



Tudományos és Műszaki Tájékoztatás

Könyvtár- és információtudományi
szakfolyóirat

56. évfolyam
HU ISSN 0041-3917

2009/10

Tartalom

Ajánló

SZÁNTÓ Péter: **Újszerű megoldások**

Cikkek

- BÁNKESZI Katalin – TAMÁSI Balázs: **Hatszáz ezer különleges könyvtári dokumentum és a Könyvtár 2.0 – Egy újszerű szolgáltatási forma terve a nemzeti könyvtárban**.....455
- LENGYELNÉ MOLNÁR Tünde: **Emberi vagy gépi kivonatolás?**466
- DRÓTOS László – KÖRÖS Kata – TERNAI Zita: **Adattármustra. Múzeumok**.....475

Beszámolók • Szemlék • Referátumok

- MAXYMUK, John: **A Google keresőjén túl** (Ref.: Drótos László).....491
- MILES, Mathew J. – BERGSTROM, Scott J.: **Van-e optimális kategóriaszám a könyvtári honlapok forrásgyűjteményeinél?** (Ref.: Drótos László).....493
- ANGEVAARE, Inge: **Digitális gyűjtemények és adatok felügyelete: „kurátorság” és más alternatívák a tudományos könyvtárakban** (Ref.: Drótos László).....495

Hírek

- A Gale új adatbázisa: Greenr • A Springer elindította az *Authormapper.com* szolgáltatást • Az Oxford University Press kiadta a Biblical Studies Online-t • Az MIT nyílt hozzáférést nyújt a tudományos cikkeihez • Az Alexander Street új terméke: Music Online Az EBSCO fogászati adatbázisa, a Dentistry • World Digital Library
(Összeállította: Szántó Péter) 499



Scientific and Technical Information

Hungarian journal of library and information science

Vol. 56 • 2009/10

Summaries

BÁNKESZI, K. – TAMÁSI, B. Special documents and Library 2.0 – setting up an innovative service at the National Library of Hungary ... 455

One of the valuable special collections of the National Széchényi Library is the Collection of Small Prints, containing nearly 3,5 million documents. Obituaries constitute a particularly interesting part of the collection. Over the past years, significant efforts were made in order to restore the documents. The next step was the digitisation of nearly 600,000 documents in collaboration with a commercial partner. However, digital documents only fulfill their mission when they are provided with descriptive data and become accessible in a searchable database. As the National Library does not have the sufficient resources for the time-consuming cataloguing of this type of documents, a decision was made to involve library users into the cataloguing process. This type of collaboration is not yet a wide-spread practice in library environment, but the innovative option was taken in order to make the obituary collection accessible for readers and researchers without further delay. This article summarises the outcomes and conclusions of the digitisation project and also outlines the plans for the

LENGYEL MOLNÁR, T.: Abstracting by people or computers? – A research project comparing the efficiency of the human creation of abstracts and automated abstracting466

The paper reports on a research project aiming to facilitate automated abstracting in Hungarian language. In the framework of the project, the author developed an offline abstracting tool in Hungarian language. The efficiency of the first version of the programme was tested in comparison with summaries made by people. 340 people participated in the project, producing abstracts of articles from a wide range of scientific disciplines. The results of the project showed that the greatest differences were observed not between the automated summaries and the abstracts made by people, but between abstracts made by different people on the basis of the same article. The paper shows how the subject of the article affects the abstracting process, and draws useful conclusions regarding the process of automated abstracting.

DRÓTOS, L. – KÖRÖS, K. – TERNAI, Z.: Scrutiny of Hungarian databases. Museums475

In the series "Scrutiny of Hungarian databases" independent experts evaluate various Hungarian online content services from the point of view of their use by libraries.

Hatszázézer különleges könyvtári dokumentum és a Könyvtár 2.0 – Egy újszerű szolgáltatási forma terve a nemzeti könyvtárban

Ma már nem szokatlan dolog, hogy a könyvtárak korszerű szolgáltatásokkal színesítik a kínálatukat: megjelennek a közösségi oldalakon, blogot írnak, Skype üzenetekben fogadják a referenzkérdéseket, sms-ben értesítik az olvasót az újdonságokról, hogy csak néhány példát említsünk. Ma már a Könyvtár 2.0 nem ismeretlen fogalom, hiszen a gyűjtemények weboldalain utat tör magának a „web2-es” szemlélet. Mégis, első hallásra merész döntésnek tűnik, hogy a nemzeti könyvtár úgy teszi elérhetővé gyűjteményének egy szeletét, nevezetesen több százezer digitalizált gyászjelentést, hogy a feltáró munkába bevonja a dokumentum iránt érdeklődő kutatókat, olvasókat.

Előszó

Az Országos Széchényi Könyvtár (OSZK) milliós nagyságrendben őriz és szolgáltat kisnyomtatványokat, többek között gyászjelentéseket is, melyekből eddig közel 600 000 darab digitalizálása készült el egy külföldi partner közreműködésével. Mielőbb szeretnénk ezt a különleges és értékes anyagot elérhetővé tenni a felhasználók számára, de ehhez az kell, hogy a kereshetőséget lehetővé tevő leíró adatok is elkészüljenek a dokumentumokhoz. Jelenleg még cédulakatalógus formájában sem áll rendelkezésre a szükséges információ, ezért előttünk áll az a nem kis feladat, hogy a dokumentumokat feltárjuk, és az adatokat elektronikusan feldolgozzuk, adatbázisba szervezzük. Figyelembe véve a digitális állomány mennyiségét és azt a tényt, hogy a fizikai gyűjtemény évekig tartó rendezése eddig is túlzott terheket rótt az OSZK kis létszámú Plakát- és Kisnyomtatványtárának munkatársaira, meglehetősen kevés esélyt látunk arra, hogy ezt saját erőből belátható időn belül megoldjuk. Tekintettel az olvasói/kutatói igényekre és a szűkös kapacitásra, úgy döntöttünk, hogy a nemzeti könyvtár weboldalain létrehozunk egy „közösségi” szolgáltatást, és így próbáljuk meg felgyorsítani a feltáró munkát, hogy az érdeklődők mielőbb használatba vehessék a gyűjteményt.

Tudomásunk szerint még nem volt arra példa, hogy egy értékes gyűjtemény feltárásának és elérhetőségének kialakításában a könyvtárosokon kívül az érdeklődő könyvtárhasználók is részt vehessenek, pedig leginkább az ő érdekük, hogy minél több feltárt információ között kereshessenek.

A szakemberek mindig tartanak attól, hogy „laikusokat” vonjanak be a szakmai munkába. De vajon laikusnak tekintendő-e az, aki kutatói szinten, vagy akár hobbiként évek óta foglalkozik az adott dokumentumtípus kutatásával? A családfakutatók sok esetben közösségekbe tömörülnek, ahol megosztják egymással tudásukat, tapasztalatukat. Ezeket a csoportokat szeretnénk felkérni a közös munkára, a gyűjtemény feltárása és hasznosítása érdekében. Hiszünk abban, hogy akik egymás között megosztják a tudásukat, hajlandók a közös tudás gyarapítására is, vagyis saját kutatói munkájuk megkönnyítése érdekében szívesen részt vesznek az adatok javításában, rendezésében, feltárásában.

Bevezető

Köztudott, hogy a digitalizálás számos problémát vet föl: az állományvédelmi, megőrzési és technológiai szempontokon kívül a leíró adatok, a szolgáltatás és az újrahasznosíthatóság lehetőségeit is vizsgálnunk kell a projektek tervezése során.

A nagy tömegű digitalizálás technikailag ma már nem okoz igazán gondot, mivel egyre fejlődnek, specializálódnak az eszközök. Az adagolós szkennernek villámgyorsan beolvassák a lapokra szedhető könyveket, a képeslapokat és más egylapos, kisebb méretű dokumentumokat.

A gép önmagában persze akkor sem tud eredményesen dolgozni, ha egy sor műveletet automatizáltak rajta a fejlesztők. Mindig szükség van digita-

lizáló munkatársakra is, akik elvégzik a tesztek, gondoskodnak a megfelelő beállításról, a digitalizálható dokumentum előkészítéséről, a beolvasás felügyeletéről. A digitalizáló munkaerő beállítása nem jelent komoly problémát, ha az előkészítést és a rendszeres ellenőrzést hozzáértő szakemberre bízuk. Több digitalizáló gépet is üzemeltethet egy-két betanított munkatárs, ha előzetesen alaposan megterveztük a folyamatot. (Természetesen ez nem igaz a védett, sérülékeny, vagy más okból speciálisan kezelendő dokumentumokra, amelyeket adagolós gépeken nem tudunk digitalizálni.)

Mindezekből következik, hogy megfelelő anyagi és technikai háttérrel a digitális képfájlok rövid idő alatt nagy mennyiségben készülhetnek el, de ekkor még csak egy mesterpéldányunk van, amely ebben a formában nem igazán alkalmas arra, hogy eljusson az olvasóhoz. Képkonverziós eljárásokkal csoportosan is megoldható a nagyméretű képekből a könnyebben kezelhető és megjeleníthető formátum elkészítése. Ha könyvoldalak szkennelt képeivel dolgozunk, a digitális képeket PDF fájlba szerkeszthetjük a szolgáltatáshoz. Ez a legegyszerűbb és leggyorsabb módja annak, hogy egy könyv elektronikus formában is eljusson az érdeklődőkhöz. Ha a PDF fájlokban szövegrészleteket szeretnénk keresni, vagy a szöveges oldalakat szerkeszthető, kereshető formában is elérhetővé akarjuk tenni, akkor már bonyolódik a helyzet, szükség lesz a szöveg szoftveres felismertetésére és esetleg korrektúrázására is.

Ha az eredeti dokumentum csak képi információt tartalmaz, vagy csak képi információt akarunk vagy tudunk kinyerni belőle, valamivel egyszerűbb (?) a dolgunk, mert a gondos szkennelés és utólagos képrekesztés (körbevágás, kontrasztállítás, esetleg retusálás stb.) után már csak a kisebb, szolgáltatható formátumot kell előállítani ahhoz, hogy a kép könnyen megjeleníthető legyen egy távoli számítógépen.

Még nem beszéltünk azonban a leíró adatokról, amelyek nélkül a mégoly gyorsan gyarapodó archívum is nehezen kezelhető digitális fájlhalmaz marad, hiszen a fájlokat nem tudjuk azonosítani, visszakereshetővé tenni. Az esetek többségében itt jönnek a gondok és a megválaszolendő kérdések:

- rendelkezésünkre áll-e egyáltalán valamilyen adat az egyes dokumentumokról?
- ha igen, az adatok milyen formában léteznek: cédulakatalógusban vagy adatbázisban?

- a leíró adatok milyen részletesek? egy konkrét dokumentumhoz kötődnek, vagy azok egy csoportjához?
- használható-e, elegendő-e a digitális fájlok azonosításához és osztályozásához a jelenleg megtalálható adatok tartalma?
- esetleg most kell feltárni az anyagot?

Mindez persze messze nem a teljes problémakör feltérképezése, csak egy áttekintő előkészítés ahhoz, amiről a jelen írás szól: egy zökkenőkel tarkított projekt tanulságairól és egy újszerű szolgáltatásra vonatkozó döntés előzményeiről. Mindenesetre egyre gyakrabban fordul elő, hogy nagyon gyorsan, nagyon sok, új digitális dokumentum áll elő, de komoly gondot okoz az azonosító adatok kérdése, a dokumentumok kereshetőségének megoldása, pedig ebben rejlik az elektronikus dokumentumok használhatóságának egyik kulcsa.

Eddig főként általános tapasztalatokról beszéltünk, most nézzük a konkrét esetet...

Az Országos Széchényi Könyvtár gyászjelentés-gyűjteménye

A széles közönség számára digitálisan szolgáltatni kívánt gyászjelentések jelenleg az OSZK egyik legszerteágazóbb gyűjtőkörével rendelkező és legtöbb dokumentumot őrző különgyűjteményi tárában, a Plakát- és Kisnyomtatványtárban található meg. Tárunkban az 1711 után Magyarországon megjelent, vagy külföldön kiadott magyar vonatkozású kisnyomtatványok (azaz legfeljebb 16 oldal terjedelmű nyomtatványok) példányszáma megközelíti a 4 milliót.

Az OSZK méltán híres gyászjelentés-gyűjteményének projektszintű könyvtári feldolgozása és kutathatóvá tétele két évvel ezelőtt, 2007 augusztusában zárult le. Hosszú és fáradságos út vezetett odáig, hogy az online szolgáltatás gondolata reálisan felmerülhetett.

A gyászjelentés-gyűjtemény átfogó és tervszerű rendezésének háttérében az ún. „mormon program” állt. A genealógiai adatokat világszerte módszeresen gyűjtő, megőrző és szolgáltató *GSU (Genealogical Society of Utah)*, amelyet a Salt Lake City-ben (Utah állam, USA) székelő *Mormon Egyház* működtet, 2003 januárjában kötött megállapodást az OSZK-val a gyászjelentés-gyűjtemény mikrofilmzéséről és digitalizálásáról. A megállapodás a következő fontos részleteket tartalmazta:

egyik részről a dokumentumokat őrző Plakát- és Kisnyomtatványtár (akkori nevén Kisnyomtatványtár) vállalta, hogy a 2002. december 31-éig megjelent gyászjelentéseket mikrofilmmezés céljából rendelkezésre bocsátja; a GSU pedig kötelezte magát arra, hogy fedezi az OSZK által elkészített felvételek költségét, továbbá vállalta a negatívok kidolgozását és azt is, hogy a pozitív mikrofilmekből és azok digitális változatából egy példányt térítésmentesen átad az OSZK-nak.

A projekt megvalósulásának alapvető feltétele a gyászjelentés-gyűjtemény rendezése, revíziója és feldolgozása volt. E munkának minden nehézsége és súlya a Plakát- és Kisnyomtatványtár jelenlegi és a már nyugdíjba vonult munkatársait terhelte. A korábban több részgyűjteményből álló anyagot, amely mint utóbb kiderült, 533 000 gyászjelentést foglalt magába, egységes elvek szerint és új szabályzat alapján, szigorú betűrendbe rendezték a kollégák. Ez a munka a nagyalakú kéziratos és a XVIII–XIX. századi szomorúlevelek, szomorújelentések leválasztása után kezdődhetett el. A rendezés során több szempontot kellett figyelembe venni, például a ragadványneveket, vagy az azonos nevűek megkülönböztetését. A nyomdai kiadásváltozatokat és a nyelvi változatokat is el kellett különíteni. Az egyik nagyon fontos munkafolyamat a duplumok azonosítása volt, hiszen a fotózáshoz számos esetben ki kellett választani a kutathatóság szempontjából legjobb minőségű példányt.

A betűrendezett gyászjelentések végleges, beszámozott palliumokba kerültek, amelyekre a dokumentumokon szereplő családneveket is rávezették. Ezután listák készültek a restaurálásra való előkészítés, illetve fotózásra átadás céljából. A statisztikai kimutatások szerint a majdnem ötéves munka során az OSZK *Restauráló Osztályán* 2007 tavaszáig 11 543 (!) gyászjelentést 20 147 (!) munkaórán belül restauráltak. Az OSZK keretein belül működő korábbi két restauráló műhely munkatársai a rendezés és előkészítés során feszített tempóban végezték el ezt a hatalmas munkát.

A tár dolgozói a többszázszézes gyűjteményt nagyjából tízezres adagokban és számozott dobozokban adták át a *Fotólabor* számára, a neveket tartalmazó listával együtt. A mikrofilmek elkészítése után, az adott tekercs számát tüntették fel a listán szereplő családnevek mellett. E listákat a későbbiekben a mikrofilmeket őrző és szolgáltató *Mikrofilmtár* is megkapta, így lehetővé vált a családnevek kutatása a mikrofilmekben.

A tár legkutatottabb dokumentumai már a mikrofilmek elkészítése előtt is a gyászjelentések voltak. A hagyományos könyvtári dokumentumnak nem nevezhető, egylapos, ritkábban többlapos nyomtatványok számos értékes adatot és információt tartalmaznak. A szomorúlevelek, szomorújelentések és a gyászjelentések a legkülönbözőbb szakterületek kutatói számára szolgálnak értékes forrással. Első helyen kell megemlíteni a családfakutatókat és a biográfusokat, de a korszakra specializálódott történészek és szociológusok is jól használhatják e gyűjtemény darabjait. A nyelvtudománnyal és a nyelvtörténettel foglalkozók számára elsősorban a korábbi, XVIII. századi szomorújelentések érdekesek. A művészettörténészek e dokumentumokat formai szempontból kutathatják, hiszen a gyászjelentések gyakran tartalmaznak grafikus elemeket, vagy grafikai alkotásokat, továbbá iniciálét, díszítő motívumokat. A különböző felekezetek elhunytjai számára készített gyászjelentések az eltérő vallási jelképek használatára vonatkozóan szolgáltatnak adatokat. A nyomdatörténészek, sőt a vízjelkutatók is gazdag forrásanyagot találnak saját kutatási területükhöz a gyászjelentés-gyűjtemény egyes részeiben.

Az itt felsorolt kutatási szempontok közül természetesen a „mormon program” végső célja nem más volt, mint genealógiai adatok összegyűjtése és hozzáférhetővé tétele. Ebből a szempontból kiemelendők az adatokban gazdagabb és formalizáltabb XIX–XX. századi gyászjelentések. Ezek elődei, a XVIII. századi szomorújelentések, formailag kevésbé voltak kötöttek, és nem egyszer irodalmi alkotásként is megállták a helyüket, de a családfakutatók számára jóval kevesebb adatot tartalmaznak.

Bár eltérő struktúra alapján, de a későbbi gyászjelentéseken feltüntették az elhunyt neve mellett az illető halálának és temetésének dátumát, helyét, az elhunyt életkorát, betöltött közéleti funkcióit, hivatalos elismeréseit. Megtaláljuk a családtagok felsorolását, akiknek rokoni kapcsolatai nem csupán az elhunyttal, hanem egymással is jól megállapíthatók. Az adatok eddig még nem ismert rokoni kapcsolatokra deríthetnek fényt. Életrajzírók, lexikoncikkírók ellenőrizhetik, és esetleg javíthatják a híres személy hosszú időn keresztül tévesen használt adatait. Hosszan folytatható lenne még a sor, hogy e jelentős gyűjtemény milyen területeken nyújthat biztos támpontot és szolgálhat új információval. Bízunk benne, hogy az OSZK-ban – az *Elektronikus Dokumentum Központ* és az *Informatikai Igazgatóság* munkatársai által – az utóbbi

hónapokban fejlesztett szolgáltatás könnyebb hozzáférést nyújt majd a kutatók és az olvasók számára, valamint lehetővé teszi, hogy a Plakát- és Kisnyomtatványtárban található gyászjelentés-gyűjtemény az interneten keresztül közkinccsé váljon.

Anélkül, hogy a megvalósuló online szolgáltatás és a mögötte található, mára jól rendezett fizikai gyűjtemény jelentőségét kisebbíteni szeretnénk, bizony le kell vonnunk a tárban elvégzett, közel öt éves előkészítő munka negatív tanulságait is.

Ahogy azt e tanulmány az ún. „mormon-projekt”-tel kapcsolatban leírja, az nem önzetlen segítségnyújtás volt az OSZK számára, habár a végeredmény kétségtelenül hasznos és látványos. A mára birtokunkban lévő hatalmas digitális gyűjteményért súlyos árat fizetett a nemzeti könyvtár, a Plakát- és Kisnyomtatványtár és annak egyes munkatársai is. Ahhoz, hogy az egész fizikai gyűjtemény teljes mikrofilmmezése elkezdődhessen és megvalósulhasson, a tár akkori vezetése a Plakát- és Kisnyomtatványtár alapfeladatainak elvégzését a minimumra csökkentette. Ennek az lett a következménye, hogy a gyarapodás során a tárba érkező, közel évi 30 000 dokumentum szétválogatása, alapszintű feldolgozása lényegében több mint négy évre leállt. A gyászjelentések szigorú rendezésével, feldolgozásával és fotózásra előkészítésével töltött évek alatt a felgyülemlett dokumentumok száma megközelítette a 120 000-et. Ez a „be nem kalkulált” lemaradás a tárat mostanáig sújtja, és napi szinten többletmunkát ró a munkatársakra. Emellett a többi részgyűjtemény korábban folyamatosan végzett feldolgozása is leállt (gondolhatunk akár a régi metszetekre, grafikai plakátokra, XVIII–XIX. századi nyomtatványainkra stb.). Hangsúlyozni kívánjuk, hogy a részgyűjtemények projektszerű feldolgozása ellen – ahogy ez a gyászjelentésekkel történt – a tár mai vezetésének nincs kifogása, hiszen a számtalan részgyűjtemény párhuzamos feldolgozása talán kevésbé hatékony, mint az egy-egyféle dokumentumtípusra koncentrált munka. Ugyanakkor a „mormon-projekt”-tel induló és napjainkban is folyó gyűjteménydigitalizálás tervezésekor nem hagyhatunk ki egy rendkívül fontos tényezőt: a digitális gyűjtemény kizárólag egy már könyvtári szempontok alapján rendezett gyűjteményről történő digitalizálással készülhet el. E rendezés és feldolgozó munka emberi erőforrást igényel, amellyel szakmailag megalapozottan és átgondoltan szükséges gazdálkodni.

A „mormon-projekt”

Amint azt a gyűjtemény ismertetőjében is olvashattuk, a nemzeti könyvtárat a Mormon Egyház (innen a projekt neve) kereste meg azzal a javaslattal, hogy ingyen elkészítené az OSZK-ban található gyászjelentések digitális másolatát, ha mikrofilmen megkaphatja a dokumentumokat. Persze, nem önzetlen segítségnyújtásról volt szó, hiszen saját honlapjukon (<http://pilot.familysearch.org>) kívánták szolgáltatni a képeket, elsősorban a családfakutatók érdeklődésére számítva. A projekt során elkészült digitális képeket természetesen a nemzeti könyvtár is szolgáltathatja, vagy bármi módon hasznosíthatja, korlátozás nélkül. Az OSZK egyébként minden esetben ragaszkodik ahhoz, hogy a könyvtárban (akár külső megrendelésre) történő digitalizálás során létrejött elektronikus dokumentumokból egy példány az ő gyűjteményébe kerüljön.

Az OSZK-ban elkészültek a mikrofilmek, amelyeket átadtunk digitalizálásra a partnercégnek. Ezt követően kezdődtek a gondok: a munka meglehetősen lassan haladt, a kapcsolattartás akadozott, ígéreteket ugyan kaptunk néha, de a digitalizált anyagokat sokáig nem láttuk. (Az első minták 2007-ben érkeztek.) Leveleinkre nem, vagy csak jelentős késéssel jött válasz. Időnként már csaknem lemondtunk arról, hogy valaha elkészülnek a szkennelések. Többszöri egyezkedés után (és több éves csúszással) végre megérkezett hat darab hordozható tárolóeszköz a digitális fájlokkal. Az eszközöknek speciális csatlakozójuk volt, de találtunk megfelelő összekötő kábelt, hogy belemézhessünk a lemezek tartalmába.

A kezdeti problémák

Megnézve az állományokat kiderült, hogy nem pont olyan minőségben készültek, mint amilyet vártunk, például nem voltak körbevágva a képek, vagy ferdén helyezkedtek el a háttérként alátett lapon (ez utóbbi a felvételek készítésénél igényelt volna nagyobb figyelmet). Igaz, az átadás feltételeit a két fél között létrejött szerződés nem részletezte kellő mértékben. A megállapodás kitért a felhasználás lehetőségeire, de az átadás minőségi feltételeit nem írta körül megfelelő részletességgel, így reklamálni is nehéz lett volna.

A megállapodás idején a mikrofilmzés úgy történt, hogy az egyes gyászjelentéseket egy A4-es, fekete lapra helyezték, és előfordult, hogy kis cédulákat is tettek a háttérlapra további információkkal, pél-

dául „color printed” felirattal. Döntés kérdése, hogy eldobjuk vagy megtartjuk ezt a háttér és a rajta lévő információkat. Vítathatatlan, hogy „elegánsabb” lenne, ha maga a dokumentum jelenne meg a képernyőn a fölösleges háttér nélkül. Létezik megoldás a hasznos terület automatikus körbevágására, a fölösleges háttér leahagyására, de a fenti körülményekre tekintettel ezt nem is olyan egyszerű megoldani. Gondoljunk csak arra, hogyan is néz ki egy gyászjelentés? Általában a szélein fekete keret van... ha a háttér is fekete, bizony elég nehéz automatizáltan (sőt manuálisan is) beazonosítani a körbevágandó méretet. Várhatóan elveszne a fekete keret is, ami ennél a dokumentumtípusnál nem igazán korrekt megoldás. Sajnos a mikrofilmezés során nem minden esetben cserélték fehérre a háttérlepet, amikor az eredeti dokumentum fekete szegélye a papír széléig kiért. Sokszor előfordul, hogy ezen a szegélyen egy minta van, például egy vékony vonalakkal rajzolt kereszt, ami adott esetben az elhunyt vallási hovatartozásáról is információt hordozhat. Ezzel vajon mit kezd az automatizált körbevágás?

Kérdés, mi az egyszerűbb: néhány próbálkozás után megközelítő pontossággal beállítani az automatikus körbevágás paramétereit, majd utána megkeresni és korigálni a rosszul sikerült példányokat, vagy inkább egyenként beállítva, „kézzel” levágni a fölösleget, ami ilyen mennyiségnél rendkívül idő- és költségigényes megoldás. Legrosszabb esetben megtehetnénk, hogy a teljes beszkenelt képet (háttérleppal együtt) használjuk fel, és azt kicsinyítjük szolgáltatható méretre. Ez egyszerűbben automatizálható lenne, de kevésbé jól használható eredményt hoz.

Mint már jeleztük, közel 600 000 kép kezeléséről van szó! Ilyen mennyiségben a „kézi munka” rendkívül erőforrás-igényes, mi pedig nem rendelkezünk a szükséges szabad kapacitással. Anyagi forrásaink nincsenek külső segítők foglalkoztatására, így marad a gépidő optimális kihasználása és a rendelkezésünkre álló emberi erőforrás bevetése, ami egyszerűbben fogalmazva saját munkatársaink fizikai és szellemi erejét, valamint véges munkaidejét jelenti. Végül úgy döntöttünk, hogy megállapítottunk egy olyan körbevágási méretet, amelyikbe mind a fekvő, mind az álló gyászjelentések belefértek, és erre a méretre állítottuk be az automatikus körbevágást.

Térjünk vissza a digitális képekre, amelyek eredeti állapotukban 1 MB körüli méretűek, szürkeárnyaltosak és 200 dpi felbontásúak. Az OSZK-ban július

elején született meg az a belső szabályzat, amely a képi dokumentumok szolgáltatásának módjáról rendelkezik. Ennek megfelelően ezt a dokumentumcsoportot 72 dpi felbontásban szolgáltatjuk. A képeket ennek megfelelően konvertáltuk, méretüket 50%-osra csökkentettük, ezzel a helyfoglalásuk is mintegy harmadára csökkent. A szolgáltatási ajánlásnak megfelelően „vízzel” is ellátjuk a dokumentumokat. A képmanipulációkhoz léteznek olyan ingyenes szoftvereszközök, amelyekkel ez a művelet egyszerre nagyobb fájlcsoporton is elvégezhető, legfeljebb szabad gép és egy kis munkaszervezés kell hozzá. A feladathoz ingyenes képkezelő programot használtunk.

Mivel a gyászjelentések teljes tartalma fontos részinformációkkal szolgálhat, ezért szükséges a szövegfelismertetés is. A gyászjelentések esetében a teljes szövegű kereshetőségnek azért van különös jelentősége, mert kezdetben csak minimális leíró adatot tudunk szolgáltatni, a szöveg pedig a kutatók számára további hasznos információval szolgálhat. A szövegfelismertetés a nagyméretű képeken végezhető el a legjobb hatásokkal. Megfelelő folyamatkezeléssel és adminisztrációval bevonhatunk több gépet is a feldolgozásba, hiszen a gépidő jobb kihasználása érdekében éjjel is futhatnak olyan előre definiált folyamatok, amelyek nem igénylik a munkatársak jelenlétét, beavatkozását.

Az eredeti dokumentum és a digitális másolat kapcsolata

Az eddig vázoltakhoz képest nagyságrendekkel komolyabb problémát jelent a képek azonosítása és kereshetővé tétele, valamint a kapcsolódó információkat tartalmazó adatbázis kialakítása és feltöltése. Azt eldöntöttük, hogy szeretnénk az olvasók számára elérhetővé tenni a gyászjelentéseket, hiszen az érdeklődés itthon is jelentős, mind a családjuk története iránt érdeklődők, mind a különféle szakterületek kutatói körében. A jól használható szolgáltatás kialakításához még kevés, ha körbevágott, kicsinyített képeink vannak. Hogyan tesszük kereshetővé a dokumentumokat? Hogyan szolgáltatjuk a kutatók számára? Mi az a minimális adatsor, amit a képekhez kell kapcsolnunk? A szolgáltatásnak csak akkor van értelme, ha a képek egyenként azonosíthatók. Az anyag kutathatóságához további adatokra és a teljes szövegtartalom kereshetőségére is szükség van.

Amint láttuk, a mikrofilmezés előkészítése során a gyászjelentéseket az elhunyt családneve alapján ábécésorrendbe rendezték, és ebben a sorrend-

ben készültek a felvételek is. A digitális másolatok folyamatos számozással, csak sorrendi információt tartalmazó fájlnevével érkeztek, filmtekercsenként egy-egy mappába rendezve. Beazonosítottuk, hogy melyik mappa melyik mikrofilmet tartalmazza, és ekkor derült ki, hogy hét film anyaga hiányzik a beérkezett tárolóeszközökről. Lehet, hogy a digitalizálás során maradtak ki, de az is lehet, hogy elkészült ugyan a szkennelés, de valami miatt a hordozóra nem került föl a digitális kép. Ennek utánajárunk, remélhetőleg viszonylag egyszerűen pótolható a hiány.

Amíg a digitalizálás készült, az OSZK Plakát- és Kisnyomtatványtárának munkatársai adatbázisban rögzítették a mikrofilm azonosítóját, valamint a kezdő és záró családneveket, és elkészült a gyászjelentéseken előforduló mintegy hetvenhat-ezer (!) családnév jegyzéke, amit szintén az adatbázisban rögzítettünk, azzal a céllal, hogy a felhasználók könnyebben eligazodjanak a nyomtatott gyászjelentések között. Ezzel lehetővé vált, hogy egy családnévre keresve kiderüljön, hogy egyáltalán van-e a gyűjteményben a keresett családra vonatkozó gyászjelentés, és ha van, melyik mikrofilmen található a másolata. A digitális képek beérkezésekor ezekhez kötöttük hozzá a mappaneveket, bár ezzel is csak a mikrofilmhez kapcsolódó információhoz tudjuk irányítani az olvasót, mert az az információ még nem állt rendelkezésünkre, hogy melyik kép konkrétan melyik nyomtatott gyászjelentéshez tartozik. Ez még mindig igen szerény visszakeresési lehetőséget nyújt. A közelebbi találathoz az olvasónak a mappában mintegy 900-1000 képet kéne átlapozni, ami rendkívül nehézkes. Ennél mindenképpen jobb feltételeket kell teremteni, ha az interneten szabadon elérhetővé és kereshetővé, kutathatóvá kívánjuk tenni a gyűjteményt.

Az egyes képek azonosítása

Két megoldást vizsgáltunk meg. Az egyik, hogy a tár munkatársainak kell beazonosítani a képeket, vagyis szakszerűen (de legalább az azonosíthatósághoz szükséges minimális adatsorral) le kell írni a dokumentumokat. Figyelembe véve, hogy a jelenleg kb. 3,5 millió egységet tartalmazó kisnyomtatvány-gyűjtemény feltárása leköti a teljes kapacitást, továbbá, hogy évente kb. 30 000 egység érkezik, a jelenlegi munkatársak már nem terhelhetők tovább. Mivel a feladat elvégzésére nincsen semmilyen anyagi forrás, külső munkaerő bevonására sincs mód, ezért elég reménytelennek látszik a helyzet. Ráadásul szeretnénk mielőbb (terveink

szerint még 2009-ben) szolgáltatathatóvá tenni a képeket, tehát kevés idő áll rendelkezésre a feladat elvégzésére és a szolgáltatás béta verziójának elindítására.

Szövegfelismertetés

A hiányzó, részletes leíró adatok átmeneti pótlására megoldásként kínálkozik a nyers szövegfelismertetés, és a teljes szövegű keresés kialakítása. Teszteket végeztünk, hogy lássuk, milyen határfokkal ismeri fel az OCR program a karaktereket, mennyire válnak kereshetővé a fő szempontot jelentő nevek, helynevek, évszámok. Az eredmény elég vegyes volt. A legnagyobb nehézség ebből a szempontból, hogy a kulcsfontosságú információ, vagyis az elhunyt neve, éppen a kiemelés miatt speciális tipográfiájú betűkkel szerepel a nyomtatványon, ami kétségesse teszi a név pontos azonosítását. Sok esetben pedig nem is teljes név szerepel az eredeti dokumentumon, hanem például ennyi: „testvérünk, János”. Mivel az elhunyt azonosítása kulcskérdés, elkerülhetetlen, hogy legalább az ő teljes nevét pontosan azonosítsuk, és a hozzárendeljük a konkrét dokumentumhoz. Úgy ítéltük meg, hogy a használható szolgáltatás érdekében ez mindennél fontosabb, tehát megpróbálunk erre valamilyen – viszonylagosan gyors és megbízható – megoldást találni. A tapasztalatok szerint a név azonosítását nem bízhatjuk kizárólag a gépre, egészen biztos, hogy emberi beavatkozásra is szükség lesz.

Az egyes gyászjelentések automatikusan felismertett szövege a kép fájlnevével azonos néven txt kiterjesztéssel, a képpel azonos mappába került. Így könnyebben megvalósítható, hogy a keresett gyászjelentés képe és szövege összehozható legyen, illetve a teljes szöveg kereshető legyen. Ez már kicsit jobb helyzet, de még nem tartunk ott, hogy az elhunyt nevére keresve el lehessen jutni a konkrét képfájlhoz.

Körvonalazódnak a legfontosabb feladatok

Célkitűzés:

- lehetőleg rövid idő alatt elérhetővé kell tenni a gyászjelentések digitális változatát,
- keresési lehetőségeket kell nyújtani
 - legalább az elhunyt neve alapján,
 - lehetőleg a teljes szövegben,
 - optimális esetben részletesebben feltárt adat-sorban,
- kutathatóvá kell tenni a dokumentumokat.

Összegezve az eddigieket, a szolgáltatás előkészítése érdekében a következő teendők körvonala-zódnak:

1. teszt jelleggel konverziót és szövegfelismertést kell végezni egy kisebb mintán, hogy fel tudjuk mérni az elvégzendő munkához szükséges erőforrásigényt;
2. fel kell mérni, hogy hogyan tudunk haladni a jelenlegi kapacitásunkkal a képek átalakításával és OCR-ezésével, szükség van-e további erőforrások bevonására?
3. a képekből szolgáltatási méreteket kell konvertálni;
4. a teljes szövegű kereséshez OCR-ezni kell a dokumentumokat;
5. meg kell tervezni a szolgáltatáshoz a leíró adatok minimális és ideális körét;
6. a képeket egyenként azonosíthatóvá kell tenni, minimálisan az elhunyt neve alapján;
7. ki kell alakítani a keresési lehetőségeket, valamint egy olyan felületet, amelynek segítségével együtt jelenhet meg a gyászjelentés képe és a felismertett szövege.

1. Az első tesztek azt mutatták, hogy megoldható a tömeges szövegfelismertetés, méghozzá úgy, hogy az elkészült TXT fájl a kép mellé, azzal azonos mappába kerül, és a képfájllal megegyező nevet kap. A folyamat meglehetősen jól automatizálható, nem igényel állandó felügyeletet, tehát a felismertetés napközben is futtatható a szabad kapacitással rendelkező gépeken, illetve éjszakára elindítható néhány gépen.

2. A képkonverzió és az OCR elvégezhető saját erőforrásokkal, ha több gépen megszervezzük az éjszakai futást is. Ez utóbbi művelet az archív képfájlokat használja, melyek a belső, korlátozott hozzáférésű tárhelyen vannak, biztonsági okokból csak kijelölt gépekről és meghatározott bejelentkezési azonosítókkal érhetőek el. Ezt a korlátozást nem is szeretnénk feloldani még átmenetileg sem, így ezt a munkafolyamatot kizárólag belső munkatársak végzik: az Elektronikus Dokumentum Központban és a Plakát- és Kisnyomtatványtárban dolgozó kollégák vesznek részt a munkában.

3. A képkonverzió párhuzamosan történhet az OCR-ezéssel. A kiválasztott, ingyenes képkézelő program jó adottságai megfelelnek az igényeinknek, egyszerre nagyon sok fájlt tud kezelni, és viszonylag gyorsan végzi el az átalakításokat. Egy menetben megtörténhet a körbevágás és a kicsinyítés. A szolgáltatási képeket az archív verzióktól elkülönítetten tároljuk.

4. Az OCR-ezést automatizáltan végezzük. Tekintve a gyászjelentések sokféle betűtípusát, és a nyomtatás eltérő minőségét, a szövegfelismertetés vegyes eredménnyel jár. Arra nincs lehetőségünk, hogy a szövegeket ellenőrizzük és javítsuk, egyelőre meg kell elégednünk egy nyers szövegváltozattal.

5. Az OSZK Plakát- és Kisnyomtatványtárának munkatársai jól ismerik a gyűjteményt és a kutatók igényeit, így meg tudják határozni a gyászjelentéseket leíró adatok körét, és ki tudják jelölni az adatsorból azokat a mezőket, amelyek minimálisan szükségesek az egyes képek azonosításához, a gyűjtemény kereshetőségéhez. A digitalizált gyászjelentések nagy száma miatt a feltárás meglehetősen időigényes munkája a napi feladatok mellett jelenleg nem fér bele a kollégák munkaidejébe.

6. A megfelelő minőségű szolgáltatás feltétele az elhunyt nevének és a gyászjelentés képének összekapcsolása, valamint felismertett nyers szöveg megjelenítése. A kutatás támogatásához javított szövegre, és adatbázisba rendezett részletesebb adatsorra is szükség lenne, de ennek az elkészítése még hosszú időt vesz igénybe.

7. A szolgáltatás kialakítása során még számos kérdésről kell döntenünk:

- megvárjuk-e, hogy teljesen feltárt legyen a gyűjtemény, vagy egy minimális adattal már elérhetővé tegyük a képeket?
- első megközelítésben megelégszünk-e az elhunyt nevének pontos beazonosításával és az erre épülő kereséssel?
- ha igen, akkor hogyan tudjuk gyorsan és viszonylag hibamentesen hozzárendelni a képekhez a teljes neveket?
- mi legyen a minimális és az optimális leíró adatsor, amit a képekhez kell rendelni a jobb kutathatóság érdekében?
- hogyan jelenítsük meg a képeket, a szöveget és a rendelkezésre álló, kereshető adatokat?
- hogyan tudnánk elősegíteni, hogy mielőbb javított szövegek és pontos adatok álljanak a kutatók rendelkezésére?

Hosszas egyeztetések után úgy döntöttünk, hogy a lehető leghamarabb közkinccsá tesszük a képeket, hiszen a korábbi partner is elérhetővé tette a digitalizált állományt és a családfakutatók még ebben az egyszerű formában is nagy érdeklődést mutatnak a gyűjtemény iránt. Bár az OSZK több információval rendelkezik a dokumentumokról, de ezek

jelenleg csak a könyvtárba betérő kutatók számára hozzáférhetők. Úgy látjuk, hogy a szerényebb szolgáltatásra is van igény, ezért több lépcsőben alakítjuk ki az interneten is elérhető keresőfelületet.

Az első lépcső, hogy az elhunyt teljes neve alapján el lehessen jutni a konkrét képhez. Kollégáink őszre kialakítanak egy olyan felületet, amelyen látható a gyászjelentés kicsinyített képe, valamint az előforduló családnevek és a lehetséges keresztnévek listája. Kattintással lehet a listákból kiválasztani a családnevet és a keresztnévet (valamint a többi névelemet), amelyek automatikusan bekerülnek abba az adatbázisba, amelyben a kicsinyített kép neve, elérési útja és a képet tartalmazó mikrofilm azonosítója is szerepel. Mivel nem lehetünk biztosak abban, hogy a névlisták tökéletesek, azért a személyneveket tartalmazó mezők a kiválasztás után még javíthatók, vagyis el lehet végezni a szükséges módosításokat, hogy lehetőleg pontos nevet tároljunk az adatbázisban. Ezzel a módszerrel viszonylag gyorsan és hibamentesen megoldható a művelet, de 600 000 kép esetében ez sem kevés munka: ha képenként fél percet számolunk, akkor is 5 000 óra munkaidőt kell rászánni. Ez öt embernek közel félévi munkája! Egy összetett név (előznévként, nemesi előnév, rang, férjezett név, leánykori név, több keresztnév stb.) leírására adott esetben a fél perc nem is elég. A munkát tovább lassítja, ha nem azonosítható egyértelműen az elhunyt neve, mert csak ennyi szerepel a név helyén: „Anyika”, és a szövegből sem derül ki a pontos név. Ilyen esetben már komoly kutatómunkát igényel, ha pontosabb adatot szeretnénk feltüntetni az adatbázisban. Be kell látnunk, hogy erre most nem tudunk vállalkozni.

A szűkös kapacitás kiegészítésére csak egy megoldás látszik: önkéntesek bevonása, pontosabban együttműködés a gyászjelentések kutatása iránt érdeklődő közösségekkel!

Az önkéntesek nagyon sokat segíthetnek abban, hogy egyre több javított szöveg álljon rendelkezésre, és ezzel a teljes szövegű keresés találati biztonsága is növekedjen. Előkészítünk adatbázismezőket is, amelyekbe a szövegek javítása során beírható vagy áttemelhető a megfelelő adat, és ezzel fokozatosan feltölthető a leíró adatbázis is. A munkát gyakorlott könyvtárosok fogják irányítani és ellenőrizni.

Már tervezzük azt a felületet, amelyen a további feltáráshoz alapvetően szükséges minimális adat-

tal, vagyis az elhunyt nevével azonosítva megjelenítjük a dokumentumot, és mellette szerkeszthetően elhelyezzük a felismertett szöveget. A keresés eleinte csak az elhunyt nevére korlátozódik, illetve a nyers szövegben lehet keresni egy szövegrészlet alapján. Böngészési lehetőség is lesz, méghozzá az elhunyt családnevének ábécésorrendjében megjelenített listák segítségével. A nagy találati halmazokban többféle navigálási módot alakítunk ki: adott találati oldalra lépést, lapozást stb.

A tervezett adatbázis az elhunyt nevével kívül tartalmaz magára az eredeti, illetve a digitális dokumentumra vonatkozó adatokat is (pl. azonosító, fájlnev, elérési útvonal, az eredeti dokumentum keletkezésére, méretére vonatkozó információk). Lehetőség szerint rögzítjük az elhalálozás és a temetés helyét és idejét, valamint az elhunyt életkorát, ha ez kideríthető. Sok esetben olvasható a gyászjelentéseken a gyászoló nevének és az elhunytéhoz való viszonyuk, ami nagyon fontos információ lehet a kutatóknak. Ezeket a neveket is felvesszük az adatbázisba, és minősítővel látjuk el (rokonsági fok vagy egyéb kötődés), hogy ennek alapján felderíthetők legyenek a szerteágazó rokoni kapcsolatok. Minél tökéletesebb a szövegjavítás, és minél pontosabbak a dokumentumról felvett adatok, annál jobban kiaknázható lesz az adatbázisban fokozatosan felépülő gazdag információhalmaz.

Megjegyzésmezőket is fenntartunk arra az esetre, ha az előre felkínált mezőkbe nem illik bele valamely fontos információ, de mégis szeretnénk azt megőrizni. Példaként említhetjük, ha egy szervezet vagy intézmény saját halottjának tekintette az elhunytat, vagy valamilyen fontos információ olvasható a tevékenységével, életével, halálával kapcsolatban.

Kérdések az újszerű szolgáltatási forma kialakítása körül

Már az előkészítő munka során, a pontos nevek hozzárendelésében is jól jönne a segítség, de mint mondtuk, ez biztonsági okokból nehezen megoldható. A felismertett szöveg javítására és a legfontosabb adatok adatbázisba emelésére végképp nem lenne esélyünk, ha nem vennénk igénybe külső közreműködőket. Mindenképpen szeretnénk ellenőrzés alatt tartani a munkafolyamatokat, de azt is tudjuk, hogy ezek a közösségi szolgáltatások akkor működnek jól, ha egyfajta bizalomra épül-

nek. Meg kell találni az egyensúlyt a bizalom és az ellenőrzött működés között. Bízunk abban, hogy minden résztvevőnek érdeke, hogy ne sérüljenek, hanem folyamatosan javuljanak a szövegek és a leíró adatok. Vannak kapcsolataink azokkal a kutatócsoportokkal, amelyeket szívesen bevonnánk a munkába. Csak regisztrált felhasználók kapcsolódhatnak be az adatjavításba, és lehetőség szerint őket is úgy választjuk meg, hogy valamelyik fent említett kutatói közösséghez tartozzanak,

ezzel is bizonyítva érdekeltségüket az adatok minőségének megőrzésében.

Természetesen mi is izgatottan várjuk, hogy ez az elképzelés hogyan válik be a gyakorlatban, hiszen még nincs tapasztalatunk ilyen típusú szolgáltatás létrehozásában és működtetésében.

A közös munka várható veszélyei (1. táblázat)

1. táblázat

Veszélyek és elhárításuk

Veszély	Megoldási javaslat
Még nem volt hasonló szolgáltatás az OSZK-ban, nincs ilyen esetre kialakított szolgáltatás politika.	Ki kell alakítani a szolgáltatás használatának biztonságos feltételeit. Béta verzióban kell tesztelni a készülő felületet. Be kell vonni a tervezésbe a leendő közreműködőket.
Előfordulhat szándékos szövegrongálás.	Mivel együtt jelenik meg a képernyőn a szkennelt kép és a felismertetett/javított szöveg, azért könnyen észlelhető, ha a kettő között jelentős eltérés van. Véglegesen megőrizzük a nyers OCR-ezett változatot, amihez kritikus esetben vissza lehet térni. A regisztrációval biztosítható a közreműködők azonosítása. Kitiltjuk azt a felhasználót, aki nem megfelelően kezeli az adatokat. Bízunk benne, hogy nem a rongálás a közösség célja, inkább segíteni fognak abban, hogy az ilyen eseteket felderítsük, és megfelelően kezeljük.
A felhasználók/kutatók nem szállnak be a várt mértékben a szövegjavításba.	Először moderálatlanul engedjük a javításokat, hogy lássuk, mennyire aktívak a felhasználók. Megfelelő kommunikációval tudatosítjuk a közreműködőkben, hogy miért hasznos a részvételük. Bevonjuk a könyvtárakat a szolgáltatás népszerűsítésébe, mert a nagyobb használati igény jelentkezése is inspirálhatja a közreműködőket. Legrosszabb esetben a szolgáltatás megmarad az első szinten, vagyis a név alapján történő keresés eredményeként megjelenik a dokumentum képe és nyers szövege. Ezt a szintet – ha lassan is – de saját erőből mindenképpen létrehozzuk.
Nem tudjuk előre felmérni, hogy milyen aktivitást tanúsítanak majd a közreműködők, és ehhez milyen szerverkapacításra lesz szükség.	A béta verziót kisebb körben teszteljük, a technikai feltételeket egy fokozatosan növekvő használatnak megfelelően szükség szerint módosítjuk.
Nem lesz egységes a javítás, a feltárt adatsor stb.	A közreműködő csoportokat felkészítjük, hogy egységes szemlélettel végezzék a javításokat. A munka menetét szervezett keretek között igyekszünk tartani, a munkamegosztásról rendszeresen egyeztetünk. Minden elvégzett javítás egy fokkal használhatóbbá teszi az adatokat, akkor is, ha ez nem egyenletesen történik az adatbázis egésze szempontjából.
Ha beindítjuk a közösségi szövegjavítást, ez feladatokat jelent a könyvtárosoknak is, de még nem tudjuk megítélni, mennyi kapacitást köt le (pl. technikai segítségnyújtás, ellenőrzés, jóváhagyás stb.).	A béta teszt idején vizsgálni kell, hogy milyen erőforrásokat köt le a könyvtárosok oldalán a szolgáltatás működtetése, az adatok ellenőrzése.
Nem tudjuk kellő mértékben bevonnani a belső munkatársakat az ellenőrzésbe, jóváhagyásba.	Megkeressük a közreműködők közül a legaktívabbakat, legtapasztaltabbakat és felkérjük őket, hogy az ellenőrzésben is legyenek segítségünkre.

A szövegjavítás, a kereshetőség és a szolgáltatás további kérdései:

- Hogyan mentjük el a változtatásokat? Hány változatot mentünk el? Melyiket tegyük elérhetővé?
- Hogyan tudjuk felmérni, hányan dolgoznak, mit csinálnak? Milyen statisztikát tudunk ehhez készíteni?
- Mérhető-e valamilyen módon a szöveg változása, a javítás eredményessége?
- Mit lásson javításkor a felhasználó: a nyers szöveget, a legutóbbi verziót, vagy mindent együtt?
- Ha sérülést, rongálást tapasztalunk, akkor melyik verzióhoz térünk vissza?
- A felhasználók végtelen számú alkalommal javíthatják a szöveget, vagy elérhet a szöveg egy végső, lezárt verziót?
- Ha egy gyengébb minőségű szöveg letárolódik, akkor minden korábbi javítás elvész?

Ezek nem könnyen megválaszolható kérdések. Tökéletesen biztonságos megoldást nem lehet kialakítani, de igyekszünk minimálisra csökkenteni a veszélyeket.

Hol tartunk?

A napokban elkészül egy olyan felület, ahol a kép kicsinyítve, de még olvasható szöveggel látszik. A képernyőn „fülek” segítségével választható az előzékek, a családnevek és a keresztnévek ábécérendes listája, ezekből kattintással lehet átemelni a keresett nevet az adatbázismezőkbe, és elemenként lehet „összerakni” belőle akár összetettebb személyneveket is. Az elhunyt nevét olyan módon rakjuk össze, ahogy az eredeti dokumentumon olvasható. A névelemekből kitüntetett szerepet kap a családnév és a keresztnév (asszonyok esetében a leánykori nevük), mert ennek alapján készül majd a böngészőlista, de a dokumentumhoz a teljes nevet hozzárendeljük.

A névmezőket szerkeszthetően jelenítjük meg, hogy javítható legyen a név, illetve új nevet is fel lehessen venni az adatbázisba, ha még nem fordult elő az elhunyt családneve. Ebben az előkészítő fázisban a nevek kijelölésével és az adatbázisba helyezésével az adott gyászjelentés pontosan azonosítható lesz.

A második fázisban már összetettebb felületre van szükség, amelyen a gyászjelentés képe és nyers szövege együtt látszik, és a szöveg szerkeszthető, javítható formában jelenik meg. A kép is nagyítha-

tó lesz, s a nagyobb verziót külön ablakban lehet tanulmányozni. Cél a szövegek javítása, és egyúttal a releváns adatok átemelése a képernyőn megjelenített adatbázismezőkbe. A leíró adatok körét a könyvtáros szakemberek határozzák meg, megjelölve azt a néhány adatot, ami az eredményes, pontos kereséshez alapvetően szükséges.

Az előkészítés során az ábécé elejétől indulva sorban haladunk a képek azonosításával. Úgy gondoltuk, hogy amint így elkészül néhány ezer rekord, azokat áttesszük a második fázisba, vagyis ezekkel már tesztelhető lesz a javító felület működése. Elképzelhető, hogy a közreműködők között felosztjuk az anyagot, például filmenként haladunk, egy-egy filmnyi mennyiséget „kiadva” valakinek feldolgozásra. Azt is lehetővé szeretnénk tenni, hogy egy keresés eredményeként létrejött találati halmazt vesz „kézbe” valaki, és ezen halad a javítással. Utóbbi megoldás talán célravezetőbb, mert jobban közelíthetünk vele a segítők érdeklődési köréhez. Lehet, hogy valaki azért szeretne részt venni a munkában, mert saját családjára vonatkozó információkat keres. Így motiváltabb lesz, és javításaival ebben az esetben is segíti az adatbázis építését, az adatok javítását. Értelemszerűen a végső cél, hogy a teljes szöveg javított és kereshető állapotban legyen, valamint a fontos kutatási adatok egy adatbázisban, rendezetten, kereshetően tárolódjanak.

A digitális állomány bővítése

Természetesen szeretnénk azt is megoldani, hogy az interneten elérhetővé tett gyűjtemény tartalmilag is gazdagodjon. A jelenleg rendelkezésünkre álló digitális képek csak a 2003-ig beérkezett dokumentumokat tartalmazzák, de a kötelempéldány-szolgáltatás keretében már jelentősen gyarapodott a nemzeti könyvtár gyűjteménye. Az utóbbi években beérkezett dokumentumokat is szeretnénk digitalizálni és hozzáférhetővé tenni. Terveink között szerepel, hogy a készülő szolgáltatás keretében lehetőséget nyújtsunk az önkéntes beadásra is, vagyis a könyvtár gyűjteményében nem szereplő (sok esetben nem is nyomdai úton sokszorosított) dokumentumok feltöltésére.

Tekintve a gyászjelentések nagy számát, komoly munka vár ránk, amelynek az idősükségletét még nem tudjuk megbecsülni. Az előrehaladás egyik fontos tényezője lesz, hogy hányan vállalkoznak arra, hogy részt vegyenek a munkában, és ki mennyi időt tud erre szánni. Az sem mellékes,

hogy a közreműködők mennyire pontosan, alaposan dolgoznak. Megkerülhetetlennek látszik, hogy az adatokat valaki ellenőrizze és jóváhagyja, erre is kell energiát, időt szánni. Célszerű lenne a felületen követni, hogy az adott rekord milyen javítottági szintet ért el, illetve hogy jóváhagyott, kész rekordról van-e szó. A jóváhagyást könyvtárosok vagy erre a feladatra kijelölt segítők végeznék megfelelő ellenőrzés után.

Kérdés, hogy melyik fázisban érdemes szabadon, mindenki számára elérhetővé tenni a gyűjteményt, és a felhasználóknak mit mutassunk meg a javítás alatt lévő adatokból, szövegből.

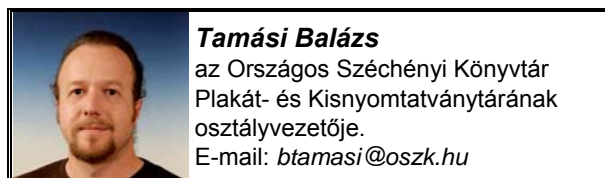
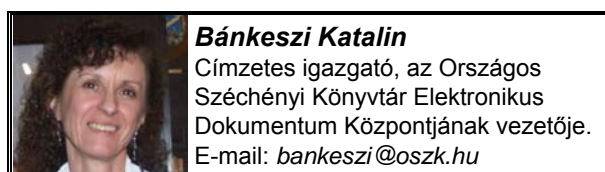
Célszerű lenne megvárni, amíg az előkészítő fázisban az elhunyt(ak) neve minden dokumentum esetében rendelkezésre áll, vagyis egyértelműen azonosíthatók a digitális képek. Amíg ez nincs így, addig a keresés csak megközelítő eredményt hoz, és lapozással lehet eljutni a keresett dokumentumhoz. Ha megvárjuk, amíg minden dokumentum egyedileg is megtalálható lesz, akkor néhány hónapos késéssel tudjuk a nagyközönség elé tárni a gyűjteményt. A javító fázis elkezdődhet előbb is, bár akkor még nem tud kiteljesedni, hiszen nem az egész állomány lesz elérhető, hanem csak a már azonosított részek. A javítással ekkor is jól lehet haladni, csak nem biztos, hogy megfelelő lesz a motiváltság, hiszen még messze nem minden név lesz kereshető.

A későbbiek során három jogosultsági szint kialakítását tervezzük. A legmélyebb hozzáférést az „admin” minősítésű közreműködők kapnak, akik az adatok ellenőrzésével és jóváhagyásával foglal-

koznak és öröködnék az adatbázis minőségén. A második szint a segítő közreműködők szintje, akik módosíthatnak az adatokon, javíthatják a szövegeket. A harmadik szint a felhasználóké, akik szabadon kereshetnek a már feltárt adatok között, illetve a teljes szövegben, de nincs lehetőségük módosítások végrehajtására. Nekik is lesz viszont lehetőségük arra, hogy elektronikus levélben, vagy magán a felületen jelezzék, ha hibát észlelnek.

A jelen cikk a tervezés előrehaladott szakaszában íródott, a fentiekben jelzett kérdések megválaszolása folyamatban van, ennek megfelelően alakul majd ki a szolgáltatás végleges koncepciója. Terveink szerint az elkövetkező hetekben megkezdődik a nevek pontos azonosítása, és még ebben az évben megjelenhet a javítófelület próbaverziója az első néhány ezer rekorddal. Reméljük, egy év múlva már beszámolhatunk az újszerű szolgáltatás bevezetésével kapcsolatos tapasztalatainkról és az első eredményekről.

Beérkezett: 2009. IX. 13-án.



A GPL-t és a Creative Commons licencet támogatja az Európai Unió

Az Európai Bizottság úgynevezett Zöld Könyvéhez kiadott állásfoglalásában az Európai Unió Gazdasági és Szociális Bizottsága egyértelműen a GPL és a Creative Commons licenc hatálya alá eső művek támogatásáról döntött.

A testület szerint a változtatások mielőbbi bevezetéséhez sürgősen módosítani kell a jelenleg hatályos szerzői jogi szabályokat és az EU szerzői jogi irányelvét. A módosítások célja, hogy megkönnyítsék a különböző könyvek és más tartalmak online kikölcsönzésének lehetőségét.

Az Európai Unió tavaly hozta nyilvánosságra Zöld Könyvét, amelynek célja a szerzői jog megreformálása, hogy az megfeleljen a 21. századi kutatás, a tudomány és az oktatás követelményeinek is. A döntéshozók azt szeretnék, hogy a tudás szabadon áramolhasson és ezáltal az eddignél több fejlesztésre kerülhessen sor az Európai Unió belső piacán. Szintén fontos egy, az EU mind a 27 tagállamában érvényes, egységes szerzői jogi környezet megteremtése. Biztosítanák azt is, hogy ha egy felhasználó nem férhet hozzá az információhoz, akkor fellebbezhessen a döntés ellen. Kiemelten támogatnák a GPL- és a Creative Commons licencű művek terjesztését, a két licenc elterjedését, valamint a GPL-licenc hatálya alá eső szoftverek alkalmazását.

Szintén kitér a dokumentum az úgynevezett elárult könyvekre, amelyeknél már nincsenek meg a szerzői jogi tulajdonosok, továbbá a felhasználók által készített tartalmakra és azok hasznosítására, illetve arra, hogy például a látássérültek vagy a vakok is az eddignél könnyebben férhessenek hozzá a különböző internetes anyagokhoz.

/SG.hu Hírlevél, 2009. október 5., <http://www.sg.hu/>

(SzP)

Emberi vagy gépi kivonatolás?*

Lehet-e olyan jó egy szöveg automatikusan létrehozott kivonata, mint az ember által készített referátum? De melyik emberé? Van-e szabályszerűség, hasonlóság a különböző területeken tanuló, dolgozó személyek releváns mondatkiválasztásában? A cikk egy nagy létszámú felmérés alapján keresi ezekre a kérdésekre a választ.

A tanulmány alapjául szolgáló kutatómunka célja a referátumkészítés automatizálásának elősegítése magyar nyelven. A kutatás része volt egy magyar nyelvű offline kivonatoló program fejlesztése is. A kvantitatív tartalomelemzés kategóriájába tartozó program egységeit a szöveg szavai képezik, az output pedig a szöveg mondataiból áll elő. Elkészült a program első verziója, melynek tesztelése nem állhat meg a működésbeli hibák kiküszöbölésénél, a sajátosságok elemzésénél; hatékonyságát az emberi kivonatokkal való összevetéssel kell tesztelni. Készült egy felmérés, amelyben a résztvevőket különböző témájú cikkek kivonatának elkészítésére kértem fel, mivel arra keresem a választ, hogy az emberek által előállított kivonat hasonló-e a gépihez, illetve ha nem, akkor mi az oka az eltérésnek.

Ehhez előbb nézzük meg, milyen módszereket alkalmaz a kivonatoló program!

A program működési elve

A kivonat előállításához vezető út első lépése a szótövek meghatározása, melyet a *Morphologic* cég „Helyeslem” szoftvere szolgáltat. A szótövek gyakorisági értékének meghatározása után történik a szignifikáns szavak megállapítása. Az outputként szolgáló mondatokat súlyozás határozza meg, ahol a mondatban szereplő szignifikáns szavakat pontozzák. A szignifikáns szavak meghatározási módját a felhasználó állítja be, aki jelenleg két módszer közül választhat:

- A *Luhn* módszerével történő meghatározás során akkor szignifikáns a szó, ha az aktuális szó-tövhöz tartozó előfordulási szám a szövegben háromnál több.
- A szótáralapú feldolgozásban akkor tekintjük szignifikánsnak a szót, ha az megtalálható a szótárban. Ebben a módszerben a felhasználónak lehetősége van saját szótár megadására, így ha

rendelkezik bármilyen gyakorisági szótárral, szakterületre jellemző specifikus kifejezések figyelembe vételével tudja elemezni a szöveget. Saját szótár hiányában a program jelen formája a köznyelvi *Szószablya Gyakorisági Szótár* első 10 000 szavából készített adatbázist használja.

A mondat súlyozott pontszámát a szignifikáns szavakon túl a benne szereplő szópárok, szóhármak és szónégyesek is növelik.¹ Az összpontszám meghatározása után átlagszámításra kerül sor, kiküszöbölve ezzel az eltérő mondathosszúságokból következő egyenlőtlenségeket.

A kutatómunka fontos része volt az emberi kivonatolás hatékonyságának vizsgálata. Törekedtem az emberek kivonatolási technikájában feltárni a szabályszerűségeket és a kapott tapasztalatokat beépíteni a programba.² A tartalomelemző elméletek szerint a mondatok cikken belüli elhelyezkedése befolyásolja a mondat szerepét, azaz kivonat készítéskor súlyozottabban kell figyelembe venni az első, illetve az utolsó bekezdésben található mondatokat. Saját felméréseim is igazolták ezen elméletek egy részét: a vizsgálat alapjául szolgáló különböző témájú cikkek esetében a 340 kitöltőnek több mint fele az első bekezdés mondatait tartotta a leglényegesebbnek, viszont elenyésző részük tulajdonított kitüntetett figyelmet az utolsó bekezdés mondatainak. Ennek eredményét be is építettem a programba, és a szöveg első bekezdésében található mondatokat dupla súllyal vettem figyelembe. A szöveg utolsó bekezdésének mondatai a szakirodalom elméletének megfelelően magasabb súlyt kapnak, de mivel saját kutatásom ezt nem támasztotta alá, ezért csak 20%-kal emeltem a pontértékeiket.

* A 2009. április 15-17. között megrendezett Networkshop konferencián elhangzott előadás alapján.

Utolsó lépésként a mondatok gyakorisági értékük alapján kerülnek rendezésre, a megjelenítendő kivonat terjedelmét pedig a felhasználó állíthatja be egy százalékos érték megadásával.

A folyamat végére előáll egy kivonat, amely a szöveg program szerinti leglényegesebb mondatait tartalmazza. *De tényleg ezek a leglényegesebb mondatok?*

A felmérés

Készítettem egy felmérést is, amely a program hatékonyságának vizsgálatát szolgálja. A felmérésben gyakorló könyvtárosok, referátumkészítő szakemberek és informatikus könyvtáros-hallgatók által elkészített kivonatokat hasonlítok össze egymással és a program output-állományával.

A felmérés alapja

Az empirikus mérés során különböző témájú szakmai cikkek kivonatának elkészítésére kértem

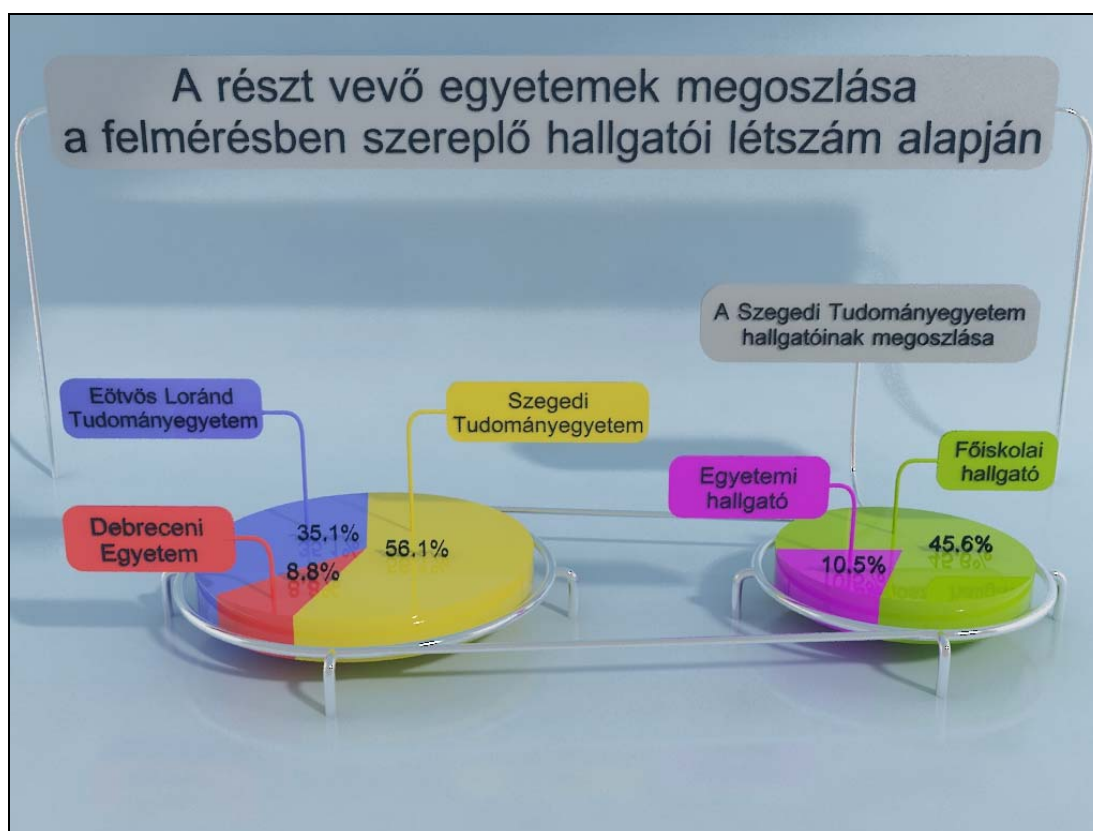
fel több felsőoktatási intézmény informatikus könyvtáros-hallgatóit.

Az alapul szolgáló cikkek kiválasztásakor két vezető szakfolyóirat aktuális számaiból választottam egy-egy cikket, így a *Könyvtári Figyelő*³, illetve a *Tudományos és Műszaki Tájékoztatás*⁴.

A felmérésben résztvevőktől azt kértem, hogy készítsék el mindkét cikk kivonatát a leglényegesebb mondataik megjelölésével és rangsorolásával.⁵ A felmérésben szereplő szakemberektől a kivonat mondatainak rangsorolását számítógépen, egy online kérdőív kitöltésével egybekötve vártam, a hallgatókkal a kivonatkészítést – megkönnyítve számukra az áttekinthetőséget – papíron végeztettem el. Számítógépes feldolgozásuk utólagos adatbevitellel történt.

A felmérésben résztvevők

A mintát egyrészt egyetemi, illetve főiskolai képzésben részt vevő informatikus könyvtáros szakos hallgatók képezik az 1. ábrán látható összetételben:



1. ábra A részt vevő egyetemek megoszlása a felmérésben szereplő hallgatói létszám alapján

Nemcsak hallgatókat kértem fel a referátum elkészítésére, hanem szerettem volna adatokat gyűjteni a szakemberek általi kivonatkészítés jellemzőiről is, ezért felkértem a KATALIST levelezőlista olvasóit és több könyvtáros szakembert az online kérdőív kitöltésére.

Mivel választ kerestem többek között arra is, hogy mennyire fontos a szaktudás a kivonatok készítésekor, egy kontrollcsoportot is bevontam a felmérésbe: az egri *Eszterházy Károly Főiskola* magyar szakos hallgatóit. Azért esett rájuk a választás, mert nekik van jártasságuk a különböző hosszúságú és témájú szövegek tömörítésében, a művek lényegének kiemelésében, azonban nincs könyvtártudományi szaktudásuk, így alapot adnak a szaktudás és a kivonatkészítésben való jártasság értékének elemzésére.

Végezetül a felmérésben a 2. ábrán látható összetétellel vettek részt a kitöltők:



2. ábra A felmérésben résztvevők megoszlása

A felmérés célja

A felmérés célja az volt, hogy megvizsgáljam, mennyire hatékony az általam készített kivonatoló program. Mielőtt hozzákezdttem a felméréshez, sejtettem, hogy rengeteg technikai-nyelvészeti problémával kell szembenéznem a program megírása során, de a legnagyobb kérdés az volt, hogy vajon jó lesz-e a program outputja. Ehhez ugyanis tudni kellene, hogy mi tekinthető „jó” kivonatnak. A felmérés eredményétől azt reméltem, hogy a kitöl-

Lengyelne Molnár T.: Emberi vagy gépi kivonatolás?

tők sok hasonló kivonatot fognak előállítani, és ez esetben lesz viszonyítási alapom.

Bemutatom a program által előállított mondatok és a felmérésben részt vevő személyek által előállított kivonatok mondatai közötti hasonlóságokat és eltéréseket, választ keresve arra, hogy létezik-e globális kivonat, továbbá, hogy van-e különbség az automatizálás alapjául választott két módszer adta output között: Luhn módszere és a szótár alapján történő kivonatolás eredménye között.

Elemzés

Az összehasonlítás alapjául a kitöltők által megadott mondatok súlyozva, mintacsoportonként kerülnek összevetésre. A súlyozásra azért volt szükség, hogy kiküszöbölhető legyen, mi számít relevánsabb mondatnak: a 10 ember által első helyre tett, vagy a 25 ember által a 10. helyre rangsorolt? Ezen felül természetesen készültek elemzések, amikor az eredeti kivonatok lettek összevetve egymással.

Mielőtt összevetnénk a program outputjával a felmérés eredményeként előállított, emberek által készített kivonatokot, nézzük meg röviden a mintacsoportok véleménye közötti hasonlóságokat és eltéréseket!⁶

Kezdjük az interdiszciplinárisabb témával foglalkozó *Koltay Tibor* cikk-kivonatainak elemzésével!

A hallgatói mintacsoportok kivonata nagyfokú hasonlóságot mutat. Az informatikus könyvtáros szakos főiskolás és egyetemista hallgatói mintacsoport súlyozással előállított kivonata csak két mondatban tér el egymástól, illetve a kontrollcsoportot képző magyar szakosok kivonatótól is. Ez 88%-os egyezőséget jelent.

A szakemberek kivonata ezzel szemben lényegesen eltér, az azonos mondatok legjobb esetben is csupán a kivonat 47%-át teszik ki.⁷

Érdeemes megvizsgálni, hogy ha nemcsak azt a 17 mondatot vesszük figyelembe, amely bekerült a kivonatba, hanem minden mintacsoportnál megnézzük, hogy az eredeti cikk összes mondata milyen súlyozott pontokat kapott, és ezt elemezzük egy korrelációs mátrixszal, akkor milyen eredményt kapunk (*1. táblázat*)!

Súlyozott pontszámokból képzett korrelációs mátrix

Súlyozott pontszámokból képzett korrelációs mátrix	Informatikus könyvtáros egyetemista hallgatók	Informatikus könyvtáros főiskolás hallgatók	Szakemberek	Magyar szakos hallgatók
Informatikus könyvtáros egyetemista hallgatók	1	0,917	0,486	0,872
Informatikus könyvtáros főiskolás hallgatók	0,917	1	0,445	0,856
Szakemberek	0,486	0,445	1	0,441
Magyar szakos hallgatók	0,872	0,856	0,441	1

A táblázat alapján jól látható, hogy a hallgatói mintacsoportok szaktól, tanulmány szintjétől függetlenül nagyon hasonló módon értékelik az egyes mondatok relevanciáját. Bár kiemelhető, hogy az egyetemista és főiskolai szintű képzésben részt vevő hallgatók esetében 0,9 fölötti nagyon szoros korrelációs kapcsolat áll fenn, azaz, ha az egyik mintacsoportba tartozó személyek előkelő helyre rangsoroltak egy mondatot, akkor a másik mintacsoport tagjai is relevánsnak tartják azt.

A szakemberek esetén már nem ekkora az egyetértés. Az ő véleményükhöz az egyetemista informatikus könyvtárosok mondatkiválasztási technikája áll a legközelebb, de ez is csak 0,486-os korrelációs értékkel. A szakemberek eltérő gondolkodásmódját jól tükrözi, hogy az 50 szakember 32 különböző mondatot rangsorolt az első helyre, és amit a legtöbben megjelöltek, az is csupán a kitöltők 12,5%-ának jelölését tudhatja magáénak, míg ez az érték a hallgatóknál 50% fölötti.

A másik, a könyvtártudományhoz sokkal jobban kötődő *Prokné Palik Mária* által írt cikk esetében előfordul olyan mondat is, amelyet az egyetemista informatikus könyvtáros kitöltők több mint 85%-a bevett a kivonat mondatai közé. Náluk és a magyar szakos kontrollcsoportnál hat mondat van, amelyet több mint a kitöltők felének kivonatában megtalálhatunk. Ez az érték a főiskolás informatikus könyvtáros-kitöltőknél hét mondat.

A szakemberek véleménye ennél a cikknél is teljesen eltérő struktúrát alkot, nincs olyan mondat, amelyben a kitöltők harmada egyetértett volna a kivonatba való beválasztás során.

A súlyozás útján előállt kivonatok hasonló képet mutatnak, mint az előző cikk esetén: az egyetemista és a főiskolás informatikus könyvtáros-hallgatók

súlyozott kivonata csupán egy-egy mondatban tér el egymástól, és 0,936-os korrelációs értékkel az összes mondat pontszáma hasonló gondolkodásmódra utal.

A szakemberek kivonata is összhangban van az előző cikknél kapott értékekkel: a 17 mondatos kivonatból csak hét-hét mondatban egyezik meg a szakhallgatók kivonatával. Ez 41%-os egyezőséget jelent.

Viszont meglepő képet kapunk, ha a kontrollcsoport kivonatát elemezzük! Míg *Koltay Tibor* cikkénél két-két mondatban tért el az informatikus könyvtáros-mintacsoportoktól a súlyozott kivonat, a diszciplinárisabb témájú cikknél már öt-, illetve hatmondatos az eltérés, miközben a könyvtár szakosok között nagy az egyetértés. Az igazán váratlan fordulatot az okozza, hogy a könyvtártudomány szakembereinek kivonata a magyar szakos kontrollcsoport kivonatával mutat a legnagyobb hasonlóságot (53%-os egyezőség, a szakhallgatókéval pedig 41%-os az egyezés).

További érdekesség, hogy a felmérésben részt vevő

- 51 szakember 37 különböző mondatot választott ki legrelevánsabb mondatnak,
- míg a 241 szakos hallgató 30 mondat közül választott,
- a 48 fős magyar szakosokból álló kontrollcsoportnál 10 mondat került az első helyre valamely résztvevőnél.

Összefoglalva: vegyes eredményekkel kezdhetjük el a program hatékonyságának tesztelését. *Míg a hallgatók között interdiszciplináris témánál magas fokú egyezőséget tapasztalhatunk, addig a szakemberek eltérő módon látják a releváns mondatokat. Ha a cikk a könyvtártudomány szakkifejezése-*

iben bővelkedik, akkor a magyar szakos hallgatók véleménye leszakad az informatikus könyvtáros hallgatókétól, a szakemberek pedig továbbra is egyéni véleményeket tükröznek.

A továbbiakban az adatokat cikkenként párhuzamosan láthatjuk a két módszernek megfelelően, továbbá a két cikk szerzője is segítette a felmérést, és elkészítette saját cikkének kivonatát, rangsorolva a kivonatba került mondataikat, így lehetőséget adtak arra, hogy az ő véleményükhöz is viszonyíthassunk.

Program kontra ember

Koltay Tibor cikkének elemzése

A program eredményeként Koltay Tibor cikkénél a 86 mondat közül Luhn módszerével történő kivonat készítés során 84 mondat kapott pontszámot, míg a Szószablya szótárának alkalmazása során valamivel kevesebb, a mondatok 91%-a, 78 mondat kapott 0-tól eltérő értéket. A 84 mondatból arra lehet következtetni, hogy a szerző a mondataiban olyan szavakat használ, amelyek a szövegben többször is előfordulnak. 47 olyan különböző szó van a cikkben – a tiltott szavakon kívül –, amely 3-nál többször fordul elő. A szövegben található 526 szótó közül 313 szerepel a Szószablya szótárban. A szópárokat és szóhármakat vizsgálva a szövegben egyetlen szónégyes, 5 db szóhármast, és 38 db szópár található. (Ezek meghatározása a szótövek alapján történik.)

A két módszerrel történő elemzés során kapott pontszámok alapján a 2. táblázat mondatait ítéli a program a legrelevánsabbnak. (Mint az eddigi elemzések során, itt is tekintsük meg az első 17 mondatot, a 20%-os kivonatot, de természetesen a program használata során lehetőség van más terjedelem listázására is.)

2. táblázat

Számítógépes output elemzése Koltay Tibor cikke esetében

Mintacsoportok	Luhn módszere alapján		Szószablya szótár alapján	
	Egyező mondatok száma	Egyezés aránya	Egyező mondatok száma	Egyezés aránya
Informatikus könyvtáros egyetemista hallgatók	3	17,65%	3	17,65%
Informatikus könyvtáros főiskolás hallgatók	4	23,53%	3	17,65%
Szakemberek	4	23,53%	6	35,29%
Magyar szakos hallgatók	3	17,65%	3	17,65%
Szerző kivonata	5	29,41%	3	17,65%

3. táblázat

A program által generált kivonat mondatai

Sorrend	A kivonat mondatai Luhn módszerével történő elemzés alapján	A kivonat mondatai a Szószablya szótár alapján
1.	1. mondat	1. mondat
2.	81. mondat	50. mondat
3.	83. mondat	21. mondat
4.	17. mondat	77. mondat
5.	38. mondat	81. mondat
6.	51. mondat	78. mondat
7.	69. mondat	38. mondat
8.	21. mondat	76. mondat
9.	76. mondat	5. mondat
10.	23. mondat	3. mondat
11.	86. mondat	82. mondat
12.	3. mondat	6. mondat
13.	5. mondat	17. mondat
14.	37. mondat	32. mondat
15.	77. mondat	51. mondat
16.	84. mondat	55. mondat
17.	22. mondat	80. mondat

A két módszerrel kapott kivonat 58,82%-ban tartalmaz azonos mondatokat: a 17 mondat közül 10-et. Az első helyre került mondat megegyezik a két módszernél, mely eredményt magyarázza a magasabb súly, ugyanis az első bekezdés mondatait magasabb súllyal veszik figyelembe.

Nézzük meg, hogy a program outputjaként előállt mondatok mennyire vannak összhangban a mintacsoportok kivonatával (3. táblázat)!

Mindkét módszert tekintve alacsony az egyező mondatok száma, a legmagasabb egyezés is csak 35%-os.

Luhn módszerét alkalmazva – ahol a gyakrabban előforduló szavak számítanak relevánsnak – a szerző válaszai alapján előállt kivonattal a legmagasabb az egyezés, a hallgatói csoportoknál pedig szinte teljesen azonos eredményt kapunk. Az egyező mondatok közül kettő minden mintacsoportnál megtalálható, de ez nem meglepő, hiszen az elemzés során a mintacsoportok súlyozott kivonatai alig tértek el egymástól. Érdekes, hogy a szerző saját kivonatával a legmagasabb az egyezés, pedig ha a szerző kivonatát vetjük össze a mintacsoportok súlyozott mondataival, akkor maximum nyolcmondatos egyezést találunk.

Koltay Tibor cikkének vizsgálatakor a Szószablya szótár alapján létrejött kivonat hasonló eredményt mutat, mint a Luhn módszerével kapott. A legmagasabb egyezés itt a szakemberek kivonatával áll fenn. A három hallgatói csoporttal ugyanazon mondatokban egyezik meg a program kivonata.

Összegezve: megállapítható, hogy Koltay Tibor cikke esetén – amely általánosabb témával foglalkozik – az egyezés nem túl magas számú a program kivonata és a mintacsoportok kivonata között.

Nézzük meg, hogy Prokné Palik Mária cikke hasonló eredményt mutat-e?

Prokné Palik Mária cikkének elemzése

Ennél a cikknél is Luhn módszerének alkalmazásával kapott több mondat pontszámot, bár a különbség kisebb, mint Koltay Tibor cikkénél. A szöveg 83%-a, 69 mondat ért el valamilyen pontértéket, ugyanis a szövegben 37 releváns szó található (amely háromnál többször előforduló nem tiltott szó). A Szószablya szótár alapján három mondattal kevesebb, a szöveg 85,5%-a kapott pontot, amelynek alapja, hogy a szöveg 487 szava közül a Szószablya szótárban megtalálható 297 szó. A szövegben 1 szónégyes, 6 szóhármass, és 40 darab szópár található.

A pontszámok alapján kialakult sorrendben az első 17 mondat eredményét a 4. táblázat mutatja.

4. táblázat

A program által generált kivonat mondatai

Sorrend	A kivonat mondatai Luhn módszerével történő elemzés alapján	A kivonat mondatai a Szószablya szótára alapján
1.	6. mondat	6. mondat
2.	22. mondat	63. mondat
3.	3. mondat	3. mondat
4.	63. mondat	36. mondat
5.	21. mondat	46. mondat
6.	7. mondat	5. mondat
7.	10. mondat	48. mondat
8.	36. mondat	11. mondat
9.	50. mondat	47. mondat
10.	19. mondat	32. mondat
11.	42. mondat	21. mondat
12.	61. mondat	22. mondat
13.	64. mondat	24. mondat
14.	5. mondat	53. mondat
15.	48. mondat	74. mondat
16.	39. mondat	34. mondat
17.	68. mondat	1. mondat

A két módszerrel előállított kivonat 11 mondatban egyezik meg, amely érték Koltay Tibor cikkénél 10 mondat volt. Az eredményt azonban befolyásolja az első bekezdés magasabb súlyozása. A cikk első bekezdése hat mondatot tartalmaz, amelyből a Szószablya szótár alapján létrejött kivonatba három mondat került be, míg Luhn elvével hozva létre a kivonatot, az első bekezdés mondatai közül öt mondat található meg a kivonatban.

Prokné Palik Mária cikke esetében a Luhn módszerével kapott kivonat mondatai szinte minden mintacsoportnál⁸ legalább kétszer annyi közös mondatot tartalmaznak, mint amennyi a szótár alapján készített kivonatban előáll.

Nézzük meg az egyező mondatok számát és arányát bemutató táblázatot (5. táblázat)!

5. táblázat

Számítógépes output elemzése Prokné Palik Mária cikke esetében

Mintacsoportok	Luhn módszere alapján		Szószablya szótár alapján	
	Egyező mondatok száma	Egyezés aránya	Egyező mondatok száma	Egyezés aránya
Informatikus könyvtáros egyetemista hallgatók	6	35,29%	2	11,76%
Informatikus könyvtáros főiskolás hallgatók	6	35,29%	3	17,75%
Szakemberek	8	47,06%	4	23,53%
Magyar szakos hallgatók	8	47,06%	5	29,41%
A szerző kivonata	4	23,53%	2	11,76%

A Luhn módszerével kapott kivonat mondatai nagyobb egyezést mutatnak a mintacsoportok kivonatával, mint ha a szerző kivonatát viszonyítjuk a mintacsoportokéhoz, illetve ha a szakemberek kivonatát viszonyítjuk a hallgatói csoportok eredményeihez.⁹

A szótár alapján készült kivonat mondatai alacsony egyezést mutatnak. A mondatok is változatosabbak, nincs olyan mondat, amely az összes mintacsoportnál megtalálható.

Véleményem szerint *a kapott eredmények oka a cikkek tartalmában, szövegében fedezhető fel.* Prokné Palik Mária cikke témáját és szövegezését tekintve is szakcikk, amelyben sok könyvtári szak kifejezést használ. Mivel a Szószablya szótár általános témájú, ezek közül kevesebb található meg benne. A Luhn módszerével 3-nál többször előforduló szavak 75%-a van meg a Szószablya szótárban. Ennek következtében nem a szakszavakat tartalmazó mondatok lesznek súlyozva, és ez magyarázza a mintacsoportok kivonatával való alacsonyabb egyezést. Luhn módszerével a szakmailag lényeges mondatok jobban előtérbe kerülnek, hiszen ha a szerző egy szakkifejezést többször használ, akkor az azt tartalmazó mondatok meg is kapják érte a pontértéket. Ellentmondásnak tűnhet a két output közötti magas átfedés. Ne feledjük azonban, hogy a szópárokért, szóhármakért, az esetleges szónégyesekért mindkét módszernél pluszpontok járnak. Ez, illetve az első bekezdés súlyozása is oka a több közös mondatnak a két kivonat között.

Mivel a Koltay Tibor cikkében található szakszavak nemcsak a könyvtártudományhoz köthetők, hanem interdiszciplinárisak, *elmondható, hogy az automa-*

tikus kivonatolás szakcikkek esetében hatékonyabban használható. Itt a leggyakrabban előforduló kifejezések 99%-a megtalálható a Szószablya szótár szavai között is, ezért nem tapasztaltunk lényeges eltérést a két módszer kivonata között.

Következtetések a felmérés eredményei alapján

A hallgatói mintacsoportok kivonataiban nagy az átfedés, a szakemberek és a szerzők kivonatával több közös mondatot találunk, de már szerkezetükben, összetételükben is eltérnek a hallgatói csoportok eredményétől.

A mintacsoportok kivonatainak elemzésénél kimutatható az első mondatok hangsúlya. Koltay Tibor cikke esetén az első három mondat mind a négy mintacsoportnál megtalálható,¹⁰ míg Prokné Palik Mária-nál az első két mondat található meg minden mintacsoportnál, és mellette még mindegyiknél van más mondat is az első bekezdésből. Ennek hatására a programba is beépítésre került, hogy az első bekezdés mondatai nagyobb súlyt kapjanak. Azonban a két alapul szolgáló cikk közül az elsőnek csak egyetlen mondatból áll az első bekezdése, így ezért a súlyozással csak az első mondat került be a számítógépes outputba, míg a másik cikknél hat mondatból áll az első bekezdés, ezért a súlyozás hatására három-öt mondat bekerült a kivonatba.¹¹ A mintacsoportoknál mindkét esetben három-három mondat származik a szöveg elejéről. Ezzel a módszerrel csak közelíteni lehetett a mintacsoportok eredményéhez, teljes mértékben nem sikerült elérni.

Kérdés, hogy szükséges-e egyáltalán ugyanolyan kivonat elérése, mint a mintacsoportoknál kapott eredmény? Beszélhetünk-e globális kivonatról, az

„abszolút lényegről”? Ennek megválaszolása nem könnyű, ezért nézzük meg a következő elemzést, amely talán közelebb visz minket a megoldáshoz!

Az automatizálás szempontjából a legfontosabb azoknak az elveknek a feltárása, amelyek az összes mintacsoportra teljesülnek. A kivonatok részletes elemzése már megtörtént, bemutattuk a csoportok kivonatainak szerkezetét, a közös mondatok számát, elhelyezkedését, csoportonkénti átfedését. Annak a lényeges eredménynek a kimutatására azonban még nem került sor, amelyben megvizsgáljuk, hogy hány olyan mondat van, amely az összes mintacsoportnál, illetve a szerzőnél is megtalálható a végső kivonatban.

Koltay Tibor cikke esetén három olyan mondat van, amely az összes mintacsoport és a szerző kivonatában is megtalálható.¹²

Prokné Palik Mária cikke esetén *nincs egyetlen olyan mondat sem, amely mind a négy mintacsoport és a szerző kivonatában is megtalálható lenne*. Hat olyan mondat van,¹³ amelyek a mintacsoportok kivonatában közösek, ebből három található meg a kivonatóló program Luhn módszerével történő előállítás során, szótáralapú előállításnál pedig egyik sem szerepel a 20%-os kivonatban.

Azt tapasztaljuk, hogy a négy mintacsoportot és a szerzőt alapul véve a különböző személyek eltérően látják a cikkek lényegét. Bár *hasonlóságról beszélhetünk, de azonosságról semmi esetre sem*.

Megvizsgáltam az egyéni eredeti kivonatokat. A 340 ember kivonata között *nincs két olyan személy, aki pontosan ugyanazt a kivonatot* – azonos mondatok, azonos sorrenddel – *hozta volna létre bármely cikket is véve alapul*. Majd a követelményekből engedve az is vizsgálat tárgya volt, hogy hány olyan személy van, akinek a kivonatába ugyanazok a mondatok kerültek, eltekintve a sorrendtől. Ennek eredménye a 6. táblázatban látható.

Koltay Tibor cikkénél van 7 olyan személy, akik ugyanazon mondatokat választották be a kivonataukba, míg a másik cikkénél 6 azonos kivonatot találunk. Ez nagyon alacsony érték! A felmérés eredménye alapján *nem beszélhetünk egyedüli, tökéletes KIVONAT-ról*. Láthatjuk, hogy a vizsgálati alanyok között alig találunk olyan személyeket, akik pontosan ugyanazt a kivonatot készítették volna el, így a számítógépes kivonatólástól sem várhatjuk, hogy megfeleljen egy adott eredmény-

nek, mert nincs ilyen mintakivonat. A súlyozással előállított kivonat jó képet ad az egyes mintacsoportok tagjainak véleményéről, alkalmas a szerkezeti összetétel elemzésére, szabályszerűségek levonására, illetve a súlyozás jól tükrözi, hogy egy mondatot a többség előkelőbb helyre tesz, vagy kevésbé fontosnak tart. Az így előállt kivonat azonban nem tekinthető globális kivonatként, hiszen amint kiderült, ezt mindenki másként látja.

6. táblázat

Egyező kivonatok alakulása

	Esetek száma	
	Koltay Tibor cikke	Prokné Palik Mária cikke
7 azonos kivonat	1	-
6 azonos kivonat	-	1
5 azonos kivonat	-	-
4 azonos kivonat	2	4
3 azonos kivonat	8	14
2 azonos kivonat	44	41
Teljesen különböző kivonat	174	149
Nem készített kivonatot	39	45

A két cikk eltérő témája szintén jól tükrözi az automatizálhatóság nehézségeit! Míg egy interdiszciplináris témánál a mintacsoportok között nagyobb egyetértéssel találkozunk,¹⁴ addig az egy szaktudományhoz kapcsolódó cikk már jobban megosztja az olvasóközönséget. A számítógépes kivonatóló program két módszerrel történő szignifikáns szövmeghatározása is eltérő képet mutat az eltérő tartalmú cikkek esetén. *Szakszöveg kivonatólása során Luhn módszere hatékonyabbnak bizonyul*,¹⁵ nagyobb átfedést mutat a mintacsoportok súlyozással létrejött kivonatainak mondataival, mint a Szószablya szótár alapján történő kivonatkészítés. *Az interdiszciplináris szövegnél nincs lényeges eltérés az általános szótáron alapuló és a Luhn módszerével történő kivonatkészítés között*.

Összefoglalva: megéri esélyt adni a kivonatólás automatizálásának. Valószínű, hogy programmal még sokáig nem lehet a művekben lévő gondolatokat visszaadni; sőt még a mondatkiválasztás eredményessége is kritizálható, de mint láthattuk, az emberi kivonatkészítés sem a sémákon alapuló egységes gondolatok tükröződése. Minden ember

maskent választja ki a relevans mondatokat, es a kulonböző szakteruleteken dolgozó szakemberek látják a legeltérőbben a cikkek lényegét. Ezért úgy vélem, megéri energiát fektetni egy magyar nyelven működő kivonatoló programba, és lesznek terulettek, ahol majd ennek hasznát veszik.

Irodalom és jegyzetek

- ¹ A tiltott szavak nincsenek figyelembe véve.
- ² Az eredményeket részletesen az alábbi cikkben ismertetem: LENGYELNÉ Molnár Tünde: A kivonat-készítés sajátosságai egy felmérés adatainak a tükrében. = Könyvtári Figyelő, 53. köt. 2. sz. 2007. p. 475–495.
<http://www.ki.oszk.hu/kf/e107_plugins/content/content.php?content.56T>
- ³ KOLTAY Tibor: Szöveg, információ, relevancia: néhány adalék a témakörhöz. = Könyvtári Figyelő, 51. köt. 3. sz. 2005. p. 514–518.
- ⁴ PROKNÉ Palik Mária: A tartalmi feltárás problémái online könyvtári katalógusokban. = TMT, 52. köt. 11–12. sz. 2005. p. 525–527.
- ⁵ Mivel a cikkeknek 20%-os tömörítését vártam el a felmérésben résztvevőktől, ezért mindkét cikk esetén 17 mondatot kellett megjelölniük.
- ⁶ Az adatok részletesen megtekinthetők az Irodalom 2. hivatkozásában.
- ⁷ Az egyetemista informatikus könyvtáros mintacsoport kivonatával az egyezés 47%; a főiskolai informa-

tikus könyvtáros mintacsoport kivonatával az egyezés 35%; a magyar szakos mintacsoport kivonatával az egyezés 47%.

- ⁸ A magyar szakosoknál 1,6-szoros.
- ⁹ A szerző kivonata a mintacsoportok kivonatával 6–6–2–4 mondatban egyezik meg (a táblázatban található sorrendnek megfelelően); a szakemberek kivonata a hallgatói csoportokkal 7–7–9 mondatban egyezik meg.
- ¹⁰ Az egyetlen kivétel: a főiskolás informatikus könyvtáros hallgatóknál a második mondat nem szerepel a kivonatban, de az első és a harmadik náluk is.
- ¹¹ Szószablya, illetve Luhn elve alapján történő kivonatolás során.
- ¹² Az 1., a 3., illetve a 26. mondat.
- ¹³ 1., 2., 7., 19., 30., 41. mondatok.
- ¹⁴ A súlyozás után létrejött kivonatokra alapozva.
- ¹⁵ Ha a hatékonyság mértékének mintacsoportokként az egyéni kivonatokból létrehozott kivonatokkal való egyezést tekintjük.

Beérkezett: 2009. IX. 14.-én.



Lengyelne Molnar Tunde

az Eszterházy Károly Főiskola Média-informatika Intézetén főiskolai docens, tanszékvezető-helyettes.
Email: mtunde@ektf.hu

Népszerűbb e-könyvek, több kalózmásolat

Az e-könyvek növekvő népszerűségével párhuzamosan egyre inkább attól rettegnek a kiadók és a szerzők, hogy ezek piacán is megjelennek a kalózkópiák.

A fájlcsere, illetve a kalózmások megjelenésének lehetősége már most komoly aggodalmat kelt a könyvkereskedők és -kiadók körében. Abban ugyanakkor a piaci szereplők egyetértenek, hogy nem akarják ugyanazokat a hibákat elkövetni, mint amiket a zeneipar korábban.

A kiadók, a kereskedők és a szerzők félelme nem alaptalan. Néhány fájlcsere-portalon már felbukkantak az első e-könyv-kalózkópiák. Egyelőre még nem lehet komoly problémáról és bevételkiesésről beszélni, de a könyviparnak már most megfelelő stratégiát kell kidolgoznia, ha nem akar a zeneipar sorsára jutni.

Az elmúlt évek nem igazán voltak sikeresek a könyvkereskedők számára. A komoly érdeklődésre számot tartó művek eladása ugyanis az Egyesült Államokban tavaly 13 százalékkal csökkent. 2008-ban az e-kiadványok csak az eladott művek 1,6 százalékát tették ki. Idén nyáron viszont már rekordokat döntött a forgalmuk az USA-ban, így ez az arány a közeljövőben jelentősen megváltozhat. Ennek oka, hogy egyrészt a korábnál több elektronikus könyvet adnak el világszerte, másrészt a gyártók is folyamatosan újabb olvasókat dobnak piacra. A szakemberek e két dolog miatt szinte biztosra veszik, hogy a kalózpéldányok száma is nőni fog.

Ed McCoy, az Association of American Publishers ügyvezető igazgatója szerint már most megfigyelhető, hogy jelentősen emelkedett a fájlcsere hálózatokban elérhető illegális e-könyvkópiák száma.

Tény, hogy alig néhány órával a rég várt alkotás megjelenése után a kalózok is lecsaptak Dan Brown új könyvére. A hangoskönyvet és az e-könyvet több tízezeren töltötték le illegálisan.

/SG.hu Hírlevél, 2009. október 6., <http://www.sg.hu/>

(SzP)

Adattármustra. Múzeumok

A sorozat egyes számaiban független szakemberek értékelik a magyar nyelvű internetes könyvtári szempontból fontos tartalomszolgáltatásait megadott szempontok szerint; ez alkalommal a múzeumi 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.

Néprajzi Múzeum – On-line adatbázisok

Honlap:

<http://www.neprajz.hu/tartalom.php?menu2=28>

E-mail: db@neprajz.hu

Fenntartó: Néprajzi Múzeum

Értékelés dátuma: 2009. június 18–22.

Értékelő: Ternai Zita

Általános értékelés

A Néprajzi Múzeum számos gyűjteménnyel rendelkezik, melyek egy részét – műtárgyakat, fotókat, kéziratokat, rajzokat és archív népzenei felvételeket – már korszerűen digitalizálták, adatokkal ellátták és gyűjteményi adatbázisokban visszakereshetővé tették. „On-line adatbázisok” összefoglaló néven a Néprajzi Múzeum honlapjáról érhető el a „Gyűjtemények” menüpontban. A szolgáltatás háttéréről és történetéről csak annyi derül ki, hogy az internetes megjelenítés két pályázat keretében (IHM-ITP-2/A és IHM-ITP-8) valósult meg, az *Informatikai és Hírközlési Minisztérium* támogatásával.

Az adatbázis-gyűjtemény két pillérből áll, hat nagyobb egységre osztva. A műtárgygyűjteményekhez két adattár tartozik, a többi az ún. *Etnológiai Archivum* köré rendeződik. Az adatbázisok az alábbi sorrendben érhetőek el:

- **Néprajzi Múzeum műtárgygyűjteménye:** a múzeum 18 különböző gyűjteményéből válogatott 30 808 tárgy adatait foglalja magában, 52 000-nél is több színes, jó felbontású fényképfelvétellel illusztrálva.

- **Beregi Múzeum, Vásárosnamény műtárgygyűjteménye:** a vásárosnaményi Beregi Múzeum néprajzi témájú műtárgygyűjteményének digitalizálásával a Néprajzi Múzeumot bízták meg. Az eredmény 4 000 körüli darabszámú adat és színes kép.
- **Fotógyűjtemények (Fénykép- és Diapozitívgyűjtemény):** „18 619, többségében 1941 és 1955 között készült, vegyes néprajzi témájú, veszeléyztetett (nitrátos) fekete-fehér archív fotónegatív” és 2 000 színes diapozitív.
- **Kéziratgyűjtemény:** a feldolgozottságáról pontos adatokat nem írnak, az ismertető szerint a legrégebbi és legrosszabb állapotú könyvek digitalizálása folyamatosan zajlik. Két blokkra bontották, két gyűjtőhöz kapcsolódóan: *Abafi (Aigner) Lajos* és *Bíró Lajos* gyűjtéseire. Ha rákattintunk az egyes témákra, a szkennelt anyag azonnal olvasható.
- **Népzene gyűjtemény és Népzene gyűjtemény – fonográf felvételek:** e gyűjtemény feldolgozása izgalmas téma. A 20. század elején fonográfot használtak népzenei felvételek készítésére, később ezeket írták át kottába. A Néprajzi Múzeumban átörökítették az utókornak a fonográf-hengereket: lejegyezték a kottáit, ennek eredményeként 3 300 digitalizált támlap jeleníthető meg és ezernél több fonográf-felvétel hallgatható mp3 formátumban.
- **Rajzgyűjtemény:** adatbázisában 1 387 néprajzi tárgyról készült rajzot és akvarellt láthatunk.

A tesztelés során talált hiányosságok a következők:

- A szolgáltatás előzményeiről, magáról a projektről, a munkában részt vevő személyekről (esetleg intézményekről) semmit nem írnak. Nem tudjuk, mikor kezdődött, illetve hogy lezárult-e a fejlesztés? Szokták-e még frissíteni az adatokat?

Nem említik, milyen rendszerben készültek az adatbázisok. Minimális súgója van, ami csak az „On-line adatbázisok” menü legalján levő hivatkozásból érhető el; a keresőben nem helyeztek el rá utalót.

- A hat különböző adatbázisban nem kereshetünk egyszerre, hiányzik egy közös keresőfelület és egy egyszerű böngésző is. Ráadásul az adatbázisok használatát nagyban hátráltatja az igen gyakori 503-as hibaüzenet: „Service temporarily unavailable”, azaz: a szolgáltatás ideiglenesen nem elérhető.
- A keresőfelület használata nehézkes, navigációja hiányos. A legnagyobb probléma, hogy a rendszer alapbeállítása alapján automatikusan csonkol előre és hátra, amit sehogy sem lehet megszüntetni. Logikai operátorokat sem enged használni – gyakorlatilag semmilyen megszokott speciális karakterre nem reagál. A találatokat nem lehet menteni.
- A honlap legtöbb része angolul is olvasható, ugyanakkor az adatbázisokhoz ki van írva, hogy: „This page is available only in Hungarian.”

Összegzés: Páratlan anyagot tett közzé a Néprajzi Múzeum. Az adatbázisokkal ismerkedvén rögtön az elején beütöttem a szülővárosom nevét és meglepetésemre elég nagyszámú találathalmaz érkezett, amelyet nagy örömmel tanulmányoztam. A digitalizált képek jó minőségűek, precízen, maximális mennyiségű adattal kiegészítve. A fonográf-felvételek nyilvánosságra hozatala kuriózumnak számít. Az, hogy valamennyit egyenként végig lehet hallgatni elejétől a végéig, felbecsülhetetlen érték. Kár viszont, hogy mindezeket egy elavult rendszer alatt futtatják, hogy egy ilyen kevésbé kidolgozott kereső áll a felhasználók rendelkezésére a példátlan gyűjtemény darabjainak előcsalogatására. A használhatóságot és a hatékonyságot nagyban növelné, ha a fejlesztők néhány napot fordítanának a rendszer korszerűsítésére (legalább a legalapvetőbb hibák korrigálására és a navigáció javítására) és a felület barátságosabbá tételére. Szükség lenne legalább egy angol nyelvű változatra is, hiszen nemzetközi viszonylatban is különleges az elérhető anyag.

Részletes értékelés

Felület, használhatóság, segítség

- Az adatbázis-gyűjtemény be van építve a Néprajzi Múzeum honlapjába. Kétféleképpen lehet elérni. Ha a www.neprajz.hu főoldalon az ún. „Gyorslinkek”-ből kattintok az „On-line adatbázisok”-ra, új lapon nyílik meg. Ennek az az elő-

nye, hogy elegendő egy kattintás, hátránya, hogy ez a lehetőség csak a nyitóoldalon szerepel. A másik útvonal ugyan a honlap bármely részéről bármikor elérhető, de sajnos eléggé el van bújtatva, nincs menübe kiemelve. Első alkalommal végig kellett nézennem egyenként a teljes menüsört. Az „On-line adatbázisok” a „Gyűjtemények” menüpont részeként szerepel, egy almenüben. Innen rákattintva viszont ugyanarra az ablakra nyitódik. (A kezdőlap forrását megnézve valószínűsíthető, hogy egy apró, de kellemetlen következetlenség miatt nyílnak különbözőképpen a hivatkozások.)

- Az adatbázisok nyitólapja általános bemutatást ad róluk: mindegyikről van egy rövid leírás, a nevükre kattintva pedig – tartalomtól függően – a keresőre, kéziratoknál az olvasható anyagok listájára, népzenenél a hallgatható zenék jegyzékére ugrik (1. ábra). A projektről magáról azonban semmit nem olvashatunk.



1. ábra A szolgáltatás nyitólapja

- Minden adatbázishoz külön kereső tartozik. Nincs közös keresőfelület, sem egyszerű böngészési lehetőség. Az egyes keresők sémája egységes, de nem látszik a nevük, így a rendszert kevésbé ismerő felhasználó, aki nem tudja fejből, melyik adatbázishoz melyik űrlap tartozik, bizony elveszítheti a fonalat, hogy melyiket is használja éppen. Hely bőségesen lenne az információ elhelyezésére. Adatbázisváltáskor mindig vissza kell lépni (a böngésző „visszanyíl” gombjával egyenként!) az ismertetőhöz, hogy ott egy másik adatbázisra kattintva, megjelenjen annak a keresője. Ezt a nehézkes folyamatot le kellene egyszerűsíteni például úgy, hogy elhelyeznének egy felsorolást az adatbázisokról, ahol kiválaszthatnánk, hogy mely(ek)ben kívánunk keresni. Léte-

zik ugyan egy adatbázis-választó gomb, de nem működik, helyette két adatbázis indexét mutatja. Ez a hiba számomra arra volt jó, hogy adatokat nyertem az alkönyvtárak utolsó módosításának időpontjáról. Mivel eddig sehol nem találtam időadatot a projektre vonatkozóan, ez a véletlen most nagyon hasznos volt. Kiderült, hogy a népzenei felvételeket 2005-ben, a műtárggyűjtemény egy részét 2009. április 1-jén módosították utoljára.

- A lap legalján található a „További információk” néven futó sűgőhöz vezető hivatkozás. Ez a nem túl beszédes név először számomra azt sugallta, hogy az adatbázisokról fogok még bővebb ismereteket kapni, de meglepetésemre egy tájékoztató van itt a keresőfelület használatáról. A sűgő nem túlságosan informatív, de a legnagyobb hibája mégsem ez, hanem az, hogy máshonnan nem érhető el. A keresőben sehol nincs rá hivatkozás! Statisztika, számláló pedig nem található.
- Az adatbázisok külleme, színvilága teljesen igazodik a Néprajzi Múzeum honlapjának arculatához. A honlap méretezése elavult, kis felbontásra tervezett. A keresőnél szintűgy: ott baloldalra csúszik az egész hasznos felület. A honlapon picik a betűk, alig látni a menűket és a nyelvválasztást szimbolizáló zászlókat. A múzeum logója (főleg az angol) olyan apró, hogy szövegrésze nem is olvasható.
- A karakterkódolás az ismertetőkből hibás, nincs idézőjel, helyette kérdőjel áll mind Mozillában, mind Explorerben. (Például: ...Mátray a 19. század legjelentősebb zenetudósa, akinek ?Magyar Népdalok Egyetemes Gyűjteménye? címmel három füzetet jelent meg.)
- A fenntartó szervezet a Néprajzi Múzeum. Az adatbázis használatával, működésével kapcsolatos észrevételeket, javaslatokat egy e-mail címre lehet elküldeni.

• Kereső funkciók, találatok

- Az adattárakat két típusra lehet osztani. Egy részük kereső használatát igényli (Műtárggyűjtemény, Beregi Múzeum, Fotó-, Népzene- és Rajzgyűjtemény) míg másoknál (Kézirat- és Népzene-fonográfgyűjtemény) egy listából kiválasztva azonnal teljes szöveggel olvasható vagy hallgatható az anyag.
- A Műtárggyűjtemény keresőjének űrlapja nyolc mezőből áll (2. ábra). A „Gyűjtemény” listája legördülő menüből nyílik, nincs egységesítve. A *gyűjtemény* szót többféleképpen is jelölik, rövidítik: pl. *Óceánia-gyűjtemény*, *Táplálkozásgyűjtem.*, *Állattartás-pásztorm. gy.* Ez utóbbinál a *pásztorm.* kifejezést nem tudtam értelmezni. A honlapon

visszakerestem az egyes gyűjtemények felsorolását és leírását, ott oldották fel: *pásztorművészet*.

2. ábra A Néprajzi Múzeum műtárggyűjteményének keresője

- A „Leltári szám”, a „Tárgyleírás és a „Keresés minden mezőre” adatmezők szabadszavasak. Ez utóbbi mező olyan – nem publikus – adatokra is működik, amelyek keresőkérdekként közvetlenül nem érhető el. A rendszer megkülönbözteti a kis- és nagybetűket, automatikusan csonkol a – megszokott – hátra mellett előre is. A sűgőban a csonkolás e módját jelezték is, egy érdekes példával illusztrálva: „akár el is hagyhatjuk a kezdőbetűket annak érdekében, hogy egyetlen találati halmazba kerüljenek a kis- és a nagybetűvel kezdődő azonos szavakhoz tartozó rekordok.” Ez mind szép és jó elképzelés, de a valóságban szinte használhatatlanná teszi a szabadon kitölthető keresőmezőket, mert semmilyen – általában e célra használt – speciális karakterrel nem tudom megszüntetni sem az előre-, sem a hátra-csonkolást! Megoldhatatlan kérdés, hogy csak azokat a találatokat adja ki a rendszer, amelyekben mondjuk a *kés* szó szerepel, mert például a *készül* kifejezés is mindig az eredmények közt lesz. Jó lenne legalább a találati listában kiemelni a keresőszavakat, mert akkor gyorsabban észrevennénk a releváns tételeket. A csupa nagybetűs próbálkozásra (*KÉS*) csak 19 találat érkezett, ebből az egyik tárgyleírásában (NM 74179 leltári számú) rengeteg a nagybetűs szó. Ezek nem fognak találatként megjelenni, ha hagyományosan: kisbetűsen keresünk rájuk. A „Leltári szám” mezőbe beírva a % jelet, 30 888 találat érkezett. Az adatbázisról írt tájékoztatóban ez a szám csak 30 808, ami elírás lehet. A sűgőban olvasható, hogy a kereső adatmezők között *ÉS* reláció van. Rengeteg sikertelen próbálkozásom után valószínűsíthető, hogy a többi logikai operátor

nem használható; feltehetően az általában erre szolgáló speciális karaktereket itt is a keresőszo részének tekinti a rendszer.

- A „Tárgynév” mező legördülő: abc-ben fel van sorolva az összes választható lehetőség. Ezt jó megoldásnak tartom, mert támpontot, kiindulási alapot adhat a felhasználónak. De a listában külön szerepelnek például az *Abronsos hálósúly* és a *Hálósúly* tárgynevek. Ha azt szeretném, hogy a keresésem során az összes olyan rekord megjelenjen, amelyben szerepel a *hálósúly*, azt csak a „Keresés az összes mezőre” rubrikában tudom megtenni, de el kell hagynom a kezdőbetűjét, mert szerencsétlen módon a tárgynevek nagy kezdőbetűvel szerepelnek a listában, ám ha a keresett szó egy összetett kifejezés második tagja, akkor már kisbetűvel írják. És ennél a kifejezésnél még szerencsém is van, mert az *álósúly* önmagában nem értelmes szó, így nem kapok egyéb hamis találatokat. Érdemes lenne ismételtetővé tenni a mezőket, hogy ne ilyen nyakatekert módon lehessen csak a találatokat kinyerni.
- A „Gyűjtés-”, „Készítés-” és „Használat helye” legördülő mezők közös adatszótárakkal rendelkeznek. Így előfordulhat, hogy nem minden keresőkérdés ad találatot. A hely meghatározása tág fogalom: falu, vármegye, kontinens stb. egyaránt előfordul a listán. Egymás alatt szerepelnek az *Afrika* és *Afrika?* kifejezések. Mivel itt sem tudom a mezőket ismételni, nem lehet egyszerre megnézni az afrikai és a feltehetően afrikai gyűjtéseket.
- A kereső űrlap alján ki lehet választani, hogy hány találatot mutasson egyszerre: 1, 5, 10, 20, összes. Alapként 10 van beállítva. Feltétlenül szükséges lenne a 20 és az összes közé köztes megoldásokat is elhelyezni, mert az előzőekben említett 30 888 találat lapozása így elképesztő türelemjáték. Ha pedig az összes-t választom, hosszú percek telnek el és még mindig csak a találatok megjelenítésén dolgozik a gép... Az „Űrlap törlése” gomb értelemszerűen funkcionál. A „Keresés indítása” a gombra kattintással és Enterrel egyaránt működik.
- Ha nincs találatunk, csak ennyit ír ki – nem ajánl fel segítséget, módosítási lehetőséget, de még a sűgőhoz sem irányít el. Ha módosítani szeretném a keresést, nincs rá mód, a rendszer nem kínál rá opciót, ami nagy hiba. Elvileg, ha a böngésző visszanyúl gombjával megyek vissza a keresőre, az utolsó állapotához kellene kerülnöm, de itt egy üres űrlap fogad. A találatok megjelenítésénél nem látszik a keresőkérdés. A lista tetején fel van tüntetve az összes találat (3. ábra). Jó

megoldás, hogy a lap tetején és alján egyaránt lapozhatunk, és hogy ki van írva, éppen hol tartunk: „Összesen 1300 találat. Ebből jelenleg a(z) 11-20 sorszámú megjelenített.” Ez a kiírás a tájékozódás szempontjából elengedhetetlen, de feleslegesen hosszú a szöveg. A találatok meg vannak számozva. Hiányzik a listaformátum, azonnal a teljes rekord jelenik meg táblázatos formában. Ez a megoldás nem praktikus, mert a rekordok leírása nagy kiterjedésű, hosszan kell görgetni a következő találatig. Például a „Leltározó” és az „Adatrögzítő” mező a laikusok számára felesleges információ, elég lenne a belső hozzáférésnél mutatni. A rendszer egyáltalán nem rugalmas, nem választhatjuk meg, hogy mely adatok és milyen sorrendben jelenjenek meg a képernyőn. Minden mező látszik, előre meghatározott sorrendben. A „Tárgynév” csak a 3. sorban szerepel, pedig szerintem ez az érdeklődőknek, kutatóknak a legfőbb adat. Fontossági sorrendben másodikként a hozzá tartozó digitális képet tenném, hogy a tárgyat be lehessen azonosítani. Ez viszont csak a leírás végén kap helyet, így nem is látjuk azonnal, mert a műtárgyak adatai nem férnek rá egy képernyőre. A fénykép kis terjedelmű, de rákattintva új ablakban, nagy felbontásban jelenik meg. Explorerben, ha az automatikus nagyságú ablakba nem fér bele a kép, nem engedi teljes képernyőre nagyítani, így egészében nem tudjuk megtekinteni a fotót, csak görgetni lehet. Néha több kép is van: más szögből, közelebről, ezek sajnos csak egyenként nézhetők meg nagy méretben, mert nincsenek

Vissza az adatbázisválasztóhoz Új keresés																																									
Összesen 30888 találat. Ebből jelenleg a(z) 1-10 sorszámú megjelenített.																																									
következő 10 találat utolsó oldal																																									
1	<table border="1"> <tr><td>Gyűjtemény</td><td>Afrika-gyűjtemény</td></tr> <tr><td>Leltári szám</td><td>NM 80545</td></tr> <tr><td>Tárgynév</td><td>Csörgő</td></tr> <tr><td>Helyi elnevezés</td><td></td></tr> <tr><td>Gyűjtés helye</td><td>Belga Kongó</td></tr> <tr><td></td><td>Belga Kongó</td></tr> <tr><td></td><td>busongo</td></tr> <tr><td>Gyűjtő</td><td>Torday Emil</td></tr> <tr><td>Gyűjtés ideje</td><td>1907-1909</td></tr> <tr><td>Készítés helye</td><td></td></tr> <tr><td></td><td></td></tr> <tr><td></td><td></td></tr> <tr><td>Készítés éve</td><td></td></tr> <tr><td>Készítés kora</td><td>1890-1909</td></tr> <tr><td>Készítő neve</td><td></td></tr> <tr><td>Használat helye</td><td></td></tr> <tr><td></td><td></td></tr> <tr><td></td><td></td></tr> <tr><td>Használat éve</td><td></td></tr> <tr><td>Használat kora</td><td></td></tr> </table>	Gyűjtemény	Afrika-gyűjtemény	Leltári szám	NM 80545	Tárgynév	Csörgő	Helyi elnevezés		Gyűjtés helye	Belga Kongó		Belga Kongó		busongo	Gyűjtő	Torday Emil	Gyűjtés ideje	1907-1909	Készítés helye						Készítés éve		Készítés kora	1890-1909	Készítő neve		Használat helye						Használat éve		Használat kora	
Gyűjtemény	Afrika-gyűjtemény																																								
Leltári szám	NM 80545																																								
Tárgynév	Csörgő																																								
Helyi elnevezés																																									
Gyűjtés helye	Belga Kongó																																								
	Belga Kongó																																								
	busongo																																								
Gyűjtő	Torday Emil																																								
Gyűjtés ideje	1907-1909																																								
Készítés helye																																									
Készítés éve																																									
Készítés kora	1890-1909																																								
Készítő neve																																									
Használat helye																																									
Használat éve																																									
Használat kora																																									

3. ábra A Műtárggyűjteményi adatbázis találati listája

nyilak, amelyekkel végig lehetne nézni egymás után valamennyit. Internet Explorerben, ha jobb egérrel a kép tulajdonságaira kattintok, nem látom a kép típusa, azaz „Nem érhető el”, de a Mozilla Firefoxban megtudható, hogy .jpg kiterjesztésűek. A „Tárgyleírás” mező pozitívuma, hogy igen aprólékos, minden részletre kiterjedő bemutatást ad. A „Gyűjtés-”, „Készítés-” és „Használat helye” rubrikák több részből állnak, a földrajzi hely megjelölése egyre tágabb: falu, város, (vár)megye és végül az ország megjelölése.

- A találatok megjelenési sorrendjének logikájára nem jöttem rá. Néhol az egyes gyűjteményen belül leltári szám szerint növekszik, de a leltári számok sem egységesek (pl.: *NM 2007.39.248*, *NM 80.18.1*, *NM 26347*). Ha a találatok között több gyűjtemény is van, a gyűjtemények sorrendje nem követi a betűrendet.
- A Beregi Múzeum műtárgygyűjteménye és az összes többi adatbázis is azonos sémával rendelkezik, mint az előző, csak más mezőkre lehet keresni. Egyetlen figyelmeztetést vettem észre: míg az első adatbázisnál „Leltári szám” nevet kapott a keresőmező, az összes többinél ez „Leltári jelzet”. A „Tárgynév” nevű legördülő mező szintén betűrendben jeleníti meg a listát. Találtam egy kifejezést, amelyet nem egységesítettek. Megtalálható *asztalfutó*; *asztali futó* és *asztalfutó*, *csipketerítő* néven is. Két helytelen, nehezen értelmezhető adatmező név is van: „Gy. Megye” és „K. megye”. Ezek feloldására rengeteg hely állna rendelkezésre. A találatok megjelenése itt annyival szerencsésebb, hogy egy rekordhoz kevesebb adat tartozik, így egy képernyőn elférnek, és látszik a hozzájuk kapcsolódó kép is.
- A Fotógyűjtemények keresőjének érdekessége, hogy a *Dokumentumtípus*nál sok lehetőséget kínál fel, ez jól szűkíthetővé teszi a találati listát. A „Készítő neve” ábécésorrendből választható. A listából kijelöltem a *Győrffy István* és *Viski Károly* készítőket, de nem érkezett találat!? Ha csak *Viski Károlyt* választom, egy találat van, de abban nem szerepel *Győrffy*. A „Készítés ideje”-nél a sugó nem jelölt meg szabályos beviteli dátumformátumot, és a találatok nézegetése közben jöttem rá, hogy nincs is egységesítve! Példák az adatfelvitelre: *1914*, *1900-as évek*, *1940-1944*, de olyan is van, hogy *1940-44*. Utóbbi kettő nem azonos számú találatot ad. Az „Ország” nevű legördülő mezőben *Budapest* is szerepel. Elírás is akad: *D. Abánia* és *D. Albánia* egyaránt szerepel, de furcsa módon ugyanaz az öt találat van mindkét esetben. A rövidítések feloldására itt is

lenne hely. A legdurvább, legszembetűnőbb hiba éppen Magyarországnál jelentkezik: négyféle írásmóddal van a listában: *Magyar*, *Magyarország*, *Magyarország*, *Magyarország*, melyekhez különböző számú találat tartozik. Feltétlenül egységesíteni kellene őket!

- A *Népzene gyűjtemény*ben a „Dokumentumtípus” adatmező neve „Dok.típus”. A Fotó- és Rajzgyűjteménynél ugyanez nincs rövidítve, ez megint egy apró figyelmetlenség. Van egy olyan mező is, hogy „Életkora”: itt oda lehetne írni, hogy kinek/minek a korára vonatkozik. A következő elképesztő hiba jó néhány napig fennállt, mert a tesztelésemkor mindig jelentkezett, azonban a cikkírás pillanatában nem. (Lehet, hogy csak túlterhelés érte az adatbázis-kezelőt és időközben újraindították a rendszert.) Egy képernyőmentést készítettem róla, érdemes megemlíteni mégis. Mozillában a találatok megjelenítésénél két mező: az „Anyag” és a „Technika” hibás, a következő szöveg íródott ki: „Warning: mssql_query() [function.mssql-query]: message: Invalid object name 'Ari03_#0404_01_Vw'. (severity 16) in C:\Intranet\emir\lari\lista.php on line 56...” (4. ábra). Ez a két mező Explorerben is szerepelt, ott csak egy kis négyzetben lévő x jelezte, hogy valaminek lennie kellene ott. Jobban megnézve vettem észre, hogy ennek a két mezőnek eleve semmi keresnivalója nem lenne itt, hiszen a Rajzgyűjteményhez tartoznak!

Viszsa az adatbázisválasztóhoz Új keresés

Összesen 5475 találat.
Ebből jelenleg a(z) 1-10 sorszámu megjelenített.

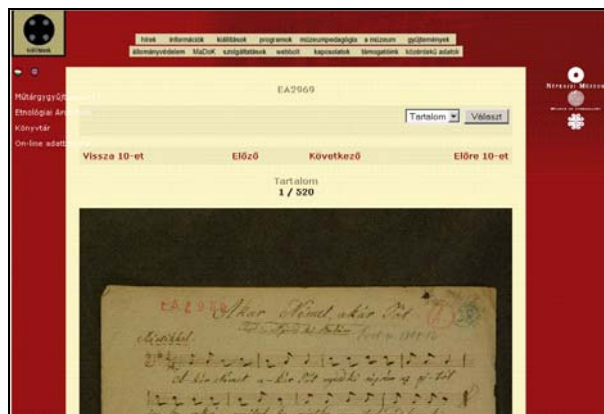
következő 10 találat utolsó oldal

1	Gyűjtemény	Népzenei gyűjtemény
	Leltári jelzet	NM MH 1
	Dok. típus	fonográf felvétel
	Gyűjtő	Vikár Béla
	Gyűjtés ideje	1896.12.
	Helynév	Mezőkeresztes
	Megye	Borsod vm.
	Népcsoport	magyar
	Adatközlő	Takács István cseléd
	Életkora	
	Előadásmód	ének
	Kezdősor	Fehér László lovat lopott
	Anyag	Warning: mssql_query() [function.mssql-query]: message: Invalid object name 'Ari03_#0404_01_Vw'. (severity 16) in C:\Intranet\emir\lari\lista.php on line 56...
	Technika	Warning: mssql_query() [function.mssql-query]: message: Invalid object name 'Ari03_#0404_01_Vw'. (severity 16) in C:\Intranet\emir\lari\lista.php on line 56...

4. ábra A *Népzene gyűjtemény* rekordjai hibásan jelentek meg

- A Rajzgyűjteményben az „Anyag” és a „Technika” mezőknél a felhasználónak kell megfogalmaznia a keresőkérdést. Ezt nem tartom jó megoldásnak, mert egy laikus nem tudja majd kitalálni, milyen technikák és anyagok léteznek, vagy mi ezeknek a szakszerű megfogalmazása. De tegyük fel, hogy tudja. Van olyan rekord, ahol a „Technika” mezőben a *ceruzarajz és vízfestmény* kifejezések együtt szerepelnek. Erre biztosan nem gondol senki, így fontos találatoktól eshetünk el. A találatok megjelenésénél ugyanaz a hibajelenség volt, mint a *Népzene gyűjtemény*-nél. A cikk írása közben már nem jelentkezik a rendellenesség, de előtte folyamatosan fennállt. Itt a bibliográfiai leírásban kétszer is szerepelt az „Anyag” és a „Technika” mező. Az első helyen ki voltak töltve, a másik helyen hibaüzenet állt: „Nem lehet kapcsolódni az MSSQL kiszolgálóhoz!”
- A Kéziratgyűjtemény teljes szövegű adatbázis. Itt az egyes egységek neveire kattintva nem a kereső ablaka nyílik meg, hanem az egyes gyűjtők és a gyűjtések ismertetői. Innen választhatjuk ki a megjeleníteni kívánt dokumentum hivatkozását. A digitalizált szövegek JPG képként azonnal olvashatók. A képek nem nagyíthatók, nem menthetők. Az a szerencse, hogy viszonylag jó minőségűek, jól olvashatók, de az akadálymentességtől nagyon messze állnak. A megjelenítésnél azonnal szembetűnik, hogy hiányzik a kézirat címe, és a fejléccen sem olvasható (5. ábra). Saját tapasztalatomból kiindulva mondhatom, hogy a 196. oldalnál félbehagyott kéziratához később visszatérve nem tudtam, mit is olvasok. A navigáció itt is hiányos, nem vezet vissza semmi a főoldalra vagy az adatbázisválasztóhoz. A lap tetején található egy betűszám kombináció (pl. EA 2968). Ez a leltári jelzet lehet, de a laikusok számára feltétlenül magyarázni illenék. Alatta zárójelben szerepel, hogy hány oldal. Ez azonban csak a Mesejegyzék kéziratban van feltüntetve, sehol máshol. A panel jobb felső sarkában egy legördülő menüben a Tartalom kifejezés van beírva. Ezért teljesen felesleges legördülő megoldást alkalmazni, mert egyenként végignéztem, az összes kéziratnál csak a Tartalmat lehet választani. A funkciója az, hogy visszaugrik az első oldalra. Az elégtelen navigáció itt jelentkezik leginkább. Az oldalak közötti lépdelés nem megoldott, ha nem előre vagy hátra akarok egyet lapozni, akkor a következő lépcső a *Vissza/Előre 10-et*. Ez rendkívül kevés lehetőség, 703 oldalon így eljutni a végére szinte képtelenség. Szükség volna egy végére ugró gombra, de leginkább egy olyan mezőre, ahová

a felhasználó írhatja be, hányadik oldalra szeretne lépni. Kis honlapszerkesztési hiba, hogy ha meg van nyitva bármelyik kézirat, a panelje belelóg a bal oldalra elhelyezett menübe, Mozilla és Explorer alatt egyaránt.



5. ábra Egy oldal digitalizált kézirat

- A Népzene gyűjtemény – fonográffelvételek adatbázisa okozhatja a legnagyobb fejtörést az avatatlan felhasználónak. Az ismertetőjénél pusztán ennyi áll: „MH 1-1012, mp3 formátumban”. Azonnal három kérdés is felmerülhet: mi az a fonográffelvétel? mi ez a sok szám? mit takar az mp3 formátum? Az adatbázis nevére kattintva egy irtatlan hosszú listát látunk, ami növekvő sorrendben betűkből és számokból áll. Ez szintén a leltári jelzet lehet, de nem jelzik. A kódok mellé elkéne egy rövid tartalmi ismertető. Ezt a hiányt a Népzene gyűjtemény adatbázisát használva tudom feloldani, ott látszik, melyik leltári számhoz mi tartozik. De pontosan ebből merül fel a következő kérdés: például az MH 10-nél több alrész is van, az MH 10 a-tól 10 f-ig. Ezek miért nem szerepelnek a fonográf felvételek listáján? Kiválasztva egyet, felugrik egy ablak, hogy megnyitni vagy menteni kívánom-e a fájlt? A kattintás után tehát automatikusan próbálja lejátszani a rendszert a hanganyagot, de el tudom képzelni, hogy nem mindenki tudja, hogyan is kell. Mindenképpen fel kellene tüntetni, hogy az adatbázis használatához a számítógéphez csatlakoztatott hangszóróra és telepített médialejátszóra van szükség. A tökéletes az lenne, ha kitennének egy hivatkozást, hogy a felhasználó onnan ingyen letölthesse a szükséges szoftvert. Az egyes fonográffelvételek elején felkonferálják a forrást. Ez nagyszerű beazonosítás a későbbi felhasználáshoz. Sajnos a felvételek minősége nem túl jó, sokszor alig érthető, serceg, de ez az eredeti hanganyagok minősége miatt lehet. A

fájlt el lehet menteni; ilyenkor automatikusan a henger számát kínálja a mentés nevének. A mindenkori felhasználási feltételek a Néprajzi Múzeum honlapjának alján elhelyezett „Copyright” menüből érhetők el.

- Az összes adatbázisról elmondható, hogy a találati lista viszonylag gyorsan megjelenik – ha nem több (tíz)ezres listáról van szó. A találatokat nem lehet menteni, nyomtatni, e-mailen elküldeni. Az egyes tételeket nem lehet kiválogatni, egymás mellé rakni, és a felhasználó által alkotott listát elmenteni. A használatot élvezhetőbbé és izgalmasabbá tenné, ha az egyes mezők adatai átjárhatók lennének, azaz egy beépített böngésző funkció is lenne. Gondolok itt arra, hogy ha az egyik tételnél a „Gyűjtési helye” pl. *Jósvafő*, akkor itt rákattintva *Jósvafő*-re az összes olyan tétel megjelenne, amelynél a gyűjtés helye ugyanaz.

Tartalom, minőség

- A Néprajzi Múzeum online adatbázisai hatalmas mennyiségű információt tárnak elénk. Látszik, hogy átgondolt munka van a háttérben. Az anyagok gyűjtése, rendszerezése, leírása nagyon alapos. A 30 000-nél is több, sokszor apró műtárgy fényképezése végeláthatatlan munka lehetett. A veszélyeztetett fotókat és a legrosszabb állapotú kéziratok egy részét, az adatbázisba kerülésnek köszönhetően archiválták. Remélhetőleg folytatódni fog ez a tevékenység. A digitalizált műtárgyak és kéziratok minősége megfelelő. A kutatómunkát jelentősen megkönnyítik, akár otthonról is el lehet kezdeni a felderítést, mert egy minden eddigi módszernél könnyebben áttekinthető, gyorsabban megtalálható, nem helyhez kötött, visszakereshető anyagot tett közzé a múzeum. Már csak a rendszert kellene finomítani, hogy a befektetett munka teljes mértékben hasznosítható legyen.
- Az elérhető tartalmak karbantartásának minőségén még javítani kell. A nagyon gyakori „átmenetileg nem elérhető” hibaüzenet miatt a tesztelésem kétszer annyi időt (és türelmet) vett igénybe, mint kellett volna. A Népzene- és a Rajzgyűjteményben helytelenül megjelenő „Anyag” és „Technika” mezők egy idő után automatikusan megjavultak. Az említett hibák talán elkerülhetők lennének, ha gyakrabban ellenőrizné a rendszergazda, hogy túlterhelt-e a rendszer és okoz-e ez fennakadásokat?

Metaadatok, bibliográfiai információk

- Mind a Néprajzi Múzeum, mind az online adatbázisok metaadattal való ellátottsága nagyon

szegényes, csak a *title* mezőt töltötték ki. Sokat lendítene a helyzeten az is, ha legalább az egyes rekordok XML formátumban elérhetőek lennének.

- Teljesen érthetetlen, hogy a bibliográfiai leírás miért hiányzik a digitalizált kéziratokról. Csak egy egészen leegyszerűsített leírást olvashatunk, ami messze áll a szabványostól.

Folklórdokumentációs Központ –
Multimédiás adatbázisok

Honlap: <http://www.hagyomanyokhaza.hu/main/mediatar/adatbazisok>

E-mail: fdk@hagyomanyokhaza.hu

Fenntartó: Lajtha László Folklórdokumentációs Központ

Értékelés dátuma: 2009. június 28.

Értékelő: Kőrös Kata

Általános értékelés

A *Hagyományok Házában* működő *Lajtha László Folklórdokumentációs Központ* az 1981-ben alapított *Néptáncosok Szakmai Háza* dokumentációs anyagainak örököse. Elsődleges célja hozzáférhetővé tenni a Kárpát-medence kulturális régió népi kultúráját rögzítő audiovizuális és szöveges dokumentumokat. Korszerű technikai megoldások alkalmazásával a szakértő és a laikus érdeklődő közönség számára egyaránt jól használható formában szolgáltatják a népzenei és néptáncgyűjtemények anyagait. Az adatbázisok között négy szolgáltatás vált hozzáférhetővé a Folklórdokumentációs Központ honlapjáról. A Folklórdokumentumok Keresőrendszere mellett a Martin Médiatár könyvtári keresőrendszere, a Hangzó kiadványok adatbázisa és egy Térképes kereső áll a látogatók rendelkezésére.

A Folklórdokumentumok Keresőrendszere egy gondosan felépített, komoly tudományos munkát magába foglaló adatbázist takar. A tartalmi feldolgozás legnagyobb egységeként a gyűjtemények során született audiovizuális és szöveges dokumentumokat a legkülönbözőbb szempontok szerint, hierarchikus időbeli és tematikus egységekre szegmentálva teszik kereshetővé. Az archívum egy egyedülálló néprajztudományi tezauruszra épül. Kidolgozását az anyagok feltárásával és az adatbázis fejlesztésével párhuzamosan végzik. A multihierarchikus fogalmi háló használata kiküszöböli a

névváltozatokból, elgépelésekből adódó hibákat, és sok szempontú összetett keresést tesz lehetővé tudományosan megalapozott szakszerű felépítésével.

A Folklórdokumentációs Központ adatbázisainak tesztelése alapján az alábbi fejlesztések ajánlhatók:

- A jelenlegi funkciók többsége csak Internet Explorer alatt használható. A keresőrendszer platformfüggetlen működésének megoldása növelné a szolgáltatás használhatóságát.
- A néprajzi tárgyú adatbázis széleskörű érdeklődésre tarthat számot, multimédiás adatbázisként pedig vakok és gyengénlátók számára is jól használható lehetne, ha készülné akadálymentes változat.
- A folklórdokumentumok sok szempont szerint kereshetők, de nincs szűkítési lehetőség a leírt média formátuma szerint, pedig a keresett tételek megtalálását könnyítené egy ilyen szempont bevezetése.
- A tezausz többnyelvűsítése bővíthetné a felhasználók körét. A magyar néprajztudomány külföldi kutatók és a magyar kultúra iránt érdeklődők számára egyaránt érdekes lehet.
- Az érdeklődés felkeltésére, a rendszerrel vagy akár a néprajztudomány fogalmaival ismerkedő felhasználók számára jó lehetőség volna a tezauszban szabad böngészést is biztosítani.
- A keresőrendszer fejlesztésével a felhasználók igényei jobban kielégíthetők lennének. Nemcsak listából beemelhető keresőkérdéseket is kombinálni képes összetett kereső kialakítása, elmenthető keresési történet, rendezhető és letölthető találati listák létrehozására alkalmas funkciók beépítése kényelmesebbé tenné a rendszer használatát.
- Más néprajzi tárgyú adatbázisokkal együttműködve a tartalmas szolgáltatás még további fejlesztésére látok lehetőséget. Az *MTA Zenetudományi Intézetének* adatbázisa vagy a *KFKI Magyar Népzenei Archívuma* megfelelő partner lehetne az együttműködésben.

Részletes értékelés

Felület, használhatóság, segítség

- A Hagyományok Háza honlapja jól felépített, logikus szerkesztésű oldal, színvonalas, igényesen összeállított tartalommal és megjelenéssel. A honlap szerkezetéből fakadóan azonban az egyes aloldalakon a navigáció néha nehézkes, mert a középső keretben, különösen az alsó ke-

ret vastagsága miatt, a hosszú listák az oldaltörek hiányában nehezen kezelhetők.

- A Folklórdokumentációs Központ aloldalán a központ történetéről rövid bemutatkozó szöveget olvashatunk, és megtaláljuk a munkatársak teljes listáját. A bemutatkozó szöveg fölött rögtön feltüntették a felhasználói visszajelzésre, további információ kérésére alkalmas e-mail címet és telefonszámot.
- A központ munkájáról az egyes menüpontok részletesen tájékoztatnak – bár feltűnő, hogy a rendezvények menüpont 2008 óta nem frissült. A honlapról elérhető az intézmény állományába tartozó Lajtha László- és Martin György-hagyaték teljes dokumentációja.
- Az Adatbázisok menüpont alatt négy, a tájékoztató szerint egyelőre tesztelési céllal közzétett keresőt találunk. A különféle szolgáltatásokról rövid ismertetőt olvashatunk, használatukat pedig – a Hangzó kiadványok adatbázisát kivéve – részletes, közérthető útmutató segíti, az egyes keresőfelületeknél pedig bővebb, tematikus sugógok hívhatók elő. A Folklórdokumentumok Keresőjéhez keresési tippeket is adnak a készítőik. Ez utóbbiak szemléletesebb leírása, a keresőfelületek képeinek megjelenítésével, még nagyobb segítséget jelentene. A Folklórdokumentumok Keresőrendszerének használati útmutatójában viszont négy menüpont is 404-es hibát mutat (regisztráció, általános felhasználási feltételek, adatvédelmi nyilatkozat, a tartalom megjelenítésével kapcsolatos műszaki információk).
- A honlapnak és a keresőknek van angol nyelvű változata, egyedül a hangzó kiadványok adatbázisa nem érhető el más nyelven.
- A hibaüzenetek érhetőek, a rendszer készítői figyelnek rá, hogy felhívják a felhasználók figyelmét, ha egy rekord a folyamatos feldolgozás miatt még nem teljes, vagy a hiányos adatolás miatt nem publikálható.

Kereső funkciók, találatok

- A Hagyományok Háza honlapjáról a Folklórdokumentációs Központ aloldalán az „Adatbázisok” menüpont alatt érhetjük el a különböző adatbázisokat.
- A *Hangzó kiadványok adatbázisa* az intézmény gyűjteményében megtalálható publikált népzene- és néptáncfelvételek leíró adatait teszi közzé (6. ábra). Az adatbázis első ránézésre egy egyszerű keresővel ellátott lista, melyben címek, előadók, műfajok, települések, vármegyék, tájegységek vagy gyűjtők adatai szerint kereshetünk, vagy a listát különböző szempontok szerint rendezve (a kereshető szempontokon túl albu-

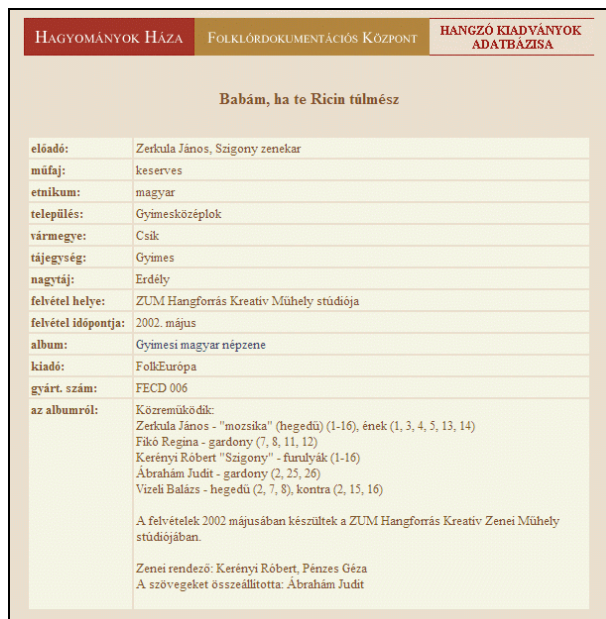
mok, előadásmódok vagy etnikumok szerint) böngészhetünk.



6. ábra A Hangzó kiadványok adatbázisa

- A böngészés során használható listák hossza sokszor nehezíti az áttekintést, és a laikus felhasználó tájékozódását nem könnyítik sem a műfajbesorolások, sem az előadásmódok kategóriái. Nem egyértelmű például, hogy mi lehet a különbség a *betlehemes* és *betlehemezés*, az *altató* és *altatódal*, az *akasztó* és *akasztós* műfajmegjelölések között, nem beszélve a nem magyar nyelvű, nagyon hasonló helyesírású kategóriákról. Egy tájékoztató segédlet a műfajokról és az előadásmódokról használhatóbbá tenné az adatbázist.
- Az egyszerű keresőbe beírt kifejezéseket kereshetjük pontosan, minden, vagy bármely szóra jellemző egyezéssel.
- A hangzó kiadványok elsődleges rendezési egyége ebben az adatbázisban az album és a felvétel. Ezért lehetséges, hogy a találati listában minden esetben csak az album vagy felvétel előadója és címe jelenik meg, hiába kerestünk más szempont alapján. Ez a táblázatos megjelenítési mód, és az, hogy nincs lehetőség a találati lista további rendezésére, a találatok szűkítésére, megnehezíti a tájékozódást.
- A találati listában a rekordok melletti *Részletek* felírra kattintva új ablakban tekinthetjük meg a kiválasztott album vagy felvétel adatlapját (7. ábra). Az adatlapon részletes leírását kapjuk a felvételnek, és az előadó, előadásmód, műfaj, etnikum, település, vármegye, tájegység, nagytáj, gyűjtő, felvétel helye és időpontja, az album címe, szerkesztője, kiadója, kiadásának éve, gyártási száma mellett megtaláljuk az forrásdokumentumra, vagyis az albumra vonatkozó legfontosabb adatokat. Egyes rekordoknál idézeteket is találunk az albumok kísérfűzetéből, és sok helyen a felvételek lejegyzett szövegei is szerepelnek. Az albumok és felvételek rekordjai között

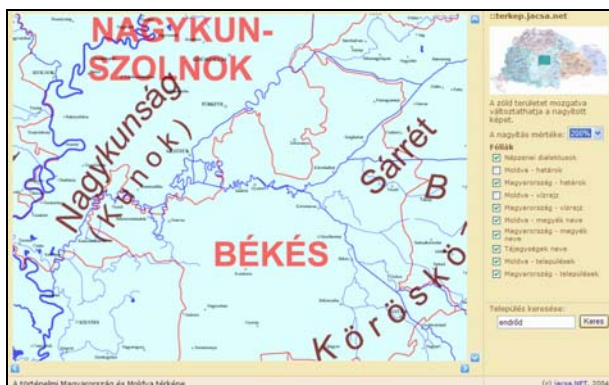
közvetlenül válthatunk az egyes címekre kattintva. Sajnos az általunk kiválasztott rekordok kijelölésére, eltárolására, lementésére nincs mód.



7. ábra Egy felvétel űrlapja

- A *Térképes kereső* egy fejlesztés alatt álló segédlet (8. ábra). Jelenleg csak Internet Explorer alatt használható, az SVG (Scalable Vector Graphics) modul telepítése után. A fejlesztők terve szerint a későbbiekben a térkép segítségével az egyes helynevekre kattintva kereshet majd a felhasználó a rekordok földrajzi név mezőiben, továbbá nyomon követheti az egyes települések összeolvadását, megszűnését, az adatok 1914-ig visszamenő feldolgozása után. Jelenlegi állapotában a történelmi Magyarország és Moldva digitális térképén, mintegy leemelhető fóliákként megjeleníthető a népzenei dialektusok kiterjedése, a moldvai és magyarországi határok, vízrajzi jellegzetességek, megyék, tájegységek és települések nevei, továbbá rákereshetünk az egyes települések neveire. Az egyes szegmensek adatlapján pedig a települések egy részénél már elérhető az a funkció, hogy a település nevére kattintva egy új ablakban megnyílik a térkép a megfelelő földrajzi helyet mutatva.
- A *Folklórdokumentumok Keresőrendszerében* két keresőfelület közül választhatunk. Az *Egyszavas keresőbe* minimum két karakter hosszú keresőkérdést írhatunk. Kereshetünk szótöredékre, többnyelvű variánsokra, a tezaurusban rokonértelmű kifejezésekre, vagy gyűjtőfogalmakra. Megadhatjuk, hogy a keresést a rendszer a tezaurusban, a szólistákban, vagy a szöveges

mezőkben is elvégezze. A rendszer tehát kezelni tudja a szinonimákat, a települések kereshetők magyar és mai román neveik alapján is. A keresőrendszer engedélyezi az ékezetek nélküli keresést, és a keresőkifejezést automatikusan csonkoltnak tekinti.



8. ábra A térképes keresési segédlet

- A találati listában a találatokat csoportokba rendezve kapjuk annak alapján, hogy mely mezőben volt találat a keresőkifejezésre (9. ábra). Kiváltható, hogy a kifejezés mint gyűjtőfogalom, szinonima, vagy pontos kifejezés érvényes az adott találatra, és hogy hány szegmensre vonatkozó találat van az adott kategóriában. A kis nyílakra kattintva haladhatunk tovább egy következő listára, amelyben gyűjtésekre lebontva szerepelnek a találatok. A megfelelő gyűjtést kiválasztva, annak adatlapjára szintén nyílak segítségével juthatunk. Ezen az adatlapon csak a keresi feltételnek megfelelő médiarészek jelennek meg, de lehetőségünk van a gyűjtéshez tartozó összes médiarész megjelenítésére is.



9. ábra Találati lista a Folklor dokumentumok Keresőrendszerében

- A gyűjtések adatlapján a gyűjtés legfontosabb adatai – helyszíne, ideje, jelzete, adatközlők ne-

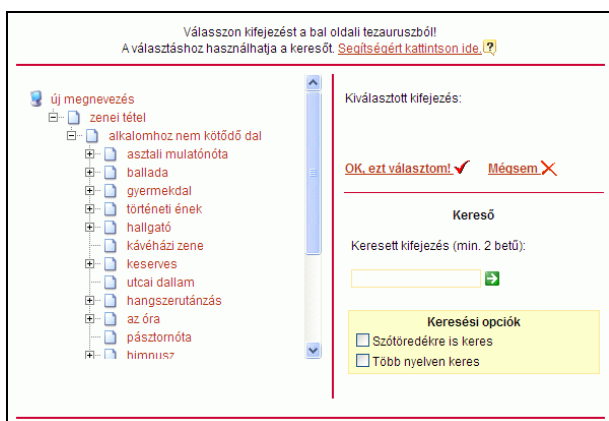
ve, további adataik (esetleg életkoruk) és a gyűjtők neve – mellett az is szerepel, ki végezte el a gyűjtés katalogizálását és az anyag szegmentálását.

- A médiarészek adatlapjára továbblépve megjelenik a médiarész szegmenseinek szerkezete: folklórszöveg, dallam, tétel, a különböző szegmenskategóriák és mellette a felvétel időrendjében egy-egy néhány percesre bontott részlet.
- A számozott szegmensekre kattintva az adott szegmens egyedi adatlapját kapjuk, és a beépített médialejátszón meghallgathatjuk vagy megtekinthetjük a hozzá tartozó felvételt (10. ábra). Lehetőségünk van egy szegmens megrendelésére is közvetlenül az adatlapról, de arra vonatkozóan, hogy pontosan milyen feltételek mellett juthatunk hozzá, a honlapon nem közölnek információt. A szegmensek adatlapja egyetlen kattintással kinyomtatható.



10. ábra Egy dallam adatlapja

- A *Tematikus keresőben* földrajzi hely, népszokás neve, fajtája, nyelv és nyelvjárás, etnikum, szereplő, előadás eszköze, előadásmód és a gyűjtés adatai alapján kereshetünk. A fogalmakat hierarchikus listákból emelhetjük be a keresőmezőbe (11., 12. ábra.) A listákban kereshetünk, akár szótöredékre, vagy többnyelvű változatra is, így könnyebb eligazodni a fogalmak hierarchiájában. Igen kellemetlen meglepetés azonban, hogy a korábbi kereséseinket a keresőúrlap megőrzi, de nem engedi kitörölni, így ha egy újabb keresésnél egy másik keresőmezőt szeretnénk használni, csak akkor tudunk módosítani a feltételen, ha egy másikat emelünk be a korábban használt mezőbe. A keresés eredményeként ennél a módszernél rögtön a keresési feltételnek megfelelő gyűjtések listáját kapjuk, és innen léphetünk tovább a felvételek, majd a szegmensek adatlapjai felé.



11. ábra Keresés a tezaurusz segítségével/1.



12. ábra Keresés a tezaurusz segítségével/2.

Tartalom, minőség

- A Folklórdokumentációs Központ adatbázisaiban nemcsak folklórgyűjtesek adatait találhatjuk: népzenei, néptánc-, népszokás- és egyéb folklór-felvételeket, hanem a folklórmozgalmak képviselői által elő- és közreadott revival felvételeket is.
- Az intézmény gyűjteményébe tartozik Martin György és Lajtha László hagyatéka, de része *Kallós Zoltán* teljes gyűjteménye, továbbá magángyűjtemények és partnerintézmények anyagai is kereshetők az adattárakban. Az adatbázisok jelenlegi formájukban tesztelési céllal váltak elérhetővé a nagyközönség számára. Tartalmuk a dokumentumok folyamatos feldolgozásával dinamikusan bővül.
- A néprajztudományi tezaurusznak köszönhetően az adatbázis rekordjait magas színvonalon tárgyszavazták és tették kereshetővé. Gépelési hibák csak elvétve fordulnak elő a szöveges mezőkben, bár néhol akadnak nem szabványos karakterkódolásból fakadó hibák, különösen a hangzó anyagok lejegyzett szövegei között. Az adatbázis tartalma összességében nagyon igényes munka eredménye.

- A tájékoztató szöveg szerint a felvételekből csak 30 másodperces részletek tölthetők le, a felhasználók viszont azt tapasztalhatják, hogy a hosszabb, 5-6 perces szegmensek is egyben hozzáférhetők. Az ellentmondást a rendszer tesztelési fázisa indokolhatja.

Metaadatok, bibliográfiai információk

- Az oldal forráskódja szinte semmilyen metaadatot nem tartalmaz. A szolgáltatás népszerűsítésére hasznos volna magukat az adatbázisokat is ellátni a megfelelő, géppel olvasható adatokkal.
- A hozzáférhetővé tett felvételek bőségesen elvannak látva látható kísérőadatokkal, azonban a metaadatok formája nem minden esetben szabványos, és lementésükre, importálásukra nincs mód.
- Más keresőmotorok vagy adatszolgáltatók számára az intézmény nem teszi elérhetővé a leíró adatait. Az együttműködés kialakításához mindenképpen jól strukturált, szabványos metaadatokra volna szükség.

Digitális Múzeumi Könyvtár

Honlap: <http://info.arcanum.hu/muzeumi>

E-mail: info@museum.hu (?)

Fenntartó: Kultúra az Interneten Alapítvány (?)

Értékelés dátuma: 2009. július 4.

Értékelő: Drótos László

Általános értékelés

A honlapján *Digitális Múzeumi Könyvtár (DMK)* feliratot viselő, de a saját sűgőjében *Múzeumi Kiadványok Adatbázisa* néven emlegetett szolgáltatás ismertetője szerint „Magyarország jelenlegi legnagyobb kulturális digitalizálási programja”: 19 megyei múzeum, valamint a *Budapesti Történeli Múzeum* és a *Közlekedési Múzeum* kiadványait teszi hozzáférhetővé az interneten. Közel 1300 kötet (nagyreszt évkönyvek) mintegy 370 ezer oldalának digitalizálására került sor 2008-ban, melyek – automatikus karakterfelismerés után – kétrétegű PDF dokumentumokként kereshetők és tekinthetők meg. Ahogy az ismertető szöveg is hangsúlyozza: a múzeumokban folyó tudományos kutatómunka eredményeit bemutató kiadványok papírfőmában sokszor nehezen hozzáférhetőek és egyébként is kevés publicitást kapnak, így egy

ilyen – ráadásul egyben és teljes szöveggel kereshető – digitális archívum sokat javíthat a helyzeten. Sajnos ez a cél még nem igazán teljesül, mert alig egy-két múzeum honlapjáról van hivatkozás a DMK-ra, a *muzeum.lap.hu* gyűjtőoldalon nem szerepel, és a Google is éppen csak a kereső űrlapját látja, már a trükkös módon megnyíló sűgőt sem tudta leindexelni – így szinte alig van esély, hogy a bennfenteseken kívül valaki rátaláljon. Igaz, a *www.museum.hu* portálon benne van a főmenüben (itt egy harmadikfajta néven: *Digitális Könyvtár*), úgyhogy csak ebből, valamint az adattár kezdőlapján levő *museum.hu*-, *Arcanum*- és *NKA-logó*kából sejthető, hogy kik és milyen támogatásból készítették ezt a szolgáltatást, mert amúgy impresszumot vagy kapcsolattartó címet nem találunk az oldalon.

A kereső és a böngésző eszköz a digitalizálást is végző *Arcanum Adatbázis Kft.* saját fejlesztésű WARCTIS programja. Keresni a teljes szövegben, a kiadványok tartalomjegyzékeiben, illetve a fájlnevekben lehet; logikai műveletek és helyettesítő karakterek egyaránt használhatók. A találati listák szövegkörnyezettel együtt jelennek meg és a keresett szavakat a PDF oldalképeken is kiemelik átszínezéssel. A gyorsabb letöltés érdekében a PDF fájlok fel vannak darabolva, így egy teljes könyv vagy tanulmány egyszerre nem nyomtatható, menthető, vagy másolható, csupán csak oldalanként. A saját ablakkezelési technika miatt hivatkozni sem lehet URL címmel ezekre a digitalizált publikációkra, ami igencsak megnehezíti a szolgáltatás tudományos célú használatát, s ezt tovább súlyosbítja a szabványos bibliográfiai leírások és a géppel olvasható metaadatok hiánya.

A tesztelés során talált három komolyabb hiányosság a következő:

1. A nem megszokott ablakkezelés és az erősen javascriptes felület megnehezíti a felhasználó életét és a keresőgépek számára is láthatatlanná teszi ezt a rendkívül értékes tartalmat. Az akadálymentes hozzáférés is jogosan elvárható lenne azok számára, akik előtt ugyan elvileg megnyílt 370 ezer oldal gépi felolvasatásának lehetősége, de a speciális kezelőfelület ugyanakkor elrejti előlük az egészet.
2. Bár a szolgáltatás nevében szerepel a „könyvtár” szó, de valójában nyoma sincs a könyvtári OPAC-oknál megszokott keresőfunkcióknak: nem lehet a kiadványok címére rákeresni, nincs külön szerző szerinti keresés vagy index (a szerzői nevek a tartalomjegyzékek egyéb adatai között található csak meg), és persze nincsen tárgyszavas vagy bármiféle egyéb temati-

kus keresési/szűkítési lehetőség sem, mivel a kiadványok digitalizálása nem társult könyvtárosi értéknövelő feltáró/leíró munkával. Sem a kiadványokról, sem a bennük levő egyes publikációkról nincsen részletes leírás, csak rövid alapadatok.

3. Hiányoznak a honlapról mindazok az elemek, amelyek nélkül nem lenne szabad közpénzből létrehozott tartalomszolgáltatást működtetni: projektleírás (angolul is), impresszum, kapcsolati címek, frissítési dátum, az adatbázis összetételét mutató statisztika, forgalmi adatok, külső hivatkozások más hasonló webhelyekre stb.

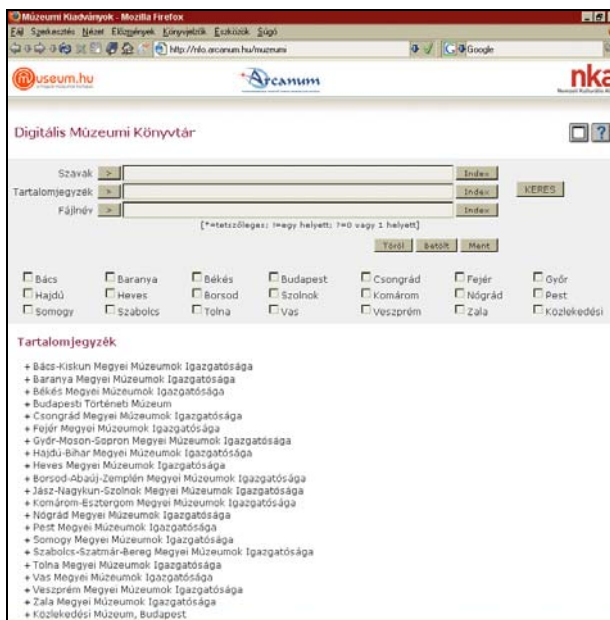
Összefoglalva: A DMK egy jelentős digitalizálási program eredménye, amely korszerű technológiával, hatalmas mennyiségű, szélesebb körben eddig szinte „láthatatlan” tudományos publikációt tesz teljes szöveggel visszakereshetővé és megismerhetővé. Nemcsak múzeumi és egyéb szakemberek számára rendkívül értékes a tartalma, hanem tanárok, diákok, újságírók, és a történelem, helytörténet, irodalomtörténet, néprajz, művészet stb. íránt érdeklődő laikusok is haszonnal lapozgathatják – amennyiben tudomást szereznek róla. Ehhez egyrészt a DMK honlapját kell ellátni a megfelelő háttér-információkkal, továbbá népszerűsíteni minden lehetséges helyen/weboldalon, és legalább a tartalomjegyzékeket indexelhetővé lehetne tenni a keresőgépek számára (pl. a MATARKA adatbázisán keresztül, azzal összekapcsolódva). Másrészt olyan hozzáadott munkát kellene befektetni (pl. MNB leírások átvétele vagy linkelése, tárgyszavazás, személynévmutató), ami valódi könyvtári szintű kereshetőséget biztosítana ennek a jelenleg csak szövegtárként használható tartalomszolgáltatásnak.

Részletes értékelés

Felület, használhatóság, segítség

- Az *nfo.arcanum.hu/muzeumi/* oldalon egy arcuat- és ismertető nélküli kereső- és böngészőfelület fogadja a látogatót (13. ábra). Az oldal három keretből áll: a felsőben a közreműködők emblémái a honlapjukra mutató ugrópontokkal, ez alatt a keresőűrlap, és legalul a kinyitogatható tartalomjegyzék. Görgetősávok csak ennél a harmadiknál vannak engedélyezve, így ha valaki 800 pixelnél keskenyebb ablakban próbálná használni az űrlapot, az a gombok egy részéhez nem fér hozzá. Igaz, a sok segédablak miatt mindenképpen nagy felbontású monitort célszerű használni, de ugyanezen okból szükséges lenne néha a keresőablakot keskenyebbre állítani. Hasonló, csak

függőleges irányú keretméret-probléma van a böngészésre szolgáló alsó frame-nél és a PDF dokumentumokat megjelenítő ablakban is: mindkettőt jó volna magasabbra növelni időnként, de a `frameborder=0` beállítás miatt a felhasználó nem veszi észre, hogy ezeknek a széle megfogható és feljebb tolható.



13. ábra A Digitális Múzeumi Könyvtár nyitólapja és egyben keresőfelülete

- A sűgő utolsó oldala szerint lehetőleg Internet Explorer és legalább Adobe Reader 7 verzió kell a rendszer használatához, valamint a felugró ablakok engedélyezése. Ezeket a tudnivalókat a kezdőlapra is ki kellene írni, mert már maga a PDF formátumú sűgő megnézése sem egyszerű. Nálam az Explorer biztonsági okokból megtagadta a megnyitását mindaddig, amíg a megbízható helyek közé fel nem vettem a szervert, vagy nem engedélyeztem, hogy a PDF dokumentumok a böngészőablakon belül nyíljanak meg. Ugyanakkor viszont gond nélkül tudtam használni a szolgáltatást Firefox vagy Opera böngészővel és 6.0-ás Adobe Readerrel is. Egy hiba viszont az ajánlott szoftverekkel is többször előjött: ha egy könyv végére érve tovább akartam lapozni, akkor egy „A fájl nem ezzel kezdődik: '%PDF-'.” hiba-üzenet jelent meg. Az egyoldalas PDF-ek közötti lapozás elég nehézkes, mert csak egyesével lépegethetünk előre vagy vissza, így egy hosszabb tanulmányon vagy könyvön „végigfutni” sokáig tart, és az különösen frusztráló, hogy nem tudjuk, hány oldalas az a dokumentum, amit éppen nézünk. Ezért, ha böngészéssel választottuk ki va-

lamilyik publikációt, akkor érdemes a tartalomjegyzékben megnézni, hogy az utána következő milyen oldalszámon kezdődik, így meg tudjuk ítélni a terjedelmét. És érdemes a Readert a böngészőbe integrálni, valamint megadni a számunkra kényelmes nagyítás mértékét a megnyitási beállításai között, mert különben az időnként részét az oldalképek méretezésével tölthetjük.

- A rendszer többféle ablakot használ és ezek eltérően viselkednek: például van amikor egymásra nyílnak és van amikor nem; van amikor önállóan bezárható, de van amikor más ablakokat is bezárnak magukkal együtt. Erről alapos és bonyolult leírás van a sűgőben, amit ajánlott át tanulmányozni annak, aki gyakran és komolyan akarja használni ezt a tartalomszolgáltatást. Ha valaki megtanulja a szabályokat, akkor kétségtelenül hatékonyabban tud dolgozni a rendszerrel, de a böngészők átláthatóbb ablakkezeléséhez szokott átlagfelhasználók elvesznek a megnyíló oldalak között, és nem találják meg a mentéshez, nyomtatáshoz, oldalon belüli kereséshez, kimásoláshoz, könyvjelző elhelyezéshez stb. szükséges jól ismert ikonokat és menüpontokat sem. Az eligazodást valamelyest segíti az az ablakkezelő kisablak, amely a sűgőt jelző kérdőjel gomb mellett van (egy fehér négyzet ikon).
- A felhasználót csak ez az `a080226.pdf` nevű sűgő segíti a használatban, egyéb tájékoztatást nem talál és kapcsolati címet sem, ahová segítségért fordulhatna. A sűgő egyébként elég jól megírt és képekkel is illusztrált, de látszik rajta, hogy nem kifejezetten ehhez az alkalmazáshoz készült, mert más adatbázisokból is vannak benne képernyőfotók; az Acrobat saját keresőfunkcióit is ismerteti, pedig ez az online szolgáltatásnál nem érhető el; több helyen is „Könyvjelzők”-ről beszél, miközben a DMK-ban ugyanez „Tartalomjegyzék” néven van; és nem hívja fel a figyelmet az adattár olyan sajátosságaira, hogy például a kiadványok címei nem lettek leindexelve, így ezek csak böngészéssel és némi szerencsével találhatók meg. Az általános használati utasításhoz utólagosan(?) hozzáírt bevezető meglehetősen futtában készülhetett, mert a mindössze négy bekezdés külalakját sem sikerült egységesíteni. A sűgő szövegében több gépelési hiba is van, ezek egy helyesírás-ellenőrral könnyen megtalálhatók (pl. *nagítás*, *kicsinyítés*, *kölnböző*).

Kereső funkciók, találatok

- A kezdőlapon az alsó keret szolgál a kiadványok múzeum, majd azon belül kiadványcím, végül szerző+tanulmány cím szerinti böngészésére. A

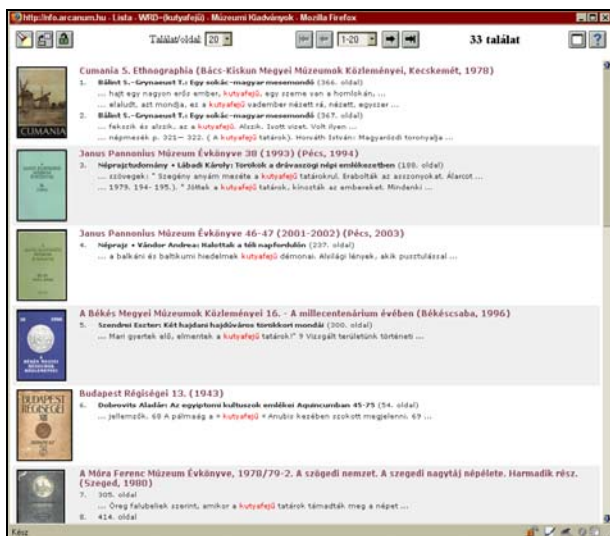
múzeumok neve csak félig-meddig van betűrendben (pl. Borsod megye Heves megye után jön), az egyes múzeumokon belül pedig a kiadványok sorrendje semmilyen felismerhető logikát nem követ (lásd pl. Békés megye alatt), így nem elég, hogy a kiadványcímekhez nincsen kereső, de még a böngészéssel való megtalálhatóságukat is sikerült megnehezíteni. (Egyfajta pótmegoldás lehet, ha ennek az alsó frame-nek a tartalmát vágólapon át kimásoljuk egy szövegszerkesztőbe és abban keresünk.) A hierarchikus tartalomjegyzék egyes ágainak kinyitását és becsukását a nehezen felfedezhető és egérrel nehezen eltalálható + és – jelekkel tehetjük meg; ha ugyanis a kötetcímrre próbálnánk kattintani, akkor azonnal a címlapja jelenik meg.

- A keresőúrlap három mezőből áll: 1. a kiadványok teljes szövegében (beleértve a tartalomjegyzéküket is) előforduló szavak, 2. csak a tartalomjegyzékek szavai, és 3. a fájlnevek kereshetők velük. (Utóbbival a kívülállóknak nem sokra mennek, mert a névkódok feloldása nem nézhető meg.) A keresőszavak kiválasztását indexablakok segítik, a korábban beírt szavak pedig a mezőnév utáni > gombra kattintva hívhatók elő, ami hasznos funkció. Teljes kereső-kifejezések elmentésére és későbbi előhívására is van mód, ezeket cookie-ban tárolja a böngészőnk. Szűkítési lehetőség csak a kiadó/közreadó múzeum szerint van (kipipálható kockák formájában), vagyis nem tudjuk például évszám, földrajzi hely, nyelv, kiadványtípus, témakör vagy egyéb, a könyvtári katalógusoknál megszokott szempontok szerint pontosítani a keresésünket.
- Az úrlap alatt egy minisúgó ismerteti a joker karaktereket (*!?), de már a használható logikai műveletek írásmódja (AND, OR, NOT, nW, nN) vagy a kereső alapértelmezett működése (OW, vagyis több szóból álló kifejezés) csak a súgó 9. és 10. oldalán derül ki – ezeket jobban szem elé kellett volna tenni. Ezek mellett érdekes lehetőség (lenne) a kétkötőjeles intervallumkeresés (pl. 1848--1849), de sok gyakorlati haszna nincs, mivel a tartalomjegyzékben nincsen évszám a cikk szerzője és címe után, a teljes szövegben pedig bárhol előfordulhatnak ezek a számok (akár az oldalszámban vagy az irodalomjegyzékben is), így időszak szerinti szűrésre sem alkalmas ez igazán.
- Az indexelés során minden szót nagybetűsre és ékezet nélkülire (kivéve: Á, É, Ö és Ü) konvertáltak – ennek előnyei (a régi szövegek eltérő írásmódjából adódó találatvesztés csökkentése) és hátrányai (bizonyos szavaknál nagy mennyiségű hamis találat) egyaránt vannak. A szavak listái

indexablakokban nézhetők meg, ezekben külön is lehet keresni, akár csonkolással is, illetve százasaival előre lapozgatni – sajnos a visszafelé lapozás gombot lefelejtették, így ha véletlenül túlhaladunk a megfelelő szón, akkor kezdetjük előlről az egészet. Az indexekben az is látszik, hogy melyik szó hányszor fordul elő az adattárban, de ez becsapós, mert a többoldalas tanulmányok címszavai minden oldalhoz hozzá vannak rendelve; így például az, hogy a tartalomjegyzék indexében 78-szor fordul elő a LEVESEK szó, nem azt jelenti, hogy 78 db. ilyen című cikk van, hanem csak egyetlen 78 oldalas tanulmány okozta ezt a magas, a gasztronómia iránt érdeklődőknek első látásra igen biztató előfordulási számot.

- A találati listák gyorsan jönnek és látványosak (14. ábra). Nemcsak a cikk szerzője, címe és oldalszáma, valamint a keresett szavak szövegekörnyezete jelenik meg, hanem az azt tartalmazó kiadvány alapadatai és címlapképe is. A felső navigációs keretben a találatok teljes száma, az oldalankénti találatszám, a lapozónyilak és -menü mellől viszont nagyon hiányzik egy-egy olyan ikon, amivel a találati listát nyomtatni vagy menteni lehetne, mert a legtöbb felhasználónak nem kézenfekvő alternatíva a jobb gombos menük vagy a billentyűparancsok használata. Mint ahogy valószínűleg az sem világos neki azonnal, hogy a listában mire is kell kattintania, mert sem a címlapképek, sem a sötétpiros kiadványcímek nem viszik a kívánt oldalra, hanem vagy az apróbb betűs cikkcímre, vagy ha az nincs, akkor a találati sorszámra vagy az oldalszámra kell kattintani – utóbbiak viszont nincsenek vizuálisan kiemelve. A lebegősúgókkal is lehetne segíteni neki, hogy melyik hivatkozás hova is vezet, de jelenleg mindegyiknél ugyanaz a „PDF fájl megnyitása” felirat jelenik meg.
- Fontos – és erre a súgó is utal –, hogy a keresési egység a könyvoldal, vagyis az számít találatnak, ha a beírt keresőfeltétel egy adott oldalnál teljesül. Tehát nem biztos, hogy a keresett szavak ugyanabban a cikkben vannak (ha egy oldalon két vagy több cikk is található); és az is megtörténhet, hogy bár a keresett kifejezés ugyanabban a mondatban van, mégsem találja meg a program, mivel a mondaton belül oldaltörés volt, így ezt két külön rekordként kezeli a kereső. Az oldalon megtalált szavakat kiemeli a rendszer: a találati listákban élénkpirossal, a PDF fájlokban pedig a Windows színsémájának „Kijelölt elemek” nevű kódjával, ami sajnos általában egy sötét szín, így érdemes ideiglenesen például világossárgára állítani a Vezérlőpult/Képernyő

funkciójával, mert különben nem fogjuk látni a keresett szavakat a fekete betűs képoldalakon.



14. ábra Egy találati lista eleje, címlapképekkel és szövegkörnyezettel

Tartalom, minőség

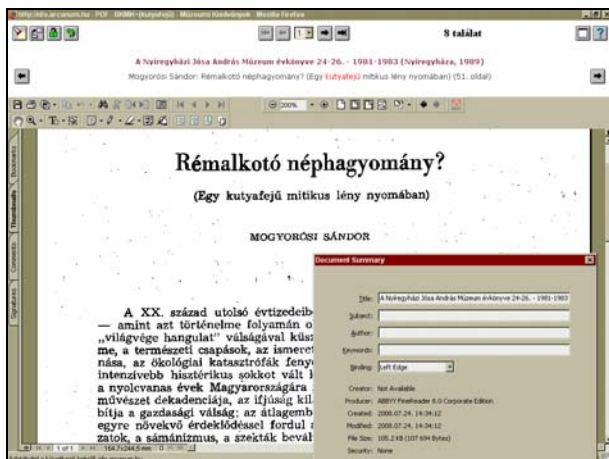
- A DMK sűgője szerint közel 1300 kiadványt tartalmaz az adatbázis. Nekem a tartalomjegyzék kimásolásával „csak” 1234 tételt sikerült összeszámolnom, de mindenképpen hatalmas és komoly tudományos értékkel bíró szöveghalmazról van szó. A digitalizálás – remélhetőleg csak első fázisa – tavaly zajlott, így még mindössze négy 2008-as kiadványt lehet találni közöttük. A legrégibbi pedig egy 1889-ben megjelent „Budapest Régiségei 1.” c. kötet.
- A digitális képek minőségéről is tájékoztat a sűgő: a szkennelés általában 300 DPI és 256 szürkeárnyalatú, apró betűs kiadványoknál 400 DPI, színes képeknél 300 DPI és true color beállításokkal történik. A szolgáltatási verziók 300 vagy 400 DPI-s fekete-fehér szövegoldalok, ahol szükséges, ott szürkeárnyalatú vagy színes képekkel. Ezek a paraméterek az OCR program számára akár 98-99%-os felismerési pontosságot, a felhasználó számára pedig gyorsan letölthető és még élvezhető oldalképeket eredményeznek. A kétrétegű PDF-ben az olvasó az eredeti kiadvány fakszimiléjét látja, amely mögött van a kereshető, kimásolható felismertett szöveg (egyres Readerekben a Ctrl/4 gombbal ezt is meg lehet nézni). A DMK esetében valóban egészen jó minőségű ez az OCR-es szöveg, még az

idegen nyelvű – sőt akár a cirill betűs – cikkek-nél is. Természetesen a nagyon régi vagy összefirkált/szakadozott kiadásoknál, az apró betűs részeken és az oldaltöréseknél elválasztott szavaknál nem ilyen jó a felismerési arány, de ezt a világ nagy digitalizáló műhelyeiben sem tudják jobban megoldani.

- A sűgőben olvasható az is, hogy a tartalomjegyzékeket minden esetben korrektúrázták, így ezek helyessége 99.9% fölötti. Ugyanakkor a tartalomjegyzék indexében könnyen lehet olyan hibákat találni, amelyeket egyszerűen ki lehetett volna szűrni és amelyek akadályozzák a keresést: például a francia és az angol szavakban maradt nyomdai aposztrófok (ezeket a billentyűzetről nem lehet bevinni, így csak az indexből választhatók ki), vagy a szintén tévesen benne maradt záró idézőjelek (ezért pl. az *odakinn* szóra nincs találat a tartalomjegyzék keresőjében, mert az hibásan, idézőjellel együtt: *odakinn* formában lett leindexelve). De érdemes lett volna a tartalomjegyzékekben olyan mintakereséseket is lefuttatni a korrektúrázás után, mint a betűt követő szám, így nem maradtak volna az indexben például ilyen szavak: *SZÁZADI9*, *SZEGEDEN1918-1920*. Ezeknél a hibás szavaknál egy fokkal kellemetlenebb az, amikor nincsen kattintható hivatkozás a böngészhető tartalomjegyzékben, például a „Nyíregyházi Jósza András Múzeum évkönyve 1. – 1958” kötetben a „Koroknay Gyula: Églises médiavales à clôtüre droite du sanctuaire du Comitát Szabolcs-Szatmár” c. írásánál, feltehetően a hibásan a cím végén maradt „[127” oldalszám miatt nem sikerült automatikusan rátenni a megfelelő PDF oldalra mutató hiperhivatkozást.

Metaadatok, bibliográfiai információk

- Metaadatok és bibliográfiai leírások tekintetében rosszul áll a Digitális Múzeumi Könyvtár, ami egy könyvtár esetén elég furcsa. A honlap fejlécében csak egy `<TITLE>Múzeumi Kiadványok</TITLE>` sor van, ami már a negyedik névvariációja ennek a szolgáltatásnak. Az egyes PDF fájlokban is csak a *Title* sor van kitöltve, mely a publikáció eredeti forrását tartalmazza (15. ábra) És nincs nyoma annak sem, hogy átadnának valamilyen géppel olvasható metaadatot az országos vagy az európai keresőknek, így ez az értékes digitális gyűjtemény kívülről láthatatlan marad.




15. ábra A PDF oldalak megjelenítő ablaka navigációs gombokkal és a fájl metaadataival

- A szemmel látható bibliográfiai adatok sem elég részletesek, többnyire csak az alapinformációkra szorítkoznak, például: „Bányáskönyv (Salgótarján, 2000)”, ami a kiadvány azonosításához elegendő lehet, de irodalomjegyzékben való hivatkozáshoz nem biztos, hogy elégséges. Legalább annyit meg kellene oldani, hogy az egyes kiadványok mellé tett gombbal át lehessen menni valamelyik OPAC-ba vagy az MNB webes felületére, ahol megtalálható a könyv teljes és szabványos leírása. Az pedig külön gond, hogy az oldalakra felszabdalt cikkrészletek lementés vagy ki nyomtatás után azonosíthatatlanná válnak, mert az adataik a PDF fájlokon kívül, a felső frame-ben voltak, ami nem őrződik meg. Ezen a prob-

lémán a képek „vízjelezésével” lehetne segíteni, vagyis minden oldalképre rátenni a kiadvány és a cikk adatait is, ha az eredeti nyomtatott kiadás oldalfejlécében ezek az információk nem szerepelnek.

Beérkezett: 2009. VII. 8-án.



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



Kőrös Kata
 a Petőfi Irodalmi Múzeum
 könyvtárosa.
 E-mail: korosk@pim.hu



Ternai Zita
 könyvtári informatikus.
 E-mail: zita-6@freemail.hu

Vook: e-könyvek videókkal

Az e-könyvek gyakran alig különböznek a nyomtatott társaiktól. Ezen a megoldáson akart változtatni valami újdonsággal a Simon & Schuster. A kiadó elhatározta, hogy ezeket az alkotásokat igazán multimédiássá teszi azáltal, hogy kisfilmeket mellékel hozzájuk. A vookok nem csak egyszerűen digitalizált és így elérhetővé tett nyomtatott alkotások, hanem videókkal kiegészített igazi multimédiás művek a számítógép-, illetve az iPhone-tulajdonosok számára.

A vállalat az új fajta termékeket vookoknak nevezte el, a mozaikszó a video és a book szavak ötvözéséből született meg. Jelenleg négy ilyen kiadvány kapható, egy szépségtanácsadással foglalkozó és egy fitnessmű, valamint két regény. Két verzióban férhetők hozzá. Az első változatok a weben, a böngésző segítségével tekinthetők meg, illetve olvashatók, míg a másik lehetőség, hogy a felhasználó a mobil készülékén (iPhone, iPod) tekinti meg őket. Előbbi kiadások 7, utóbbiak 5 dollárba kerülnek.

/Sg.hu Hírlevél, 2009. október 6., <http://www.sg.hu/>

(SzP)



A Google keresőjén túl

Az „Oxford English Dictionary” szócikke szerint a „google” szót először egyfajta speciálisan csavart ütésre használták a krikett és a tollaslabda sportban. De az 1998 szeptemberében indult Google kereső nem innen, hanem egy matematikai szlengszóából kapta a nevét: a „googol” a 10-nek a századik hatványát jelenti. Ezzel a gigászi névvel az új keresőgép azt sugallta, hogy képes lesz megbirkózni a web elképzelhetetlen méreteivel. A gyors sikert az is jelzi, hogy egy év múlva már igeeként is használták az új kereső nevét a *Usenet* hírcsoportokban, és az akkori vezető keresőgépek: a Lycos, az AltaVista és a Yahoo! gyorsan háttérbe szorultak. A 21. század első évtizedét egyértelműen a Google uralja és már nemcsak a webkeresőjével, hanem az „istállójába” tartozó egyéb szolgáltatásokkal is, mint például a Google Images, a Google Books, a Google Scholar, a Google Maps stb.

Az amerikai *Rutgers University* tájékoztató könyvtárosa cikkében sorra veszi azokat az alternatívákat, amelyeket a Google helyett használhatunk, ha másfajta módon szeretnénk keresni, vagy valamilyen speciális témában kutakodunk az interneten, vagy amikor weblapok helyett más típusú dokumentumokat szeretnénk megtalálni.

Eltérő megoldások

A Google általános és szabadszavas keresőjének egyik alternatíváját az emberi közreműködéssel készülő válogatások jelentik, ahol is szakértők által összegyűjtött, minőségi információforrásokban tudunk keresni, melyek relevánsabb találatokat adnak. Ezek egyik jó példája a brit *Intute* (www.intute.ac.uk) adatbázisa, ami már több mint 100 ezer, oktatási, illetve tudományos jellegű forrást tartalmaz – jellemzően kormányhivatalok, kutatóintézetek és szervezetek honlapjait. A keresést leszűkíthetjük nagyobb témákra: művészetek és humán területek, egészségügy és élettudományok, mérnöktudományok és technika, valamint

társadalomtudományok. Egy másik érdekes oldal az önmagát „tematikus indítóállvány”-ként definiáló *Pinakes* (www.hw.ac.uk/libwww/irn/pinakes/pinakes.html). Az *Alexandriai Könyvtár* katalógusáról elnevezett oldal nem tartalmaz saját keresőt, hanem csak ugrópontokat a legfontosabb tematikus webkatalógusokhoz, olyan változatos témákban, mint például a jog, az építészet, az űrhajózás, a történelem, a zene, a várostervezés.

Egy másfajta megközelítést jelent a természetes nyelvű keresés. Az *Ask.com* (www.ask.com) – korábbi nevén: *Ask Jeeves* – oldalán a szabad szavas keresés mellett kérdő mondatokat is beírhatunk, és azután a rendszer további opciókat kínál a találatok pontosításához. Kevésbé megbízható válaszokat ad az *Answers.com* (www.answers.com), ahol szintén lehet angol mondatokkal keresni, de a szótárak és lexikonok szócikkei mellett az internetező közönség által összeállított, kétes minőségű WikiAnswers oldalai is megjelennek a találatok közt.

Egyes keresők a vizuális megközelítést választották: A *Quintura* (www.quintura.com) például címkefelhőt készít a beírt keresőszavakhoz, így további releváns szavakra kattintgatva tudjuk finomítani a keresésünket. Ennél is látványosabb megoldást használ a *KartOO* (www.kartoo.com), amely Flash formátumú tématerképet jelenít meg találati lista helyett, és az eredmények közötti kapcsolatot is mutatja. Bár a cikk szerzője nem igazán látja a hasznát és értelmét, de a vizuális beállítottságú embereknek ez lehet, hogy áttekinthetőbb.

Egy további eredményprezentálási technika a klaszterezés. Erre a *Clusty* (clusty.com) a legjobb példa. Ha ezt a keresőt használjuk, akkor a találatok altémakörökbe csoportosítva jelennek meg. Például a globális felmelegedésre keresve egy oldalsávban ilyen témacsoportok tűnnek fel: „ég-hajlatváltozás”, „hogyan állítsuk meg a felmelegedést”, „széndioxid-kibocsátás” stb., és ezekre kattintva finomítható a találati lista. A *Grokker*

(www.grokker.com) szintén ezt a kézre álló megoldást használja, de ez csak a Yahoo! és a Wikipédia keresőjére épül, míg a Clusty több keresőgépet is lekérdez.

Végül ott vannak az általános keresők számára láthatatlan „mélyweb” elérését biztosító eszközök. A *Complete Planet* (aip.completeplanet.com) saját állítása szerint 70 ezer adatbázist és speciális keresőt kérdez le. Az *OAIster*-hez (www.oaister.org) hasonlóan mélyen belenyúl a nagy tudományos és kormányzati adattárakba és webhelyekbe. Az *OAIster* közel ezer szolgáltatótól gyűjti be a metaadatokat és így több mint 16 millió, másként nehezen hozzáférhető adatrekord között kereshetünk vele.

Speciális témák

A legjobb válaszokat általában az adott témakörre specializált források nyújtják. Gazdasági témákhoz például ilyen a *BizSeer* (bizseer.ist.psu.edu), amivel oktatási, tudományos, céges és kormányzati webhelyeken tudunk cikkeket, műhelytanulmányokat, jelentéseket és más hasonlókat keresni. Ráadásul a *Penn State University* által fejlesztett *BizSeer* idézetségi keresésre is alkalmas. Ide kapcsolhatók még az álláskereső oldalak is: ezek közül a két legnagyobb a *Monster* (www.monster.com) és a *Career Builder* (www.careerbuilder.com). A tudományos és műszaki témák fontos forrása az *Elsevier* által fenntartott *Scirus* nevű kereső (www.scirus.com). Ez már több mint 450 millió folyóirat-cikket, preprintet, szabadalmat, tudományos weboldalt és más hasonlókat indexel le. Az eredmények három kategóriába osztva jelennek meg: előfizetéses folyóiratok, kiemelten fontos webhelyek, és a web többi része. A *ChemSpider* (www.chemspider.com) egy ennél is specializáltabb kereső, amely molekulaszervezetek adatait és kísérő információit gyűjti egy hatalmas adatbázisba. Orvosi témákban kitűnő keresőgép a *Healia* (www.healia.com): a weben, a Pub-Med adatbázisban, valamint a klinikai vizsgálatok között kereshetünk vele. A nagyközönség számára ennél is érdekesebb a nagyon népszerű *WebMD* (www.webmd.com), amivel mindenféle egészségügyi kérdésre kaphatunk választ.

A kormányzati weboldalak leghatékonyabban az erre a célra létrehozott szolgáltatásokkal kereshetők. Az Egyesült Államokban ilyen a szövetségi és a tagállami hivatalos honlapokat és dokumentumokat egyaránt leindexelő *USA.gov* (www.usa.gov). Nagy Britanniában a *Direct.Gov* (www.direct.gov.uk) szolgál hasonló célra. A kormányzatok egyik erőssége a statisztikagyártás, de a megfelelő

adatsorokat nem mindig könnyű megtalálni. Ebben segít a *Census Bureau* által fejlesztett *DataFerret* (dataferrett.census.gov). Itt egy Java alkalmazás segítségével kereshetünk az USA kormányzata által fenntartott demográfiai, gazdasági, környezeti, egészségügyi és más jellegű adattárakban.

Különleges típusok

A YouTube a legnagyobb és legnépszerűbb videogyűjtőhely a weben. A *Truveo* (www.truveo.com) keresővel nemcsak a YouTube, hanem a Google Images, az AOL és a MySpace videói is megtalálhatók. Az Internet Archive *Moving Image Archive* nevű gyűjteménye (www.archive.org/details/movies) nem annyira a „csinálj saját műsrot” mozgalmat támogatja, hanem inkább az ismeretterjesztő és a kulturális filmekre helyezi a hangsúlyt. Ugyanez a helyzet a hangfelvételek esetében is: míg a *Podcast Alley* (www.podcastalley.com) elsősorban az önkifejezési célból létrehozott podcast-ok között keres, addig a *LearnOutLoud* (www.learnoutloud.com) inkább az információs és az oktatási jellegű felvételekre összpontosít, mint például a beszédközpontú rádiók és a hangoskönyvek. A blogok is az önkifejezés egyik új formáját jelentik, és természetesen ezeknek is megvannak a speciális keresők; közülük a legjobb a *Technorati* (technorati.com).

Képek esetében a Google és a Yahoo! képkeresői mellett ott van a *Ditto* (www.ditto.com), amely magát a legnagyobb vizuális tartalomkeresőnek definiálja. Hangokat és effekteket a *FindSounds* (www.findsounds.com) és a *Soundsnap* (www.soundsnap.com) segítségével találhatunk legkönnyebben: állathangok, repülő nyíl surrogása, nevetés, almaharapás hangja, szerszámok és járművek zaja, és még milliónyi egyéb hang között kereshetünk itt. És végül, ha világtérképre van szükségünk, akkor az egyik jó hely erre a *National Geographic* honlapján a *Map Machine* (maps.nationalgeographic.com/map-machine).

Bár most a Google dominálja az internetes keresést, de mindig jönnek új versenytársak újfajta megoldásokkal, ami kikényszeríti a folyamatos fejlesztést. Ez az állandó versengés különösen hasznos azoknak, akik az információkeresésből élnek.

/MAXYMUK, John: Searching beyond Google. = The Bottom Line: Managing Library Finances, 21. köt. 3. sz. 2008. p. 97–100./

(Drótos László)

Van-e optimális kategóriaszám a könyvtári honlapok forrásgyűjteményeinél?

A felsőoktatási könyvtáraknak komoly kihívást jelent, hogy hogyan vonzzák a honlapjukra a felhasználókat. Egy 2007-es amerikai felmérés [1] kimutatta, hogy a hallgatók közel háromnegyede a könyvtár helyett már inkább az internetet használja információkeresésre; nem mindegy tehát, hogy a könyvtári weboldalak mennyire felhasználóbarát felületet és tartalmat kínálnak. Ezeknek a honlapoknak az egyik tipikus szolgáltatása az internetes források tematikusan rendezett címgyűjteménye. A könyvtárosok által összeállított forráslisták eléggé hasonló témakörneveket használnak, ugyanakkor a témák száma jelentősen változik: néhol csupán csak öt kategória van, máshol viszont akár száznál is több. A cikkben ismertetett kísérlet során a szerzők arra próbálták választ kapni, hogy a tematikus kategóriák száma mennyiben befolyásolja a keresés hatékonyságát?

Szakirodalmi áttekintés

McGillis és Toms 2001-ben publikált kutatásuk [2] során azt találták, hogy a felhasználóknak gondot okozott megkeresni egy adatbázist a könyvtári honlapon való navigálással: nehezen tudták eldönteni, hogy a honlapon felsorolt kategóriákból melyiket válasszák, hogy eljussanak hozzá. Jeng 2006-os disszertációjában [3] egy sor tesztet elemzett a digitális könyvtárak használhatóságával kapcsolatban. Többek között az alábbi megállapításokat tette:

- A felhasználók „elveszetteknek” érzik magukat: nem látják át a digitális könyvtár struktúráját.
- Félreérthető terminológia: a szakszavak és a fogalmak helytelen megválasztása volt a felelős a használhatósági problémák 36 százalékáért.
- A folyóiratcikkek és a tematikus adatbázisok megtalálása komoly nehézséget okozott a felhasználóknak.

Más, nem könyvtári témájú publikációk is érdekesek a jelen kutatás szempontjából, mindenekellett Miller híres tanulmánya [4] a rövidtávú emberi memória 7 ± 2 tételes kapacitásáról. Ezt a felfedezést néha tévesen úgy interpretálják, hogy nem ajánlott 9 elemnél hosszabb menüket használni a weblapokon. Valójában a későbbi kutatások kimutatták, hogy a „tömbösítés” (*chunking*) módszerét használva – vagyis amikor az elemeket csoportokba foglaljuk úgy, hogy az ugyanazon csoporton

belüli elemek szorosabb kapcsolatban legyenek egymással, mint más csoportok elemeivel – a rövidtávú memóriával is sokkal nagyobb halmazt lehet kezelni annál, mint amit Miller kimutatott.

A menük hosszára és mélységére vonatkozó vizsgálatok némileg ellentmondásos eredményeket adtak. Larson és Czerwinski [5] például arra a következtetésre jutottak, hogy akár az egyes menüsintek kiterjedését, akár a szintek számát növeljük, a felhasználó reakcióideje, hibázási gyakorisága és komplexitásérzése egyaránt növekedni fog, ezért közepesen hosszú (8-9 elemű) és lehetőleg nem túl mély menük használatát ajánlották. Ugyanakkor Snowberry, Parkinson és Sisson [6] azt találták, hogy a tesztelt 2^6 , 4^3 , 8^2 és 64^1 struktúrák közül (ahol pl. a 2^6 egy kételemű és hatszintű menürendszert jelent) a kategóriákba csoportosított 64^1 bizonyult a legelőnyösebbnek mind a keresési sebesség, mind pedig a pontosság szempontjából. Zaphiris is megerősítette azt, hogy a weben minél mélyebb egy menürendszer, annál lassabban találja meg a felhasználó, amit keres. [7] Azt is megmérte, hogy a kinyitható menüknél a felhasználók válaszideje 50 százalékkal hosszabb, mint az egymást követő menük esetében. [8]

A következőkben ismertetett vizsgálat csak az egyszintes menükre terjedt ki, és azt próbálta meghatározni, hogy az ezekben felsorolt témakörök száma és jellemzői hogyan befolyásolják a keresés hatékonyságát.

A tesztelés ismertetése

A kutatáshoz 30 angolszász könyvtári weblapról választottak ki tematikusan rendezett forráslistákat, majd ezek számát 12-re szűkítették, melyek jól reprezentálták a többit is, egymáshoz eléggé hasonló témakör-megnevezéseket tartalmaztak, és méretük a kicsitől (5 téma) az egész nagyig (72 téma) terjedt. A tesztalanyokból véletlenszerűen ugyancsak egy tucatnyi csoportot alakítottak ki és minden csoport egy-egy ilyen tematikus listát kapott – kinyomtatva, ábécérendben és egységes külalakkal. A résztvevők 90%-a diák volt, a többiek tanszéki alkalmazottak. Mindegyiküknek ugyanazt a 11 kérdést tették fel, és az illetőnek ki kellett választania a kapott menülistából azt az egy menüpontot, amely megítélése szerint legvalószínűbb

ben elvezetné a válaszhoz. A kérdések az aztékokra, a gördeszkázásra, az ismétlődő mikrotraumára, a francia forradalomra, a farkasok visszatelepítésére, egy üzletlánc marketingstratégiájára, az atomenergiára, Emmanuel Kantra stb. vonatkoztak – vagyis a legkülönbözőbb témákra, amelyek egy könyvtári kutatás során előfordulhatnak. A tesztelést ketten felügyelték: egyikük feltette a kérdést és feljegyezte a megjelölt témakört, a másikuk pedig mérte azt az időt, amire a tesztalanynak szüksége volt a témakör kiválasztásához.

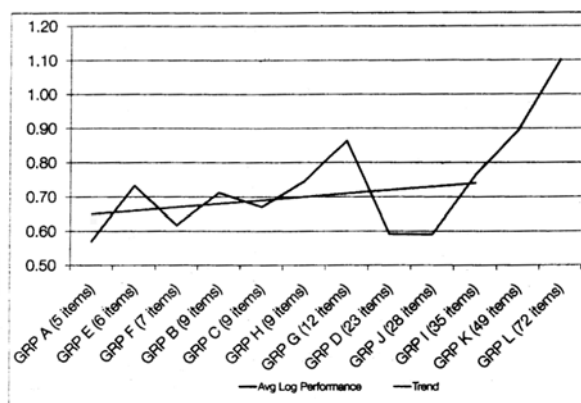
Előzőleg minden lehetséges kérdés-válasz kombinációhoz két könyvtáros relevanciaértékeket rendelt, vagyis eldöntötték, hogy egy adott tematikus listában egy-egy témakör mennyire releváns az egyes kérdések esetében. A vizsgálat során azt szerették volna megtudni, hogy a hasonlóan magas relevanciájú válaszokat összehasonlítva van-e különbség a válaszidőben a különböző tematikus listák között?

A vizsgálat eredményei

A mérési adatokat egyváltozós általános lineáris modellanalízis segítségével értékelték ki, külön-külön mindegyik kérdésre. A modellben a függő változó a válaszidő volt, a magyarázó változó (kovariáns) pedig a válasz relevanciája, vagyis hogy mennyire illeszkedett egymáshoz a kérdés és a válaszként megjelölt téma. (A „Gazdaság” témakör pl. igen magas relevanciaértéket kapott az aktuális munkanélküliségi rátára vonatkozó kérdéssel kapcsolatban, ugyanakkor egyáltalán nem volt releváns a Szaturnusz négy holdjának nevét firtató kérdés esetében.) Az analízis főfaktora az a témalista volt, amit a tesztalany használt. A statisztikai modell így foglalható össze: válaszidő = lista + relevancia + (lista x relevancia) + hiba.

Az általános lineáris modell használhatóságához két feltételnek kell teljesülnie: egyrészt az adatokat véletlenszerűen kell kiválasztani egy normális sokaságból, másrészt a csoportok varianciájának azonosnak kell lennie (homoszkedaszticitás). A tesztelés során mindkét feltétellel voltak problémák, ezeket a szokásos logaritmikus transzformációval mérsékeltek. A 11 kérdésből így már csak három nem felelt meg a követelményeknek, ezért az ezekre adott válaszokat a további elemzésből kizárták. A maradék 8 kérdés adatait feldolgozva és grafikonon ábrázolva – ahol a vízszintes tengelyen az egyes, változó elemszámú listák vannak, a függőleges tengelyen pedig a válaszidők logarit-

musa –, nem látható semmilyen trend, ami azt mutatná, hogy a listák elemszámának növekedésével a válaszidő is változik. Úgy tűnik, hogy inkább a kérdés és maguk a kategóriák befolyásolják a válaszidőt, és nem annyira a kategóriák száma. Egy 28 témakörös lista hasonló válaszidőt eredményezett ugyanazon a relevanciaszinten, mint egy 6 témakörből álló. Ha statisztikailag hasonlítjuk össze az egyes tesztcsoportok válaszidejének középértékét ugyanazon kérdéseknél, akkor is ezt kapjuk: nincs semmi jelentős eltérés közöttük. Ha viszont a 8 kérdés válaszidejét átlagoljuk, és ezt ábrázoljuk az egyes listák elemszámának függvényében, akkor némileg tisztább képet kapunk (1. ábra). A trendvonalat is megrajzolva azt látjuk, hogy a válaszidő fokozatosan növekszik az elemszám növekedésével nagyjából az 50 tételes listáig, majd ezután hirtelen megugrik.



1. ábra Az egyes tesztcsoportok 8 kérdésre adott átlagos válaszideje és a trendvonal (zárójelben a csoport által tesztelt témalista elemszáma)

A válaszidők és a relevancia között viszont semmilyen kapcsolatot nem sikerült kimutatni: a gyors válasz nem feltétlenül jelentett releváns választást, és ugyanígy: a lassú válaszidő sem hozott irrelevánsabb eredményeket. Kétféle viselkedésmintát lehetett megfigyelni. Voltak, akik gyorsan és könnyen döntöttek, miután átfutották a választható témaköröket, de ez nem mindig volt releváns választás. Más esetekben viszont a tesztalanyok elbizonytalanodtak, ha elsőre nem sikerült dönteniük, ilyenkor próbálták átértelmezni a kérdést, hogy egy megfelelő kategóriát tudjanak hozzá találni. Bár a szemmozgásukat nem regisztrálták műszerrel, de úgy tűnt, hogy elkezdik újra átfutni oda-vissza a listát, hogy találjanak egy illeszkedő kategóriát. Ez a fajta késlekedés két esetben fordult elő: amikor egyik témakör sem tűnt megfelelőnek, illetve amikor a listák kezdtek hosszúvá válni.

Feltételezhető, hogy ha a lista elemszáma túl nagyra nő, akkor ez az átfutás egyre jobban elassítja a választási folyamatot.

A kategórianévek érzékelhetően nem befolyásolták a tesztalanyok teljesítményét. Három olyan témakörlista is volt, amelyek mindegyike 9 elemből állott, és ezek különböző elnevezéseket használtak ugyanazon témakörökhöz, de a válaszdők nagyon hasonlóak voltak mindhárom esetében. Ám elképzelhető, hogy egy későbbi vizsgálat, amely több, azonos elemszámú, de különböző témanéveket használó listát hasonlít össze, kimutathat majd eltéréseket a reakcióidőkben. Ugyancsak másfajta eredményeket hozhatnának más, esetleg szerencsésebben megválasztott tesztkérdések, mert bár az ismertetett vizsgálat során a teszt tervezői igyekeztek minél változatosabb témájú és eltérő nehézségi fokú kérdéseket feltenni, de végül a tizenegyből hármat el kellett vetniük, mert nem feleltek meg a statisztikai modell követelményeinek.

Irodalom

- [1] JONES, Steve: The Internet Goes to College. Ed. Mary Madden. Washington, D.C.: Pew Internet and American Life Project, 2002. p. 3.
- [2] MCGILLIS, Louise – TOMS, Elaine G.: Usability of the Academic Library Web Site: Implications for Design. = *College & Research Libraries*, 62. köt. 4. sz. 2001. p. 361.

- [3] JENG, Judy H.: Usability of the Digital Library: An Evaluation Model. PhD dissz., Rutgers University, New Brunswick, New Jersey, 2006. január, p. 38–42.
- [4] MILLER, George A.: The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information. = *Psychological Review*, 63. köt. 2. sz. 1956. p. 81–97.
- [5] LARSON, Kevin – CZERWINSKI, Mary: Web Page Design: Implications of Memory, Structure and Scent for Information Retrieval. Los Angeles, ACM/Addison-Wesley, 1998. p. 25
- [6] SNOWBERRY, Kathleen – PARKINSON, Mary – SISSON, Norwood: Computer Display Menus. = *Ergonomics*, 26. köt. 7. sz. 1983. p. 705.
- [7] ZAPHIRIS, Panayiotis G.: Depth vs. Breath in the Arrangement of Web Links.
<http://www.soi.city.ac.uk/~zaphiri/Papers/hfes.pdf> (letöltve: 2007. nov. 1.).
- [8] ZAPHIRIS, Panayiotis G. – SHNEIDERMAN, Ben – NORMAN, Kent L.: Expandable Indexes Versus Sequential Menus for Searching Hierarchies on the World Wide Web
<http://agrino.org/pzaphiri/Papers/expandableindexes.pdf> (letöltve: 2007. nov. 1.).

/MILES, Mathew J. – BERGSTROM, Scott J.: Classification of library resources by subject on the library website: Is there an optimal number of subject labels? = *Information Technology and Libraries*, 28. köt. 1. sz. 2009. p. 16–20./

(Drótos László)

Digitális gyűjtemények és adatok felügyelete: „kurátorság” és más alternatívák a tudományos könyvtárakban

A digitális anyagok sérülékenyek. Bár ez a nyomtatott dokumentumokra is elmondható, de azok helyes kezelésének módszereit kidolgoztuk az elmúlt évszázadok alatt, míg a digitális információ megőrzéséről sok mindent meg kell még tanulnunk. A „digitális kurátorság” egyre ismertebb fogalom: azoknak a tevékenységeknek a gyűjtőneve, amelyek a digitális adatok és tudományos eredmények karbantartására és használhatóságának megoldására irányulnak, azok teljes életciklusa alatt, a jelenben és a jövőben egyaránt. Magában foglalja a digitális archiválás és megőrzés folyamatát, de az adatok létrehozásának és menedzselésének helyes gyakorlatára is kiterjed, valamint azokra az értéknövelő eljárásokra is, melyekkel új információ és új tudás állítható elő. A hollandiai *Digitális Megőrzés Koalíció* koordinátora cikkében

azt vizsgálja, hogy a tudományos könyvtárak milyen típusú digitális adatokat kezelnek, és hogy ezeket hogyan igyekeznek megőrizni:

- átadják-e ezt a terhet az erre szakosodott intézményeknek;
- vagy maguk vállalják fel a digitális megőrzés felelősségét, de a tényleges munkát kiszervezik és mással végeztetik el;
- vagy pedig teljesen magukra vállalják a feladatot és saját digitális repozitóriumot építenek ki az intézményen belül?

Elektronikus publikációk: a raktározástól a hozzáférésig

A digitális kor egy sor új kihívást hozott a szakkönyvtárakba, ezek egyike az áttérés a nyomtatott

könyvekről és folyóiratokról az elektronikus változatokra. Ez a váltás másfajta beszerzési gyakorlatot eredményez, például licenctárgyalásokat – jellemzően nagybani – előfizetésekről. És változik a hozzáférés módja is, mert többnyire az interneten át, online érhető el a tartalom. Ezáltal szinte észrevétlenül eltűnt a könyvtárak egyik alapfeladata: az információ raktározása. A könyvtárak számára meglehetősen nyugtalanító érzés, hogy nem birtokolják többé a tudományos publikációkat, hanem meg kell bízniuk a kiadókban, illetve a közvetítőkben, hogy betartják a szerződés „folyamatos hozzáférés”-re vonatkozó pontját. A kiadók pedig egyre inkább a szaporodó digitális őrzőhelyekkel szerződnek, mint amilyen például a Holland Nemzeti Könyvtár e-Depot szolgáltatása, vagy a Portico és a LOCKSS; gyakran akár többel is, hogy az eltérő megőrzési technikákkal minél jobban csökkentsék az adatvesztés kockázatát. Más esetekben törvényi szabályozás alapján történik a digitális dokumentumok biztonságba helyezése – többnyire a nemzeti könyvtárakban.

Leginkább jól szerkesztett, PDF formátumban tárolt publikációkról van szó, melyek archiválása és a jövőbeli hozzáférés biztosítása nem különösebben nehéz műszaki feladat. Emellett az ilyen elektronikus kiadványok raktározása nem is elsősorban a tudományos könyvtárak felelőssége, hanem a kiadóké és a digitális megőrzéssel foglalkozó nemzeti vagy nemzetközi szervezeteké.

Helyi digitalizált gyűjtemények

A megőrzendő anyagok másik csoportját a digitalizált gyűjteményrészek jelentik. Mivel a könyvtárak elkezdtek digitalizálni állományuk egyes darabjait, hogy így könnyebben hozzáférhetővé tegyék őket az olvasóknak, logikusnak látszik, hogy egyúttal azt a felelősséget is felvállalták, amit ezeknek a – sokszor jelentős költséggel létrehozott – számítógépes állományoknak a hosszú távú megőrzése jelent, hogy ezek a jövő generációi számára is használhatók maradjanak. Ám valójában ez nincs mindig így. A cikk szerzője számos olyan előadást hallott az éves LIBER-konferenciákon, amikor az előadó ezzel zárta a mondandóját: „Digitalizáltuk a gyűjteményt és ezáltal megőriztük a jövőnek. Köszönöm a figyelmet!” Az élet azonban nem ilyen egyszerű. Nem lehet eleget hangsúlyozni, hogy a digitalizálás nem ugyanaz, mint a digitális megőrzés. Ha valamit számítógépes formára alakítunk, az csak az első lépés – keletkezik egy digitális

objektum, ami ki van téve a jól ismert veszélyeknek:

- a hordozóeszköz tönkremenetele,
- a hardver vagy a szoftver elavulása,
- a szervezet megszűnése, és végül, de nem utolsósorban:
- emberi hibák.

Senki nem ismeri a végleges válaszokat ezekre a problémákra és valószínűleg nem lesz tökéletes megoldás később sem. Viszont a kurátorság már most is képes minimalizálni a kockázatokat egy olyan mozgó környezetben, ahol a digitális adatok folyamatosan özönlenek és a számítástechnika pedig állandóan változik. A gyakorlatban egy biztonsági mentés házon kívüli tárolása megfelelő kockázatcsökkentő megoldásnak látszik, de valójában csak egy óvintézkedés, ami nem nyújt védelmet az itt felsorolt veszélyek mindegyike ellen. Ha azt akarjuk, hogy a digitalizált anyag fennmaradjon, akkor folyamatosan gondozni kell. Itt nem létezik a „jótékony hanyagság”, mint a nyomtatott művek világában, amikor is örökre eltűntnek hitt antik könyvek csodálatos módon egyszer csak újra felbukkannak poros padlásokon. A digitális információ életben maradása teljesen a megfelelően működő hardver- és szoftverkörnyezetre függvénye. Az *Ausztrál Nemzeti Archivum* ezért is vezette be a „teljesítőképesség-mutató” (performance model) fogalmát arra az esetre, amikor egy digitális objektumot elő kell hívni: objektum+(hardver+szoftver)=teljesítőképesség. Ennek a „teljesítőképességnek” a megvalósítása többet jelent annál, mint hogy biteket tárolunk egy szerveren. Olyan szervezetet igényel, amely nemcsak arra ügyel, hogy ezek a bitsorozatok ne sérüljenek, hanem folyamatosan figyeli a világban zajló technológiai változásokat, amelyek hatással lehetnek a megőrzendő anyagra és olyan stratégiákat dolgoz ki, amelyekkel megoldható, hogy ezek a digitális objektumok a következő generációs számítógépeken is megjeleníthetők legyenek. Ilyen jól ismert stratégiák például a migrálás, az emuláció/szimuláció és a normalizálás – ezek mind meglehetősen bonyolult műszaki eljárások. Kulcsrakész és mindenre kiterjedő termékek még nem léteznek a piacon ezekre a feladatokra. Vagyis: minden olyan könyvtárnak, amely saját maga akarja felügyelni és megőrizni a digitalizált gyűjteményét, bizonyos fokig késznek kell lennie arra, hogy részt vegyen az alkalmazott műszaki megoldások továbbfejlesztésében is, különösen, ha egy már meglévő hozzáférési rendszert szeretne fenntartani.

Jó hír, hogy a saját digitalizált gyűjteményrészeknél rendszerint csak néhány fájlformátum fordul elő (PDF, JPG és TIFF), melyek kurátorsági szempontból viszonylag könnyen megőrizhetők hosszabb távon is. Egy másik fontos tényező, amely megkülönbözteti ezeket a digitális anyagokat a következő fejezetben tárgyalt tudományos kutatási adatoktól, hogy ezeket rendszerint maguk a könyvtárak hozzák létre, így feltehetően megfelelően strukturált és részletes metaadatok is tartoznak hozzájuk, ami megkönnyíti a velük való munkát és a megőrzést. De ezen könnyebbségek ellenére a hosszú távú megőrzés ebben az esetben is komoly vállalkozás egyrészt a költség miatt (ami nem annyira a tárolóeszközök árát jelenti, mert azok egyre olcsóbbak, hanem a szervezeti háttér fenntartását), másrészt a szükséges szakértelem miatt (mert bár már egyre több üzleti vállalkozás is megjelenik ezen a területen, az archívum üzemeltetése és összekapcsolása a könyvtárban működő információs rendszerekkel, sok helyi és számítástechnikai ismeretet igényel). Kérdés ezek után, hogy érdemes-e a tudományos könyvtáraknak felvállalni ezt a feladatot és felelősséget, különösen most, hogy a költségvetés egyre szűkebb és a helyzet a belátható jövőben sem lesz jobb? A nemzeti könyvtáraknak nem nagyon van más választási lehetőségük: többnyire törvény írja elő, hogy a nyomtatott és a digitális anyagok őrzőhelyként kell működniük. De a szakkönyvtáraknak több alternatívájuk is van:

- Egyszerűen házon belül tárolják a digitális gyűjteményüket és remélik a legjobbakat. Talán ez a leggyakoribb megoldás, de kockázatos. Viszont ha az eredeti állomány épségben marad, akkor lehet azzal érvelni, hogy hosszú távon olcsóbb lesz újradigitalizálni egyes tönkrement fájlokat, mint gondosan őrizni a teljes digitális gyűjteményt. Ám ez csak akkor fogadható el, ha valóban a veszélyek alapos megfontolása után döntöttek így; vagyis például garantálni tudják, hogy a fizikai hordozókon levő eredeti dokumentumok sértetlenek maradnak a jövőben.
- Keresnek egy külső intézményt, akár egy nemzeti szervezetet, akár egy fizető céget, és ott tárolják a digitális állományt. A tudományos könyvtárak egy részének ez járható út lehet, különösen amelyek túl kicsinek tartják magukat ahhoz, hogy saját digitális raktárt üzemeltessenek. Fontos viszont, hogy valóban megbízható őrzőhelyet találjanak, olyat, amely tényleg képes a szükséges felügyeletet ellátni, és amely gondoskodni tud arról, hogy a tárolt anyag integrálható legyen a könyvtári rendszerbe. Az adattárolásra szakosodott vállalkozások valójában gyakran csak biz-

tonsági mentést képesek nyújtani, miközben hosszú távú őrzőhelyként hirdetik magukat. (Hasonló ez ahhoz, mint amikor úgy reklámoznak egy optikai tárolóeszközt, hogy ezer évig is megmaradnak rajta az adatok – ami jól hangzik, de ez semmi védelmet nem jelent a hardver és a szoftver elavulása ellen. Hiába maradnak rajta a bitek ezer évig, ha nem lesz olyan számítógép, amely fel tudná dolgozni őket.)

- A harmadik lehetőség egy saját digitális archívum kiépítése. Ez a leginkább nagyratörő és mélyreható választás a lehetséges alternatívák közül.

Digitális kutatási adatok

Miközben a szakkönyvtárak megszabadultak a kiadói e-folyóiratok őrzésének felelősségétől, egy másik feladat jelent meg náluk: a kutatási adatok tárolása és gondozása. Ez igazi kihívás, tekintettel arra, hogy a tudományos adatok a legkomplikáltabb formáját jelentik a megőrzendő digitális információknak. Mind az adatok, mind az előállítók, rendkívül változatosak, és olyan komplex digitális objektumokat is tudni kell kezelni, mint az „élő” adatbázisok. Ugyanakkor a Google-szerű vállalkozások korában, amikor úgy tűnik, hogy a szakkönyvtárak lassan elveszítik az összes tradicionális életcéljukat, itt egy lehetőség arra, hogy új, mással nem pótolható helyet vívjanak ki maguknak – egyenesen az oktatási vagy kutatási anyaintézményük információs rendszerének közepében.

Jelenleg elég kevés egyetemi könyvtárnak van olyan digitális repozitóriuma, amely kellően feltöltött ahhoz, hogy valódi segítséget jelentsen a tudományos kutatásban és az oktatásban. Mindazonáltal a szakkönyvtáraknak legalább három olyan alapvető sajátosságuk van, amelyek különösen alkalmassá tehetik őket a tudományos kutatási eredmények kurátorságára:

- a küldetésükben benne van a hosszú távú megőrzés,
- strukturális támogatást kapnak,
- részei a tudományos közösségnek.

Igaz, hogy a második érv megkérdőjelezhető, mert a könyvtárak finanszírozása ritkán elégséges, viszont ők legalább költségvetésbe beépített támogatást kapnak. Mióta a tudományos kutatás projektrendszerben és határozott időre szóló pénzügyi támogatásokkal folyik, a folyamatos finanszírozás hiánya miatt a kutatás során összegyűjtött adatok gyakran odavesznek, miután a projektek lezárul-

nak. Ennél még a könyvtárak szerény, de legalább biztos költségvetése is nagyobb garanciát jelent a hosszú távú megőrzéshez.

Ami azt az előnyt illeti, hogy a könyvtárak részei a tudományos közösség hálózatának: ez nem vehető biztosra. Ehhez a könyvtári és adatkezelési szolgáltatásokat speciálisan az adott kutatói kör igényeire kell(ene) szabni, és elsődleges prioritást adni az ő szükségleteiknek és kéréseiknek. Nem véletlen, hogy a digitális megőrzés területén leginkább használt Open Archival Information System keretrendszer kidolgozói kiemelt helyet és külön szakkifejezést rendeltek a „célközönséghez”.

Ezek a magas elvárások arra a döntésre is vezethetnek, hogy a tudományos könyvtár végül is nem száll be ebbe a folyamatba. Főleg, hogy a könyvtári szféra még inkább csak küzd a digitális világ különböző kihívásaival, miközben egyes tudományágak kutatói idejekorán felismerték a digitális jövőt és ügyesen megszervezték a maguk információs hálózatát, akár európai, sőt akár világméretűekben is. Ezek a nagy hálózatok általában jól megalapozottak és az adott tudományág területén dolgozók igényeire vannak optimalizálva. Így a tudományos könyvtáraknak nem sok szerep jut bennük, főleg mert a legnagyobb problémájuk a stabil és jelentős méretű pénzügyi háttér megszerzése, amiben a könyvtárak nem tudnak segíteni.

Tanácsok

Befejezésül néhány gyakorlati, „tedd” és „ne tedd” típusú jó tanács azoknak a könyvtáraknak, amelyek azon gondolkodnak, hogy hogyan kezeljék a digitális gyűjteményüket és az általuk szolgált kutatók digitális információkra irányuló igényeit:

- Találjuk meg a „célközönségünket”.
- Üljünk le, írjunk egy tervet, és alakítsuk ki a célközönségünkkel kapcsolatos politikánkat. Sokszor már ennyi is elég, hogy felismerjük az erősségeinket és gyengeségeinket, és helyes sorrendbe tegyük a prioritásokat.
- A tervben vegyük számba azokat a digitális gyűjteményeket, amelyeket gondoznunk kell, és próbáljuk felbecsülni, hogy milyen problémákat okozhatnak a belátható jövőben.
- Beszéljünk a célközönségünk tagjaival az információs igényeikről és építsük be őket a tervünkbe.
- Selejtezzünk, és azután selejtezzünk újra. Bár néhány archiváló szakember még mindig az „őrizzük meg mindent” elvet hirdeti, mert „kik vagyunk mi, hogy eldöntsük, mire lesz szükség a jövőben?”, de a legtöbben már egyetértenek abban,

hogy „mindent” megőrizni nem lehetséges és nem is célszerű.

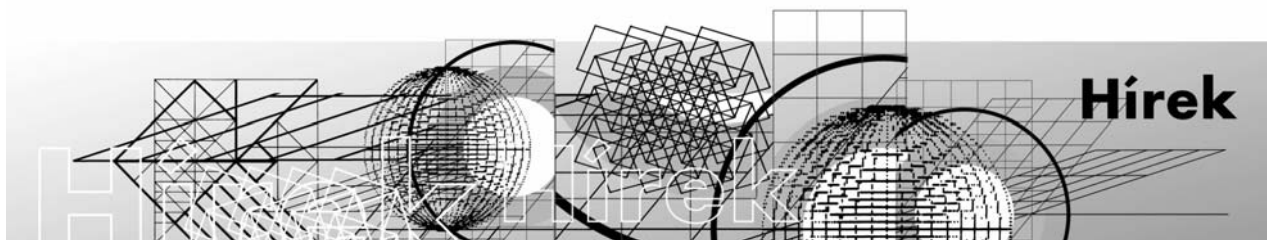
- A méret számít. Az első digitális objektum, amit a kurátorságunk alá veszünk, rendkívül költséges lesz, de a milliomodik már csak fillérekbe kerül majd. Ha nem vagyunk elég nagyok ehhez, akkor válasszuk a következő pontot.
- Keressünk partnereket, lehetőleg hozzánk hasonló területen és mérettartományban. A gyakorlat azt mutatja, hogy az együttműködés leginkább a hasonló partnerek között jön létre, és egyre nehezebbé válik, minél távolabb vannak a szervezetek egymástól a hagyományaikat és a céljaikat tekintve.
- Válasszunk egy szakértői hálózattal rendelkező ernyőszerületet. Ez lehet regionális, országos vagy nemzetközi.
- Kényeztessük el a célközönségünket. Engedjük, hogy a kutatók kutatók maradhassanak, például ne erőltessük rájuk a mi metaadat-sémáinkat, hanem inkább igyekezzünk nekik olyan adatkezelő eszközöket és megoldásokat kínálni, amiket könnyen be tudnak építeni a munkafolyamataikba.
- Döntsük el, hogy miben vagyunk jók, és ennek alapján keressük meg a helyünket az adatkuratori szervezetek hálózatában.

De ha ez nem megy, akkor legalább ennyit tegyünk meg:

- Soha ne tároljunk fontos digitális információt floppy lemezen, CD-n, vagy lokális PC-n, hanem töltsük fel megbízhatóbb hardverre.
- Rendszeresen mentsük az adatainkat és állapotunk meg egy kollégánkkal, hogy vigyázunk egymás back-up állományaira.
- Készítsünk listát a felügyeletünk alá tartozó digitális anyagokról és tartsuk naprakészen.
- Keressünk megbízható őrzőhelyet az adatainknak. Ez lehet a nemzeti könyvtár, vagy egy nemzeti archiváló szervezet, vagy akár egy olyan kolléga, aki digitális repozitóriumot üzemeltet. Ha nem áll módunkban hivatalosan meggyőződni arról, hogy egy archiváló hely mennyire megbízható, nézzük meg az intézmény egészét és tegyük fel magunknak a kérdést: valószínű, hogy ez a szervezet még ötven vagy száz év múlva is létezni fog?
- Az erőforrásainkat a hozzáférés kérdésére összpontosítsuk, mert hosszú távon csak az fog számítani. Minden más csak eszköz ahhoz, hogy a hozzáférés lehetőségét megőrizzük.

/ANGEVAARE, Inge: Taking care of digital collections and data: 'Curation' and organisational choices for research libraries. = LIBER Quarterly, 19. köt. 1. sz. 2009. p. 1–12./

(Drótos László)



A Gale új adatbázisa: Greenr

A Gale új adatbázisa, a *Greenr (Global Reference on the Environment, Energy and Natural Resources)* a környezet, az energia és a természeti erőforrások terén nyújt megbízható referenz tartalmat. Híreket, háttér-információkat, videókat, kommentárokat, elsődleges dokumentumokat, statisztikákat tartalmaz olyan fontos témákban, mint például az energiarendszerek, az egészségügy, az élelmiszer, a klímaváltozás, a népességnövekedés és a gazdasági fejlődés. A keresés hasonlít a népszerű weboldalakban történő keresésre, lehet böngészni vagy keresőkifejezésekkel feltenni a kérdést.

A *Greenr interaktív térképeket* is kínál, amelyekben böngészhetünk ország vagy tárgykör szerint, vitafórumokat, lektorált esettanulmányokat, a helyi erőforrások és események hirdetési lehetőségét, web 2.0 eszközöket közösségépítés céljából.

/Advanced Technology Libraries, 38. köt. 4. sz. 2009. p. 7-8./

A Springer elindította az AuthorMapper.com szolgáltatást

A *Springer Science+Business Media* elindította az *AuthorMapper.com* (www.AuthorMapper.com) ingyenes online elemző rendszerét, amellyel a kutatások trendjeit lehet feltárni. Jelenleg az összes Springer-folyóiratból több mint 3 millió folyóiratcikket hasznosít a rendszer, de a többi tudományos, műszaki és orvosi kiadó metaadatait is be dolgozzák a továbbiakban. Komplex kérdéseket lehet feltenni tárgyszó, tudományterület, intézmény, folyóirat, szerző szerint. Eredményül időgrafikonokat, statisztikai táblákat kaphatunk a legújabb vagy a történeti jellegű tudományos trendekről, megállapíthatjuk a szakirodalmi irányzatokat, feltárhatunk tágabb tudományos összefüggéseket és fellehetjük a kutatási terület további szakértőit. Például időgrafikonon szemlélhetjük meg, hogy egy adott kutatási téma felfutó vagy lecsengő fázisban van-e. Megállapíthatjuk, hogy egy adott

kutatási témában melyek a legtermékenyebb kutatóintézetek és össze is hasonlíthatjuk őket. Megnézhetjük azt is, hogy lakóhelyünkhöz legközelebb hol dolgoznak a témában szakemberek.

/Advanced Technology Libraries, 38. köt. 4. sz. 2009. p. 7-8./

Az Oxford University Press kiadta a Biblical Studies Online-t

Az *OUP (Oxford University Press)* megjelentette az *Oxford Biblical Studies Online*-t, amely elérést nyújt a Biblia hat alapvető Oxford-kiadásához, valamint Biblia-kutatásokhoz (kommentárok, annotációk), referenzanyagokhoz és elsődleges dokumentumokhoz.

Az adatbázisban megtalálható az *Oxford Study Bible*, a *New Oxford Annotated Bible 3/e*, az *Access Study Bible with Apocrypha* és az *Authorised King James Bible with Apocrypha*. A referenzművek között szerepel az *Oxford Bible Commentary*, az *Oxford History of the Biblical World*, az *Oxford Encyclopedia of Archeology in the Near East*, az *Encyclopedia of Dead Sea Scrolls* és a *How to Read the Bible*.

Az 5000 szócikket tartalmazó adatbázisban a tudományos leírások mellett térképek, illusztrációk, interaktív idősorok, bibliográfiák és szójegyzékek is találhatóak. A kutatás valamennyi szintjének szól, hat havi frissítése megoldja az újabb tartalmak folyamatos felvételét.

/Advanced Technology Libraries, 38. köt. 4. sz. 2009. p. 8./

Az MIT nyílt hozzáférést nyújt a tudományos cikkeihez

Annak érdekében, hogy az *MIT* tudományos kutatásainak eredményi szélesebb körben legyenek elérhetőek, a *Massachusetts Institute of*

Technology egyhangúan megszavazta a döntést, hogy a kutatásairól szóló cikkeket ingyenesen hozzáférhetővé teszik a weben a *nyílt hozzáférés* keretében. Az egyetemi oktatók nem kizárólagos engedélyt adtak folyóiratcikkeik nyílt hozzáférése érdekében, hanem a MIT és a *Hewlett Packard* által kidolgozott *Dspace* platformon keresztül. Az MIT az *első* olyan egyetem az USA-ban, amely egyetemszintű elérést tesz lehetővé, mivel a *Harvard* és a *Stanford* csak egyes részlegei tekintetében kínálja azt. Az új kezdeményezésnek köszönhetően az egyetem oktatói által publikált, évente több ezer folyóiratcikk válik elérhetővé a weben és kereshetővé a Google-val és más keresőmotorokkal.

/Advanced Technology Libraries, 38. köt. 5. sz. 2009. p. 2./

Az Alexander Street új terméke: Music Online

Az *Alexander Street* elindította a *Music Online*-t, a létező *legátfogóbb* adatbázist a klasszikus, jazz, világ- és amerikai zene terén. Sajátossága, hogy audio- és videofelvételeket, teljes szövegű referenzanyagokat, kottákat, életrajzokat és képanyagokat egyaránt tartalmaz. A felhasználó különböző, a későbbi meghallgatásra kijelölt tételek listáit (playlist) állíthatja össze, válogatva akár a *Music Online*-ból, akár a weben, annotálhatja a tételket, állandó url-en tárolhatja őket saját használatra vagy más előfizetőkkel való megosztásuk céljából. Például, listát készíthet egy adott zenemű különböző felvételeiből, hozzárendelve a videót, amely a zeneművet tartalmazza, a webről letöltött vonatkozó publikációkat, a szerző életrajzát és

fényképét. A könyvtárak *előfizethetik*, illetve *megvehetik* a teljes adatbázist, vagy egyes részeit, ezek pedig együttesen is kereshetők (cross-searching).

/Advanced Technology Libraries, 38. köt. 5. sz. 2009. p. 4-5./

Az EBSCO fogászati adatbázisa, a Dentistry

Az *EBSCO* kiadta a *Dentistry and Oral Sciences Source* adatbázist *EBSCOhost* szolgáltatása keretében. A világon az *első fogászati tárgyú* adatbázis 160 folyóiratot és több tucat monográfiát indexel és referál teljes körűen (cover-to-cover). Emellett 130 teljes szövegű fontos folyóiratot és monográfiát is tartalmaz.

/Advanced Technology Libraries, 38. köt. 5. sz. 2009. p. 5-6./

World Digital Library

A *Kongresszusi Könyvtár* és az *UNESCO* 32 további partnerintézménnyel közösen létrehozta a *World Digital Library* weboldalt, amely a *világ könyvtárainak és levéltárainak* unikális kulturális anyagaihoz nyújt ingyenes hozzáférést a nyilvánosság számára, beleértve kéziratokat, térképeket, ritka könyveket, filmeket, hangfelvételeket, fényképeket. A weboldal hét nyelven lesz kereshető és több tucat nyelven írt dokumentumot kínál.

/Advanced Technology Libraries, 38. köt. 5. sz. 2009. p. 12./

Kedves Olvasóink !
A következő alkalommal
összevont lapszámmal jelentkezünk.

Avt
szerkesztősége