

Drótos László

Országos Széchényi Könyvtár MEK Osztály

A MEK 2.0 verziójára vonatkozó fejlesztések összefoglalása*

A Magyar Elektronikus Könyvtár (MEK) a hazai internet egyik legrégebb és legismertebb „képződménye”, a legnagyobb magyar nyelvű tudományos és szépirodalmi szöveggyűjtemény a hálózaton. Népszerűsége részben állandóságának és egyszerűségének is köszönhető: a dokumentumok évek óta ugyanott találhatóak, és egy egyszerű menürendszer, illetve katalógus segítségével böngészhetők, kereshetők. A digitális dokumentumok leírása és szolgáltatása terén időközben elterjedt szabványos megoldások, valamint a növekvő olvasói igények miatt azonban szükségessé vált a MEK szerkezetének újragondolása és újjászervezése.

Történet

A Magyar Elektronikus Könyvtár 1994-ben jelent meg az interneten, először egyetemi gopher szervereken elosztott információforrásként, majd 1995-ben központi szolgáltatásként – mint az NIIF egyik alprojektje – a *helka.iif.hu* nevű gépen. Egy évvel később készült el a webfelület a *http://mek.iif.hu* címen, és mind a mai napig lényegében ez a MEK központi szolgáltatása (a „hátrányos helyzetű” felhasználók kedvéért a gopher és az anonim FTP-elérés is megmaradt). Ez az „1.0-s verzió” mára már több szempontból is megérett a változtatásra: egyrészt közben megjelentek és elterjedtek azok a nemzetközileg is szabványos megoldások az elektronikus dokumentumok leírására és tárolására, amelyek a MEK ötletének megszületésekor még nagyon hiányoztak, másrészt a könyvtár gyűjteménye (több mint 4300 dokumentum) és forgalma (havi 50-60 ezer látogató) meghaladta azt a mértéket, amihez elegendő volt az egyszerű fájlarchívum struktúra. 1999 őszén az Országos Széchényi Könyvtár révén komoly szakmai, szervezeti és pénzügyi háttérrel kapott az Elektronikus Könyvtár, így 2000 elején elkezdték megtervezni a MEK 2.0 verziót, amely egy több tízezer darabos gyűjteménnyel rendelkező, korszerű és szabványos, a jelenleginél látványosabb és könnyebben használható digitális könyvtár felépítését jelenti. A fejlesztés kezdetben elég lassan haladt, mert egyrészt meg kellett teremteni a technikai feltételeket (saját szerver és szoftverkörnyezet), másrészt kevés szabad munkaerő és pénz jutott az új rendszer kialakítására, hiszen közben a jelenlegi szolgáltatást is folyamatosan bővíteni, fejleszteni kellett. 2001 végén indult be igazán az érdemi munka,

mára már nagyjából elkészült az adatfelviteli rész, valamint az új webfelület, amelyből egy demonstrációs változat már megnézhető a *http://mek.oszk.hu* oldalon (1. ábra).

Előfeltételek

A tervezéskor a következő szempontokat tartottuk fontosnak. Csak nyílt és ingyenes szoftvereket használjunk az alapszolgáltatáshoz, de egyes kiegészítő funkciókhoz (pl. fejlett teljes szövegű keresés) kereskedelmi szoftverek is szóba jöhetnek. A rendszer működtetői és felhasználói oldalról egyaránt könnyen megtanulható és esztétikus legyen, valamennyi fontos menedzselési funkció webfelületen át történjen. Kompatibilisnek kell lennie a vonatkozó nemzetközi szabványokkal, és adatcsere szinten együtt kell működnie más könyvtári adatbázisokkal, internetes szolgáltatásokkal. A rendszer teljes egészében vagy egyszerűsített formában szabadon átvehető lesz olyan nonprofit szervezetek számára, amelyek a MEK-et vagy annak részhalmazát saját szerverről szeretnék szolgáltatni (pl. határon túli tükrözések), illetve saját elektronikus könyvtárat építenek.

Technikai környezet

A MEK 2.0 verzió az OSZK-ban egy Compaq Proliant típusú szerveren épül (600 MHz-es Pentium III processzor, 256 Mb-ot RAM és 36 Gigabájt

*Az OSZK-ban a *Digitális könyvtári kezdeményezések* c. szakmai napon (2002. május 24.) elhangzott előadás alapján.

1. ábra

winchester). A gépen Debian Linux operációs rendszer van, a fejlesztés pedig MySQL adatbázis-kezelővel és PHP nyelven folyik. A későbbiekben a Zebra nevű, XML keresést és Z39.50-es elérést is biztosító szoftvert és – valószínűleg – a Search++ nevű teljes szövegű keresőprogramot is integráljuk a rendszerbe. (Mindkettő ingyenes szoftver.) A webszolgáltatást egy Apache szerver biztosítja. A csoportmunkához szükséges közös levelezés egy Mailman listakezelő segítségével történik.

A MEK 1.0-s és 2.0-s szolgáltatása jó pár évig párhuzamosan fog működni – előbbi a helkán, utóbbi a mekmesteren –, és a két gép egymás tükörszervereként szolgál majd. (Ez az egyik irányban már most is így van.) A gopher szolgáltatást megszüntetjük, a legfontosabb felhasználói körnek – a vakoknak – egy egyszerűsített webfelület készül. Az anonim FTP-hozzáférést a 2.0-s verzióban is biztosítjuk majd, elsősorban a részleges vagy teljes tükörzések és az FTPmail szolgáltatást használók érdekében.

Gyűjtőkör, minőség, copyright

A MEK megcélzott felhasználói köre továbbra is az oktatási és tudományos szféra, valamint az otthoni internethasználók és a magyar kultúra iránt érdeklődő külföldiek. Ennek megfelelően meg kell próbálni minél teljesebben szolgáltatni a kötelező és ajánlott irodalmat, tankönyveket és jegyzeteket, az elektronikusan publikált szakirodalmat, valamint a nemzeti kulturális örökségnek azokat a darabjait, melyeknek megléte jogosan elvárható egy, az OSZK neve alatt működő digitális archívumtól. Ugyancsak jó volna elérni a hazai és a határon túli kortárs irodalom csaknem teljes körű bemutatását, legalább egy-két művel minden író és költő esetében. Külön hangsúlyt kell fektetni a magyar irodalom idegen nyelvű fordításainak közzétételére. Ki lehetne bővíteni a gyűjtött dokumentumok körét is: a kották és térképek mellett szóba jöhetnek majd muzeális könyvek faksimile másolatai, képzőművészeti alkotások, régi fotók, hangfelvételek és zenék (komoly- és népzenei művek) is.

Az alapvetően szöveges művek közül a MEK 2.0-ba csak olyan dokumentumtípusokat veszünk fel, amelyek egy hagyományos könyvtár polcaira is felkerülhetnének: vagyis nem gyűjtünk már rövid önálló verseket, novellákat, cikkeket stb., csak verses- és novellásköteteket, valamint hosszabb tanulmányokat. A rövidebb, kevésbé értékes vagy gyorsan elavuló tartalmú írások részben a továbbra is üzemelő 1.0-s MEK-ben maradnak, részben a kifejlesztés alatt levő EPA-ba (Elektronikus Periodika Archivum) kerülnek majd.

A dokumentumok (szöveg)minőségének javítása is fontos szempont. A MEK továbbra sem működhet kiadóként, de mivel az elektronikus publikálásnál gyakran nincs szerkesztőségi vagy kiadói kontroll, legalább a szépirodalmi műveknél meg kell oldani a korrektúrát minden esetben, mielőtt felvennénk a gyűjteménybe. A többi szöveges anyagnál legfeljebb a helyesírás ellenőrzése lehet reális cél. Lehetőség szerint mindenhol az ISO 8859-2 kódtáblát használjuk a magyar ékezetekhez, egyéb nyelvek speciális karaktereinek kódolása az Unicode szabvány szerint történik. A fontosabb szövegformátumokhoz (HTML, Word, PDF) ajánlások készülnek, és az önálló tipográfiával nem rendelkező publikációk külalakját ezekhez igazítjuk. A nem szöveges dokumentumoknál a számítógépen való elfogadható minőségű megjelenítés, illetve lejátszás a fő szempont, nem szükséges olyan nagy méretű és kiváló minőségű digitális másolatok szolgáltatása, amelyek az eredeti művek helyettesítésére is alkalmasak. A gyűjtemény gyarapításában egyre nagyobb arányban kell a megbízhatóbb kiadói forrásokat igénybe venni.

A 2.0-s verzióba már csak olyan művek kerülhetnek, amelyeknek a szerzői jogi helyzete tisztázott. Vagy van írásos (e-mail is elegendő) engedély a MEK-ben való archiválásra, vagy az Artisjuson keresztül jogosítjuk a műveket (jelenleg a szépirodalmat).

A saját gyűjtemény mellett továbbra is gyűjtjük az elektronikus folyóiratok, újságok és magazinok elérési címeit (ez az adatbázis majd beépül az EPA-ba), valamint az interneten található hazai és fontosabb külföldi könyvtári és könyves információforrások URL-jeit. Nem foglalkozunk viszont már egyéb szervereken levő elektronikus dokumentumok linkelésével, ezt a feladatot más szolgáltatók (pl. a Neumann-ház) időközben átvették.

A dokumentumok tárolása

A MEK-be kerülő dokumentumok változatossága (tudományos és szépirodalom, lexikonok, bibliográfiák stb.), különböző eredete (szerzőktől/kiadóktól, más webserverekről, CD-ROM-kiadványokról, szkennelt könyvekből stb.) és a felhasználás eltérő módjai (keresés, kinyomtatás, felolvasatás stb.) miatt nincs olyan egységes és optimális formátum, amibe minden egyes dokumentumot – reális költségekkel – konvertálni lehetne. Ezért a dokumentumot eredeti formátumának és feltételezett felhasználási módjának függvényében általában többféle, lehetőleg nyílt formátumban tároljuk. Ajánlatos egy alapvetően online böngészésre és egy alapvetően letöltésre/nyomtatásra szánt formátumot is feltenni. A támogatott formátumok: HTML, XML, PDF, RTF, PostScript, TeX DVI, JPEG, PNG. Megtűrt típusok: Microsoft Word 6.0, Word 97 és Word 2000, valamint a különböző e-book formátumok és a GIF képfarmátum. Egyszerű ASCII textállományokat már nem szolgáltatunk, kivéve az online felolvasásra szánt könyveknél, de valószínűleg azoknál is inkább speciális tagekkel ellátott XML formátumot használunk majd. A letöltésre szánt állományokat (ha érdemes) ZIP tömörítéssel csomagoljuk.

Az elektronikus dokumentumokat tartalmazó fájlok tárolása egy kétszintes, sorszámozott directory-struktúrában történik. Az első szint legfeljebb 1000 directoryt tartalmaz, melyek mindegyike további 100 alkönyvtárra oszlik. Így maximum 100 000 dokumentum helyezhető el az új MEK-ben. Minden dokumentumhoz tartozik egy *index.html* nevű fájl, amely egyfajta „borítólap”-ként szolgál, és számos funkciója van (többek között tartalmazza a metaadatokat). Lásd például: <http://mek.oszk.hu/html/00012>. A dokumentumot különböző formátumokban, szabványos fájlnevekkel (pl. 00012.rtf, 00012.pdf stb., illetve 00012rtf.zip, 00012html.zip stb.) szintén ide töltik. További kiegészítő állományok is kerülhetnek ezekbe a számozott directorykba, például a könyv szkennelt címlapja vagy a szerző arcképe, egy rövid fűlszöveg, a megjelenítéshez szükséges speciális fontkészletek, az online felolvasatható text vagy XML verzió stb. Ugyancsak ide kerül egy *index.xml* nevű fájl, amely az adatcserére és a Zebrával való visszakeresésre alkalmas XML formátumban tartalmazza a dokumentum bibliográfiai adatait. Ezeket a metaadatokat – kiegészítve a MEK rövid ismertetésével és a felhasználásra vonatkozó szabályokkal – a letölthető ZIP fájlunkba is belesomagoljuk

(mek.html néven), hogy együtt „utazzanak” a dokumentummal.

Katalogizálás

A dokumentumokról a jelenleginél részletesebb, de alapvetően egy egyszerűsített könyvtári címléírás készül, és ezeket az adatokat egy adatbázisban tároljuk. Legfeljebb 18-féle adatszoportot rögzítünk (pl. Cím, Szerző, Dátum, Kiadó, Tárgyszó), ezek többsége további adatelemekre bomlik (pl. Főcím, Alcím, Sorozati cím). Valamennyi adatszoport ismételhető szükség esetén. Egyes adatelemek értéke (pl. Nyelv, Típus) egy kötött listából választható ki, mások (pl. Szerző, Tárgyszó) szabadon kitölthetők, de ajánlott megnézni a már felvitt adatokat tartalmazó indexet, és az egységesítés érdekében lehetőség szerint onnan átvenni az adatot. Nem építünk kötött, besorolási (authority) állományokat sem a név-, sem a tárgyszóelemeknél. A dokumentumok téma szerinti osztályozása kétféleképpen történik: egyrészt mindegyiket besoroljuk legalább egy főtemakörbe, témakörbe és altémakörbe (ezek listája kötött), másrészt mindegyik kap legalább egy tematikus tárgyszót, esetleg egy földrajzi és idő tárgyszót is. A tárgyszavakat lehetőség szerint az OSZK Tezauruszából választjuk (ha abban nincs megfelelő, akkor a Matrikszból: <http://www.matriksz.hu> vagy a MEK régi tárgyszólistájából). A bibliográfiai adatok részben megfeleltethetők a Dublin Core-ajánlásnak, és ezek a metaadatok bekerülnek a dokumentumhoz tartozó index.html és index.xml fájlokba is a keresőrendszerek számára. Továbbá exportálhatók, illetve megjeleníthetők lesznek majd az ISBD, USMARC, HUNMARC, XML, valamint a magyar mezőneveket mutató „címkés” formátumokban. A tervek szerint a MEK gyűjteményének legjavát – kibővített leírással – az OSZK saját, Amicus rendszerében is katalogizálják majd, az egyszerűsített bibliográfiai rekordok pedig bekerülhetnek az ODR, a Közélat, a MOKKA, a WebKat.hu és más hasonló, nagy adatbázisokba is.

Szolgáltatások

Az új MEK felhasználói felületét az olvasóktól kapott visszajelzések és az eddigi tapasztalatok alapján alakítjuk ki. A középpontba a böngészhető és kereshető katalógus kerül, a kezdőlapról egyetlen kattintással a MEK minden része elérhetővé válik (erősen lecsökkentjük a menük „mélységét”). Egy egyszerű keresés már a nyitóoldalról is

kérhető, és a dokumentumok „borítólapjáról” is indíthatók majd katalóguskeresések, az összetett, minden szempontra kiterjedő kereséshez külön oldal készül (itt repülő ékezetek, Boole-operátorok, böngészhető indexek is használhatók majd). A találati listák ékezethelyesen rendezhetők lesznek cím, szerző és dátum alapján. A HTML (és ha megoldható, akkor az RTF és PDF) formátumú dokumentumok teljes szöveggel is kereshetők lesznek, a találati listákban szöveggörnyezetben jelennek meg a keresett szavak. A dokumentumok „borítólapjáról” keresések indíthatók hagyományos könyvtári és könyvkereskedői adatbázisokban.

A webfelület minden eleme kétnyelvű lesz, általában a magyar feliratok felett jelenik meg az angol fordítás. Külön hangsúlyt kell fektetni arra, hogy a vak felhasználók is könnyen tudjanak navigálni a rendszerben. Fontos a gyors letöltődés is, a felhasználói felületet alkotó weblapok egyike sem lehet nagyobb 100 kb-otnál (képekkel együtt).

Néhány további szolgáltatás: vendégkönyv, olvasói vélemények az egyes művekhez, részletes letöltési statisztikák, sikerlisták, hírlevél, irattár, 3 dimenziós VRML-modell, virtuális kiállítások, ajánlott információforrások minden témakörben, LibInfo (korábban: MIT-HOL) tájékoztató szolgálat.

Lehetséges egyéb fejlesztések: online felolvasó szoftver, a dokumentumok hitelesítése digitális aláírással, stabil DOI/URN címek használata (az OSZK tervezett URN-szervere segítségével), a MEK böngésző és kereső funkcióinak integrálása az EPA-val (egyidejű keresés a könyvek és a folyóiratok között), közös keresőfelület más elektronikus könyvtárakkal, linkek és egyéb funkciók működésének automatikus ellenőrzése (minőségbiztosítás), nyilvántartó rendszer az országban folyó digitalizálási munkákról, olvasói fórumok, automatikus konvertálók a formátumok között (főleg az e-book készülékekhez), a webszerver forgalmának auditálása.

Jelenleg a dokumentumok átválogatása, egységesítése és katalogizálása folyik, valamint készül a visszakereső rendszer. A 2.0-s verzió valószínűleg a jövő év elején nyílik meg a nagyközönség számára, de még további hónapok kellene a teljes kiépítéséhez. Ha minden tervbe vett fejlesztés elkészül, akkor világviszonylatban is korszerű és tartalmas lesz a Magyar Elektronikus Könyvtár. Mi pedig elkezdhetjük a MEK 3.0 tervezését.

Beérkezett: 2002. VI. 10-én.