

Moldován István

Három az egyben

Közös kereső az elektronikus könyvtárban
<http://mek.oszk.hu/kozoskereso>

Az internetes szolgáltatások természetes jellemzője egyrészt a sokféleség, a sokszínűség, másrészt ennek következtében állandósul az integrációs folyamat, a különböző szolgáltatások összekapcsolása, „egyablakossá” alakítása.

Az Országos Széchényi Könyvtárban több mint egy évtizede már nemcsak bibliográfiai és katalógusadatok, hanem teljes szövegek és képi dokumentumok is felkerülnek az internetre. Ezek egy része saját hagyományos dokumentumainak digitalizált változata (például corvinák¹, térképek), más részük elektronikus formában gyarapított, eredeti digitális dokumentum. A digitális szolgáltatások különböző projektek, pályázatok támogatásával, időben elosztva jöttek létre, ezért ezek még egy tapasztalt könyvtárosnak is nehezen áttekinthetőek, nemhogy az idelátogató olvasónak vagy a távoli felhasználónak. A könyvtár nemrég megújult honlapján megjelent ugyan egy összefoglaló *Digitális könyvtár*² oldal, de ez csak ismertetőket, linkeket tartalmaz a sokféle tartalom egy részéhez. Régóta érezzük már a hiányát egy olyan eszköznek, amely több forrásban egyszerre tesz lehetővé egyszerű vagy összetett keresést a felhasználók számára. A feladat nagyja még mindig előttünk van, de a közelmúltban egy újabb jelentős lépést tettünk a tervezett cél felé. A reális megvalósítás érdekében, a szűkös erőforrásokkal gazdálkodva egyelőre csak néhány adatbázist kapcsoltunk össze. A szegedi Contenta³ rendszerhez hasonlóan nem egy nagy közös adatbázisba kényszerítettük speciális gyűjteményeinket, hanem azok egyedi szolgáltatásait megtartva, egyfajta metafelületet készítettünk hozzájuk.

Első lépésben három adatbázis vált így közösen kereshetővé, amelyek létrejöttét, fejlesztését a MEK Egyesület⁴ is támogatja. A közös kereső kifejlesztésére nem volt külön könyvtári forrásunk, ennek megszületését is az Egyesületnek köszönhetjük, amelyik az erre szánt összeget az Internet Szolgáltatók Tanácsától⁵ kapta.

Elsőként nézzük át röviden, milyen gyűjteményekről is van szó, milyen tartalmakról.

* * *

Magyar Elektronikus Könyvtár (MEK)

<http://mek.oszk.hu>

A MEK 1994-ben indult el az NIIF⁶ támogatásával, és 1999 óta a nemzeti könyvtár szolgáltatásaként működik. Alapvetően könyvjellegű, eredeti digitális vagy papírról digitalizált, nyilvánosan terjeszthető dokumentumokat gyűjt és szolgáltat. A könyvek többféle formátumban (pl. HTML, PDF, DOC, RTF, EPUB, PRC, LIT, DjVu, MP3) találhatóak, és a részletes metaadatok is különböző szabványok szerint (pl. ISBD, MARC, DublinCore, OAI) nézhetők meg, illetve vehetők át más rendszerekbe.

Elektronikus Periodika Archívum és Adatbázis (EPA)

<http://epa.oszk.hu>

A MEK kiegészítéseként indult 2004-ben. Első lépésben az online e-folyóiratok és egyéb időszaki kiadványok katalogizálását célozta meg, egy szűkebb körben pedig ezek fizikai archiválását is, vagyis egyszerre nyilvántartási és megőrzési funkciót is betölt. A leíró információk között az előbbieket **T**(ávoli), az utóbbiakat **A**(rchivált) betűkkel jelöljük. Metaadatokat az adatbázisban csak a kiadványok szintjén hozunk létre. Magát az archivált tartalmat pedig számonként egy hierarchikus fájlstruktúrában tároljuk, jellemzően PDF vagy HTML formátumban.

Digitális Képtár (DKA)

<http://dka.oszk.hu>

2007-ben indult digitális képi dokumentumok gyűjtésére és szolgáltatására. Vannak benne eredeti digitális fényképek (főként közösségi és képmegosztó oldalakról, illetve közvetlenül a fotósoktól), valamint könyv- és folyóirat-illusztrációk a MEK és az EPA állományából. A fájlok nagyrészt JPEG, kisebb részben pedig TIFF formátumúak. Újabban a gyűjtőkör kiegészült az animációkkal is, egyelőre PowerPoint és Prezi formátumú prezentációkkal. Elsőként főként a könyvtártudomány területéről igyekszünk nyilvánosan szolgáltatható prezentációkat, oktatási és konferencia-anyagokat gyűjteni. A szolgáltatás jellemzője, hogy a gyűjteményben található képek egyenként vannak feldolgozva: részletes, DublinCore-alapú metaadatokat készülnek hozzájuk. Ennek köszönhetően a könyvekben, folyóiratokban és újságokban megjelent illusztrációk és az internetről származó egyéb képi dokumentumok sokféle tartalmi és formai szempont szerint visszakereshetővé válnak.

Mindhárom szolgáltatás technológiai alapja közös. Ingyenes szoftvereszközök (pl. Apache webservert, MySQL adatbázis-kezelő, PHP szkriptnyelv) segítségével, saját fejlesztésként jöttek létre az adatbázisok. A közös keresést megkönnyíti, hogy a három gyűjtemény tartalmi feltárása is azonos módszerrel történik. A MEK-ben már 2003-ban elkezdtek használni az OSZK Köztauruszát⁷, és ma már mindhárom rendszerben ez alapján folyik a dokumentumok tárgyszavazása. Emellett a kezdetektől fogva fontosnak gondoltuk, hogy ne csak keresni lehessen a digi-

tális gyűjteményekben, de böngészni is főbb témacsoportok szerint. Ezért a nemzetközi osztályozási rendszerek nyomán kidolgoztunk egy leegyszerűsített, három szintű tematikus struktúrát, amelyet mindhárom adatbázis egységesen használ. A fentieknek köszönhetően a közös kereső ráépülhet az egyező tartalmi feltárára.

Az EPA adatbázisunkban pár évvel ezelőtt elkezdtük egységesíteni az archivált digitális folyóiratok tartalomjegyzégeit, XML formában készítve el azokat. Ezeket a fájlokat folyamatosan átadjuk együttműködő partnereinknek, akik cikkszintű visszakeresést biztosítanak a MATARKA⁸ és az EHM⁹ szolgáltatások keretében. A tartalomjegyzékekből ezért saját keresőt nem készítettünk, így a most kifejlesztett közös kereső is csak a kiadványszintű metaadatokban, valamint a cikkek szövegében tud keresni.

A <http://mek.oszk.hu/kozoskereso> oldal (1. ábra) az egységes tartalmi feltárási elvek alapján egyrészt közös keresést biztosít digitális dokumentumaink *metaadataiban*: Alkotó, Cím és Téma szerint. Utóbbi nemcsak a tárgyszó, de a dokumentumtípus és a témakör szerinti keresést is magába foglalja. A metaadatok mellett keresni lehet a digitális dokumentumaink *teljes szövegében* is. Mivel a MEK-ben egy dokumentum többféle formátumban is előfordul, ezért kiszűrtük a duplikátumokat, vagyis csak egy formátumot indexelünk le: amennyiben van HTML változat, akkor azt, ha nincs, akkor az RTF/DOC vagy a PDF fájlt. A DKA-ban pedig a teljes szövegű kereső az összes metaadat mellett a képaláírásokban és a részletes ismertető szövegekben is keres.

1. ábra. A közös kereső kezdőlapja: <http://mek.oszk.hu/kozoskereso>

A közös osztályozásnak köszönhetően a három adatbázis egyszerre *böngészhető* is témakörök és azon belül altémakörök szerint. Egyszerre több altémát is kijelölhetünk, ha a Ctrl gombot lenyomva kattintunk a nevékre.

A találati listákban a legfontosabb metaadatok (pl. azonosító, URL, cím, alkotó, téma, típus, felkerülés dátuma, szöveggörnyezet) láthatók, valamint egy bélyegkép is megjelenik (2. ábra). Maga a dokumentum pedig egy új lapon nyílik meg, amennyiben az URL címére (a teljes szövegű kereső esetében a „Fájlnév:”-hez tartozó URL címre) vagy a kis képre kattintunk. Így a böngészőablak vagy -fül bezárásával könnyen visszatérhetünk a listához, továbbá egyszerre több dokumentumot is meg tudunk nyitni. A találati listák felett további szűkítési, valamint rendezési lehetőségeink is vannak (például gyűjtemény, témakör, dátum).

MEK-EPA-DKA közös kereső

Keresőkérdés: farkaspók
 Találatok száma körülbelül: 31 Megjelenített találatok száma maximum: 1000

1

DKA-24307
<http://keptar.oszk.hu/024300/024307>
NAPOZÓ FARKASPÓK / KALLOJAZMIN
 állatkép, vadon élő állat, izeltlábú, pók | Magyarország, Kálló
 Biológia, etológia / Állattan
 fénykép
 2011-07-05
 Fájlnév: <http://keptar.oszk.hu/024300/024307/meta.html>

... F ő c í m : Napozó **farkaspók** B e s o r o l á s í c í m : Napozó **farkaspók** A L K O T Ó S z e r e p : fényképész

2

DKA-24506
<http://keptar.oszk.hu/024500/024506>
ÜREGBEN LAKÓ FARKASPÓK / KALLOJAZMIN
 állatkép, vadon élő állat, izeltlábú, pók | Magyarország, Kálló
 Biológia, etológia / Állattan
 fénykép
 2011-07-21
 Fájlnév: <http://keptar.oszk.hu/024500/024506/meta.html>

... F ő c í m : Üregben lakó **farkaspók** B e s o r o l á s í c í m : Üregben lakó **farkaspók** A L K O T Ó S z e r e p : fényképész

3

MEK-604
<http://mek.oszk.hu/00600/00604>
MADARAINK / HERMAN OTTÓ, CHERNEL ISTVÁN
 madarak, ornitológia
 Biológia, etológia / Etológia; Állattan
 szakkönyv
 2003-03-09
 Fájlnév: <http://mek.oszk.hu/00600/00604/html/kotet00000/fejzet00036/h100415/h200538/h300542/h400543/h500544.htm>

... legkeményebben daczolt a hajnali fagygyal; lombozatja meg sem igen fonnyad, meg sem igen sárgul, csak leválik, zölden egészében. A **farkaspók** elröpítette selymes fonalát, felmosolyog búcsúzóul a szelíd napsugara, hogy ezernyi fényjátékban törjön meg a rét fűszálain megakadt, harmattól csillogó -szönyegen. Száraz avaron hőmplyget a szél, a füvesen

4

EPA-143
<http://epa.oszk.hu/00100/00143>
KÖNYVTÁRI FIGYELŐ
 információtudomány, könyvtár, könyvtári tájékoztatás, könyvtárügy
 Könyvtáratan, információtudomány / Könyvtáratan általában; Információtudomány általában; Tudománytörténet
 időszaki kiadvány, szakmai
 2004-06-24
 Fájlnév: <http://epa.oszk.hu/00100/00143/00069/161.htm>

... volt például a Lycos, melyet a Carnegie Mellon University tanára, Michael Mauldin hozott létre, és amelynek neve a **farkaspók**-félék latin nevéből (Lycostia) származik. Ezek azok a pókfajták, melyek nem a hálóban várják a zsákmányt, hanem aktívan keresik áldozataikat. A keresés technikájában a Lycos több újdonságot hozott: a kifelé mutató

2. ábra. Egy teljes szövegű keresés találati listájának részlete
 DKA, MEK és EPA dokumentumokkal

A dokumentumok teljes szövegének indexelése és visszakeresése a Xapian¹⁰ nevű, nyílt forráskódú programmal történik. A szoftver fejlett keresési képessé-

gekkel rendelkezik (például szótövezés, ékezet nélküli keresés, logikai műveletek). A találatokhoz tartozó adatok egy, a közös kereső számára létrehozott SQL adatbázisból emelkednek át a találati listába. Ez az adattábla minden éjjel frissül, így az előző nap felkerült új tételek metaadatai másnap már kereshetők. A teljes szövegű index pedig havonta bővül az újonnan felkerült dokumentumokban talált szavakkal.

Bízunk benne, hogy új szolgáltatásunk hasznos lesz a felhasználók számára, mert az integrált felület egyszerűbbé teszi az egy adott téma, személy, hely vagy szakkifejezés iránti kutakodást, s hogy az eddig kevésbé ismert gyűjteményeinkre és dokumentumainkra is felhívja a figyelmet.

JEGYZETEK

- 1 <http://www.corvina.oszk.hu>
- 2 http://www.oszk.hu/digitalis_konyvtar
- 3 <http://contenta.bibl.u-szeged.hu>
- 4 <http://mek.oszk.hu/egyesulet>
- 5 <http://www.iszt.hu>
- 6 <http://www.niif.hu>
- 7 <http://mek.oszk.hu/adatbazis/thes.htm>
- 8 <http://www.matarka.hu>
- 9 <http://ehm.ek.szte.hu/ehm>
Vö.: Burmeister Erzsébet: Az EHM (EPA–HUMANUS–MATARKA) közös kereső létrejötté. = Könyv, Könyvtár, Könyvtáros, 2011. 10.sz. 25–31. p.
- 10 <http://xapian.org>