

Vince Mátyás

Nyitott archívumok – az MTI az interneten elérhetővé tette régi híreinek nagy részét

A Magyar Távirati Iroda Zrt. (MTI) 2009 decemberében sajtótörténeti jelentőségű lépésre szánta el magát: megnyitotta XIX. és XX. századi hírchívumának jelentős részét a nyilvánosság előtt. Az MTI internetes honlapján keresztül mintegy ötmillió digitalizált hír között böngészhetnek ingyenesen az érdeklődők. A tájékozódást korszerű keresőszoftverek segítik. A kormányzati céltámogatásból finanszírozott 100 millió forintos digitalizálási és állagmegóvási projektben az MTI partnere az Országos Széchényi Könyvtár (OSZK) és a Magyar Országos Levéltár (MOL) volt.

A hírchívumi és fotóarchívumi digitalizálás terén régóta megoldatlan feladatokra az MTI Zrt. 2009-ben 100 millió forint céltámogatást kapott a 2167/2008. (XII. 4.) számú kormányhatározat alapján. A támogatásról intézkedő kormányrendelet a következő célokat rögzítette: „a Magyar Távirati Iroda Zrt. részére a digitális archívum közérdekű használhatóságának javítására, keresési lehetőségeinek bővítésére, a fényképarchívum és a hírchívum egy részének digitalizációjához, az így digitalizált adatok – nem kereskedelmi célú felhasználás esetén ingyenesen hozzáférhetővé tétellel megvalósuló – internetes közzétételéhez, valamint az 1989 és 2004 közötti hírchívum – nem kereskedelmi célú felhasználás esetén ingyenesen hozzáférhetővé tétellel megvalósuló – internetes közzétételéhez”. A digitalizálásra szánt állami pénz felhasználására egy MTI-s projekt keretében került sor, amely a „Nyitott archívumok” elnevezést kapta.

A Magyar Nemzeti Múzeum még 1993-ban védett muzeális gyűjteménnyé nyilvánította az MTI archívumait (*Sajtóadatbank és Mikrofilmarchívum – együttesen Hírchívum és Fotóarchívum*). Ennek a gyűjteménynek a megóvási kötelezettségéről – az MTI-ről, mint nemzeti hírügynökségről szóló – 1996. évi CXXVII. törvény rendelkezik. A törvény 2. § (1) bekezdésének j) pontja szerint ugyanis a nemzeti hírügynökség közszolgálati feladata, hogy „a tevékenysége során birtokába került kulturális értékek és történelmi jelentőségű eredeti dokumentumok tartós megőrzéséről és védelméről archívumában gondoskodik, azokat szakszerűen összegyűjti, tárolja, gondozza, és az azokhoz való hozzáférhetőséget biztosítja.”

Amikor az MTI vezetése az elmúlt évtizedben folyamatosan kérelmezte a kormányzatnál, hogy külön forrásokat kapjon e muzeális archívum állagmegóvására és digitalizálására, támogatási igényét azzal indokolta, hogy a tulajdonában lévő fotó- és szöveges archívumok állagromlása a papír-adathordozók és a fotónegatívek estében veszélyes méreteket öltött. Az állomány veszélyeztettségét külső szakértői anyagok is igazolták. Ugyanakkor egyre növekedett az igény az archívumi tartalmak digitalizálására annak érdekében, hogy a felhasználók korszerű keresőrendszerek segítségével gyorsan és hatékonyan megtalálják az MTI képeit (összesen 13 millió fotónegatív van az MTI birtokában), illetve az MTI korábbi híreit (128 év több tízmillió híreről van szó).

Az MTI fennállásának 128 esztendejéből több mint száz éven át papíron rögzítette folyamatosan készülő és legépelt napi híreit. (Az előfizetők a legváltozatosabb módon kapták meg a számukra fontos MTI-információkat; volt, hogy futár vitte a legutóbbi néhány óra stencilezett híreit a szerkesztőségekbe, volt, hogy a „központból” telefonon olvasták be a legfrissebbeket.) A számítógépes világ persze az MTI-t is meghódította: 1988 óta a már számítógépen készült MTI-hírek digitális archívumban gyűlnek és érhetők el, az MTI által kiadott fotók pedig 2001. január óta csak digitális formában kerülnek archiválásra.

A „Nyitott archívumok” projekt részben a már digitalizált MTI-s hírek, illetve képek egy részének internetes nyilvánosságra hozatalát szolgálta, részben pedig a papíron és mikrofilmen megőrzött hírek digitalizálását. (Az MTI archív fotóanyagának részleges digitalizálását saját és egyéb külső for-

rásokból – NKA, Norvég Alap – fedezi a hírügynökség.)

Az MTI híreiből mostantól öt korszakban külön-külön lehet keresni a <http://www.mti.hu> oldalon, az MTI internetes nyitóoldalán. Az érdeklődők ingyenesen búvárkodhatnak az 1887 és 1919 közötti évekből fennmaradt, és az Országos Széchényi Könyvtárban megőrzött MTI-hírek között; ezek sajnos csak töredékek, nem maradtak teljes évfolyamok az utókorra. Az 1920 és 1945, valamint az 1945 és 1949 közötti időszak már teljesebb; itt a német nyelvű és a közgazdasági kiadás kivételével minden fellelhető MTI-híryanagot digitalizáltunk és szövegfelismerő szoftveren átfuttattunk, majd elhelyeztük az adatbázisban. Érdeklődésre tarthat számot, hogy mindkét időszakból az adatbázisba tettük az úgynevezett bizalmas anyagokat is: ezek az MTI szerkesztőségében előállított hírek csak szűk kör számára voltak elérhetők. Az 1956-os év teljes anyagát is digitalizáltuk, ez is külön kereshető az adatbázisban. Külön adatbázisban gyűjtöttük össze és tettük kereshetővé az 1988 és 2005 közötti időszakban kiadott, már eredetileg is digitálisan készült MTI-híreket. Ez utóbbi adatbázis az MTI-híreken és a napi bizalmas kiadványokon kívül az *Országos Sajtószolgálat (OS)* és az *OTS*, valamint az MTI korabeli rádiófigyelő szolgálatának – a rendszerváltásig bizalmas – anyagait (pl. a *Szabad Európa Rádió*, a *BBC*, az *Amerika Hangja* híryanagát) is tartalmazza leírt formában.

A most kinyitott MTI-archívumok híryanagainak lelőhelye nem csak a Magyar Távirati Iroda: alapvetően a Magyar Országos Levéltárban és az Országos Széchényi Könyvtárban – könyvmegos és mikrofilm formában – őrzik az MTI történelmi híryanagainak szinte teljes egészét. Az MTI-hírek őrzői természetes partnerei voltak e hatalmas digitalizálási projektnek: a MOL és az OSZK hatékony közreműködése és segítsége nélkül ezt a programot nem lehetett volna ilyen rövid idő alatt végrehajtani.

Az anyagok digitalizálását a Magyar Országos Levéltárban jelentős állományvédelmi előkészítés előzte meg. A MOL-nál végzett munka során keletkezett digitalizált oldalak száma meghaladja a 710 ezret! Az OSZK-nál tárolt XIX. századi és XX. század elejei MTI-híreket az OSZK maga digitalizálta, s így kerültek az adatbázisba.

A digitalizálás folyamata minden esetben a következő volt: a dokumentumról digitális kép készült fényképfelvétel, vagy szkennelés során. Az így

előállt TIFF fájlokat JPG fájlkként mentették el. A következő lépésben történt meg a JPG fájlok OCR-ezése (digitális szövegfelismerése) majd a PDF fájlok létrehozása.

Az OSZK-ban található MTI-s hírállomány digitalizálására közbeszerzési pályázatot írt ki az MTI, amelyet az OSZK nyert meg. (Az OSZK a teljes digitalizálási folyamatot maga végezte el.) Az OSZK-nál az eredetileg 10 100 oldalra becsült digitalizálandó oldalszám végül 30 907 oldal lett.

A Magyar Országos Levéltárnál lévő anyag feldolgozására szintén közbeszerzési eljárást indított az MTI Zrt., amelyet az *Arcanum Adatbázis Kft.* nyert meg. Az Arcanum a teljes digitalizálási folyamatot elvállalta az OCR-ezést is beleértve. (Az eredeti terveinkben és az Arcanum vállalatában is szerepelt mikrofilm-digitalizálás. A tesztek és az első digitalizálási, majd OCR-ezési kísérletek bebizonyították, hogy várakozásunkkal ellentétben az 1944–1949-es és az 1956-os évközből származó mikrofilmekről készült képek OCR-ezése gyenge eredményt produkált. Ezért úgy döntöttünk, hogy az OSZK tulajdonában lévő negatív mikrofilmek feldolgozása helyett az eredeti, papíralapú nyomtatványokról végezzük a hírek digitalizálását.)

A MOL két levéltárában, a budapesti Bécsi kapu téri és a Hess András téri levéltárakban folyt a hírek digitalizálása. A Bécsi kapu téren az Arcanum Kft. által digitalizálásra került a MOL K-428 törzsszámú Magyar Távirati Iroda fondjából 1248 kötet, amely összesen 565 626 oldal. A digitalizált anyag a következőket foglalja magába: 1920. október 1. és 1944. június 30. közötti időszak: az MTI napi tudósításai, a napi tudósításokkal kapcsolatos közlemények, Rádiófigyelő, Lapszemle, Bizalmas külpolitikai szemle, Külföldre adott hírek, Bizalmas értesítések, Heti kiadás, Budapesti értesítő, Egészségügyi hírek, Ki nem adott közlemények, Pro Domo, Házi tájékoztató, A frank-per hírei, a *Magyar Országos Tudósító (MOT)* hírei.

A Hess András téren a XXVI-A-14 törzsszámú anyagból az Arcanum Kft. 280 kötetet digitalizált, összesen 151 637 oldalt. A digitalizált évkörök a következők: az 1944. július 1. – 1949. december 31. közötti időszak teljes napi külpolitikai és belpolitikai hírközlése, valamint a Kulturális hírek és a Sporthírek, ezen kívül egy különkiadás *Sztálin* születésnapjára. Az 1956-os évből digitalizálták a külpolitikát, a sporthíreket, és a bizalmas értesítéseket; végül pedig a Magyar Országos Tudósító háború utáni két évfolyamát, egészen a megszű-

néség, azaz az 1947. március 1. és az 1948. június 30. közötti időszak anyagát. A MOL által őrzött MTI könyvművekből összesen 1528 kötetet digitalizáltunk, amelyek összterjedelme 717 263 oldal.

A teljes projekt során a MOL és az OSZK által őrzött MTI könyvművek közül 1619 kötetet, azaz összesen 748 170 oldalt digitalizáltak. Két történelmi korszak MTI-hírei várnak még digitalizálásra: az 1950–1955-ös korszak, illetve az 1957–1987 közötti periódus. Remélhetőleg ezek digitalizálására is lesz finansziális lehetőség az elkövetkező években.

Az úgynevezett „könyvműve” híryanag digitalizált változata nem csupán az MTI Zrt. honlapján keresztül vált ingyenesen megtekinthetővé, hanem – az érintett intézményekkel kötött megállapodás alapján – a Magyar Országos Levéltár és az Országos Széchényi Könyvtár is hasonlóképpen digitalizálva és keresésre alkalmas módon teszi lehetővé a széles nagyközönség számára az MTI-hírekhez a hozzáférést.

A projekt részeként sor került az adatbáziskeresőrendszer, egy felhasználóbarát keresési rendszer kifejlesztésére. Az MTI szükségesnek látta, hogy könnyen kezelhető, a szűkítési lehetőségeket támogató, és jó találati lehetőséget nyújtó keresőmotort adjon az adatbázisokhoz. Ezt az *Ovitas Magyarország Kft.*-vel együttműködve alakították ki. Az átadott rendszer olyan webalkalmazás, amely az MTI-ben működő *Automatikus Kódoló Rendszer (AKR)* által metaadatokkal ellátott anyagok és a tudásbázisban lévő fogalmak és kapcsolatok felhasználásával valósítja meg a keresőszolgáltatást az 1988–2005 évek közötti MTI-híryanagra. Ez az alábbi követelmények kielégítését jelenti:

- Hagyományos teljes szöveges keresés.
- Fogalomalapú keresés: fogalmak azonosítása a keresőkérdésben, dokumentumokon lévő fogalmak és kódok felhasználása a keresés során, fogalmi kapcsolatok súlyozott kihasználása (pl. földrajzi taxonómia, személy-intézmény kapcsolatok stb.) a keresés során, fogalmi konfliktusok kezelése (elsősorban személyek).
- A keresőmotornak kezelnie kell a magyar nyelvi specialitásokat úgy, ahogy az AKR rendszer esetében is. Elvárt megoldás a MorphoLogic szótóvesítő modul használata, amelyhez az MTI kivételszótárát kell hozzáadni.
- A találati listában a keresés finomítását segítő fogalom/kód és dátumalapú statisztikák készítése, valamint grafikus megjelenítése a felhasználók számára. Ez gyakorlatilag az adott találati

halmazban lévő dokumentumok különböző metaadat-típus (pl. dátum, fogalom, témakód) szerinti eloszlását mutatja, amely így áttekinthető nyújt a felhasználónak a teljes találati listáról, illetve könnyen tovább tudja finomítani a kívánt irányba a keresését.

A „Nyitott archívumok” projekt keretében az MTI saját internetes oldalán külön aloldalt hozott létre a rendszerváltás időszakának hír- és fotóanyagainak bemutatására. Az oldal „időutazás” jelleggel napról napra nyitja meg a korabeli tartalmakat; elérhetősége: <http://1989.mti.hu>. Az aloldal kialakításakor arra törekedtünk, hogy egyszerre mutassuk be, tárjuk a széles nyilvánosság elé az MTI értékes szöveges és fotóarchívumi kincseit digitalizált formában, s egyúttal az interneten olvasók számára érdekes, hasznos, forrásértékű tartalomszolgáltatást nyújtunk a közszolgálati jegyében. Az oldal technológiai kialakításának munkálatait részben saját erőből, részben a céltámogatás terhére végeztük el, a folyamatos, napi frissítés, a sok forrásból érkező kiegészítő anyagok felhelyezése külön e feladattal megbízott szerkesztő munkája.

A napi hírek a rendszerváltás oldalon – egyelőre az 1989-es évre kiterjedően – a következő rovatokon keresztül érhetők el a maguk teljességében (ez vonatkozik egy adott hír tartalmára és a napi hírkiadásra is, vagyis minden hír teljes terjedelmében nyilvánossá vált): Belpolitika, Gazdaság, Sport, Panoráma (színes hírek), Országos Sajtószolgálat (OS). (Az MTI médiatörténeti jelentőségű „találmánya” az Országos Sajtószolgálat, amely a rendszerváltásnak szinte tömegkommunikációs szimbóluma lett, hiszen első alkalommal vált lehetővé, hogy társadalmi szervezetek, pártok úgy közöljék mondanivalójukat a széles nyilvánossággal, hogy közleményüket az MTI teljes terjedelmében, változtatás nélkül juttatja el az előfizetőkhez.)

Az oldal kialakításába külső partnereket is bevontunk. Ennek eredményeként egészíthettük ki a korabeli MTI-híreket és -fotókat más tartalmakkal: például a *Magyar Televízió*hoz és a Szabad Európa rádióhoz betelefonáló emberek üzeneteivel, külföldi magyar nyelvű rádiókban elhangzott interjúk, tudósítások hanganyagával. Ugyancsak közöljük (első alkalommal) azokat az MTI raktárból előkerült, s e projekt keretében digitalizált dokumentumokat, amelyek az MTI egykori rádiófigyelő szolgálatának termékei, s a külföldi magyar nyelvű rádiók adásai (BBC, Szabad Európa, Amerika Hangja, Vatikáni Rádió stb.) egy részének leírt verzióját tartalmazzák. A „Nyitott Archívumok” pro-

jekt 2009. június 30.-iki lezárása után az 1989-es aloldal további működtetését az MTI Zrt. saját forrásból oldja meg.

Az MTI fotóarchívumáról – az MTI csak előfizetők által látogatható digitális fotóbankja jelenleg közel egymillió digitalizált fotót tartalmaz, a negatívarchívum pedig kb. 13 millió fotót – keresztmetszeti képet kíván őrizni a „Nyitott Archívumok” projekt keretében megnyitott fotóadatbázisban. A Fotómozai elnevezésű internetes felületen egyfajta ízelítőt adunk az MTI teljes képvagyonából: a nyilvánosság számára 2009 decemberétől elérhetővé tettünk több mint 20 ezer kisebb méretű képet – ingyenesen a 2001-től 2005-ig terjedő korszakból. Ezt az állományt a hírügynökség 2010 során több tízezer archív képpel folyamatosan bővíti (a cél 70 ezer kép), és szélesíti majd a tematikát is.

A „Nyitott Archívumok” projekt keretében részben az állami céltámogatásra alapozva, részben saját, MTI-s forrás igénybe vételével a hír- és fotódigitálizáláson kívül az MTI az archívumában rejtőző fotók nyilvánosságra hozatalára is költött: így töb-

bek között az ötvenes évekből előkerült MTI-s fotókból a hírügynökség válogatást jelentetett meg egy fotóalbumban „Elfelejtett képek” címmel. A könyv anyagát több helyen kiállítás formájában is láthatta a hazai nagyközönség. Szintén részben a céltámogatás segítségével sikerült kiadni az MTI fotóarchívumi felvételeire alapozva az MTI „Film – Színház – Muzsika” sorozatának második és harmadik kötetét, amelyek *Törőcsik Mari* és *Kállai Ferenc* életútját mutatják be.

Beérkezett: 2010. I. 4-én.



Az Opera a leggyorsabb böngésző

A norvég cég vezérigazgatója szerint az *Opera* a leggyorsabb böngésző. A legújabb, 10.50-es változat a 10.10-esnél akár hétszer gyorsabb.

Az új böngésző elsősorban a Javascript-motorja és az új grafikai könyvtára miatt különbözik elődeitől, valamint jóval hatékonyabban használja ki a rendelkezésre álló képernyőfelületet és privát böngészési üzemmódot is kínál.

A norvég szakemberek a bevált klasszikus menüsört egy menügombbal helyettesítették, így jóval több hely maradhat a képernyőn, például a honlap-előnézeti ablakoknak. A szoftver lehetővé teszi, hogy valaki akár egy ablakban is privát böngészést indítson, anélkül, hogy ehhez be kelljen járnia a korábban megnyitott oldalakat. A Firefox például bezárja a korábbi ablakokat, ha a felhasználó aktiválja a privát böngészés üzemmódot. Az újdonságok közül kiemelendő a Windows 7, a HTML5 és a CSS3 eddiginél jobb támogatása.

Az új Opera 10.50 egyelőre csak a Windows operációs rendszerekre érhető el, de hamarosan letölthetők lesznek a Mac OS X-et, a Linuxot, a FreeBSD-t és a Solarist támogató verziók is.

/Sg.hu Hírlevél, 2010. március 4., <http://www.sg.hu/>

(SzP)