

Borsos Ildikó

Digitalizált Törvényhozási Tudástár – az Országgyűlési Könyvtár digitalizálási projektje

„A digitális könyvtár álma nem azért kerül újra és újra előtérbe, mert izgalmas jövőt ígér, hanem azért, mert egy, a múlthoz teljesen hasonló, ám annál jobb és gyorsabb jövőt ígér.”¹ Többek között ez, és a múlthoz való gyorsabb és jobb hozzáférés megoldásának szándéka vezette az Országgyűlési Könyvtár vezetőit is akkor, amikor megálmodták a Digitális Törvényhozási Tudástárat; egy olyan digitális gyűjteményt, amely gyorsan, hatékonyan, az érdeklődők széles köre számára egy időben tud hozzáférést nyújtani jogforrásokhoz, parlamenti dokumentumokhoz, szakirodalomhoz, megkönnyítve ezzel a jogalkotási tevékenységet, az oktató- és kutatómunkát, a könyvtári tájékoztató tevékenységet, vagy akár a Magyar Országgyűlés tevékenységével kapcsolatos érdeklődő olvasgatást.

Előzmények

Az Országgyűlési Könyvtár 2001 óta foglalkozik programszerűen könyvtári dokumentumok digitalizálásával. A kiemelten védett könyvtári állományrészek digitalizálása mintegy 600 000 oldalnyi országgyűlési forrásdokumentum feldolgozásával kezdődött meg. 2009-re elkészült

- az 1860–1990 között megjelent hivatalos magyar, nyomtatott parlamenti dokumentumok (naplók és irományok),
- az 1884 és 1949 között 18 kötetben megjelent országgyűlési almanachok, valamint
- az 1848 és 1990 közötti időszakra vonatkozó országgyűlési házszabályok

köteteinek digitalizált változata, és digitalizálták több, az Országgház történetével foglalkozó művet is².

A DTT projekt célkitűzései

A Digitalizált Törvényhozási Tudástár (DTT) EKOP-2.1.5-09-2009-0001 elnevezésű projekt (1. ábra) az Új Magyarország Fejlesztési Terv (ÚMFT) Elektronikus Közigazgatás Operatív Programja (EKOP) keretében, az Európai Unió támogatásával és az Európai Regionális Fejlesztési Alap társfinanszírozásával, 2010. január 4. és 2011. június 30. között valósul meg. Célja mintegy 2 millió oldalnyi történeti országgyűlési dokumentum, jogforrás (közlönyök, döntvények), valamint jogi, törté-

nettudományi, politikai tárgyú könyv és időszaki kiadvány digitalizálása, adatbázisba szervezése és integrált felületű portálon való szolgáltatása.



1. ábra A DTT projekt honlapja: <http://www.ogyk.hu/dtt>

A létrejövő parlamenti tudástár hatékonyan kívánja segíteni az országgyűlésben folyó jogalkotási munkát, a jogalkotás háttérében folyó szakmai tevékenységet, az oktató- és kutatómunkát, és támogatni kívánja a könyvtári tájékoztató munkát.

A szerzői jogi szempontból szabad felhasználásúnak tekinthető dokumentumok teljes körűen és térítésmentesen hozzáférhetővé válnak majd az interneten. A szerzői jog védelme alatt álló dokumentumok dedikált könyvtári hálózaton keresztül lesznek elérhetők a magyarországi könyvtárak és felhasználók számára.

Tervezés

Az alábbiakban, mintegy esettanulmányként, saját példánkon keresztül megpróbáljuk bemutatni egy könyvtári digitalizálási projekt tervezési szakaszának egyes mérföldköveit, az átgondolásra váró kérdéseket, szempontokat³.

A digitalizálandó dokumentumok kiválasztása

A kiválasztás általános alapelveinek meghatározása

A tervezési folyamat egyik első lépéseként meghatároztuk a digitalizálásra javasolt művek *általános kiválasztási szempontjait*. Esetünkben ezek a tudományterületek, a dokumentumtípusok, a nyelv, az idő, a kiadások, a példányok, valamint a dokumentumok teljessége:

- a) **Tudományterületek:** főszabályként a könyvtár fő gyűjtőkori szakterületeiről történt a dokumentumok kiválasztása. Ehhez ún. szakterületi munkacsoportokat állítottunk fel. A munkacsoportok először meghatározták a dokumentumok kiválasztásának adott szakterületi szempontjait, majd összeállították a dokumentumlistákat. A digitalizálandó dokumentumok a jogtudomány, a politikatudomány és a legújabb kori történettudomány területéről kerültek ki.
- b) **Nyelv:** alapértelmezésben magyar nyelvű dokumentumokat választottunk ki, amelyek azonban tartalmazhatnak latin, német, francia, angol nyelvű szövegrészeket is.
- c) **Idő:** az 1945 előtt megjelent szakirodalomból igyekeztünk minden, általunk fontosnak ítélt szakirodalmi tételt a listákra tenni, míg az 1945 után megjelent szakirodalomból válogattunk.
- d) **Dokumentumtípusok:** könyvek (monográfiák, tanulmánykötetek, lexikonok, adattárak stb.), időszaki kiadványok, jogszabályok. A dokumentumtípusok különös ágát képezik az országgyűlési dokumentumok (almanachok, lakáskönyvek, költségvetések és zárszámadások, közös ügyi iratok stb.)
- e) **Kiadások:** fő szabály szerint minden mű utolsó kiadását digitalizálja a könyvtár; a művek előző kiadásai tartaléklistára kerültek. A változatlan kiadások közül a példányok állapota alapján az ún. *Állapotfelmérő munkacsoport* tett javaslatot arra, hogy melyik kötetet digitalizálja a könyvtár.
- f) **Példányok:** a művek bibliográfiai ellenőrzése során megtörtént a fellelhető példányok és kiadások azonosítása, majd a munkacsoport ja-

vaslatot tett arra, hogy a könyvtárban elérhető különböző példányok közül melyiket digitalizáljuk.

A kockázati terv részeként a szakterületi munkacsoportok tartalék listákat is készítettek, amelyekre a szakmailag javasolt, de valamiért nem digitalizálható művek kerülhettek, olyanok például, amelyekből célszerű lenne később másik példányt beszerezni, vagy a legutolsó kiadást megelőző kiadások.

A dokumentumlisták kialakításának további szempontjai:

- digitalizálták-e már valahol a projekt keretében feldolgozni javasolt műveket;
- van-e az eljárásra javasolt dokumentumok között olyan kötet vagy dokumentumrész, melynek indokolt és célszerű a speciális feldolgozása (l. adattárak, lexikonok, kronológiák, jogszabálymutatók stb.);
- külön listán célszerű összegyűjteni azokat a dokumentumcsoportokat és egyedi műveket, amelyek szétszedhetők, felvághatóak. A könyvtár összes példányának egybevetését követően egyedi döntés szükséges egy-egy dokumentum lapokra vágásáról. Ez rendkívül kényes feladat, mert fontos művek egyes példányainak jövőjéről kell felelős döntést hozni. A döntés kihat az adott mű példányának sorsára, de egyben a mű valamennyi további példányát – országosan is – védi;
- össze kell gyűjteni azokat a dokumentumcsoportokat és egyedi műveket, amelyeket nem célszerű (nem szabad) elszállítani a könyvtárból, és helyben is csak szigorúan meghatározott módon, feltételekkel szkennelhetők.

Szakterületi listák

Országgyűlési forrásdokumentumok

Az országgyűlési forrásdokumentumok jelentős részének digitalizálása már korábban (2008–2009) megtörtént (l. naplók, irományok, házszabályok, almanachok). A projekt keretében a következőket digitalizálták:

- a közös ügyi – ún. delegációs – dokumentumok (jegyzőkönyvek, naplók, irományok, 1867–1918),
- költségvetések és zárszámadások (közös ügyi 1868–1918, Magyarország 1968–1988),
- az országgyűlés képviselőinek lakáskönyvei.

Országgyűlési szakirodalom

Az országgyűlésre vonatkozó monografikus szakirodalmat a DTT projekt keretében teljességre törekvően kívánja digitalizálni a könyvtár. Az országgyűlési szakirodalmi lista összeállításának szempontjai a következők voltak:

- átfogó országgyűlés-történeti művek, a magyar parlamentarizmus történeti korszakainak irodalma;
- az országgyűlésekről szóló művek speciális műfajai (országgyűlési beszédek, kortesbeszédek, a választók előtt mondott programok, képviselői beszámolók, emlékiratok, visszaemlékezések);
- az országgyűlési pártokról szóló művek;
- az országgyűlési tisztségviselőkről, és a tisztviselőkről szóló művek;
- kiemelkedő törvényjavaslatok története (választójog, felsőház reformja);
- az országgyűlés résztvevőiről szóló (szép)irodalmi művek.

Állam- és jogtudomány

Szakirodalom

A jog- és az államtudományi szakirodalom kiválasztásának legfontosabb szempontja az volt, hogy a kiválasztott dokumentumok mutassák be a magyar jogtudomány előzményeit, annak fejlődését, és jeles képviselőinek munkássága által fejezzék ki a jogi nyelv és a jogász gondolkodásmód folyamatosságát. Az Országgyűlési Könyvtár jogi gyűjteményének teljességére figyelemmel ezek leginkább a két világháború közötti jogi irodalomban, jogi érvelésben és fogalomhasználatban öltének testet, ezért tartalom szempontjából ezek digitalizálása a legindokoltabb. A jogi szakirodalmat az adott szakterületi munkacsoport az egyes jogágak szerint készítette elő digitalizálásra. Válogatásuk alapelve az volt, hogy a témakör átfogó, összefoglaló műveit gyűjtsék össze, ezért a kisebb terjedelmű, esetlegesen aktuálpolitikai hatásoktól befolyásolt írások kimaradtak a válogatás jelen szakaszából. A jogi szakirodalmi dokumentumlista által felölelt jogágak a következők: közigazgatási jog, közjog, alkotmányjog; magánjog, polgári jog; büntetőjog; igazságügyi szervezet (a bíróságokra, az ügyvédségre, az ügyészségre és a közjegyzőkre vonatkozó legfontosabb, elsősorban átfogó jellegű jogi szakirodalom).

Jogforrások

A magyar hivatalos közlöny történeti fejlődése végig követhető lesz a készülő adatbázisban:

az 1848. június 8-án induló *Közlöny* című kiadványtól kezdve a mai digitalizált *Magyar Közlöny*ig. A tervezett digitális gyűjtemény tartalmazza majd a minisztériumok hivatalos közlönyeit is 1874-től az 1950-es évek végéig.

Döntvénytárak

A döntvénytárak közül a XVIII. századtól 1952-ig megjelent köteteket digitalizálják (teljességre törekvően).

Időszaki kiadványok

Az 1945 előtt megjelent jogi időszaki kiadványok teljességre törekvően, míg az 1945 utáni kiadványokból csak a legjelentősebbeket választottuk ki.

Politika- és történettudomány

E szakterületeken fő válogatási elvként a politikát *elemző igényvel vizsgáló* művek kerültek a listára (államelmélet, hatalomelmélet, demokrácia, választójog és a közigazgatás-elmélet). A cél minden digitalizálásra kerülő magyar politikatörténeti vonatkozású dokumentum esetében az, hogy az adott mű bemutassa az adott korszak parlamenti munkáját, légkörét; a parlamenti képviselők, illetve jelentősebb politikusok tevékenységét, életútját; a parlament törvényhozási tevékenységét. A dokumentumok tartalma mellett jelentősége van a szerző személyének is: az adott kor politikai életében jelentős szerepet játszó személyiségek politikatörténeti vonatkozású művei nagyobb arányban kerültek be a kiválasztott dokumentumok közé.

Műfajukat tekintve a kiválasztott művek között – terjedelemtől függetlenül – tudományos igényű szakmunkák, visszaemlékezések, korrajzok, életrajzok, jellemrajzok, röpiratok és a publicisztikák is találhatóak.

A digitalizálásra kiválasztott dokumentumok listáját a <http://dti.ogyk.hu/lapok/mit.aspx> oldalon tettük közzé.

A kötetek előzetes állapotfelmérése

A dokumentumok nagy része 1851 és 1945 között jelent meg, ennél fogva a kötetekben gyakran fellelhetők dőlt betűs, gót betűs és egyedi betűtípusú szövegrészek, egy oldalon különböző betűméretű, vastagságú, eltérő szövegirányú, többhasábos szövegek, hivatkozások is. A kötetek fényképeket, térképeket, táblázatokat, grafikonokat is tartalmaz

hatnak, és előfordulnak közöttük kihajtható, a kötetből eltérő méretű vagy különálló mellékletek is.

A digitalizálás folyamatát ezek a jellegzetességek nagyban befolyásolják, ugyanúgy, mint a kötetek nyithatósága (a kötetek egy része nem nyitható 180 fokban); az átnyomódás egyes szövegoldalakon; a fátolypapírral fedett oldalak; aláhúzások a szövegben; olvasói bejegyzések; tévesen elhelyezett oldalak, a folyóiratokban helytelen sorrendben bekötött borítók, mellékletek stb.

A kötetek állapotfelmérésének vizsgálatára hoztuk létre az *Állapotfelmérő munkacsoportot*, amelynek fő feladata az, hogy megvizsgálja és jellemezze a kiválasztott kötetek állapotát. A munkacsoport számos alapvető határozott meg, amelyek az állományvédelemre, a kötészeti adatokra, a szövegtükör és a könyvtest vizsgálatára vonatkozóan tartalmaztak alapvető szempontokat, amelyeket egy ún. *Állapotfelmérő adatlapban (Kísérőlap)* állítottak össze. Ez tartalmazza a dokumentumok terjedelmi, szerkezeti és az átadás idején jellemző állapotának legfontosabb ismérveit, az alkalmazni javasolt digitalizálás technológiát, valamint könyvek esetében az adott mű szerzői jogi vonatkozásait is. A könyvek esetében egyedi; az időszaki kiadványok, a közlönyök és az országgyűlési forrásdokumentumok esetében csoportos állapotfelmérő kísérőlap készült.

Az így összegyűjtött információkból egy *háttéradatbázist* hoztunk létre, amelynek – a nyilvántartáson túl, – számos egyéb haszna is van: például könnyen leválogathatók belőle az azonos módon digitalizálható dokumentumok, használatával egyszerűbbé válik az átadás-átvételi jegyzőkönyvek előkészítése.

A példányok vizsgálata

Az állapotfelmérés következő lépéseként a kiválasztott művek könyvtárunkban található példányait gyűjtöttük és hasonlítottuk össze egymással a hiányos, rongált, elveszett stb. kiszűrése és az adott mű digitalizálendő példányának kiválasztása céljából.

A kötés vizsgálata

Megvizsgáltuk, hogy a kötés eredeti vagy újrakötött, ép vagy sérült. Vizsgáltuk a kötetek nyithatóságát és a belső margó méreteit is. Erre egyrészt azért volt szükség, hogy meg tudjuk határozni a digitalizálás módját (ha nem nyitható az adott kö-

tet, csak könyvbölcsőben lehet digitalizálni), másrészt hogy lássuk, digitalizálás után mely köteteket kell javításra, esetleg újrakötésre küldeni (a belső margó mérete meghatározza, hogy a mű újraköthető-e vagy sem). Archiválási szempontból ugyanakkor – a szöveg megőrzése mellett – fontos a könyvkötés művészeti, kultúrtörténeti jelentősége is. Erre tekintettel úgy döntöttünk, hogy az eredeti borítót, kötetst minden esetben színesben, színhelyesen kell szkennelni.

A könyvtest vizsgálata

Az állapotfelmérés során rögzítettük a kötetek terjedelmi adatait is, hiszen a mennyiségi adatok meghatározók a további feladatok és a tárolási kapacitások tervezésében, az archiválásban, a digitalizált állományok fájljainak elnevezésében, továbbá az ellenőrzéseknél. Adatokat gyűjtöttünk a dokumentumok méretére vonatkozóan is: fontos tudnunk, hogy milyen számban szerepelnek listáinkon az átlagostól eltérő méretű dokumentumok vagy eltérő méretű oldalakat (pl. kihajtható mellékleteket) tartalmazó kötetek.

A későbbi digitalizálás szempontjából kiemelten fontos, hogy teljes, hiánytalan példányok kerüljenek a végső listákra; kiválogatásuk viszont időigényes feladat (a dokumentumok teljességének ellenőrzése, a hiányok összegyűjtése, pótlása, a téves oldalszámok, téves kötészeti sorrend feltárása stb.). Az itt befektetett munka viszont később sok kellemetlenségtől kímélheti meg a digitalizálókat.

A dokumentumokban – a szövegben és a mellékletekben is – gyakoriak a színes vagy fekete-fehér illusztrációk (fénykép, táblázat, grafikon, térkép, grafika). Feltárásuk, megjelölésük azért fontos, mert ezekre fokozottan figyelnie kell a digitalizálást végző munkatársnak.

Külön figyelmet igényelnek a kötetekben előforduló mellékletek, amelyek jelzése (számozása, elnevezése), terjedelme, helye, mérete rendkívül eltérő lehet a digitalizálásra kiválasztott kötetekben. Nagyon bonyolult szerkezetű kötet esetében szükség lehet akár a kötet ceruzás végigszámozására is. A mellékletek helyét úgy adtuk meg, hogy a kötet helyreállításakor eredeti helye egyértelműen meghatározható legyen.

A papír vizsgálata

A papír minősége és állapota további szempont a digitalizálás módjának meghatározásához. Kiemel-

ten fontos a további roncsolódás megakadályozása. Ennek kapcsán vizsgáltuk, hogy a kötés szálirányban történt-e, milyenek a papír szélei, mennyire töredezett a papír, lehet-e lapozni a kötetet, mennyire törnek, sérülnek a lapok, van-e a papíron bármilyen elszíneződés, mekkora ezek aránya a kötetben stb.

A szövegtükör vizsgálata

A digitalizálás teljes munkafolyamatában alapvetően fontos, hogy karakterfelismerést követően a szöveg szavaira keresni lehessen, a szöveg minősége pedig alapvetően befolyásolja a karakterfelismerés eredményét. Ezért nagy hangsúlyt fektettünk a szövegtükör vizsgálatára, mivel minőségi elvárásainkat csak az egyes kötetekből kiindulva lehet megfogalmazni. Vizsgáltuk a kontrasztot: ha a betűk kontúrja elmosódott, ha a kötetben található nagyon halvány szövegrészek, ha zavaros a szöveggép, ha a szövegtükör átnyomódott a következő oldalra, vagy ha az oldalak egy része japán fátypapírral kasírozott, akkor a karakterfelismerés minősége romlik. Befolyásolja (nehezíti) a karakterfelismerés minőségét az is, ha a szövegtükör az átlagostól eltérő méretű betűt (sokkal nagyobb vagy sokkal kisebb betűket) tartalmaz, esetleg eltérő szövegírányú szövegrészek (pl. függőleges feliratozás) fordulnak elő a szövegben. Megnéztük, hogy az oldalakon van-e folt, lyuk vagy szakadás, ha igen, mekkora, s hogy lehet-e, érdemes-e javítani. Az állapotfelmérő lapon rögzítettük, ha a kötetekben bejegyzéseket találtunk: mekkora a kézírást szövegrészek aránya; tartalmaz-e a kötet tulajdonosi (possessor) bejegyzéseket a szöveg között. Az aláhúzások és a firkált szövegrészek esetében vizsgáltuk, hogy tollal vagy ceruzával készültek-e, milyen mértékben, eltávolíthatók-e (radírozhatók-e), esetleg szükség van-e a kötet tisztítására.

A kötetekről megállapított minden fontos információt rögzítettünk az állapotfelmérő lapon, majd a háttéradatbázisban.

Az informatikai háttér tervezése

A projekt informatikai környezetének kialakításában meghatározó fontosságú a megfelelő szoftver(ek) kiválasztása, a hardverszükséglet tervezése és az archiválás részleteinek kidolgozása. A digitalizálandó dokumentumok paramétereinek ismeretében lehetőség van az informatikai eszközrendszer, a szükséges háttérkapacitások meghatá-

rozására. A feladatok fontos része a beszerzéseket érintő közbeszerzési eljárások részeként az ún. *műszaki specifikáció* pontos kidolgozása.

A digitalizálási folyamat lényeges eleme, a digitalizált állomány archiválási és reprodukálhatósági, azaz hosszú távú megőrzési és szolgáltatási feladatainak megoldása. Ennek során megterveztük az archiválandó állomány fájlstruktúráját, kialakítottuk a tárolási fájlformátumokat, meghatároztuk az adathordozók típusát, és kialakítottuk a digitális raktározás szabályait. Az archiválási eljárás kialakítása során kidolgoztuk azt a mentési technológiát is, amely hosszú távon alkalmas a digitális képek és a kétrétegű PDF fájlok archív és munkaállományainak megőrzésére, illetve újbóli előállíthatóságára.

A dokumentumok mozgatása, szállítása, újrakötése

Már a tervezés időszakában is, a projekt végrehajtása során pedig több alkalommal több ezer kötetet kell megmozgatni: levenni a polcra, az adott művelet helyszínére szállítani, esetleg tovább egy újabb helyszínre, majd visszatenni a polcra. Ennek előkészítése és koordinálása külön logisztikai tervezést igényel, melynek átgondolásakor a következőkre voltunk figyelemmel:

- A kötetek mozgatására több alkalommal szükség lesz: a digitalizálásra javasolt művek kiválasztása, az állapotfelmérés elkészítése az elszállításra való előkészítés idején, míg végül a kötetek visszakerülnek eredeti helyükre.
- Az el- és a visszaszállítások tervezhetősége és ellenőrizhetősége, a teljes folyamat nyomon követhetőségének megoldása érdekében célszerű egy *átadás-átvételi jegyzőkönyvet* készíteni és használni. (Ez különösen akkor fontos, ha a digitalizálást végző vállalkozó külső helyszínen dolgozik.) A jegyzőkönyvek mellékletei azok a *kísérőlapok*, amelyek tartalmazzák az egyes kötetekre vonatkozó terjedelmi, szerkezeti, valamint állapotinformációkat. A visszaszállítást követően a dokumentumok állapotát a kísérőlap tartalma (megállapításai) alapján lehet ellenőrizni.
- A szállításokhoz, a dokumentumok külső helyszínen történő tárolásához és digitalizálásához állományvédelmi szabályokat alkottunk, melyek fontosságát az is mutatja, hogy előírásaink a közbeszerzési dokumentáció részét képezték.
- Meghatároztuk az újrakötendő művek kötészeti adatait is (mivel bizonyos típusú dokumentumok szkennelésének feltétele a kötetek lapra vágása), vagyis pontosan előírtuk, hogy hogyan és

milyen anyagok felhasználásával kérjük a könyvtári kötetek újrakötését. (Ez is a közbeszerzési dokumentáció része lett.)

Szerzői jogi kérdések

A projekt során felmerülő szerzői jogi problémák feltárására és azok kezelési elveinek kidolgozására egy ún. Szerzői jogi munkacsoportot állítottunk fel. Első feladatuk az volt, hogy megvizsgálják, a különböző dokumentumcsoportok esetében (monográfiák, tanulmánykötetek, folyóiratok, közlönyök, forráskiadások, szöveggyűjtemények, lexikonok stb.) milyen típusú szerzői jogi problémák merülhetnek fel. Ezt követően kezdtek hozzá az egyes művek jogi státuszának meghatározásához:

- ki a szerző (szerző, társszerző, szerkesztő, fordító stb.),
- mi a mű keletkezésének időpontja,
- a védelmi idő alapján fennáll-e a szerzői jogi oltalom, s ha igen, meddig.

Az egyes művek szerzői jogi vonatkozásait az adott dokumentum kísérőlapján, illetve a háttéradatbázisban is rögzítették.

A digitalizálás

Az eddig ismertetett feladatok elvégzése után a digitalizálásra kiválasztott köteteket – kísérőlapjukkal együtt – az elszállításra előkészítjük.

A dokumentumok digitalizálását és adatbázisba szervezését nyílt közbeszerzés keretében kiválasztott vállalkozóval végeztetjük. A vállalkozó által elvégzendő munkafolyamat a következő lépésekből áll össze:

1. A folyamat során először szkenneléssel digitális képek készülnek a kötetekről.

Ehhez meghatároztuk a szkennelés általános követelményeit, elsődlegesen azt, hogy a szkennelés borítótól borítóig tart, azaz az adott kötet összes oldalát rögzíteni kell, az üres oldalakat is beleértve. Az oldalkép minőségére vonatkozóan számos előírást rögzítettünk: egy kép csak egy dokumentumoldalt tartalmazhat, a kép nem lehet szerkesztett, összeillesztett, a képnek a teljes oldalt kell tartalmaznia stb.

A *digitalizálási technológia* megválasztásában saját dokumentumaink maximális állapotvédel-

mére törekedve háromféle digitalizálási módot határoztunk meg:

- a dokumentum *fizikai egységének megőrzésével* történő digitalizálás: a dokumentumot állománykímélő technológia alkalmazásával, kötetben kell digitalizálni, a nyithatóság minél kisebb dőlésszöge mellett;
- a dokumentum *fizikai egységének részleges megőrzésével* történő digitalizálás: a dokumentum kiemelhető a kötéséből, de ívei nem vágathatók fel;
- a dokumentum *fizikai egységének megőrzése nélkül* történő digitalizálás: a dokumentum lapra szedhető, laponként digitalizálható.

2. Az így elkészült képeket OCR (*Optical Character Recognition = optikai karakterfelismerés*) technológiával szöveggé, vagyis a betűket és számokat szövegfájlá kell alakítani.

A közbeszerzési dokumentáció egyik legfontosabb része az OCR-rel szemben támasztott minőségi követelmények meghatározása. A szoftveres karakterfelismerés eredményeként létrejött szöveg emberi beavatkozással történő további javítása ugrásszerűen megnöveli a költségeket. Az elmúlt időszak digitalizálási tapasztalatai azt mutatták, hogy nem lehet általánosságban minőségi követelményeket meghatározni: a digitalizálásra átadott dokumentumok minősége alapvetően meghatározza a szövegfelismerés során elérhető minőséget. Ezért minimálisan dokumentumtípusra, de egyes esetekben a konkrét dokumentumok szintjéig meg kell határoznunk a szövegfelismerés minőségére vonatkozó követelményeinket. (Mi bizonyos dokumentumtípusok egyes adatainál közel 100%-os szövegfelismerést írtunk elő).

3. A képek és a szöveg összekapcsolásával *kétretegű PDF* állományok készülnek.

Ez a dokumentumok teljes szövegében való visszakeresés lehetőségét teszi lehetővé. A „*felső rétegben*” az eredeti dokumentum képe lesz látható, míg az „*alsó rétegben*” (mintegy „*mögötte*”) az optikai karakterfelismeréssel kereshetővé tett szöveg.

4. A digitális dokumentumok adatbázisba szervezése.

A könyvtár a könyvtári feladatok ellátására jelenleg az ALEPH 500 V.18.01 verzióját használja; annak katalogizálási, beszerzési, WEB-OPAC,

folyóirat-kezelő, kölcsönzési és digitális gyűjteménykezelő modulját alkalmazza. Az integrált könyvtári rendszer RED HAT Enterprise Linux operációs rendszer alatt fut.

A projekt keretében létrejövő digitális rekordok adatbázisba szervezését és a digitális gyűjtemény kezelését a *DigiTool* szoftver alkalmazásával kívánjuk megoldani. A *DigiTool*, amely szöveges, audio- és videofájlok kezelésére is alkalmas, minden szükséges eszközt rendelkezésre bocsát a digitális gyűjtemények létrehozására, archiválására, keresésére, menedzselésére és megjelenítésére.

A digitális dokumentumokat tehát a vállalkozó rögzíti a *DigiTool*-ban, majd a Dublin Core ajánlásai alapján a *Meta-munkacsoportunk* által kialakított formában elkészíti a metaadatokat. (Figyelembe véve az ESE [Europeana Semantic Elements] V.3.2.2⁴ előírásait – mivel a pályázati anyagban is vállalt szándékunk, hogy a digitális gyűjteményt arathatóvá tesszük az Europeana számára.)

A digitális rekordokat bekötjük az Aleph katalógusba (könyvek), illetve az ugyancsak Aleph környezetben elérhető szakirodalom adatbázis rekordjainál is rögzítjük a digitális objektum adatait.

Tesztelések, ellenőrzések

A projekt egyes feladatait mennyiségi és minőségi szempontból több ponton is ellenőrizni kell:

- a digitalizált képeket (mennyiségük és minőségük);
- a digitális szövegeket (az eredeti és a felismert szöveg azonossága) mintavétellel, dokumentumcsoportonként;
- az oldalak sorrendjét az egyes kötetekben (= fájlokban);
- a fájlneveket az egyes kötetekhez kapcsolódóan;
- az adatbázisba szervezés eredményeit;
- a visszaszállított kötetek állapotát, az új köteteket;
- végül a portál funkcióit (komplex tesztelés: keresés, böngészés, megjelenítés, letöltés, nyomtatás, jogosultságok stb.).

Tartalomszolgáltatás

A digitális gyűjteményt várhatóan 2011 nyarán tesszük közzé. A rekordok gyűjtemény(ek)be szer-

vezését, a tartalomszolgáltatás formáját, a keresési lehetőségeket és módokat, a létrejövő *Parlament* *Tudástár* portál szerkezetét és felhasználói interfészét az Országgyűlési Könyvtár munkatársai alakult *Tartalomszolgáltatási munkacsoport* készíti elő.

A jogvédelem alatt nem álló dokumentumokat az új portálon ingyenesen, bárki számára hozzáférhetően tesszük közzé az interneten.

A gyűjtemény szerzői jog által védett tartalmát – a szerzői jogi törvény adta lehetőséget kihasználva – *dedikált hálózaton* lehet majd elérni. Ehhez várhatóan nyilvános szolgáltatásokat nyújtó könyvtárak, oktatási intézmények, múzeumok, levéltárak, kép- és hangarchívumok csatlakozhatnak. A dedikált gépeken elérhető dokumentumok esetében nem engedélyezhetők: a letöltés, a nyomtatás, a mentés, az e-mailben történő küldés, csak az olvasás.

Zárszó

Az európai kultúrában és társadalmakban már régóta jelentőségének megfelelő hangsúllyal kezelik azt a tényt, hogy a parlamentek történetéhez, működésük szabályrendszeréhez és funkcióihoz kapcsolódó ismeretanyag és tudás a nemzeti történelem és a politikai kultúra szerves része. Ezért nemcsak a múlt megértése, hanem a jövő formálása szempontjából is megkülönböztetett figyelmet fordítanak arra, hogy ez az ismeret- és tudáshalmaz hozzáférhető legyen, reális esélyt biztosítva ezzel arra, hogy ezen információk – a család, az oktatás, a képzés, a kultúra közvetítésével – befolyást gyakoroljanak az egyének és a közösségek életére. Most induló projektünkkel erre törekszünk mi is.

Jegyzetek és hivatkozások

¹ James J. O'Donnell: *Avatars of the Word*. Idézi: *Abby Smith: Why digitize? Council on Library and Information Resources. Washington, D.C. February 1999.*

<http://www.clir.org/pubs/reports/pub80-smith/pub80.pdf>

² Elérhetőségük:

[http://www.ogyk.hu/e-](http://www.ogyk.hu/e-konyvt/mpgy/dokumentumok.html)

[konyvt/mpgy/dokumentumok.html](http://www.ogyk.hu/e-konyvt/mpgy/dokumentumok.html),

illetve

<http://mpgy.ogyk.hu/>

³ Terveink szerint, később, a projekt előrehaladásával párhuzamosan, a digitalizálás, majd az adatbázisba-szervezés és publikálás egymásra épülő lépéseit, szintjeit is hasonló cikkben követjük nyomon.

⁴ Europeana Semantic Elements specifications v.3.2.2:
<http://version1.europeana.eu/web/guest/technical-requirements/>

Irodalom

ANGEVAARE, Inge: Digitális gyűjtemények és adatok felügyelete: „kurátorság” és más alternatívák a tudományos könyvtárakban. = TMT, 56. köt. 10. sz. 2009. p. 495–498.

BAKONYI Géza –KOKAS Károly: Bevezetés a könyvtári informatika alapjaiba. Szeged, JATEPress, 2006. (Habent sua fata libelli, 2.) ISBN 963-482-764-0, p. 15–29, 167–180.

Calimera Digitális Útmutató. A Calimera project az Európai Unió IST programja keretében létrehozott együttműködési kezdeményezés. =
<http://www.ki.oszk.hu/old/calimera/>

CHAPMAN, Stephen: A szövegdigitalizálás menedzselése. = TMT, 50. köt. 8. sz. 2003. p. 325–330.

GOLDEN Dániel – TÓTH Tünde –TURI László: Virtuális örökkévalóság: objektumok a digitális könyvtárban. = Palimpszeszt 10. sz. 1998.
http://magyar-irodalom.elte.hu/palimpszeszt/10_szam/20.htm

HORVÁTH Tibor – PAPP István (szerk.): Könyvtárosok kézikönyve. 4. kötet. Határterületek. Budapest, Osiris, 2002. p. 176–180.

JANTZ, Ronald – GIARLO, Michael J.: Digitális megőrzés: egy megbízható digitális adattár architektúrája és technológiája. = TMT, e-tmt, 53. köt. 10. sz. 2006.

Javaslatok a kulturális örökség megőrzése érdekében digitalizálandó dokumentumok körére, a digitalizálás országos összehangolására és a projektek nyilvántartására. Budapest, OSZK, Neumann Ház, Könyvtári Intézet, 2005.

KOLTAY Tibor: Virtuális, elektronikus, digitális. Elméleti ismeretek a 21. század könyvtárához. Budapest, Typotex Kft., 2007. 9. fejezet. =
<http://www.tankonyvtar.hu/konyvtartudomany/virtualis-elektronikus-080906-82>

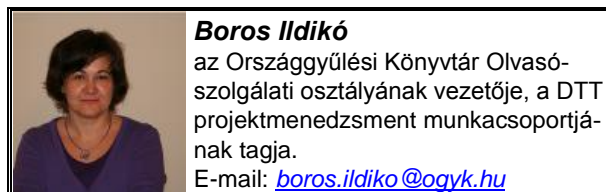
LIU, Yan Quan: Jó példák, szabványok és módszerek a könyvtári dokumentumok digitalizálásához: állapotjelentés az amerikai könyvtárak digitalizálási gyakorlatáról. = TMT, e-tmt, 53. köt. 6. sz. 2006.

Országos Könyvtári Digitalizálási Terv: 2007–2013. Budapest, NOKI Plusz Bizottság, 2005.

Sikerés digitalizálás lépésről lépésre. 1.2 Gyakorlati útmutató. MINERVA Projekt. Budapest, OSZK, 2004. MINERVA Projekt. =
http://mek.oszk.hu/minerva/html/dok/goodpractice_hun.htm

TÓSZEGI Zsuzsanna: A szövegdigitalizálás döntési folyamata. = Könyvtári Figyelő, 2006. 3. sz. p. 245–260.

Beérkezett: 2010. V. 28-án.



A sikerkönyv-sorozatokban bízik az e-könyvpiac

A tavalyi utolsó negyedév csaknem 60 millió dolláros amerikai e-könyvforgalma az idei első negyedévben 91 millió dollárra nőtt a digitális kiadványokkal foglalkozó International Digital Publishing Forum szerint. A *Financial Times*-ban most megjelent elemzésből viszont kiderül, hogy az e-könyveladás még mindig csupán 5-6%-át teszi ki az amerikai könyvforgalomnak, Angliában pedig alig éri el az egyszázalékos részesedést.

A közkeletű vélekedésekkel ellentétben ugyanis a könyvkiadásban nem a nyomda és a szállítás viszi el a sok pénzt: a szerzők bevezetése és reklámozása rója a legnagyobb terhet a kiadókra. Ezt a költséget pedig az e-könyvek estében sem lehet megtakarítani. Az sem lendít a bolton, hogy a digitális könyveket olcsóbban kínálják.

Ilyen körülmények között a kiadók egyszerűen nem vállalhatják a kockázatot, biztos befutókra van szükségük, ha talpon akarnak maradni. A könyvkiadók is azt teszik, amit a filmstúdiók, az egyszer már bizonyított közönségsikereket bővítik szériákká.

Még mintegy öt év kell ahhoz, hogy az e-piac betöltse a nyomtatott könyvek forgalmában keletkezett űrt. Műfajilag a regények iránt a legnagyobb az érdeklődés a digitális kiadás terén, de az e-könyv olvasók és a grafikai megjelenítés rohamos fejlődésével a gyermekkönyvek és a szakkönyvek iránt is megnő majd a kereslet.

/SG.hu Hírlevél, 2010. június 28., <http://www.sg.hu/>

(SzP)