



## Országgyűlési dokumentumok adatbázisa – internetes hozzáféréssel

Az Országgyűlési Könyvtár egyik alapfeladata az országgyűlési dokumentumok gyűjtése és szolgáltatása. Ebben a tevékenységben hozott forradalmi változást az országgyűlési dokumentumok adatbázisa, mely a <http://mpgy.ogyk.hu> címen érhető el.

Az Országgyűlési Könyvtár több éven keresztül saját maga digitalizálta az országgyűlési naplókat, majd az *Arcanum Kft.* bevonásával, kibővítve a képviselőházi irományokkal, adatbázisba szervezte az 1861 és 1918 közötti időszak anyagát. A képviselőházi naplók 322, a képviselőházi irományok 401, a főrendiházi naplók 46 kötetét tartalmazza jelenleg az adatbázis. Az összesen 769 kötet közel 312 000 oldalnyi anyagában kétrétegű pdf technológia segítségével lehet keresni. A felső – látható – réteg az eredeti dokumentumoldal jó minőségű fekete-fehér digitális képe, az alsó – mögöttes, láthatatlan – réteg pedig ugyanezen oldal optikai karakterfelismerő (OCR) programmal feldolgozott szövege. Ez a technológia egyesíti a vagy képként, vagy szöveggként digitalizált dokumentumállományok előnyeit, így az egyik csökkenti a másik hiányosságait. A létrehozott elektronikus dokumentum az eredeti hű mása, de a szövegben szavasan keresni, a találati oldalt megjeleníteni, sőt szövegrészletet szövegszerkesztőbe áttemelni is egyszerűen lehet.

Az adatbázisban kétféleképpen kereshetünk: szavasan, különféle szűkítő mezők segítségével a dokumentumokban, és hagyományos módon, a naplómutatók segítségével is, ugyanis valamennyi megjelent mutatókötet (képviseelőházi 1861 és 1939, illetve országgyűlési 1947 és 1949 között, főrendiházi 1861 és 1910 között, felsőházi 1927 és 1944 között) rögzítésre került az adatbázisban.

Az Országgyűlési dokumentumok digitalizált anyaga egyrészt időben és térben kitágítja a hozzáférést, mert a nap 24 órájában, a gyűjtemények nyitva tartásától függetlenül, Magyarországon és határainkon túl is könnyen elérhető, másrészt jelentős értékmentést jelent, mert a hordozóanyagában veszélyeztetett, savas papíron lévő forráseggyűttest megmenti az utókor számára.

Az Országgyűlési Könyvtár lehetőségeihez mérten folytatja a digitalizálást; célunk az, hogy 1990-ig a nyomtatásban megjelent valamennyi napló és iromány, képviselői almanach, országgyűlési házszabály adatbázisba kerüljön.

Villám Judit  
(Országgyűlési Könyvtár)

---

## Nyílt hozzáférés, nyílt forráskód és digitális könyvtárak: a jelenlegi trendek a világ egyetemi könyvtáraiban

Az indiai szerző cikkében rövid áttekintést ad a nyílt hozzáférés és a nyílt forráskódú szoftverek mozgalmairól, és felhívja a figyelmet arra, hogy ezek különösen hasznosak a tudományos célú digitális gyűjtemények esetében.

Az 1995-ben létrejött Digitális Könyvtárak Szövetsége (*Digital Library Federation*) meghatározása szerint a digitális könyvtárak olyan szerveződések, amelyek megfelelő erőforrásokkal (pl. szakemberekkel) rendelkeznek ahhoz, hogy digitális művek gyűjteményeit kezeljék: vagyis kiválogassák, rendszerezzék, leírják, terjesszék ezeket a dokumen-

tumokat, megőrizték az eredeti tartalmukat, gondoskodjanak az időtállóságukról, egyszerű és gazdaságos hozzáférést nyújtsanak egy vagy több felhasználói kör számára.

Az informatika fejlődése a tudományos és felsőoktatási könyvtárakban is új lehetőségeket nyitott az ismeretek kezelésére. A nyílt forráskódú szoftvereknek köszönhetően a világ sok országában jöttek létre digitális könyvtárak; a nyílt hozzáférés mozgalom pedig megteremtí a keretet ahhoz, hogy a kutatók és a diákok szabadon megismerhessék a legfrissebb szakirodalmat és a kutatási eredményeket.

### Nyílt hozzáférés

A tudományos kutatásban és a felsőoktatásban résztvevőknek az egyik legnagyobb akadályt az érdeklődési körükben megjelent publikációkhoz való hozzáférés jelenti, mivel ezeket sokszor olyan magas előfizetési díjú folyóiratokban közlik, amelyek túlságosan drágák a könyvtárak többsége számára. A nyílt hozzáférés mozgalom ezt az akadályt szeretné megszüntetni azáltal, hogy kiáll a szakirodalom szabad elérhetősége mellett az interneten, hogy „bárki elolvashassa, letölthesse, lemásolhassa, kinyomtathassa, továbbadhassa ezeket a publikációkat, továbbá keresni tudjon bennük, ugrópontokat tehessen rájuk, begyűjthesse a teljes szövegüket egy adatbázisba, illetve bármilyen egyéb legális célra felhasználhassa őket pénzügyi, jogi vagy technikai akadályok nélkül – eltekintve a magához az internethez való hozzáféréshez szükséges feltételektől”. Ezt a felhívást a 2001-ben Budapesten tartott találkozó résztvevői fogalmazták meg ([www.soros.org/openaccess](http://www.soros.org/openaccess)). A BOAI (*Budapest Open Access Initiative*) kezdeményezésnek azóta csaknem ötezer aláírója van, közöttük a legfejlettebb és a fejlődő országok tudományos könyvtárai és könyvtári szövetségei egyaránt megtalálhatók. A budapesti találkozót több hasonlóan fontos összejövétel és felhívás követte, mint például a 2003-as *Bethesda Statement on Open Access Publishing* és a *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, vagy az IFLA által 2004 februárjában kiadott közlemény a szakirodalomhoz való szabad hozzáférésről ([www.ifla.org/V/cdoc/open-access04.html](http://www.ifla.org/V/cdoc/open-access04.html)). Az Open Access mozgalomban kulcsszerepet játszott az angliai *University of Southampton* professzora, *Stevan Harnad*, illetve maga az egyetem is, ahol az elsők között hoz-

tak létre egy ilyen nyílt archívumot. Az Egyesült Királyságban a tudományos kutatásokat támogató egyes szervek ma már kötelezik a kutatókat arra, hogy a támogatás fejében az eredményeket összefoglaló munkáikat ilyen nyílt repozitóriumokban vagy digitális könyvtárakban is elhelyezzék. Folyóiratcikkek esetében ez történhet a publikálás előtt vagy után.

*John Willinsky* egy 2003-as cikkében a nyílt hozzáférés megvalósulásának alábbi módjait különböztette meg: a szerzők által működtetett pre- vagy post-print-archívumok, a szabadon hozzáférhető e-folyóiratok, a nyomtatva és nyilvános elektronikus változatban egy időben megjelenő kiadványok, a késleltetve szabaddá váló e-folyóiratok, illetve a csak részleteikben (pl. csak egyes cikkek vagy csak a kivonatok esetében) nyilvános kiadványok, a szerzők vagy a részt vevő tagok által finanszírozott nyílt hozzáférés, valamint az egy főre eső nemzeti jövedelem alapján ingyenessé tett szolgáltatások.

A legfontosabb nyílt archívumokat az *Open Directory of Open Access Repositories* ([www.openoar.org](http://www.openoar.org)) szervezet próbálja meg nyilvántartani. A cikk írásakor, 2007 novemberében, az ilyen szolgáltatásokat működtető szervezetek száma 734 volt, ezeknek több mint a fele európai. A különböző gyűjtemények együttműködését, közös kereshetőségét az *Open Archives Initiative* ([www.openarchives.org](http://www.openarchives.org)) által kidolgozott szabványok teszik lehetővé, mint például a metaadatok összegyűjtését végző OAI-PMH.

### Nyílt forráskódú szoftver

Az OSS (*Open Source Software*) kezdeményezés az 1980-as évek közepén indult, és azóta nagyon fontossá vált a könyvtári világban is. Az *Open Source Initiative* ([www.opensource.org](http://www.opensource.org)) meghatározása szerint az OSS nem egyszerűen ingyenes szoftvert, illetve a program forráskódjának szabad hozzáférhetőségét jelenti, hanem előírásokat is a kód jövőbeli felhasználására, módosítására és a származtatott szoftverekre – de személyekre, csoportokra vagy alkalmazási területekre vonatkozó megkülönböztetések nélkül. Ezeket a fejlesztéseket gyakran GNU licenc alatt teszik közzé, amely megengedi a programozóknak, hogy megváltoztassák, továbbfejlesszék és terjesszék a szoftvert, azzal a feltétellel, hogy a változtatásokat is szabadon hozzáférhetővé teszik mások számára. Mivel

elég nagy a terminológiai zavar az *open source*, a *freeware*, a *shareware* és a *public domain* szoftver fogalma között, a cikk az [opensource.org/docs/osd](http://opensource.org/docs/osd) oldalon található 10 ismerv és *Richard Stallman*, a GNU és az OSS mozgalom egyik vezéregyéniségének négyponos meghatározása alapján igyekszik tisztázni ezt a kérdést. A lényeg, hogy az OSS szoftverek licenc alatt terjesztett alkotások, amelyeket szabadon futtathatunk bármilyen célból, tanulmányozhatjuk, és az igényeinkhez alakíthatjuk a kódjukat, továbbfejleszthetjük és továbbadhatjuk őket. A nyílt forráskód modellje a hagyományos copyrightnak egy olyan kiegészítése, amely a nagyközönség számára szabadon hozzáférhetővé és felhasználhatóvá teszi a programkódot, egy közösségi együttműködésen alapuló programozási környezetet hozva így létre.

A nyilvános könyvtárak és a nyílt szoftverek eszméje sokban hasonlít. A könyvtárak gyűjteményei, személyzetük szaktudása, sőt még a könyvtárak fizikai terei is szabadon hozzáférhetőek egy széles közönség számára, non-profit módon, közösségi finanszírozással. A könyvtárak nagyon gyakran használnak nyílt forrású szoftvereket, még ha nincsenek is esetleg tudatában annak, hogy milyen sok fontos könyvtári szolgáltatásuk épül rájuk.

A *Yale Medical Library* 1999-ben egy külön webhelyet indított a könyvtári szempontból érdekes nyílt szoftverek jobb megismertetése céljából ([www.oss4lib.org](http://www.oss4lib.org)), és ugyanebben az évben kezdődött el Új-Zélandon az első, OSS-alapú könyvtári rendszer, a *Koha* fejlesztése ([www.koha.org](http://www.koha.org)), amelyet azóta már a világ minden részén használnak (Európában pl. 37 alkalmazást tart nyilván a *Koha Wiki*). Az UNESCO is elkötelezte magát a nyílt szoftverek mellett, és a honlapján ([www.unesco.org/webworld/portal\\_freesoft](http://www.unesco.org/webworld/portal_freesoft)) 13 ilyen rendszert mutat be a *Digital Library* menüpont alatt.

### **Digitális könyvtári szoftverek**

A digitális könyvtárak működtetése egy sor, egymásra épülő munkafolyamatból áll: szerzeményezés, katalogizálás, tárolás, visszakeresés, hozzáférés szabályozás stb. Ma már jó néhány OSS megoldásból válogathatunk ezekhez a feladatokhoz: az egyetlen részfeladatot (pl. felhasználási statisztika) megoldó egyszerű szkriptektől kezdve a minden munkafázist támogató nagy, integrált rendszerekig. A cikk ezekből három, széles körben

elterjedt szoftvert ismertet: a *Dspace*, az *E-Prints* és a *Greenstone* rendszereket.

A *Dspace* ([www.dspace.org](http://www.dspace.org)) eredetileg egyetemi környezetre készült, egy tudományos műhely (intézet vagy tanszék) szellemi termékeinek összegyűjtésére, archiválására és visszakeresésére szolgál. A *Hewlett Packard* és a *MIT* könyvtárainak közös fejlesztése. A Java nyelven írt és PostgreSQL adatbázis-kezelőre épülő rendszer három „rétegből” áll: az alkalmazói szint tartalmazza a felhasználói és a kezelői felületet, valamint a kötegelt betöltő modult; a munkafolyamat réteg a tartalomkezelő, az adminisztrációs, a kereső és a böngésző elemekből áll; a tároló modul pedig az adatbázis kezelését végzi. Mindegyik modulhoz egy jól dokumentált csatolófelület (API) tartozik, így könnyen lehet hozzájuk új programrészeket illeszteni; és vannak is már ilyenek, például a *Jena* (a HP Labs által hozzáadott RDF eszköztár) vagy az *OAI Cat* (az OCLC-nél fejlesztett OAI-PMH szerver). A *Dspace*-nek több mint 240 alkalmazása ismert már világszerte, a legtöbb az USA-ban.

Az *E-Prints* ([www.eprints.org](http://www.eprints.org)) szintén az intézményi repozitóriumok számára kidolgozott nyílt szoftver. A *University of Southampton* fejlesztése és eredetileg pre-print-archívumokhoz szánták, de manapság már mindenféle más célra is használják: például kutatási jelentések, konferencia-előadások és egyéb digitális anyagok gyűjteményeihez. A honlapon levő ismertető szerint „egy olyan rugalmas platform, amelyre minőségi és értékes dokumentumtárak építhetők”. Az egyik legkönnyebb és legegyszerűbb módja annak, hogy kutatási eredményeket, tudományos adatokat, téziseket és jelentéseket, vagy akár multimédia-anyagokat tartalmazó archívumokat hozzunk létre. Jelenleg 201 ilyen szolgáltatásról lehet tudni, többségük Nagy-Britanniában van.

A *Greenstone Digital Library* ([www.greenstone.org](http://www.greenstone.org)) rendszer születési helye az új-zélandi *University of Waikato*, ahol a 2000-es évek elején kezdtek el egy programkönyvtárat fejleszteni digitális gyűjteményekhez használható OSS komponensekből. A fejlesztést ma már az UNESCO és a *Human Info* nevű szervezet is támogatja. Az immár 3. verziójánál járó, Java-alapú rendszer Unix és Windows környezetben egyaránt működik, az egyes modulok XML nyelven kommunikálnak egymással. A *Greenstone* által kezelt gyűjtemények metaadatok és teljes szöveg szerint egyaránt kereshetők, a találatok relevancia és egyéb szempontok szerint rendezhetők. Egy digitális könyvtár

létrehozása öt lépésből áll: az anyag összegyűjtése, az adatok leírása, a gyűjtemény definiálása, majd felépítése, és végül a kész gyűjtemény megtekintése. A GSDL-nek is egyre több alkalmazása van már világszerte, olyan országokban is, mint például Argentína, Kirgízia, a Fülöp-szigetek, vagy éppen Szudán.

A szerző a cikk összefoglalásában hangsúlyozza a „nyílt” szó fontosságát, hiszen az *open access* mozgalom, az *open source* szoftverek és a *nyilvános digitális könyvtárak* mind ugyanazt a „nyílt

modell” követik, amely a szellemi tulajdon megosztásán alapul, és a társadalmak fejlődését és gazdagodását kívánja elősegíteni.

**/KRISHNAMURTHY, M.: Open access, open source and digital libraries: A current trend in university libraries around the world. = Program: electronic library and information systems, 42. köt. 1. sz. 2008. p. 48–55./**

(Drótos László)

## Archiválás a nagyméretű digitalizálás korában (Fehér könyv)

A *Könyvtári és Információs Erőforrások Tanácsa (Council on Library and Information Resources = CLIR)* megbízásából Oya Y. Rieger, a *Cornell Egyetem* könyvtárának munkatársa készítette ezt az összefoglaló tanulmányt. Azt a problémát vizsgálja, hogy vajon a digitalizált anyagok minősége és fontos metaadatokkal való ellátása miként valósul meg, most, amikor könyvek millióit digitalizálják olyan hatalmas projektekben (large-scale digitization initiatives = LSDI), mint a *Google Book Search* vagy a *Microsoft Live Search Book*. Az elkészítés során széleskörűen – kérdőívek módszerével is – tájékozódott a könyvtáros közösségben, valamint a szakirodalomban. Ezek a projektek vegyes érzelmeket keltettek egyes kutatókban és könyvtárosokban, akik sokat tettek a nyomtatott gyűjtemények jó minőségű digitális szurrogátumaiért. Hasonlóan kérdések vetődnek fel a nagy digitális gyűjtemények és elérhetőségük hosszú távú fennmaradásával kapcsolatban. A CLIR célja a tanulmány elkészítésével az volt, hogy stimulálja a kérdések megvitatását és az együttműködési szándékot.

A bevezetést (1. fejezet) követően a tanulmány további 5 fejezetre és egy függelékre tagolódik.

A másodikban megismerkedhetünk négy nagyméretű digitalizálási projekttel és digitalizálási stratégiákkal. A már említett két kereskedelmi vállalkozás mellett két nem profit célú projektet mutat be a szerző: az *OCA-t (Open Content Alliance)* és a *Million Book Projectet*.

Előjáróban tárgyalja azokat a motivációs tényezőket, amelyek a projektekben részt vevő számos

könyvtárat, illetve üzleti és nem profit vállalkozásokat az együttműködésre készítette.

A könyvtárakat három fő érv készíti a részvételre: az első alapvető küldetésükhöz kapcsolódik, fejleszteni a hozzáférést gyűjteményeikhez, mégpedig világszerte. Míg korábban a nagyobb egyetemi könyvtárak az Egyesült Államokban 5-10 000 könyvcímet voltak képesek digitalizálni évente, a *University of Michigan–Google LSDI* ma 30 000 kötetet hetente, így a könyvtár teljes minősített gyűjteményét öt év alatt feldolgozzák. A második szempont az állomány megőrzése. A kiadók gyakran nem tartanak meg példányokat kifutott kiadványaikból, így az archiválás a hosszú távú megőrzést szolgálja. A harmadik motivációs cél a könyvtár saját kutatás-fejlesztési munkájának támogatása, akár gyakorlat megszerzése nagy digitális állományok kezelésében, akár például új keresőeszközök fejlesztése.

A vállalkozások céljai eltérőek. A *Google Book Search* program teljes könyveket digitalizál, mind szabad használatúakat (public domain), mind jogvédeletteket. Ez utóbbiakból csak kis részleteket dolgoznak fel. A végső eredmény egy kereshető index lesz. 2007 augusztusában 27 könyvtár vett részt a munkában. A *Google* ezen kívül több mint 10 000 kiadóval működik együtt.

A 2005-ben indult *Microsoft-projekt* először az *OCA*-val, majd további könyvtárakkal szabad felhasználású, 1923 előtt kiadott könyvek teljes szövegű adatbázisát fejleszti.

A két projektben eltérő a részt vevő könyvtárak helyzete, mások a saját, illetve a közös anyagok

további használatának feltételei. Kiegészítésként a Microsoft kiadókkal is végez munkát a Live Search Books Publishers Program keretében. A cég saját interfészt készített a szövegekben végzett kereséshez.

Az OCA-projektet az *Internet Archive* és a *Yahoo* szervezte nem profit és kormányzati, valamint kulturális és technológiai intézmények részvételével. Célja szabad hozzáférésű digitális gyűjtemények létrehozása és elérhetővé tétele az *Internet Archive* és a *The Open Library* keretében. Az OCA valódi digitális könyvtárat kíván létrehozni több nyelven, multimédiás dokumentumokat is beleértve. Ezek a tartalmak kereshetők is a főbb keresőgépekkel. Az állományok megtalálhatók az *Internet Archive*-ban a Microsoftnál és a *Library of Alexandria*-ban, a jövőben más repozitóriumokban is. Jelenleg szabadon használható anyagokra összpontosítanak, de tárgyalnak kiadókkal is. A projektet a Microsoft és az Adobe is támogatja.

A Million Book Projectet a *Carnegie Mellon University School of Computer Science* és az *University Libraries* szervezik; magja a digitális könyvtárakra vonatkozó kutatási program. Az elmúlt hat évben az USA *Nemzeti Tudományos Alapjának* támogatásával és nemzetközi kapcsolatokkal (26 intézmény) több mint 1,4 millió könyvet szkenneltek Kínában, Indiában és Egyiptomban. Partnerük volt még az *Internet Archive* is. A projekt 2007-ben befejeződött, de a résztvevők folytatni kívánják a munkát. A részt vevő országok saját anyagaikat maguk kívánják tárolni, a hozzáférésről még nincs megegyezés.

A harmadik fejezet a digitalizálás és az archiválás fontosabb adminisztratív és technikai kérdéseit tekinti át. Ez a leírás ugyan nem lehet teljes, azonban érinti, pontosabban felteszi a legfontosabb kérdéseket.

A négy témakör: kiválasztás digitalizálásra és maga a digitalizálás; a tartalom előállítás, ezen belül a kép minősége; az archiválási metaadatok, ezek leíró és szerkezeti típusai, a minőség ellenőrzése; a műszaki infrastruktúra, végül a szervezési infrastruktúra.

A kiválasztás sok kérdést vet fel. Tisztázni kell azonban, hogy a digitalizálás fogalma nem azonos az archiválást célzó (digitális) átalakítással. Az első kérdés mindjárt az, hogy minden digitalizált anyagot meg kell-e őrizni hosszú távon? Közismertek az olyan adatok, hogy a gyűjtemények 10–

20%-a adja a cirkulált tételek 80–90%-át. Nyilván, a korlátozott forrásokat célszerű a használati gyakoriság szerint felhasználni. De vajon a digitalizálás, és ezért a könnyebb hozzáférés, nem módosítja-e a használat gyakoriságát? Bár az említett nagy projektekben ritkán szerepelnek értékes, ritka vagy speciális gyűjteményrészecskék, és a könyvdigitalizálás módszerei is finomodtak az elmúlt időszakban, mégis vizsgálni kell, hogy fenyegeti-e sérülés a kiválasztott tételeket.

Mint a digitalizálásnál mindig, most is gondosan kell vizsgálni a jogvédelemhez kapcsolódó kérdéseket.

A digitális archiválás számos döntést és előkészületet feltételez. Szabványos elvek és eljárásmodok alapozzák meg a sikeres programokat. A futó LSDI-k legtöbbje nem felel meg ezeknek a kritériumoknak. A tanulmány táblázatokban foglalja össze egyrészt azokat a folyamatokat, amelyeket végig kell járni a digitális archiválás során, másrészt az elemzett projektekben elfogadott műszaki specifikációkat. Emellett itt is bemutatja a legfontosabb vitatott kérdéseket. Az például egy évtized óta elfogadott szakmai megállapodás, hogy archiválási célra a veszteségmentesen tömörített TIFF formátumot használjuk. A TIFF-nek számos előnye van, de 1992 óta nem változott. A nagyméretű digitalizálás programjaiban azonban sokan fordulnak a JPEG2000 felé, amely ISO szabvánnyá vált, és sok előnyös műszaki tulajdonsága van, amellett, hogy a metaadatok beépíthetők a fájlba. Lassú terjedésének az a fő oka, hogy a böngészők nem támogatják megfelelő módon.

Kritika illeti azt a gyakorlatot, hogy ma is a kilencvenes évek technikájával hajtjuk végre a digitalizálást. Példaként említi a rögzített értékű felbontást és bitmélységet, anélkül, hogy a digitalizáló eszköz minőségi paramétereit tesztelték volna. Vizsgálatok mutatják, hogy azonos felbontással két különböző szkennelvel végzett digitalizálás különböző képeket mutathat. Nem szabad eltekinteni a be rendezések gondos tesztelésétől.

Külön alfejezetek foglalkoznak az archiválási metaadatokkal. Ezek létrehozási stratégiáival a *PREMIS (PREservation Metadata: Implementation Strategies)* dokumentum foglalkozik. Felhasználható még a digitális képekre vonatkozó, 2006-ban kiadott *ANSI/NISO Z39.87 (Technical Metadata for Digital Still Images)* szabvány is, de a munkákat megkönnyítő műszaki metaadatok szerepe még nem jól dokumentált.

A futó programokban a keresést és dokumentumkezelést elősegítő leíró metaadatok a helyi OPAC-ok bibliográfiai rekordjaiból származnak, míg a strukturális metaadatokból csak kevés található. Törekvés látható arra, hogy globálisan egységes állandó azonosítókkal lássák el a digitalizált könyveket.

A minőségirányítási alfejezet áttekinti az egész munkafolyamat során teendő ellenőrzési és szabályozási lépéseket, és azt veti fel az egyik legfontosabb kérdésként, hogy az ilyen nagy tömegű digitalizálási munkában lehetséges-e teljes körű ellenőrzés, mert ez nagyon idő- és költségigényes feladat. Bár bizonyos technikai elemek automatikusan is ellenőrizhetők, mégis emberi vizsgálatra marad a paraméterek többsége. A vizsgált projektek eltérő módszereket követnek, és például a Google-kezdeményezés kapott már kritikákat a digitalizálás minősége miatt.

A műszaki infrastruktúrával foglalkozó rész számos sérülékeny elemet említ meg, és kiemeli a tároláshoz kapcsolódó kérdéseket. A technikai fejlődés miatti avulás is sok problémát okoz, mégis a könyvtárak számára – a felmérések szerint – a hatalmas tömegű anyag kezelése jelenti a nagyobb gondot.

A tanulmány ismerteti a *Getty Research Institute* nemrégiben készült felmérését, amely szintén arra a következtetésre jutott, hogy az ilyen programokban részt vevő könyvtáraknak újra kell gondolniuk az infrastruktúra- és adattárolási modelljeiket.

Egyedül a technológia azonban nem tudja megoldani az archiválási problémákat, az intézményi politikák, stratégiák és finanszírozási modellek is fontosak. Bár egy évtized óta folytatnak vitákat különböző fórumok a digitális archiválás mikéntjéről, alig maroknyi könyvtárnak van olyan programja, amely támogatná a nagy tömegű archiválást. A tanulmány felsorolja az alapvető szervezési követelményeket, és röviden bemutat néhány szabványt és ajánlást, mint az *Open Archival Information System (OAIS)*, a *Trustworthy Repositories Audit & Certification (TRAC)*, a *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)*, és a *Defining Digital Preservation*.

A negyedik fejezetben a tanulmány megkísérli előrejelezni az LSDI-k hatását a nyomtatott gyűjteményekre.

Mivel sok könyvtár súlyos helygondokkal küzd, valószínűleg nyomásnak lesz kitéve, hogy a digitalizálás után a nyomtatott eredeti példányokat vegye ki az állományból. Azonban a szakirodalom arra is rámutat, hogy a digitális állományok hatására még növekedhet is az igény a nyomtatott változatokra. Az olvasók közül sokan maradnának inkább ezek mellett, különösen a testesebb könyvek esetében. Történtek már kísérletek a „nyomtatás igény esetén” típusú megoldásra, persze csak a szerzői jogszabályok betartásával.

Az ötödik fejezet ajánlásokat fogalmaz meg az archiválási stratégiákhoz, amelyeknek az az elsőrendű célja, hogy segítse az LSDI-ben érdekeltet közötti megbeszéléseket.

A 13 ajánlás öt téma köré csoportosítható:

1. A digitalizálás, mint archiválási módszer. Újra kell gondolni az archív képekre vonatkozó másfél évtizede született digitalizálási követelményeket. Új módszerek, új műszaki megoldások, új olvasói igények születtek azóta. Hajlékony minőségbiztosítási programot kell kidolgozni. Kompromisszumot kell keresni az archiválás és a hozzáférés követelményei között.
2. Tartós hozzáférés. Javítani kell a hozzáférést a digitalizált tartalomhoz, mind a hagyományos módon, mind a legújabb technikák használatával. A hosszú távú használat feltétele az is, hogy világosan értsük a szerződéses korlátozások hatását az archiválási felelőségekre, arra, hogy mit tehetünk és mit nem azokkal a digitalizált anyagokkal, amelyek a kereskedelmi partnerekkel közösen készültek.
3. Az archiválás menedzsmentje. Újra kell gondolni a nyomtatott állomány tárolását, különösen a ritkán használt vagy redundáns tételekét. Meggondolandók a földrajzilag megosztott tárolás lehetőségei. Támogatni kell a digitális mesterpéldányok regiszterének használatát.
4. Digitális archiválási stratégiák. Körvonalazni szükséges egy archiválási akciótervet a nagyméretű digitalizálási kezdeményezések számára. Miután minden könyvtári anyagot nem lehet digitalizálni, ki kell dolgozni a digitális archiválás szintjeit kijelölő elveket.
5. A kutatókönyvtárak stratégiái. A könyvtáraknak szemmel kell tartani és egymás között meg kell osztani a költséginformációkat. Újra kell gondolniuk könyvtári prioritásaikat és stratégiáikat. Gyorsabban kell válaszolniuk a

kihívásokra, és el kell mozdulniuk egy mozgékonyabb és nyílt tervezési modell felé. Újra kell gondolniuk állományfejlesztési elveiket.

Végül a hatodik fejezet következtetései döntően a kulturális intézmények együttműködésének szükségességéről szólnak. Egyetlen könyvtár sem képes a Google vagy a Microsoft léptékű vállalkozás végrehajtására, azonban másokkal együtt megfelelhetnek egy ilyen típusú projekt kihívásainak. Az erők egyesítésének van helye, ha ez költséghatékonyabb és minden együttműködő partner nyer a közös akcióban.

A kulturális intézmények azonban a tudományos örökség kezelői, és erről nem mondhatnak le akkor sem, ha a kereskedelmi érdekek számukra is nyújtanak bizonyos előnyöket. Alapvető szempontnak kell maradnia a tartós hozzáférés fenntartásának és a költséghatékonyaságnak.

A kutató-könyvtáraknak módosítani kell szerepüket és programjaikat azért, hogy kielégíthessék a 21. század olvasóinak igényeit.

A tanulmány jellegénél fogva széles olvasóközönségnek szól a digitalizálásban részt vevő intézmények munkatársaitól a gyűjteményeket használó kutatókon át a támogatást nyújtó alapítványok vagy kormányzati szervek vezetőiig. Az anyag komoly értéke, hogy gondos és terjedelmes lábjegyzeteiben sok-sok hivatkozás található, amelyek a további tájékozódást megkönnyítik az érdeklődő olvasónak.

/RIEGER, Oya Y.: *Preservation in the age of large-scale digitization. A white paper.* = <http://www.clir.org/activities/details/lstdi.pdf>

(Horváth Péter)

---

## Melyik a „csúcskereső”?

A folyóirat munkatársai az Egyesült Királyság hat legnépszerűbb internetes keresőjét tesztelték oly módon, hogy a találatok száma helyett azok relevanciájára helyezték a hangsúlyt. Kiindulópontjuk szerint az átlagos kereső az első, általában tíz találatot (plusz a szponzorált csatolókat) tartalmazó oldalról nem lép tovább, ezért elsősorban azt próbálták megállapítani, hogy a valóban releváns, keresett oldal a találatok „élmezőnyében”, az első tíz találat között foglal-e helyet?

A tesztelés során a következő keresőkérdéseket fogalmazták meg:

- „Apple Pie” – „Apple Computers” – egyszerű logikai (Boole-féle) keresőkérdés;
- („Microsoft Office Online” OR „MS Office Online”) NEAR (Outlook – „Outlook Express”) – összetett logikai (Boole-féle) keresőkérdés;
- Plumbers in New Cross – a lokalizációs funkció tesztelése;
- Raleigh – több jelentésű, kétértelmű kifejezések kereshetőségének tesztelése (csak a Raleigh típusú kerékpárokra voltak kíváncsiak);
- What time is it in Bangalore? – kérdésértelmezés tesztelése;
- iPAQ hw6900 – termékinformáció keresése;
- 08700 100 222 – számok kezelésének tesztelése (közvetlen elérési pontok keresése a BBC panaszzonalához);

- 22/11/1963 – dátumértelmezés tesztelése (a tesztelők a *John Fitzgerald Kennedy* elleni merénylet részleteire voltak kíváncsiak);
- pas de bourre – ritkán használt idegen kifejezések keresése (egy a baletthez kapcsolódó szak kifejezésről van szó);
- Information World Review – egyszerű, minden nap használatos szavakból felépülő, kulcskifejezés-értékű szókapcsolatok kereshetőségének tesztelésére (idézőjelek használata nélkül).

### AOL

<http://search.aol.com>

Találati relevancia: 64%

A Google által működtetett AOL az ötödik legnépszerűbb keresőmotor az Egyesült Királyságban. Az összetett keresés lehetőségei meglehetősen behatároltak. Cím, URL, szöveg, csatolók, nyelv és dokumentumforma szerint lehet keresni, valamint dátum alapján (három hónapra, hat hónapra vagy egy évre visszamenőleg).

A felnőtt tartalom szűrésére szolgáló *Safesearch* opció alapértelmezésben be van kapcsolva. Egy találati oldalon alapértelmezésben 10 találat jelenik meg, s ezen a beállításon nem is lehet (pontosabban a cikk írásakor nem lehetett) változtatni. A

szponzorált csatolók alulról és felülről kapcsolódnak a találati listához, a „Sponsored links” felirat ugyanakkor nem szembeötlő.

A keresőnek több, a tesztelők tetszését elnyerő különleges funkciója van: teljes nézet (a kép, video-, audiotalálatok és a kereskedelmi ajánlatok egy oldalon való integrálása); kulcsszavakat tartalmazó „okos doboz” („Smartbox”); keresési előzmények tárolása. Ezeket azonban valamilyen okból csak az amerikai felhasználók érhetik el.

A tesztelés időpontjában a keresési eredmények a következőképpen alakultak:

- az első („almás pite”) keresés meglepő módon nem hozott szponzorált csatolókat találatként;
- a második keresésnél az „Outlook Express” – helyesen – nem szerepelt a találatok között;
- a harmadik keresésnél az első találatban a következő szöveg volt olvasható: „We have no plumbers in New Cross”, a második találat pedig a Brent Crossban található vízvezeték-szerelők listáját tartalmazta;
- a negyedik keresőkérdés problematikus, hiszen honnan is tudhatná a keresőmotor, hogy a kérdést feltevő a Raleigh típusú kerékpárokra kíváncsi, és nem az azonos nevű városra? A tesztelőknek a kérdés finomításában az „okos doboz” sem volt segítségére, mivel a benne szereplő kulcsszavak között a „bicycle” nem található;
- a bangalóri helyi időre vonatkozó kérdésre az első négy találat megadta a választ, a további találatok azonban más vonatkozású információkat tartalmaztak;
- a következő keresés (iPAQ hw6900) eredményei lenyűgözték a tesztelőket: mind a tíz találat hasznos információkkal szolgált az adott termékről, a teljes nézetben megjelenő kereskedelmi ajánlatok pedig arról tájékoztattak, hogy hol vásárolhatják meg az adott terméket.

Összegezve: a keresőmotor előnyére válik a multimédiás keresési találatokat integráló „teljes nézet” megjelenítés, az „okos doboz” és a keresési előzmények raktározása. Hátránya, hogy mind ezen szolgáltatások csak az USA-ban érhetők el.

### Ask

[www.ask.co.uk](http://www.ask.co.uk)

Találati relevancia: 41%

Az Ask keresési algoritmusának különlegessége, hogy a találatok rangsorolásakor figyelembe veszi, hogy az adott témájú oldalra hány olyan csatoló

mutat, amely ugyanolyan témájú oldalakon található. A tesztelőket a kereső olyan finomságai nyűgözték le, mint amilyen például az egyes találatok mellett megjelenő távcső-ikon, amely fölé helyezve a kurzort, megjelenik a weboldal kicsiny képe (thumbnailje). A szponzorált csatolók azonban az AOL esetén tapasztaltakhoz hasonlóan nem különböznek el feltűnően a többi találattól.

A kereső jól teljesített a testreszabás terén: egyfelől szabályozható az oldalanként megjelenítendő találatok száma, másrészt az egyes találatok külön ablakban is megnyithatók. Az összetett keresés az AOL-éhoz hasonló beállítási lehetőségeket tartalmaz, a dátum szerinti visszakeresés azonban nagyobb mértékben finomítható.

A felnőtt tartalom szűrése alapértelmezésben ki van kapcsolva.

A keresési eredmények alakulása:

- az első keresőkérdésre egyetlen „alma pite” recept vagy egyéb releváns tartalmú oldal sem érkezett találatként, a szponzorált csatolókat leszámítva;
- a new crossi vízvezeték-szerelőkre vonatkozó kérdésre két releváns válasz érkezett;
- a negyedik keresőkérdésre („Raleigh”) hozott találatok között első helyen szerepelt a kerékpárgyártó hivatalos honlapja; a keresőfeltételek szűkítése („Narrow Your Search”) eszköz pedig az angol kerékpár szó (bicycle) több formáját, szinonimáját is felsorolja keresési javaslatként;
- jóllehet elviekben az Ask egy „tégypel egy kérdést – megkapod a választ” elven alapuló szolgáltatás, az egyetlen direkt kérdésre („What time is it in Bangalore?”) mindösszesen egy helyes válasz érkezett; a bájos a dologban, hogy a válasz ugyancsak korrekt módon, teljes mondatban nyert megfogalmazást a találati oldal tetején, vagyis: „The time in Bangalore is...” a mondat alatti keresőmezőt használva megtudhatjuk, hogy más településeken mennyi a helyi idő;
- a JFK merénylet dátumát adott formában beírva (22/11/1963) megtudható, hogy a képlet eredménye: 0.0010188487; ha magáról a merényletről szeretnénk bővebb információ birtokába jutni, mondatszerűen kell fogalmaznunk: „What happened on 22/11/1963”;
- a kilencedik keresőkérdésre („pas de bourre”) hat releváns találat érkezett, ebből négy angol nyelvű.

Jóllehet az Ask esetén a találati referencia alacsony, a keresési finomítások jól működnek, s



gyakorta kapjuk meg a találati oldal tetején a választ a feltett kérdésünkre.

*Előnye* tehát: a keresés finomítását szolgáló eszközök. *Hátránya*: tolakodó szponzorált ugrópontok, alacsony találati relevancia, valamint hogy a „felnőtt tartalom” szűrése alapértelmezésben ki van kapcsolva.

## Google

[www.google.co.uk](http://www.google.co.uk)

Találati relevancia: 70%

A keresőmotorok piacvezetőjeként számon tartott Google a maga mintegy 20 billió oldalt tartalmazó adatbázisával bevételei egy részét ma már a keresési szolgáltatásai bővítésére költi. Az Information World Review munkatársai a tesztelés során a cég hagyományos szolgáltatására, a webkeresésre koncentráltak.

A testreszabást jónak találták: egyfelől meghatározható az oldalanként megjelenítendő találatok száma, a felnőtt tartalom szűrésének három fokozata van: a középső csupán a képes tartalmat szűri, a szövegeset nem<sup>1</sup>. A nyelvi beállítás kiválóan működik: nem csak a felhasználói felület nyelve változik meg, hanem a keresés nyelv és ország szerinti szűkítése is. A Google a Föld számtalan lakójához „szól az anyanyelvén”, valamint egy webfordító szolgáltatás is a részét képezi. A szponzorált ugrópontok feltűnően elkülönülnek.

A tesztelés tapasztalatai:

- az első keresés (alma pite) az első oldalon kilenc releváns találatot hozott;
- a második (összetett logikai) keresőkérésre az elsőhöz hasonlóan kilenc releváns találat érkezett;
- a vízvezeték-szerelőkről szóló kérdés ugyanolyan eredményt hozott, mint az AOL esetén;
- a Raleigh kerékpárokra vonatkozó keresés három releváns találatot eredményezett;
- a bangalóri helyi időre vonatkozó kérdésünkre megkapjuk a választ az oldal tetején, a további öt releváns találattal egyetemben;
- a keresett termékről a tizből nyolc oldal nyújtott releváns információt, valamint lehetőség nyílt összehasonlítani az oldal tetején felsorakoztatott kereskedelmi ajánlatokat;
- a tizből nyolc találat azonosította a BBC panaszvonalát;
- a JFK elleni merényletre vonatkozó kérdés hat releváns találatot hozott;

- az utolsó keresőkérés (Information World Review) tíz releváns csatolót hozott találatként.

A Google tehát jól teljesített, és ami a találati relevanciát illeti, ha nem is jelentősen, de megelőzte a Yahooot. *Előnyei*: az adatbázis mérete, az intelligens sorrendezési algoritmus, a keresés szűkítésének lehetősége. *Gyöngéje*: a lokalizációs keresés.

## Microsoft

[www.live.com](http://www.live.com)

Találati relevancia: 52%

Az MSN Search utódjának számító Microsoft Windows Live Search, ha lehetséges, még a Google-nél is letisztultabb megközelítést kínál.

A könnyen áttekinthető, egyszerű külső könnyen áttekinthető, egyszerű belsőt takar. Az első keresés előtt csupán a nyelvi beállításon lehet változtatni. A keresés lefolytatása után azonban a „Beállítások”-ban lehetőség nyílik a testreszabás finomítására: meghatározható, egy oldalon hány találat legyen, és az azonos oldalról származó találatok közül hány jelenjen meg, illetve ez utóbbi találatok csoportba rendezhetők. A felnőtt tartalom szűrése háromfokozatú, és alapból „enyhére” van állítva.

A „Speciális” menüpont (angol verziónál: Advanced search) a keresési feltételek sokrétű finomítását teszi lehetővé ugyancsak egyszerű, könnyen áttekinthető módon. Az egyes tényezők (frissítés gyakorisága, népszerűség, egyezés pontossága) csúszkák segítségével súlyozhatók.

A szponzorált ugrópontok diszkréten jelennek meg, akár a Google-ben.

A tesztkeresések eredményei a következőképp alakultak:

- az első keresés (almapite) sajnálatosan csak egyetlen releváns találatot hozott;
- a második keresés viszonylag jó eredményt hozott annak ellenére, hogy a NEAR operátort a kereső a keresendő szöveg részeként értelmezte;
- a kereső két vízvezeték-szerelőt talált New Crossban;
- a negyedik keresés tíz találatából hat vonatkozott a Raleigh típusú kerékpárokra;

---

<sup>1</sup> Ez a funkció magyar nyelvi beállítás esetén nem működik! (A ref.)

- a bangalóri helyi időre vonatkozó kérdésre kapott három korrekt válasz közül egyik sem került az első néhány találat közé;
- a JFK-merénylet dátuma mindössze két releváns találatot hozott.

Mindent összevetve, a tesztelés alul múlta a várakozásokat. Ahogy a szerző fogalmaz: ha ez minden, amire a Microsoft képes, akkor a Google-nek és a Yahoónak nincs mitől tartania.

A keresőmotor *előnye*: átlátható, letisztult felhasználói felület, keresési feltételek súlyozása. *Hátrány*: kevés releváns találat, hiányzik a keresés finomításának lehetősége.

### Orange

<http://search.orange.co.uk>

Találati relevancia: 28%

Az Orange egyik legnagyobb problémája a szponzorált csatolók kezelésével kapcsolatos. Ezek egyfelől nem különülnek el látványosan a többi találattól, másfelől általában túlsúlyt képeznek az első találati oldalon. A gond, hogy ezek az ugrópontok nem az oldalon megjelenített tizenkét találat mellett, hanem azok között jelennek meg.

Az egy oldalon megjelenítendő találatok száma nem határozható meg. A keresett kifejezés kiemelésére van lehetőség, ugyanakkor a felnőtt tartalom alapértelmezett szűrése helyett csupán az „engedélyezés kérése a felnőtt tartalom megjelenítése előtt” opció van alapértelmezésben beállítva. Egyéb alapértelmezett beállítás: az Orange nem keres PDF, PowerPoint vagy Excel típusú fájlokra, amíg ezt külön nem engedélyezzük neki.

A teszt alakulása:

- az első keresőkérdésnél a tizenkettőből hat találat szponzorált csatoló volt, és mind a komputer-cégre vonatkozott, holott a keresőkérdés szerint olyan oldalakat kerestek, amelyekben nem szerepel az „Apple Computers” szókapcsolat; a további hat találat közül is csak három volt releváns;
- az összetett logikai keresőkérdés értelmezése komoly problémát jelentett a keresőmotorok: egyetlen vonatkozó ugrópontot hozott találatként, ugyanakkor szponzoráltak nem szerepeltek a találati oldalon;
- a harmadik keresőkérdés (vízvezeték-szerelő) nem hozott releváns találatot;
- a Raleigh kerékpárokra vonatkozó kérdés meghozta a maga gyümölcsét: a vonatkozó négy

nem szponzorált cím mellett a kereskedelmi ajánlatok a kerékpárvásárlás terén is eligazították a keresőt;

- a bangalóri helyi időre vonatkozó kérdés zavarba hozta a keresőt: egyetlen találatot se kaptak rá;
- a BBC panaszvonalának telefonszáma négy releváns találatot hozott, ezekből kettő – meglepő módon – duplikátum;
- a Kennedy-merénylet dátuma a második helyen hozott releváns találatot, de az oldalon több vonatkozó találat nem volt, csupán még egy helyes dátumértelmezés, amely egy másik eseményre vonatkozott;
- a francia balett-terminusra egyetlen vonatkozó találatot sem dobott ki a kereső;
- az Information World Review keresőkérdésre kapott hét találatot a cikk szerzője egyszerűen „ijesztőnek” titulálta.

A kereső *előnye*: a keresett kifejezések kiemelése a találatokban. *Hátránya*: tovakodó szponzorált csatolók, nagyon gyenge keresési eredmények, a konfigurálhatóság hiánya.

### Yahoo

<http://uk.search.yahoo.com>

Találati relevancia: 68%

A Yahoo elsősorban webtár, keresőmotorjának azonban 2004 óta saját adatbázisa van, és támogatja mind a logikai keresést, mind a helyettesítő jelek használatát. Bővelkedik a keresés szűkítésének lehetőségeiben, valamint van webfordító szolgáltatása is (Babel Fish).

A szponzorált csatolók egyfelől a többi találat fölött és alatt, halványkék háttérszínnel elkülönítve jelennek meg, másrészt az oldal jobb szélén, egy külön panelban. Alapértelmezésben a felnőtt tartalom szűrése ki van kapcsolva. Ha szeretnénk, hogy a felnőtt tartalom szűrése automatikus legyen, amikor egy 18 éven aluli felhasználó jelentkezik be a számítógépre, a „SafeSearch” funkció aktiválását elmenthetjük, de ehhez be kell jelentkezünk az azonosítónkkal.

A Yahoo nagy előnye, hogy ha van érvényes előfizetésünk, több adatbázisban is kereshetünk (pl. Factiva, LexisNexis).

A kereső egyik különlegessége a „Shortcuts” funkció. Ennek segítségével közvetlenül kereshetünk információkat egy település időjárásáról, a helyi időről, egy valuta árfolyamáról, egy mértékegység átszámításáról stb.

A teszt eredményei:

- az első akadályt (almapite) jól vette a Yahoo: a keresés nyolc releváns találatot hozott;
- az összetett logikai keresést az elsőhöz hasonlóan jól értelmezte a kereső;
- a harmadik keresés viszonylag sikeresnek mondható: a lap tetején található „Local Results” feliratú ugrópont három olyan vízvezeték-szerelő elérhetőségét tartalmazza, akik New Cross negyedről kevesebb, mint egy mérföldre található, és mindezt térképpel együtt; a többi csatoló között egy releváns volt található;
- a bangalori helyi időt az oldal tetején, kiemelt helyen található feliratból tudhatjuk meg (Local time in Bangalore, India is ...); a keresés további öt releváns választ hozott;
- kellemes csalódást tartogatott a tesztelők számára a termékinformáció (iPAQ hw6900) keresése: a tízből kilenc találat relevánsnak bizonyult, s egyetlen reklám sem próbálta a keresőt

rávenni, hogy megnézzék, az adott termék hol mennyibe kerül;

- a számok kezelése nem okozott gondot a Yagoonak: a tízből hét találat a BBC-hez irányította el a keresőt;
- a dátumértelmezés sem jelentett problémát: hat találat vonatkozott a Kennedy-gyilkosságra;
- a keresőt kicsit zavarba hozta a francia terminus: a hét releváns találatból mindössze három volt angol nyelvű;
- a folyóirat címe (Information World Review) kilenc releváns találatot hozott.

A Yahoo *előnyei*: a rugalmassága, az előfizetéses adatbázisokban való keresés lehetősége és a logikai keresés támogatása. *Hátránya*: kevésbé pontos, mint a Google.

**/WINDER, Davey: The search engine that really hits the spot. = Information World Review, 234. sz. 2007. április, p. 19–24./**

(Dancs Szabolcs)

---

## Miért szűnik meg a folyóiratok kötetése a Hongkongi Baptista Egyetem Könyvtárában?

A *Hongkongi Baptista Egyetem (HKBE)* 10 ezer hallgatója több campuson tanul az előkészítő két-éves képzéstől kezdve az összes szinten, a doktori fokozat megszerzéséig bezárólag. A könyvtár eddig több mint 3 ezer folyóiratot járatott nyomtatott formában (ebből 1800 angol és egyéb nyugati nyelvű, a többi kínai). A növekvő hallgatói létszám elhelyezése egyfelől, az emelkedő előfizetési árak másfelől, új megközelítést igényelnek. A könyvtár, tapasztalva az elektronikus változat iránti növekvő érdeklődést, 2006 tavaszán felmérést végzett a felhasználók körében a nyomtatott, illetve az e-változat népszerűségének megismeréséért. Az előző években a nyomtatott folyóiratok előfizetése már folyamatosan csökkent (az angol nyelvűek körében pl. a korábbi 3 ezres nagyságrend közel 40%-kal), miközben rohamosan nőtt az e-változatok száma (11,5-ről 24 ezerre). A hongkongi egyetemek közös tárolókönyvtár létesítéséhez kezdtek (JURA-projekt), minthogy mindenütt hasonló gondok jelentkeztek.

Ezek a körülmények együttesen kellően indokolták, hogy alaposan átgondolják az eddigi folyóirat-kötetési folyamatot, amely mindig külső kötészetnél történő megrendelésre épült. Természetesen

az e-szolgáltatás is egyre drágább. Emellett érdemes tudatosítani, hogy a kötetés elég számottevő belső technológiai munkát, vagyis időráfordítást (éves szinten 1200 munkaórát) igényel, és ennek költségét is szem előtt kell tartani.

A munka első szakaszában gondosan áttanulmányozták a vonatkozó szakirodalmat, hogy világosabban lássák a kötetés – vagy annak elmaradása – esetén jelentkező pozitív és negatív (állományvédelmi, szolgáltatási stb.) következményeket. Különösen érdekes volt e szempontból a *Nevadai Egyetemi Könyvtár* radikális döntését – a további kötetésről való lemondást – indokló tanulmány *R. Anderson* tollából (*The Serials Librarian, 2003. 3–4. p. 255–260.*). Megvizsgálták a kötés kiváltására alkalmas egyéb, olcsóbb technológiákat is (műanyag fóliázás stb.).

Legérdekesebbek azok az elemzések voltak, amelyek az elektronikus környezetben felmerülő megoldásokat és költségeiket összehasonlították a hagyományos papíralapú dokumentumok hosszabb távon történő megőrzésével és szolgáltatásával. A *CLIR (Könyvtárak és Információs Források Tanácsa = Council on Library and Information*

*Resources*, Washington) nevű kormányzati tanácsadó testület által fémjelzett elemzés különösen mélyrehatóan vizsgálta a digitális gyűjtemények irányába történő átmenet kérdéseit, a hagyományos és/vagy elektronikus változattal járó tényleges kiadásokat (*The nonsubscription side of periodicals: Changes in library operations and costs between print and electronic formats. 2004, 66 p.* – hálózati címe: <http://www.clir.org/pubs/reports/pub127/pub127.pdf>).

A HKBE szakemberei gondosan tanulmányozták a kötetést kiváltani képes alternatív megoldásokat, és különösen előnyösnek ítélték az egyes számokat gyűjtő kartondobozt mint könnyen realizálható lehetőséget. Ezek után még meg kellett határozni, mely időszaki kiadványoknál érdemes erre a megőrzési alternatívára áttérni. Itt a használatelemzés módszere kínált racionális megközelítést. Kiderült, melyik kiadvány bekötött évfolyamait hányszor használták három év alatt.

A folyóiratok nagyobbik felét éves szinten egyszer sem kérték (az egyes években az 1260-ból 680,

772, és végül 889 címet); az öt vagy annál többszöri igénybevételt mindössze a címek 15–12–9%-a vonzotta 2003-ban, 2004-ben, illetve 2005-ben. A három év alatti használatot összesítve kiderült, hogy a címek negyedét egyetlen kézbevitelre sem méltatták. Így végül arra a következtetésre jutottak, hogy elég a viszonylag gyakrabban, vagyis legalább ötször használt címek (az összesnek alig 9%-a, azaz 109 kiadvány) esetében a kötetést a dobozos módszerrel felváltani. A használóknak nem kell ezekre sem várni egy-két hónapot, míg régebben ennyi volt a kötetési idő, hiszen a dobozos anyagot rögtön el lehet helyezni a raktárban, és minden megszakítás nélkül szolgáltatható (másolásnál is kényelmesebb e módszer a vastag, kötött folyóirat-évfolyamok használatánál).

**/FROST, Janet–WOO, Wing: Binding dilemma: Changing horses mid stream – Or why we stopped binding periodicals at HKBU. = Library Collections, Acquisitions & Technical Services, 31. köt. 2. sz. 2007. p. 85–95./**

(Sonnevend Péter)

## Közös Cseh és Szlovák Digitális Parlamenti Könyvtár

Széles körben elfogadott az a nézet, hogy az állampolgárok és a jogalkotó szervek kapcsolatát jelentősen meghatározza a parlamenti könyvtárak által nyújtott különböző információs szolgáltatások színvonala – függetlenül attól, hogy nyilvános vagy a nyilvánosság számára zárt könyvtárról van-e szó. Ezt a tételt ismerte fel a *Cseh Parlament Könyvtára* is, amikor 1995-ben a *Cseh Parlament Digitális Könyvtára* létrehozásáról döntött. A projekt célja az volt, hogy megteremtse a különböző parlamenti dokumentumok digitális adattárát, legyen szó kurrens vagy történeti anyagokról – ez utóbbiak esetében 1861-től kezdődően. 1997-ben az adatbázist (amely akkor 1918-tól tartalmazott dokumentumokat) elérhetővé tették az interneten is. A projekt jelentősen növelte a Cseh Parlament ismertségét és elismertségét; biztosítja a parlamenti dokumentumok fennmaradását bármely, előre nem látható természeti vagy más katasztrófa esetén; s megkönnyítette a parlamenti képviselők jogalkotó, döntéshozó tevékenységét a dokumentumokhoz való egyszerű hozzáférés révén.

A projekt hazai sikere arra ösztönözte a könyvtárat, hogy nemzetközi szinten is hírt adjon az új

adatbázisról. A *Parlamenti Kutatások és Dokumentumok Európai Központja (European Centre for Parliamentary Research and Documentation = ECPRD)* segítségével először 1998-ban szerveztek nemzetközi szemináriumot, ahol 15 intézmény mutatta be hasonló projektjét, és próbált közösen választ keresni a felmerülő problémákra. A további találkozók során merült fel egy nemzetközi projekt létrehozásának ötlete: először az osztrák parlamenti könyvtárral terveztek közös digitalizálási programot, végül a szlovák intézménnyel állapodtak meg.

A közös cseh és szlovák projekt alapját, természetesen, a közös múlt és a csehszlovák állam 1918–1992 közötti fennállása adta. Politikusok, informatikusok és könyvtárosok több évig tartó egyeztetése után végül 2002-ben írta alá a cseh és a szlovák parlament azt a megállapodást, amely egy közös digitális parlamenti könyvtár létrehozását mondta ki. A megállapodás értelmében a közös digitális könyvtárnak tartalmaznia kell az összes parlamenti dokumentumot (javaslatok, döntések, határozatok, jegyzőkönyvek stb.) 1848-tól napjainkig.

A közös projekt alapötlete az volt, hogy mind a cseh, mind pedig a szlovák parlamenti honlapok ugyanazt a tartalmat kínálják. Ennek érdekében először a Cseh Parlament bocsátotta a szlovákok rendelkezésére az 1995 óta digitalizált parlamenti anyagokat (a cseh nemzeti történeti dokumentumokat és a csehszlovák időszak dokumentumait egyaránt); a szlovák oldal pedig hozzákezdett a szlovák nemzeti anyagok, a hiányzó csehszlovák dokumentumok és a XIX. század végén született cseh történeti dokumentumok digitalizálásához. Ezt azután átadták a cseh partnernek.

A projekt sikeres megvalósítását jelentősen könnyítette a cseh és a szlovák nyelv hasonlósága (nem volt nyelvi probléma); több megbeszélés tárgya lett viszont a rendszer szerkezetének kialakítása, hiszen a különböző történeti időszakok nem feltétlenül estek egybe a választási időszakokkal, különösen a háborús években. Végül a választási időszakok szerint építették fel az adatbázis struktúráját.

A digitalizálási munkálatok elvégzésére közbeszerzéssel a szlovák NUPSESO (<http://www.nupseso.sk/>) céget választották ki, amely az ABBYY FineReader 7.0 nevű szoftver fő disztribútora. A szoftver képes a legkülönbözőbb, különleges karakterek felismerésére is (l. a XIX. századi dokumentumokat az Osztrák–Magyar Monarchia idejéből). A másik, a projektben részt vevő szlovák cég az *exe-Information Technologies* (<http://www.exeit.sk/>), amely az adatbázisszoftver fejlesztéséért felelős, beleértve az export/import, keresés stb. funkciókat, valamint az adatvédelemmel kapcsolatos kérdéseket egyaránt.

A dokumentumokat HTML és PDF formátumban is elérhetővé teszik, így módon azok eredeti formában, küllemben is olvashatók.

A Cseh Parlament 1993-tól, a Szlovák Parlament pedig 2003-tól valamennyi új dokumentumot elektronikus formában is elérhetővé tesz, és automatikusan betöltik őket a digitális könyvtárba.

A további tervek között szerepel a még hiányzó történelmi időszakok olyan dokumentumainak feldolgozása – más intézményekkel (pl. *Szlovák Központi Levéltár*) való együttműködés révén –, amelyek nem találhatóak meg a Parlamenti Könyvtárban vagy a Parlamenti Levéltárban.

### A projekt fő adatai

Időtartam: 2002-től napjainkig;

Résztvevők: Cseh Parlament, Szlovák Parlament;

Tartalom: a cseh, a szlovák és a csehszlovák parlamentek dokumentumai teljes szöveggel, digitalizált formában, 1848-tól napjainkig;

Digitalizált oldalak száma: kb. 2 millió;

Projektköltség: eddig 970 000 USD;

Elérhetőség: a cseh és a szlovák parlament honlapján egyaránt.

### További információ:

- Angolul: Joint Czech and Slovak Digital Parliamentary Library: [http://www.psp.cz/cgi-bin/eng/kps/knih/a\\_elknih.htm](http://www.psp.cz/cgi-bin/eng/kps/knih/a_elknih.htm)
- Cseh nyelven: <http://www.psp.cz/eknih/>
- Szlovákul: <http://www.nrsr.sk/dk/>

/MALACKOVA, Eva–SOSNA, Karel: *The Joint Czech and Slovak Digital Parliamentary Library*. = *IFLA Journal*, 33. köt. 3. sz. 2007. p. 251–257.)

(Koreny Ágnes)

---

## A katalogizálás végnapjai közelegnek?

A *British Library* (BL) vezető szakembere – egyben számos nemzetközi szakértői csoport résztvevője – szerint a katalogizálás (feltárás) számos, ráadásul egyre növekvő kihívással szembesül. Hogyan értelmezzük, s főként hogyan kezeljük e kihívásokat, hogy túléljük szorongatásukat?

A jelen dolgozatban az elemzés érdekében – széles értelmezésben – a katalogizálás/feltárás lényegéről az alábbi megállapítások mondhatók el:

- a forrás olyan leírása, amely alkalmas annak azonosítására, egyben más forrásoktól való megkülönböztetésére,

- a hozzáférési (elérési) pontok azonosítása és azok kezelése,
- a más forrásokhoz fűződő kapcsolatok azonosítása és kezelése,
- a forrás tartalmi elemzése,
- a tárgyszavak, indexelési kifejezések felsorolása,
- az osztályozási jelzetek felsorolása.

Növekvő kihívásként a következő tényezők említendők:

- egyre növekvő feldolgozandó input (egyedül az Egyesült Királyságban évente 160 ezer új monografikus kiadvány, évi közel 10%-os folyamatos növekedés mellett),
- új információforrások garmadája (a BL már a hálózati forrásokat is regisztrálja),
- a könyvtár számára fokozódó versenyhelyzet más szolgáltatásokkal,
- a feltárás olyan közfelfogása, amely drágának, viszont csekély értéket teremtőnek tartja,
- pénzügyi korlátok (pl. a BL a normál működéshez szükséges pénzügyi források háromnegyedét kapja a kormányzattól, ami mindössze évi átlagos 0,75%-os növekedést mutat),
- csökkenő munkaerőalap.

A világháló gyökeresen átalakítja a publikálás eddigi jellegét. Korábban a nyomtatás, továbbá a terjesztés és marketing bonyolultsága, költséges jellege azt eredményezte, hogy viszonylag kis számú szereplő érvényesült ebben az iparágban. Ma kis túlzással bárki, aki elszánja rá magát, s rendelkezik számítógéppel meg hálózati kapcsolattal, saját maga kiadójává válhat. A könyvtári világban kellene ugyanakkor biztosítania az új műfajok – ideértve a preprinteket, blogokat, honlapokat és egyébeket –, és az örökké változó időszakos kiadványok stb. gyűjtését, feltárását, megőrzését. 2006-ban már 4 millió „.uk” végű domént ismertek.

Korábban a katalógusok (bibliográfiák stb.) közvetítették az új kiadványokról tudósító adatokat az érdeklődőkhöz. Ma számos *versengő szolgáltatás* ismert. Legelterjedtebbnek a *Google* és társai nyújtotta böngészés nevezhető, de említhető – számos más mellett – az *Amazon*, miközben ezek híre egyértelműen az, hogy a nehézkes és bonyolult könyvtári feltárásnál egyszerűbbek, vagyis használóbarát módon szolgáltatnak.

A feltárás tényleg költséges művelet. Egy 2004-re vonatkozó amerikai felmérés arról számol be, hogy a mintegy 120 észak-amerikai *tudományos könyvtár* 239 millió – ebből egyedül a *Kongresszusi Könyvtár (Library of Congress = LC)* 44 millió –

USD-t költött az idevágó munkákra. A BL esetében a 2005-tel kezdődő pénzügyi évben közel 6 millió fontot, a teljes éves kiadás 5%-át költötték feltárára.

Lesz-e képzett munkaerő? Az amerikai adatok azt mutatják, hogy az e területen dolgozók harmada a következő 5–6 évben nyugdíjba megy (az LC-ben még fenyegetőbb a helyzet). Ráadásul a felsőfokú képző helyeken a tematika veszít súlyából, és a feltárást oktató tanárok jó része is egyre korosabb.

Mindez együtt kellően indokolja az alapvető újragondolást. Ennek két iránya van:

- releváns-e még a mai hálózati környezetben a (könyvtári) feltárás, és a belátható jövőben is az marad-e;
- ha igenlő választ adunk a fenti kérdésre, akkor hogyan kell változnia, hogy tényleg releváns maradjon, tehát a kihívásoknak megfeleljen?

Az első kérdésre – rövid vagy középtávon feltétlenül – igenlő válasz adható, hiszen a hagyományos kiadványok (nyomtatott könyvek stb.) mennyisége még mindig növekszik. Az előrejelzések főként az újabban feltörekvő gazdaságok (Kína, India, Brazília stb.) esetében biztosak és jelentős mennyiségi gyarapodást mutatnak, és párhuzamosan, az eddigi kiadási „nagy hatalmak” (USA, Egyesült Királyság stb.) is növelik outputjukat. Az átlagolvasó egyelőre még idegenkedik az e-könyvektől, de nem állítható, hogy ez akár 5–10 év múlva is így lesz (vannak „optimista” előrejelzések, miszerint öt év múlva minden új könyv elérhető lesz a hálóról). A nyomtatott könyv éppen úgy eltűnhet, mint a korábbi lemez, kazetta, – és újabban a CD-k? Az e-könyvek pedig a hálón akár egy-két tárgyszó révén könnyedén megtalálhatók lehetnek, s akkor mire a könyvtári feltárás? Más becslések szerint akár az egész nyomtatott könyvkincs 20 éven belül digitalizálva lesz. Egyszerű tárgyszavas keresés elégséges lesz az ebben való eligazodásra és keresésre? Sem a Google, sem a Microsoft nem így gondolja, hiszen a tömeges digitalizálásnál felhasználják a partnerkönyvtárak katalógizálási adatait.

A feltárás sokkal több, mint egyszerűen a forrás leírása. Emellett segít felderíteni azt a kontextust, amely érdekes szellemi gyarapodást kínál a használónak. Például egy szerző milyen más munkákat írt, mi jelent meg tőle esetleg álneveken, hol fordították le munkáit más nyelvre. A feltárást végző szakember munkájaként olyan „térkép” jön létre, amely segíti az olvasót az emberi faj intellektuális

eredményeinek megismerésében, a rögzített tudás felderítésében. Vannak már fejlett webtechnikák, amelyek jól ötvözik a hagyományos katalógus és a mai OPAC-ok sajátosságait a világ tudástengerében történő navigáláshoz.

Az utóbbi években az LC, a BL és más könyvtárak is keresik az új helyzettel adekvát megoldásokat. Először is tudatosítani kell, hogy korábban sem dolgoztak fel a könyvtárak mindent. Ezt a megközelítést érdemes a jövőben is alkalmazni: külön választva a gyűjteményszervezést (archiválást) és a feltárást. A hálón lévő anyagok jelentős része nem érdemel bibliográfiai leírást, megőrzésük viszont tényleges feladat. Az LC-ben és az IFLA szakértői munkacsoportjában azon dolgoznak, hogy hogyan lehetne az önfeltáró anyagokat kellően definiálni, egyben a könyvtári feldolgozásra jelölteket elválasztani. Elemi érdek a könyvtárak egymás közötti, a kiadókkal, valamint más kibocsátó/terjesztő szervezetekkel való érdemi együttműködése. Egészében az állapítható meg, hogy

az eddigi „kisipari” módszerekről, a fragmentált erőfeszítésekről át kell térni a „nagyipari” megoldásokra. Ez várhatóan azzal is jár, hogy a feltárást terjedelmes és folyton változó szabályait áramvonalasítani kell. A „mindent” akarása helyett a végrehajtható megoldásokat célszerű preferálni. Azt kell szem előtt tartanunk, hogy fő tennivalónk az egyes használó specifikus igényeit kielégítő megoldások elősegítése. Óvatos optimizmussal nézhetünk a jövőbe. A web sem lehet meg „katalógusok” nélkül (a sikeres példák közül érdemes megemlíteni az *Internet Movie Databases* bibliográfiai eszköztárát). Megjelent a bárki számára saját katalógizálást segítő eszköz (LibraryThing). Bizonyos értelemben egyszer mindnyájan katalógizálók leszünk.

**/DANSKIN, Alan: „Tomorrow Never Knows”: the end of cataloguing? = IFLA Journal, 33. köt. 3. sz. 2007. p. 205–209./**

(Sonnevend Péter)