foglal el a magyarországi levéltárak sorában. Önkormányzati levéltárként gyűjtőkörébe elsősorban a fővárosi köziratképző intézmények (minisztériumok, önkormányzatok, oktatási és jogszolgáltatási intézmények, egészségügyi és szociális intézetek, közigazgatási szervek, pénzügyintézetek) dokumentumai tartoznak, emellett iparvállalatok, egyházi szervek, családok és személyek anyagait őrzi legnagyobb részt a 19-20. századból. Az intézmény meghatározó szerepet vállal a magyar levéltárügy informatikai fejlesztéseiből is. 21 ezer folyóméternyi iratanyagához közel negyven adattárat, elektronikus segédletet készítettek az elmúlt években, amelyek egy részét napjainkig dinamikus fejlesztik. A kutatási segédletek többsége Microsoft Access adatbáziskezelővel készült, elsősorban belső használatra, de a levéltár szolgáltatásai között egyedi fejlesztésű adattárakat is találunk, amelyeknek egységes központi adatbázisba való integrálása folyamatban van. Ezek közül jelenleg öt érhető el a levéltár honlapján: a *Registrum*, a budapesti *Törvényhatósági Bizottság* jegyzőkönyvei, *Klősz György* városfotói, a *Budapesti lakcímejegyék 1900* és a *Társadalomtörténeti Adattár*.

A *Registrum* (bparchiv.hu/magyar/registrum/hu_bfl.html) a BFL saját fejlesztésű, ISAD/G (*General International Standard Archival Description*) alapú nyilvántartási rendszere, amelyben a levéltári dokumentumok strukturált leírása a fondfőcsoporttól egészen a tétel szintig megvalósítható. Az adatbázis tartalmazza az iratok fond- és állagnyilvántartását, valamint raktári jegyzékeit.

A Budapest főváros törvényhatósági bizottsága közgyűlési jegyzőkönyveinek adatbázisa (nfo.arcanum.hu/szkj) a bizottság üléseiről 1873 és

1949 között született jegyzőkönyvek teljes szövegét teszi elérhetővé PDF formátumban. Az 1881 előtti jegyzőkönyveket mai helyesírás szerinti átíratásban, a későbbi dokumentumokat az eredeti nyomtatott jegyzőkönyvek szkennelt oldalaként találjuk meg az adatbázisban.

A Klósz György által 1870 és 1910 között készített fényképfelvételek archivált gyűjteménye a fotókra jellemző legfontosabb paraméterek rögzítésével vált kereshetővé egy online adatbázisban (klosz.bparchiv.hu/klosz/app/index.fm).

A Budapesti lakcímgjegyzék 1900 (www.bparchiv.hu/cgi-bin/lakas/lakas.pl) a Budapesti cím- és lakjegyzék, illetve a Franklin-társulat által megjelentetett Lakás-jegyzék alapján készült adattár az 1900-as év budapesti lakosait cím- és névadatok, továbbá foglalkozás, rang és nemesi előnév alapján teszi kereshetővé.

A levéltári adatbázisok adatállományainak egységes szerkezetbe foglalását hivatott megvalósítani a LEAR-projekt, amelynek keretében a Társadalomtörténeti Adattár már hozzáférhető. Az integrált adatbázis az iratanyagoként eltérő adattartalmú metatárakra épül, amelyekhez külön objektumtárakban kapcsolódnak a dokumentumok digitális verziói. Az adatok közötti eligazodást különböző névterek segítik (személy- és családnevek, szervezetek neve, tárgyszavak).

A Társadalomtörténeti Adattár (Budapest Főváros Levéltára Központi Kereső) tesztelése alapján az alábbi fejlesztések ajánlhatók:

- Elengedhetetlennek tartom a projekttel kapcsolatos dokumentáció közzétételét, a tartalmat bemutató szövegek megírását, tájékoztatást a rendszer fejlesztési terveiről. Ebben a formában a felhasználó nem tudja, milyen elvárásokkal fordulhat a szolgáltatás felé.
- Legalább ennyire fontos volna egy keresési útmutató kidolgozása, a legfontosabb levéltári fogalmak, illetve elsősorban a keresésnél használható kategóriák, lehetőségek részletes, példával ellátott ismertetésével.
- A tesztelés során többször lépett fel szerveroldali hiba, mely megnehezítette az értékelést. Javítani kellene a rendszer üzembiztonságán.
- Az adatok minőségének ellenőrzése, különösen ilyen nagy mennyiségű adat esetén, körültekintést igényel. Egységesített tárgyszó- vagy kifejezésjegyzék segíthetné a magas színvonalú adatbevitelt, és kiküszöbölné a helyesírási hi-

bákból, rövidítésekből fakadó következetlenségeket.

- A kereső funkciók következetlen működése nehezíti a releváns információ megtalálását. Az operátorok sztenderd használhatósága és a keresőmotor optimalizálása megszüntetné ezt a problémát.

Jelenlegi állapotában a levéltár központi keresője ígéretes vállalkozásnak tűnik, de a keresési segédletek hiánya és a használat bonyolultsága miatt elsősorban olyan felhasználók számára megfelelő, akik pontosan tudják, hogy milyen dokumentumot keresnek. Ahhoz, hogy a rendszer szélesebb körben is használhatóvá váljon, olyan alapvető útmutatókkal kellene ellátni ezt az eleve kutatási segédletnek készülő keresőrendszert, amelyek a rendszert nem ismerő felhasználók számára is átláthatóvá teszik a szolgáltatást.

Részletes értékelés

Felület, használhatóság, segítség

- A Budapest Főváros Levéltára honlapjáról (www.bparchiv.hu) az adatbázisok menüpont alatt a *Társtört. adattár* címre kattintva érhetjük el a *Központi Kereső* nyitóoldalát (8. ábra). A levéltár honlapján sem a LEAR-projektre, sem a Társadalomtörténeti Adattárra vonatkozóan nem találunk érdemi információt. Kizárólag a *Publikációk* menüpont alatt levő, *Levéltári adatbázisok összefoglaló elemzése* című PDF dokumentumba betekintve van néhány adat a projektről és a levéltár eddigi informatikai fejlesztéseiről. Csak ebből a dokumentumból derül ki az is, hogy a Társadalomtörténeti Adattár egy nagyszabású terv első állomása. A Központi Kereső nyitóoldalon már nem szerepel a Társadalomtörténeti Adattár elnevezés, és az adatbázisról szóló további információk között sem találunk semmilyen utalást erre az összefoglaló névre. A nyitóoldalon mindössze az olvasható, hogy a kereső egy elektronikus levéltári segédlet, amely különböző szempontok szerint tesz kutathatóvá-hozzáférhetővé iratokat. A Központi Kereső cím ezért teljesen félreérthető. Ezen az oldalon semmi sem jelzi, hogy milyen fejlesztési tervei vannak az adatbázis készítőinek. A *További információk* címre kattintva egy legalább két éve nem frissített szöveget olvashatunk, amely vázolja az adatbázis tartalmát, utalva a részadatbázisok eltérő feltöltöttségi szintjére és a folyamatos fejlesztésre, de nem tájékoztat az összes elérhető tartalomról.



8. ábra A Budapest Főváros Levéltára Központi Keresőjének nyitóoldala

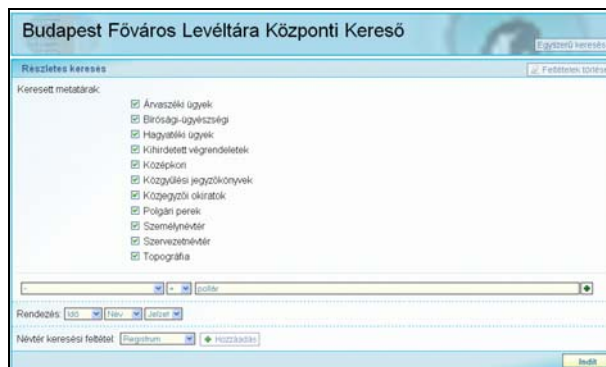
- A Központi Keresőhöz tartozó oldalak megjelenítése egyszerű, kevés grafikai elemet tartalmaz, jól áttekinthető. Nem készült viszont akadálymentes változat a keresőfelülethez, és a magyar nyelvű felületen kívül nincs más nyelvű információ az adatbázisról. A kereséshez nem áll rendelkezésre súgó, a felhasználó csak találgatni tudja az egyes funkciók működését, a listákból kiválasztható kategóriák jelentését.
- A kereső nyitóoldaláról hivatkozás vezet a BFL központi oldalára, felhasználói visszajelzésre csak itt van mód. Az adatbázissal kapcsolatos kérdésekre nem tartanak fenn külön e-mail címet, de a központi címre küldött kérdésekre gyors és udvarias válasz érkezik.

Kereső funkciók, találatok

- A BFL Központi Kereső nyitóoldalán egyszerű keresés áll rendelkezésünkre. A részletes keresés egy újabb kattintással érhető el erről az oldalról. A keresőürlapok közötti váltást a felső menüsorban állandóan megjelenő *Egyszerű keresés* és *Részletes keresés* címkék teszik lehetővé.
- Az egyszerű keresés során a beírt kifejezést a keresőmotor az összes adatmezőben keresi. A rendszer a keresőkifejezések között automatikusan ÉS kapcsolatot feltételez. Pontos kifejezésre idézőjelek segítségével tudunk keresni. Első ránézésre használhatunk Boole-operátorokat angol vagy magyar nyelven, de a tesztelés azt mutatja, hogy a rendszer nem működik következetesen: a találati lista nem lesz azonos az egyenértékű jelek, kifejezések használatánál. A „*Fodor András*”, *Fodor+András* és a *Fodor|András* kifejezésre is ugyanaz a 9 rekord a találat; a *Fodor András* keresésre 10 152, a *Fodor and András* kifejezésre 10 187, a *Fodor és András*-ra 10 152, a *Fodor*

vagy *András*-ra 10 936, a *Fodor or András* kifejezésre viszont 10 156 találatot kapunk.

- A keresőkifejezésben helyettesítő karakterek használata megengedett: ?-lel és *-gal egyszerre egy karaktert helyettesíthetünk. A kifejezések csonkolására nincs mód, bár bizonyos keresőkérdéseket automatikusan csonkoltnak tekint a kereső. Következetlen működésre utal, hogy a *Papa* keresőkérdésre megkapjuk a *Papa* vezetéknévű személyek mellett a *Papák* nevéket is, de a *Papak* keresésre már nincs találat. A keresőrendszer a kis- és nagybetűk között nem tesz különbséget, ékezetek nélküli keresésre legtöbbször nincs mód.
- A részletes keresés űrlapján külön is, vagy tetszőlegesen kijelölve kereshetők az egyes részadatbázisok adatállományai (9. ábra). A keresőmezők száma egy kis ikon segítségével szabadon növelhető, a keresést így nagyon specifikus szempontok kombinálásával szűkíthetjük. A keresőmezők összekapcsolásánál az űrlapon ÉS, VAGY és ÉS NEM operátorok közül választhatunk.



9. ábra A részletes keresés űrlapja

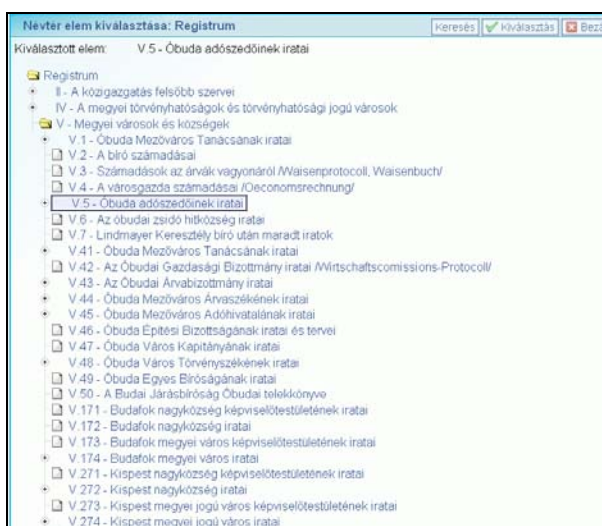
- Kereshetünk különböző azonosítók: a metatár neve, az ügyazonosító, a rekord egyedi azonosítója, vezérnév, ügyszám, jelzet, név és idő alapján. Megadhatjuk az ügy kiterjedésére vonatkozóan a résztvevők születési, illetve halálozási dátumát, a születés helyét, az ügyben érintett ingatlan budapesti helyrajzi számát, az érintettek lakhelyét, az esemény időpontját. Kereshetünk valásra, foglalkozásra, ügytípusra, haláleset-felvételi jegyzőkönyvre, tárgyszavakra, státuszra, hagyatéki leltárra, végrendeletre, pertárgyra, illetve perértékre. Megadhatjuk a kapcsolódó rekord azonosítóját, a leírás dokumentumát. Kereshetünk a dokumentum terjedelme vagy formátuma szerint, az iratanyag nyelvére, fondszámra, az őrző levéltár nevére, tartalomra, megjegyzés-

re és többek között sok más, a rekord létrehozásával kapcsolatos technikai adatra. Széles skálája ez a keresési lehetőségeknek, és jól strukturált adatállományt, mélyen feltárt dokumentumokat sejtet, azonban keresési segédlet nélkül a sok lehetőség kihasználására alig van esély. Nem egyértelmű, hogy milyen kifejezések kerülhetnek az egyes adattípusokba. Példa hiányában nem lehet eldönteni, hogy a *halálesetfelvételi jegyzőkönyv*, vagy a *hagyatéki leltár* címszavak alatt mi kereshető. Hogy nem egyértelmű listával találkozunk itt, arra utalhat magában a lenyíló listában az egyes kifejezések után található kérdőjel. A keresések alapján még az *Esemény időpontja* és a *Keletkezés dátuma* sem egyértelmű, hiszen rengeteg dokumentumnál ez utóbbi a korábbi.

- Bizonyos keresőkérdéseknél használható relációs operátor, például a nehezen meghatározható terjedelempre vonatkozó kategóriában, máshol viszont, ahol logikusnak gondolnánk, például idő megadásakor, a rendszer hibaüzenetet ad: „Az operátor ezen minősítő esetén nem megengedett.” Sokféle hibaüzenetet kaphatunk, ha a keresőkérdés nem megfelelő. Egy részük iránymutatásnak jól használható. Például: „A dátum formátuma nem megfelelő. Elvárt formátum: éééé.hh.nn”. Máskor viszont érthetetlen okból kapunk különböző hibaüzeneteket. Ilyen például: „A kért művelet végrehajtása sikertelen. A keresés adatai elvesztek. Adja meg újra a feltételeket!” Vagy ha egy több ezres listában próbálunk lapozni, gyakran kapjuk azt a hibaüzenetet, hogy „A program futása során hiba történt. Hiba az eredmény feldolgozásakor”, amelyet a részleteket megtekintve egy hosszú hibakódlista követ. A rendszer láthatóan nem képes egyszerre kezelni túl sok találatot.
- A részletes keresés űrlapján megadhatjuk a találati lista rendezési sorrendjét. Három adattípus szerint rendezhetjük egyszerre a rekordokat, *Név*, *Jelzet*, *Idő*, vagy *Fond* kategóriák közül választva. A rendezés mindig növekvő sorrendben lehetséges. A lista utólagos átrendezésére már nincs módunk.
- A kereséshez hozzáadhatunk még tetszőleges számú „Névtér keresési feltételt”. A *Személynévtér*, *Szervezetnévtér*, *Tárgyszónévtér*, *Objektumtár* és a *Registrum* elemei emelhetők be így a keresésbe. A legördülő listából kiválasztott tárat hozzáadva – egy új ablakban – az adott névtér elemei között kereshetünk. Sajnos a hiányos dokumentáció miatt a laikus felhasználó nem is

sejtheti, mire kell keresnie például az *Objektumtárban*. Sokkal könnyebb volna, ha ezeket a kereséseket lenyíló listák is segítenék, például, ha az entitás típusának egységesített neveit a *Szervezetnévtérben* egy indexből lehetne kiválasztani.

- A *Tárgyszónévtér* és a *Registrum* elemeit a felső sarokba kattintva hierarchikus listában is kereshetjük. Előbbiek esetében a lista valójában nem hierarchikus, pusztán egy betűrendes jegyzéke a használt tárgyszavaknak, ráadásul hosszúsága nehezíti a tájékozódást. Egy – akár ábécé szerint – oldalakra tördelt megjelenítési mód segíthetne eligazodni. Jól használható viszont a *Registrum* logikusan felépített struktúrája, amellyel a keresést az iratképzők és iratok típusai alapján, fond- és állagszinten lehet szűkíteni (10. ábra).



10. ábra A *Registrum* keresési feltételként beemelhető elemeinek hierarchikus listája

- A találati listákban az adattár neve, jelzet, név, az irat típusára vagy tartalmára vonatkozó kifejezések és a keletkezés ideje jelennek meg (11. ábra). A lista megjelenítésén nem változtathatunk. A keresési előzmények és a találati lista elmentésére nincs lehetőségünk, pedig egy jegyzékek, listák készítésére, kiválasztott rekordok tárolására, mentésére, exportálására alkalmas funkció felhasználóbarátabbá tehetné a keresőrendszert. Problémát okoz a hosszabb találati listákban a tájékozódás is. Egy oldalon egyszerre csak 50 találatot látunk, de nem tudunk a találati lista adott oldalaira ugrani.

Budapest Főváros Levéltára Központi Kereső							
						Egyszerű keresés	Részletes keresés
Találatok száma: 14							
<	Adattár neve	Jelzet	Név	Tipus/tartalom	Hely	Idő	
1.	Árvaszéki ügyek	02075	Lanz Györgyné ifj.	hagyatéki ügy		1878	Megtekintés
2.	Árvaszéki ügyek	02793	Bohus József	hagyatéki és örökbefogadási ügy	Cegéld,	1878	Megtekintés
3.	Árvaszéki ügyek	00748	Motskay (Mutskay) Jakab	hagyatéki ügy		1878	Megtekintés
4.	Árvaszéki ügyek	02993	Reger Jakab	hagyatéki ügy	Pest, Hoch Strasse 14..	1878	Megtekintés
5.	Árvaszéki ügyek	02993	Pap György	hagyatéki ügy		1878	Megtekintés
6.	Árvaszéki ügyek	02993	Schobloch Sztaniszló	hagyatéki ügy	Párizs,	1878	Megtekintés
7.	Árvaszéki ügyek	02993	Schobloch Maria	hagyatéki ügy	Párizs,	1878	Megtekintés
8.	Árvaszéki ügyek	02075	Lanz János György id.	hagyatéki ügy		1878	Megtekintés
9.	Árvaszéki ügyek	02075	Lanz János Györgyné (Schneider Anna)	hagyatéki ügy		1878	Megtekintés
10.	Árvaszéki ügyek	02993	Reger István	hagyatéki ügy		1878	Megtekintés
11.	Árvaszéki ügyek	02075	Lanz György legifj.	hagyatéki ügy		1878	Megtekintés
12.	Árvaszéki ügyek	00907	Szlaveszky József	gyámsági ügy		1879	Megtekintés
13.	Árvaszéki ügyek	02993	Pap Adolf	hagyatéki ügy		1878	Megtekintés
14.	Árvaszéki ügyek	02075	Lanz György ifj.	hagyatéki ügy		1878	Megtekintés

11. ábra Találati lista az Esemény időpontja = 1920 keresőkérdésre

Budapest Főváros Levéltára Központi Kereső							
						Egyszerű keresés	Részletes keresés
Metarekord Találati listában: 93.							
						Előző rekord	Következő rekord
						Találati lista	
Állapota	nem ellenőrzött						
Iratanyag nyelve	magyar						
Lakhely							
Ország	Megye/tartomány	Helység	Kerület/városrész	Utca	Házszám	Emelet/ajtó	Lakhely/ingatlan minősége
		Budapest	V.	Lipót kt.	1.		
Formátum	Papíralapú dokumentum						
Metatár neve	Polgári perek						
Státusz	felperes						
Egyedi azonosító	oai.polgper:62180						
Keletkezés dátuma	1908						
Levéltár	Budapest Főváros Levéltára						
Pertárgy/Perérték	2617 K 20 f						
Ügyazonosító	HU BFL - VII.2.c - 1908 - I.1957						
Név	Gaál Bertalan						
Ügyszám	I.1957						
Terjedelem (pagina)	17						
Fond	VII.2.c Budapesti (Pesti) Királyi Törvényszék peres iratok						
Foglalkozás	műépítész						
Ügytípus	vagyon per						
Módosítva	2009.03.12 09:22:16						

12. ábra Egy polgári peres irat rekordja

- A rekordok melletti *Megtekintés* gombra kattintva juthatunk az egyes rekordok űrlapjára, amely adattartalmanként változó felépítésű, de az iratok azonosításához szükséges legfontosabb adatokat tartalmazza (12. ábra). Az űrlapok felépítése az egyes irattípusokra jellemző különféle mezőket tünteti fel. Minden űrlapon megtaláljuk az iratok azonosító adatait, az ügyszámot, az iratok

keletkezési dátumát, a fond jelölését és az ügytípusra vonatkozó adatokat, továbbá a dokumentum formátuma is leolvasható. Speciális mezőként a Megjegyzés, vagy a Tartalom rovatban találunk bővebb információt az iratanyag tartalmára vonatkozóan. Ide kerülnek például az árvaszéki ügyekben a leszármazási listák, a tartalmazott különböző ügyek vonatkozásai, vagy a

középkori rekordoknál például a szövegek tartalmának lényege. Az iratokhoz kapcsolódó nevek és helyek több kapcsolódó adattal együtt jelennek meg. A rekordok egy részénél például a bírósági és a polgári peres ügyeknél a foglalkozást és a vallást is feltüntetik, de minden névadat esetében találunk születési helyet, időt és „anya neve” adattípust. Ahol szükséges, az űrlapok jelölik a résztvevők státuszát: hogy egy személy örökösként, gyámként vagy például felperesként szerepel-e egy ügyben. Tárgyszavakkal csak a közgyűlési jegyzőkönyvek rekordjai között találunk. E rekordokhoz minden esetben kapcsolódik a dokumentumok digitális változatának elérési útvonala is. A középkori rekordoknál találunk információt az iratanyag budapesti vonatkozásairól; arról, hogy milyen formában: átiratban, másolatban, vagy eredetiben őrzik; és a szövegkiadásokról, amelyekben a dokumentumok szövegei hozzáférhetők. Ez utóbbinak viszont csak a szígláját találjuk meg az adattárban, feloldásukhoz segédletet nem nyújtanak.

Tartalom, minőség

- A részadattárban mára többszázezer rekord kereshető. Az Árvaszéki ügyek metatárban jelenleg 12 426 rekordot találunk, döntően 1872 utáni iratok rekordjait. A bírósági-ügyészeti ügyek adattárban a nagyrészt 1890 utáni rekordok száma 275 013. A hagyatéki ügyek adattárban a budapesti járásbíróóságokon, illetve magán- és közjegyzői fondokban fennmaradt, 1914 és 1955 között lezajlott hagyatéki eljárások iratait tartalmazza, közel 81 ezer ilyen irat kereshető. A közjegyzői okiratok között az 1875 és 1950 közötti budapesti magánközjegyzők előtt kötött jogügyletek kereshetők, az 1900 előtti iratok legnagyobb része már elérhető, a rekordok száma a tesztelés időpontjában már meghaladja a 200 ezret. A polgári perek metatár a különböző budapesti bíróságokon tárgyalt ügyek 65 ezernél is több rekordját tartalmazza, a kihirdetett végrendeletek között pedig 18 ezer ügy adatait kereshetjük. Találunk továbbá 4048 középkori iratra vonatkozó rekordot, és 7165 közgyűlési jegyzőkönyvként feltüntetett rekordot, amelyek a Budapest főváros törvényhatósági bizottsága közgyűlési jegyzőkönyveinek adattárba beépített rekordjai, bár a kereső oldaláról csak annyit tudhatunk meg, hogy ezek közgyűlési jegyzőkönyvek. A Topográfia néven kereshető adattár rész funkciója pedig dokumentáció hiányában kikövetkeztethetetlen.
- A különböző adattartalmú rekordok feltöltöttsége eltérő. A rekordok nagy részénél sok adatmező

még üres. Ez jellemző a szervezetnévtér elemeire is, ahol a 38 szervezet közül mindössze a Fővárosi Bíróság adatlapja van kitöltve.

- Az adatok minősége változó, az adattárban találunk hibásan írt neveket és tárgyszavakat, továbbá feltűnő, hogy – feltehetőleg adatbeviteli hibából fakadóan – az összes lakhelymegjelölés házszámadata után a találati listában két pont jelenik meg.
- A tárgyszavak nem egységesek, a vallás- és foglalkozásnevek között nem egyet találunk, amelyet különböző formákban vettek fel. Jellemző, hogy a vallások a legtöbb rekordban csak rövidített formában jelennek meg és csak így kereshetők (pl. *róm.kat.*, *ref.*).

Metaadatok, bibliográfiai információk

- A BFL metatárban a személyek és a szervezetek névtérleírása az ISAAR (CPF) 2nd szabványon alapul. A metatárban az adatokat Dublin Core metaadat-szabvány szerint rögzítik, melyeket az adattár XML sémákban tárol. A fellelhető dokumentáció szerint a LEAR-projekt egyik célja a szabványos adatcsere megoldása, de a kereső oldalain ennek nem sok nyomát látjuk. Közvetlenül a keresőn keresztül XML fájlokat exportálni nem tudunk.

DigitArchiv

Honlap: <http://www.digitarchiv.hu>
E-mail: digitarchiv@mol.gov.hu
Fenntartó: Magyar Országos Levéltár
Értékelés dátuma: 2009. május 24.
Értékelő: Drótos László

Általános értékelés

A Magyar Országos Levéltár (MOL) 2003-tól digitalizálja szisztematikusan a modernkori iratanyagait, és 2007-től ezek egy részét az interneten is közzéteszi a DigitArchiv rendszer keretében. A szolgáltatás 2008 nyarán kapott nagyobb publicitást, a Nagy Imre-per iratainak és hanganyagainak elektronikus feldolgozása kapcsán. A www.digitarchiv.hu adattárban jelenleg az MSZMP-jegyzőkönyvek (1956–1989), az MDP-jegyzőkönyvek (1948–1956), a Minisztertanácsi jegyzőkönyvek (1867–1944), a Miniszterelnökség iratai – segédkönyvek (1867–1910), a Külügyminisztérium repertórium (1965–1990) és a Legfelsőbb Bíróság anyagaiból a Nagy Imre-per

dokumentumai, valamint a *Levéltári közlemények* publikációi (1923–2002) kereshetők vissza. Hogy ezeket a dokumentumhalmazokat az adott évi intervallumokon belül milyen arányban digitalizálták már, az sajnos nem derül ki, mint ahogy az sem, hogy a DigitArchiv online szolgáltatásában hány százalékuk érhető el teljes szöveggel, illetve szkennelt oldalképként. Azt ugyanis az egyes őrzási egységek kezelői döntenek el, hogy mely fondokat tesznek nyilvánossá, sőt a tapasztalatok szerint még egyes dokumentumokon belül is vannak olyan oldalak, amelyeket a személyi adatok védelme miatt nem jelenít meg a rendszer. Az iratok az *Egységes Elektronikus Levéltári Nyilvántartó Rendszer* (www.e-archivum.hu) honlapjáról megnyitható Bodza adatbázisban (193.224.149.3:8080/mol) is visszakereshetők, de ha itt egy találatra kattintunk, akkor minden esetben egy „Nem megfelelő paraméter!” hibaüzenet jön, úgyhogy innen még a nyilvános dokumentumokat sem tudjuk megnézni. A DigitArchiv rendszeren belül a képpoldalak megjelenítését a Microsoft *Silverlight* nevű technológiája végzi. Ehhez előzőleg telepíteni kell egy kiegészítő szoftvert, és utána kényelmesen navigálhatunk vele a jó minőségben szkennelt képek között. A keresőhöz és a felülethez van részletes súgó, de magáról a projektről nincsen semmi háttérinformáció az adattár honlapján.

A tesztelés során talált három komolyabb hiányosság a következő:

1. Ha a keresővel megpróbálunk a tartalomjegyzékek mellett az OCR-es szövegben és/vagy a hanganyagok szövegében is keresni, akkor minden alkalommal 0 találat van, még akkor is, amikor a tartalomban valójában lenne találat. Ez nagyon zavaró így, és nincs is értelme ezeknek a szűkítési feltételeknek mindaddig, amíg a nyilvános felületen amúgy sem érhető el a szöveges anyagok.
2. A keresőben – úgy tűnik – nincs csonkolási lehetőség, ami már a tartalomjegyzékek keresését is megnehezíti, hiszen be kell írni a lehetséges ragozott és képzett szóalakokat is vesszővel elválasztva, de ha majd a teljes szövegben is megnyílik a keresési lehetőség, akkor végképp elfogadhatatlan lesz, hogy nem lehet szórészleteket, illetve betűket jokerkarakterekkel helyettesíteni.
3. A szkennelt képpoldalak közül sok minden titkosítva lett, így a felhasználót folyamatosan frusztrálja egy piros „Adatvédelmi okokból az oldal nem kutatható!” felirat, vagy egyszerűen nem kattinthatók számára a találati lista tételei.

Bár az érthető, hogy aktuálpolitikai megfontolások, személyes adatok védelme, nemzetbiztonsági érdekek és más okok miatt nem minden levéltári irat lehet publikus, de egyrészt olyan tételek sem jeleníthetők meg, amelyeknél a védelmet semmi nem indokolja (pl. *Borsa Iván* cikke: „A Magyar Országos Levéltár Diplomatiikai Levéltára egyes adatainak gépi segítséggel történő feldolgozása” – *Levéltári közlemények* XLII. évfolyam, 1971.), másrészt semmilyen tájékoztatást nem kapunk arra vonatkozólag, hogy ilyen esetben kinek, hol és hogyan lehet mégis hozzáférni az adott dokumentumhoz (akár digitális, akár eredeti formájában). Úgy tűnik, hogy még a nagy hírveréssel beharangozott Nagy Imre peranyagból sem publikus semmi: még néhány mintaoldal, borítólappal, tartalomjegyzékkel, vagy Nagy Imre nyilvános beszédeinek szövege sem nézhető meg.

Összefoglalva: A 20. század magyar politikatörténetének dokumentumai iránt komoly hazai és külföldi érdeklődés van, amit kiválóan ki lehet (ne) elégíteni a dokumentumok digitalizálásával és online szolgáltatásával – vagy nyilvánosan, vagy regisztrációhoz kötve, vagy zárt levéltári/könyvtári hálózaton át. A DigitArchiv sajnos inkább elveszi a kedvét a múlt kutatóinak: a 10 percnyi várakozás után a kapcsolatot megszakító, csak magyar nyelvű, nem akadálymentes, nem minden platformon használható plugint igénylő felület; a kistudású kereső és a nehézkes nyelvezetű súgó; az 50 tételben maximált méretű és egyenként kinyitogató találati lista, a képeken keresztbeírt nagyméretű vízjel, mind rontja a „felhasználói élményt”. Pedig az adattár már jelen állapotában is sok érdekes és láthatóan gondosan digitalizált dokumentumot szolgáltat, úgyhogy érdemes volna rövid távon legalább az indokolatlanul szigorú technikai korlátozásokon enyhíteni, az adatvédelmi korlátozásokkal kapcsolatban pedig több információt adni az érdeklődőknek: vagyis leírni, hogy mi érhető el csak mutató szinten és mi teljes egészében, mi várható még, és ami pedig nem nyilvános, annak mi az oka, s milyen más módon és feltételekkel lehet esetleg hozzáférni. Hosszabb távon pedig a kereső képességeit fejleszteni kellene és megfontolni egy alternatív megjelenítő felületet is, ahol a dokumentumok – lehetőleg metaadatokkal ellátva – egyszerű JPG vagy PDF formátumban nézgethetők, menthetők, nyomtathatók, könyvjelzővel elláthatók, más weboldalakról belinkelhetők, keresőrobotokkal leindexelhetők, mert jelenleg a DigitArchivból a kezdőlapon és a súgón kívül más nem látszik kívülről.

Részletes értékelés

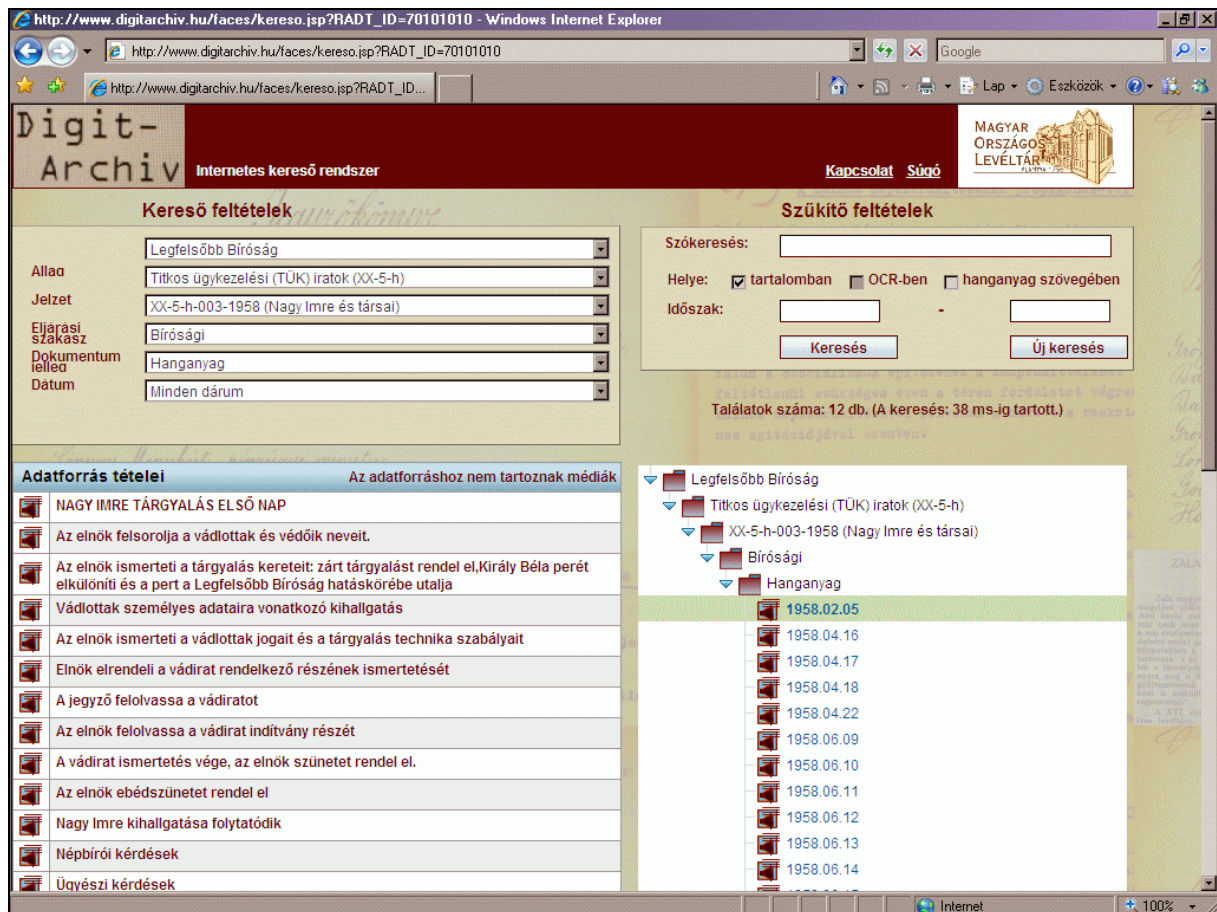
Felület, használhatóság, segítség

- A DigitArchiv a MOL honlapján (*mol.gov.hu*) az Adatbázisok menüpont alatt található. Ismertető helyett itt is csak az adattár részegységei vannak felsorolva (egy betűhibával: „Legfelőbb Bíróság”). Magának a *www.digitarchiv.hu* szolgáltatásnak önálló doménneve és arculata van. A nyitólapon a DigitArchiv és a MOL kétszer is kitett logója (utóbbiak közül az egyik visszavisz a Levéltár honlapjára), néhány dokumentum képe és háttérképként is használt montázs, valamint egy meglehetősen kicsi „Belépés” gomb fogadja a látogatót (13. ábra). A fejlécben levő „Kapcsolat” (ami egy e-postacím) és „Súgó” menüpont a keresőoldalon is megmarad, akárcsak az elég fölöslegesnek tűnő „Internetes kereső rendszer” felirat. Fontos még – a kezdő névelő nélkül írt – „Teljes funkcionalitáshoz Silverlight kompatibilis böngésző és minimum 1024x768 felbontás szükséges” figyelmeztetés a kezdőlap alján, amelyet még további technikai követelmények egészítenek ki a súgóban. A rendszernek fontos sajátossága, hogy 10 percnyi várakozás után kiléptet és visszavisz a kezdőlapra, ám ez csak a súgóból derül ki – jó volna kiírni a nyitóoldalra is. (De még jobb volna megnövelni, mert nem biztos, hogy például egy nyomtatást el tud intézni a felhasználó ennyi idő alatt, a szerver túlterheltsége pedig aligha indokolja ezt a rövid inaktivitási időt. Vagy legalább annyit megtenni, hogy az ajánlott képernyőfelbontásnál ne a belépési gomb fölé tegye ki a Firefox az újbóli belépésre felszólító pirosbetűs üzenetet, megakadályozva ezzel a visszajutást a keresőfelületre.)
- Mind a kezdő- és keresőlapot, mind pedig a dokumentumok nézegetésére szolgáló képernyőt a levéltári környezethez illő barnás paszellszín jellemzi, ezt a színvilágot élénkebb kék és bordó gombok, fejlécek és ikonok teszik változatosabbá. Az oldalak rajza, betűtípusa és elrendezése kellemes, jól áttekinthető. Kisebb olvashatósági hiba csak a Nagy Imre peranyagokhoz vezető menüknél van, mert itt a két sorba törő Eljárási szakasz és Dokumentum jelleg feliratok és a kis sortávolság miatt egymásba lógnak a betűk. (Ennél a menünél még egy elírás is rontja az összképet: „Minden *dárum*”, l. 14. ábra.)

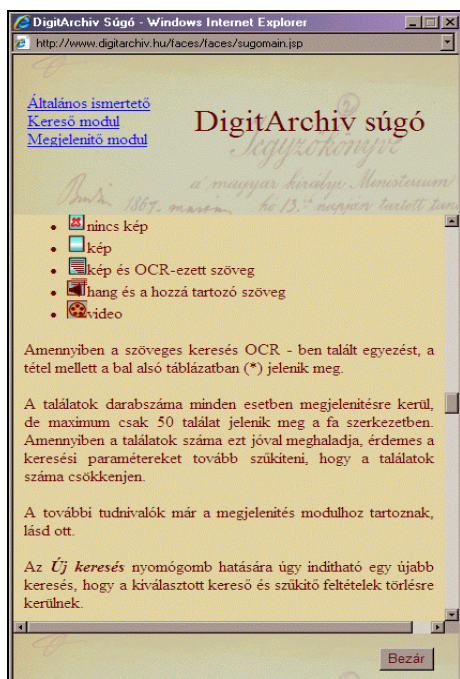


13. ábra A DigitArchiv nyitólapja

- Jó dolog, hogy a DigitArchiv azon ritka hazai adattárak közé tartozik, amelyekhez van hosszabb kezelési útmutató. Nem jó viszont, hogy ez a súgó olyan keskeny ablakban jelenik meg, amely nem méretezhető át (így a benne levő képernyőfotók és táblázatok elég torzak, ráadásul az Adobe Reader letöltő oldala is ugyanebben a kisablakban próbál megnyílni), nincs menüsora (így nehezen lehet rá hivatkozni, vagy ki nyomtatni), megnyitás után azonnal eltűnik a főablak mögött, és az egyes fejezeteire, illetve a Miniszterelnökség iratainak saját súgójára hivatkozó ugrópontok pedig nem működnek benne (15. ábra). Az sem jó, hogy a programozókra jellemző nehézkes nyelven íródott (pl. „A megjelenítésnél a média nagyobb felületen, átlátható módon történő megjelenítés érdekében a műveleteket lehetővé tevő gombok és egyéb elemek áttetsző formában jelennek meg.”, vagy: „Ha a tételeket azonosító információk rögzítése megtörtént, akkor a linkként viselkedő tétel megnevezésre kattintva érhető el a megjelenítés.”), és hogy gépelési és nyelvtani hibákat is tartalmaz (pl. *képerenyő*, *gögetősávval*, illetve: *nyíllak*, *egy hosszúságú szavak*).
- Csak magyar nyelvű felület van; az egyetlen angol felirat az időnként megjelenő „System Messages” kezdetű figyelmeztetés – igazán nem sokból állt volna e helyett „Rendszerüzenet”-et írni. A sok javascriptes menü és az oldalképek megjelenítéséhez használt Silverlight miatt természetesen akadálymentes hozzáférésről szó sincs, sajnos ez gyakori probléma a hazai közintézményi tartalomszolgáltatásoknál.



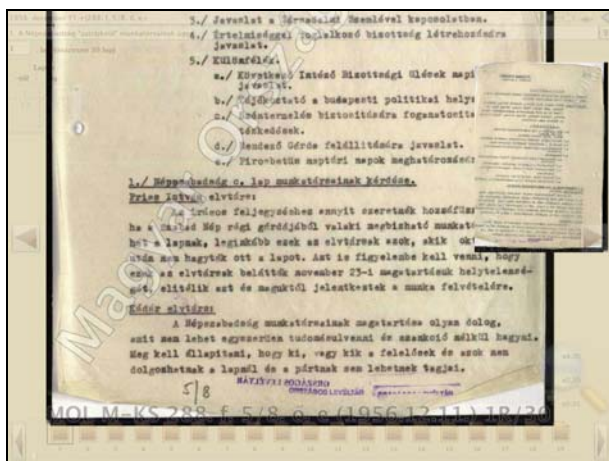
14. ábra A Nagy Imre peranyag hangfelvételeinek listája



15. ábra A súgó ablaka és a keresési tudnivalók egy részlete

- A Microsoft-féle Silverlight platform, amely az interneten igen elterjedt Adobe-tulajdonú Flash technológia versenytársának tekinthető, látványos és kényelmes megjelenítési és navigálási lehetőséget nyújt nemcsak képek, hanem egyéb multimédia tartalmak számára is. A DigitArchiv ezzel mutatja meg a szkennelt oldalképeket, és bár a súgó kétoldalas nézetről, oldalkép és OCR-es szöveg egyidejű megjelenítéséről, sőt hanganyag és szövegének szinkronban való lejátszásáról is beszél, a jelenlegi nyilvános tartalmak között nem sikerült ilyenekre példát találni a tesztelés során. A súgóban szintén említett PDF-generálás sem működik. A Silverlight felületen a képek méretezését és a lapozást halvány, de az egérkurzor hatására élessé váló ikonok és skálák segítik. Ezzel a megoldással nagyobb képernyőfelület jut a képdokumentum számára, hiszen a navigációs eszközök csak akkor jelennek meg a kép fölött, amikor szükségünk van rájuk. A hasznos felület még tovább növelhető, ha teljes képernyősre váltunk az erre szolgáló ikonnal, de ilyenkor az egérgöggövel való méretezés sajnos nem működik, és ebből a nézetből csak az Esc

gombbal tudunk kijönni, mert a másik – a teljes képernyőshöz nagyon hasonlító, de fordított nyilakat mutató ikon – nem erre szolgál, hanem csak a nagyítást állítja alapállapotba. A halványan, de mégiscsak mindig látszó navigáló eszközök, valamint a képernyő alsó és felső részére ugyancsak halványan felírt dokumentumkód és egyéb metaadatok, továbbá a minden képen végighúzó nagy vízjel szinte olvashatatlanná tesznek bizonyos sorokat. (Érdemes ilyen esetben a nyomtató ikonra kattintani, mert ekkor a kép egy normál böngésző ablakba kerül át, ahol rendesen megnézhető, sőt el is tudjuk menteni.) Egyes képeknél furcsa módon egy kicsinyített, de vízszintesen tükrözött kép is ráúsztható az eredetire, ennek értelmére nem jöttem rá (16. ábra). Bár a Silverlight a ma leginkább elterjedt Windows verziókkal és böngészőkkel együttműködik, de nem minden rendszerre telepíthető fel. Ez elég nagy ár a látványos felületért, amelyen jelenleg valójában csak egyszerű JPG képek jelennek meg, mert a multimédia platform egyéb lehetőségeit még nem használja ki a DigitArchiv nyilvános webes változata.



16. ábra A dokumentumnézegető képernyő, egy tükrözéssel megjelenő kisképpel

Kereső funkciók, találatok

• A DigitArchivban egy „Kereső feltételek” és egy „Szűkítő feltételek” doboz segítségével állíthatjuk össze a keresőkérdésünket. Szokatlan, hogy utóbbiba került maga a szöveges keresőmező is, amit más adattárak nem szűkítésként, hanem elsődleges kereső feltételként kezelnek. De itt valójában azt várják a felhasználótól, hogy előbb döntse el, hogy mely dokumentumhalmazban és lehetőleg azon belül is mely részegységben akar keresni, majd ezt a részhalmazt szűkítse tovább időszak és esetleg egy-két szó megadásával. Ez a fajta keresés azonban inkább a gyűjteményt jól

ismerőknek kézenfekvő; a kívülálló felhasználó már az első menü „Minden fond” menüpontját sem érti. Igaz, a sűgőben elolvashatja, hogy mi ez a szakkifejezés, de valójában szakmailag sem helyes ez a kiírás, mert például a *Levéltári közlemények* folyóirat anyaga nem fondokban van tárolva (és jobb lenne a menü végére tenni, nem pedig a jegyzőkönyvek közé). Az „OCR-ben” opcióhoz is célszerűbb volna a közérthetőbb „szövegben” feliratot írni, de ezzel a szűkítő feltétellel ennél sokkal nagyobb baj is van: ugyanis ha a tartalomban való keresés mellett ezt is bekapcsoljuk, akkor többé semmilyen kérdésre nem kapunk találatot. Ugyanez a helyzet, ha a hanganyagok szövegében próbálunk keresni – úgyhogy ez a két, első látásra nagyon izgalmas lehetőség használhatatlan.

- A kereső működését a használati útmutató elég jól elmagyarázza és a beviteli mezőkhöz is vannak lebegő sűgők. A „Szókeresés” mezőnél maximum négy, szóközökkel elválasztott szókapcsolat (AND művelet) írható be vesszővel elválasztva (OR művelet). Ennél bonyolultabb keresőfeltételeket (pl. kifejezések, távolsági operátorok, szótöredékek, zárójeles műveletek) nem tud a rendszer, ami különösen akkor lesz gond, ha valóban lehet majd keresni a dokumentumok teljes szövegében is. A sűgő szerint vannak tiltott szavak (névelők, gyakori kötőszavak stb.), de valójában az itt felsorolt stopword-ök közül többet is elfogad a kereső és találatokat ad rájuk (pl. hogy, pedig, egy). A keresőkérdések és a találati listák mentésére nincs mód.
- A találatok elég gyorsan jönnek, sőt a számuk mellett még az is megjelenik, hogy hány millisekundumig tartott a keresés (ami a rendszergazdán kívül mást aligha érdekel). A tesztelés során egy alkalommal egy „unhandled exception” hibaüzenetet küldött a Java szerver, de ezt a hibát később már nem tudtam reprodukálni. Csak az első 50 találat nézhető meg, és ezek sorrendjét sem tudjuk befolyásolni: a főmenü – eleve nem logikus – sorrendjében és azon belül pedig évszám szerint jelennek meg a találatokat tartalmazó fondok. Az 50-es korlát teljesen indokolatlan és nagyon megnehezíti a digitalizált állományt nem ismerő kívülálló számára a keresést, hiszen szinte pontosan tudniuk kellene, hogy mit hol találnak. További nehezítés, hogy a megszokott találati lista helyett egy hierarchikus menüt kapunk, amelynek még egyenként ki kell nyitogatnunk az ágait (néha egész mélyen), majd a bal oldalon megjelenő tartalomjegyzékben meg kell keresnünk a pirossal kiemelt keresőszavainkat tartalmazó sort, és csak ezután nézhetjük

meg végre magát a dokumentumot (amennyiben digitalizálták és nem titkos) (17. ábra). Így az ember ideje nagy részét kattintgatásokkal és görgetésekkel kénytelen tölteni. Felhasználói szempontból sokkal jobb volna a webes keresőknél megszokott, „egysoros” tételeket tartalmazó, beállítható tételszámú és sorrendű, lapozható találati lista, amelyből egyetlen kattintással elérhető a részletes adatrekord vagy maga a teljes dokumentum.

Tartalom, minőség

- Az adattár honlapja nem tartalmaz statisztikai adatokat a digitális archívum méretéről: az adatrekordok számáról, a digitalizált dokumentumok/oldalképek mennyiségéről és ezen belül a nyilvános és nem nyilvános tételek arányáról. A lefedett időszakokról is csak úgy tudunk meggyőződni, ha sorra lenyitogatjuk az egyes állományrészek menüit. Ha az időszak szerinti szűkítésnél kezdőévként 1867 előtti évszámot állítunk be, akkor 22 340 találat van, vagyis leg-

alább ennyi adatrekordból áll a DigitArchiv, de mivel nem biztos, hogy minden rekordnál van évszám adat, ezért a tényleges szám ennél több lehet. És természetesen nincs minden tétel mögött digitalizált dokumentum sem, mert vannak például olyan részek, amelyek csak mutatókból állnak. A digitális dokumentumok sok esetben adatvédelmi okokból nem nézhetők meg, bár sikerült olyan MSZP jegyzőkönyvet is találni, amelyben az egyik oldal a találati listából nem volt megnyitható, de a következő oldalt megnyitva, majd visszalapozva mégis megjelent a „titkos” oldal is (18. ábra).

- Az oldalképek jó minőségűek, színesek vagy szürke árnyalatúak, látható vízjellel védettek. A papírdokumentumok mellett a Nagy Imre és társai elleni peranyagban hangfelvételek is vannak, de ezek sem hallgathatók meg az interneten át, csak a címük látható. Az „Audiovizuális” dokumentumtípust kiválasztva pedig még ennyi sem jelenik meg, csak egy „No items found” üzenet.

The screenshot shows the DigitArchiv search results page. At the top, there is a search bar with the text 'fock jenő ndk'. Below the search bar, there are two columns of search filters: 'Kereső feltételek' (Search criteria) and 'Szűkítő feltételek' (Refinement criteria). The search results are displayed in a tree view on the right side, showing a hierarchy of folders and documents. The first folder is 'MSZMP jegyzőkönyvek', which contains several sub-folders: 'Központi Bizottság', 'Politikai Bizottság', and 'Titkárság'. The 'Központi Bizottság' folder is expanded, showing a list of documents. The first document is highlighted in green and is titled '1970. február 18-19. (kibővített ülés)+(288. f. 4/104-10...'. Below the search results, there is a table of 'Adatforrás tételei' (Source records) with six entries, each with a checkbox and a brief description. The first entry is highlighted in green and matches the search criteria.

17. ábra Egy kinyitogatott találati lista és balra egy pirossal kiemelt találat



18. ábra A találati listából nem megnézhető, de lapozással megjelenő „titkos” oldal

- A tartalommutatókban elő-előfordulnak gépelési hibák (pl. *kérdéseeiről*, *ideaiglenes*, *bartásági*), ezeket egy helyesírásellenőrzéssel könnyen ki lehetett volna szűrni, hogy ne rontsák a keresés pontosságát. Betűhiba még a lenyitható menükben is van: az MSZMP jegyzőkönyvek / Elnökség alatt: 1989. *zeptember*. Hogy a Minisztertanácsi jegyzőkönyvek menüjének évlisztája miért nincs mindenhol növekvő sorrendben, az rejtély, de lehet, hogy itt is adatrögzítési probléma van a háttérben, vagy pedig egyszerűen elfelejtették rendezni ezt a listát. Mivel OCR-es szöveget nem sikerült találni az adattárban, ezért annak minőségéről nem tudtam meggyőződni.

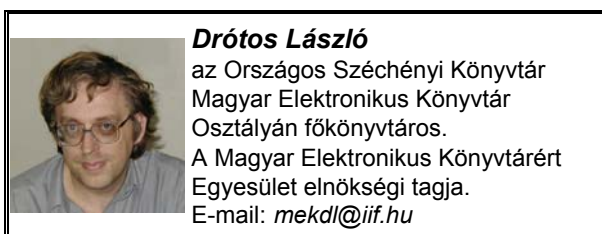
Metaadatok, bibliográfiai információk

- Az adattár honlapja csak egy üres <title></title> sort tartalmaz, és magán az oldalon is csak képként jelenik meg a szolgáltatás és a fenntartó neve. Fontos volna, hogy legalább a közintézmények lássák el szabványos Dublin Core metaadatokkal az internetes tartalomszolgáltatásuk nyitóoldalát (lehetőleg angolul is), mert ez olyan, mintha egy nyomtatott kiadványt címlap és impresszum nélkül publikálnának.
- A kezdőlap és a kereső űrlap forrásában az látszik, hogy a használt karakterkódolás sincs beállítva a fejlécbe, és a szövegben vegyesen van

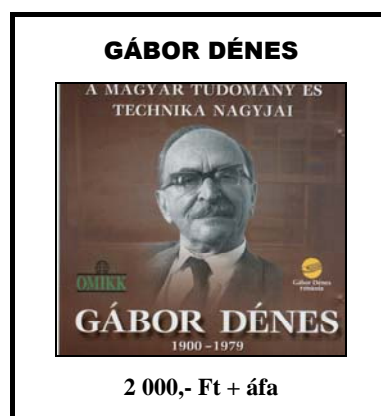
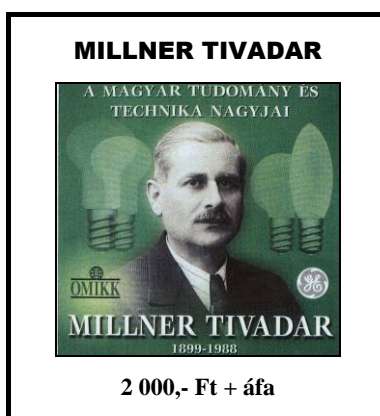
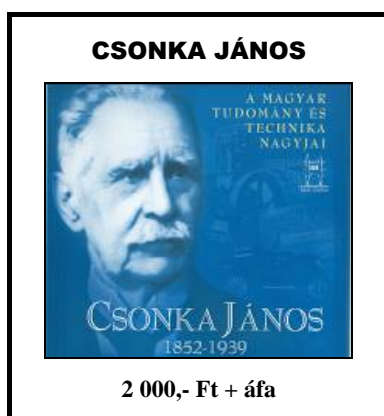
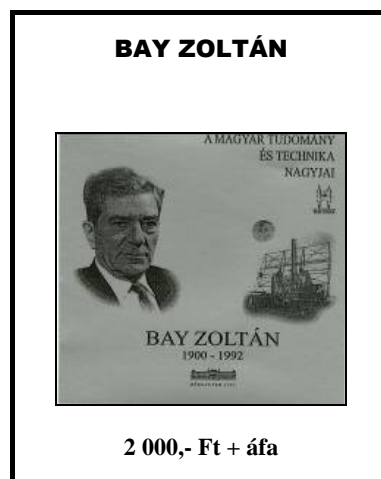
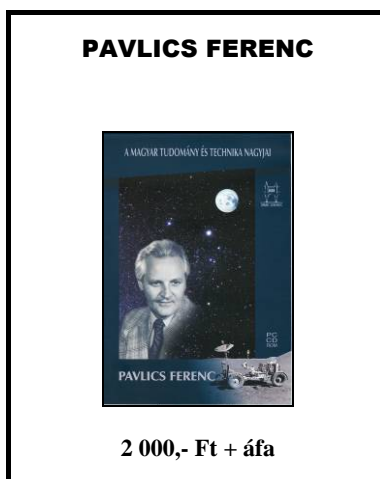
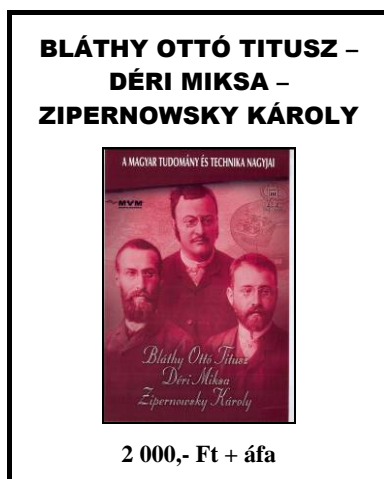
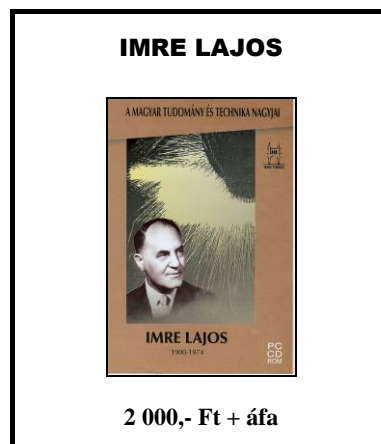
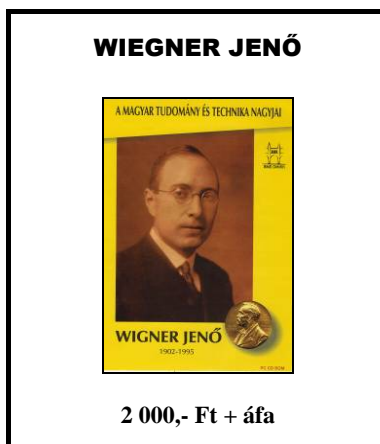
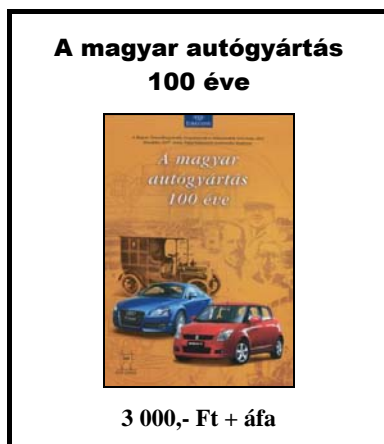
nak „acute”, Latin-2 és decimális Unicode kódolású magyar karakterek.

- A találati listákban a hierarchikus menüszerkezet mindig mutatja, hogy melyik dokumentum melyik állományrészéből és azon belül melyik fondból származik, de a képnézegető oldalon már csak az őrzési egység és a dokumentum címe látszik halványan (amíg rá nem visszük az egeret), magán az oldalképen pedig már csak a belső azonosító kód van felírva (néhol a sötét háttér miatt rosszul olvashatóan), így ha egy ilyen iratoldalt lementünk vagy kinyomtatunk, akkor például az 1989. július 31-i MSZMP elnökségi jegyzőkönyvben található „Módosítások a párt belső életével kapcsolatban” c. dokumentumoldal forrásáról már csak ennyi információk marad: *MOL M-KS 288. f. 59/4. ó. e. (1989.07.30) 2R/42*. Ezen a problémán, vagyis hogy a metaadatok és a digitális képek mentéskor vagy nyomtatáskor elszakadnak egymástól, például azzal lehetne segíteni, ha a kiválasztott oldalképeket PDF fájlba tudná menteni a felhasználó, amely szemmel látható és beágyazott formában is tartalmazná a forrásra vonatkozó és a technikai metaadatokat. Jelenleg ugyanis csak JPG formátumban lehet menteni és a felajánlott 60 karakter hosszú, szinte csak számokból álló fájlnev sem hordoz a laikus számára értelmezhető információt.

Beérkezett: 2009. V. 25-én.



**A BME OMIKK gondozásában megjelent
A magyar tudomány és technika című CD/DVD sorozat tagjai**



Megrendelhető:

a TMT Szerkesztőségénél, tel./fax: 463-2446
vagy tmt@omikk.bme.hu; acsubak@omikk.bme.hu e-mail címen.

Jelentkezési felhívás segédkönyvtáros-tanfolyamra

A Budapesti Műszaki és Gazdaságtudományi Egyetem Országos Műszaki Információs Központ és Könyvtár (BME OMIKK) segédkönyvtáros tanfolyamot hirdet.

A tanfolyam **2010. januárban**, keresztféléves képzési formában indul.

A képzés időtartama két félév.

A foglalkozásokat heti egy alkalommal, csütörtökönként, minden hónap utolsó hetében szerdán és csütörtökön 8-tól 17 óráig tartjuk.

Az oktatás elsősorban gyakorlati jellegű, amely a vizsgakövetelményekben is érvényesül.

A képzés emelt szintű középfokú tanfolyamnak minősül. A végzett hallgató munkaköre segédkönyvtáros.

Részvételi díj a két félévre 150 000,- Ft, a vizsgák költsége előreláthatólag további maximum 50 000,- Ft.

Felvételi vizsga nincs, a beiratkozás feltétele az érettségi bizonyítvány bemutatása.

A tanfolyam jegyzeteit, segédkönyveit kölcsönzés formájában biztosítja a szervező intézmény.

A képzésre azoknak a jelentkezését várjuk, akik a könyvtári munka gyakorlatát rövid idő alatt kívánják elsajátítani, és a számítógép használatában négy ECDL modul megismerésével jártasságot akarnak szerezni.

Jelentkezésüket **2009. december 15-ig** írásban kérjük elküldeni a

BME OMIKK
Segédkönyvtáros-képzés
1111 Budapest, Budafoki út 4-6.
címre vagy a
gylengyel@omikk.bme.hu
e-mail címre.

További felvilágosítás a **463-3534**-es telefonszámon Lengyel Gyöngyitől kérhető.

