

3. A fentiekkel kapcsolatban a könyvtárak elismerik annak szükségszerűségét, hogy meghatározott prioritásokhoz és felelősségekhez kell tartaniuk magukat. Nevezetesen:
- a nemzeti könyvtárak az általános kulturális és tudományos értékű dokumentumok,
 - az országos szakkönyvtárak az általuk képviselt tudományág vagy ágazat témaköreiben megjelent művek,
 - a regionális könyvtárak a hon- és helyismereti munkák,
 - a felsőoktatási könyvtárak az adott felsőoktatási intézmény profiljába tartozó irodalom digitalizálásáért viselik a felelősséget.
4. Ha egy könyvtár dönt valamely dokumentum digitalizálásáról, egyeztet más könyvtárakkal: nincs-e szándékukban ugyanazt a dokumentumot digitalizálni. A végdöntésben az az intézmény győz, amelynek legjobb feltételei vannak mind a digitalizálásra, mind pedig a hasznosításra. Ezenkívül, ha a digitalizált dokumentum eredetije valamely körülmény folytán (pl. margináliák vannak benne) különösen értékes, akkor annak további őrzése is kötelező marad.

5. A könyvtárnak törekednie kell arra, hogy a digitalizálás folyamán betartsa a vonatkozó szabványokat és konvenciókat, illetve jogtisztá másolatokat állítson elő.
6. A könyvtárak felelősek a digitalizált vagy eleve digitálisan született dokumentumok tartós megőrzéséért, ha azok eminens kulturális vagy tudományos értéket képviselnek. Ez a kötelezettségük addig áll fenn, amíg az elektronikus kiadványok hazai archívuma meg nem születik.
7. A könyvtáraknak tiszteletben kell tartaniuk a másutt hozott digitalizálási döntéseket, illetve nem szabad konkurenciaharcokba bocsátkozniuk az elektronikus kiadványok piacán.
8. A könyvtárak minél előbb, a folyamat életbevágó információs fontosságának megfelelően, hozzákezdenek a szükséges szabályozó-normatív munka elvégzéséhez.

/BARYŠEVA, O. V.–MAJSTROVIČ, T. V.: Formirovanie nacional'nogo raspredelennogo bibliotečnogo fonda elektronnyh dokumentov. = Naučnye i tehničeskije biblioteki, 8. sz. 2003. p. 35–41./

(Futala Tibor)

Elektronikus források begyűjtése és archiválása Litvániában: úton egy virtuális könyvtár felé

Az utóbbi években az elektronikus információ mennyiségének növelésével igény keletkezett ezek hosszú távú megőrzésére. A digitális környezet, amelyben ezek az információk keringenek, lehetőséget ad dokumentumok bizonyos fokú mobilitására és cserélhetőségére. Ez a helyzet csak romlik, mivel a létrehozatal és megőrzés közötti időt egyre csökkenti a technológiai fejlődés. A webdokumentumok átlagos élettartama 44 nap, és emellett tartalmuk könnyen módosítható, anélkül, hogy a változásokat észrevennénk.

A hosszú távú megőrzést leginkább azok a könyvtárak vállalják, amelyek arra törekszenek, hogy digitális könyvtárrá váljanak. Újfajta objektumokat kell megőrizniük, mint pl. adatbázisok, e-folyóiratok, e-könyvek, weboldalak, mivel az eredeti kiadók nem mindig képesek a nagy terjedelmű archív másolatok kezelésére.

A nemzeti könyvtárak sokszor különleges helyzetben vannak, mert az állami szabályozás értelmében az ő kötelességük az elektronikus dokumen-

tumok nemzeti szintű megőrzése. Ilyen indíttatású a Litván Nemzeti Könyvtár *Elektronikus Források Archívuma* (*Archive of Electronic Resources = AER*) is, célja a Litvániában kiadott elektronikus dokumentumok, valamint a Litvániával kapcsolatos külföldi kiadványok gyűjtése.

Az Elektronikus Források Archívuma projekt

A LIBIS (Lithuanian Integrated Library Information System = Litván Integrált Könyvtári-Információs Rendszer) elektronikus források alrendszere egyik, 2001-ben befejezett projektje az Elektronikus Források Archívumának vízióját volt hívatott bemutatni, megtervezve a megvalósítás szakaszait, és a szükséges pénzügyi forrásokat. A projekt az Elektronikus Kiadványok Letéti Rendszerének modelljére és a NEDLIB (Network European Deposit Libraries = Letéti Könyvtárak Európai Hálózata) projekt dokumentumaira épül. A NEDLIB egy EU által támogatott törekvés, célja a nemzeti könyvtárak együttmű-

kódésében a digitális publikációk európai letéti könyvtári hálózatának létrehozása. Javasatai között szerepel az archívumok beépítése az integrált könyvtári rendszerekbe, és a Nyílt Archívum Információs Rendszer modelljeinek adaptálása.

Az ER (electronic resources = elektronikus források) alrendszer ugyanazon az elveken működik, mint a többi LIBIS modul, melynek céljai:

- a funkciók duplázódásának kiszűrése a meglévő LIBIS modulok és alrendszereik használatával;
- a bibliográfiai és besorolási rekordok egyszerű létrehozása és multifunkcionális használata;
- ugrópontok biztosítása az Elektronikus Források Archívuma és egyéb LIBIS modulok között.

A LIBIS Elektronikus Forrás Alrendszere virtuális könyvtárként nemcsak az archívumot fogja bővíteni, hanem az egyéb LIBIS termékeket és szolgáltatásokat is, köztük a kurrens nemzeti bibliográfiát, az információszolgáltatást, a teljes szövegű elektronikus források szolgáltatását, és különböző bibliográfiai kiadványokat.

Az ER Archívum jogi alapját „a kiadványok kötelese példány-másolatainak és egyéb dokumentumok elosztásának rendjéről” szóló 1996. évi kormányrendelet jelenti. Mivel ez a rendelet nem fedi le az elektronikus forrásokat, ki kellett dolgozni az ER Archívumokra vonatkozó elveket; ezek a következők:

- a tartalomnak a Litván Köztársasághoz tartozása – több jelző határozhatja meg, pl. a földrajzi hely, amelyet az „lt” domainnév azonosít az internetcímeiben, a kiadók címében, a szerzők lakhelyében;
- a forrás státusa – prioritást kapnak a hivatalos kiadók és egyéb szervezetek információs termékei;
- a kiadó által jóváhagyott és hivatalosan bejegyzett összes elektronikus kiadvány.

Nem gyűjtik a következő forrásokat:

- intézmények belső használatú adatbázisai;
- magánszemélyek befejezetlen és nem hivatalos dokumentumai;
- az interneten folyó nyilvános kommunikáció termékei (pl. e-mailek, levelezőlisták, informális kommunikáció, internetes játékok).

Az elektronikus források fajtái

Az elektronikus forrásoknak több fajtájuk van, amelyeket lehetőség szerint az ER Archívum különböző alrendszereiben külön kell kezelni:

1. Az archívumban szereplő elektronikus dokumentumok nagyrészt az interneten megjelent statikus oldalak. Ezeket gyűjtési technikával könnyen össze lehet szedni.
2. Belső adatbázisokhoz kötött dinamikus weboldalak és egyéb tartalmak. Ezek gyakran tűzfal mögött vannak, vagy a kiadók védik őket, ezért nem érhetők el közvetlenül. Az oldalak összegyűjtéséhez megállapodás szükséges a könyvtár és a kiadók között a kötelese példány-törvény alapján.
3. Különböző hordozókon (CD, MO lemezek) kiadott digitális dokumentumok. Ezek a hagyományos papír alapú kiadványokra hasonlítanak azzal a különbséggel, hogy használatukhoz számítógépes hardverre és szoftverre van szükség. Archiválásuknál az a gyakorlati megközelítés, hogy eredeti formájukban őrizték meg őket, vigyázva az élettartamukra. Egy lehetséges megoldás az információ átvitele más hordozóra (pl. szerver merevlemezére) könnyebb hozzáférés és megőrzés végett.

Webarchiválás

A nagy léptékű webarchiválás eltérő megoldásokat igényel, mint amelyeket a könyvtárak általában használnak. A fő különbség az, hogy a webdokumentumok nincsenek ellenőrzés alatt. Általában nincs pontosan jelölve a metaadat, jelenlétük az interneten bármely pillanatban megszűnhet, és a tartalom folyamatosan változhat (portálok esetében). A helyzetet nem szabályozza rendelet, és csak a hivatalos kiadók vethetők alá a kötelese példány-törvénynek.

Az államot átfogó webarchiválásnak több szintje létezik. Össze lehet gyűjteni a nemzeti domain név alatt létező dokumentumokat, vagy válogatott kritériumok alapján végzett keresésekkel beszerezni a külföldieket. A döntést nagyban befolyásolják a meglévő technológiai lehetőségek. A legelterjedtebb megközelítés az információgyűjtés. A webgyűjtés (web harvester) olyan rendszer, amely weboldalakat tölt le, új URL-eket szűr ki a meglévő HTML kódokból, és egy olyan listába teszi őket, amelyek további letöltésre várnak, ha még nincsenek meg. A rendszer engedélyezi a meghatározott webhelyek összegyűjtését több sikeres kezdő URL után, viszont nem garantálja, hogy minden meglévő online forrást összegyűjtött. Ezért a gyűjtési folyamathoz szükség van az ellenőrzésre és új URL-ek lehetséges hozzáadására.

Több olyan szoftvercsomag van, amelyet különböző országokban átfogó vagy szelektív webhelygyűjtésre használnak. Európában a legtöbben a NEDLIB és COMBINE gyűjtőket alkalmazzák. Ezek képesek nagyméretű gyűjtésre, és elég erősek ahhoz, hogy folyamatos használatban legyenek. A NEDLIB gyűjtő egy freeware, amelyet a Helsinkii Egyetemi Könyvtár fejlesztett tovább. Legnagyobb előnye, hogy kimondottan webarchiválások létrehozására alakították.

A webdokumentumok archiválásának másik megközelítése a kiadókkal való együttműködés, melynek során a kiadók megküldik kiadványaikat a könyvtáraknak. Ez lehet időszakos, amikor a weboldalt frissítik. Az együttműködés alapja lehet a kötelességdöntvény, vagy ennek megfelelője, amelynek nyomán a folyamat hasonlít a papír alapú dokumentumok beszerzésére. Nyilvánvaló hátránya ennek a megközelítésnek, hogy csak a hivatalosan bejegyzett kiadók vonhatók be, és rengeteg kiadvány nem kerül be az archívumba. Előnye, hogy a kiváló minőségű metaadat megkönnyíti a katalógizálást.

Gyűjtési problémák

Több probléma merülhet fel akkor, amikor csak egy automatikus webgyűjtési módszert alkalmazunk:

- A létező webgyűjtőket nem tervezték rugalmasra (a gyűjtési időket nehéz ütemezni, mivel nem tudnak különbséget tenni a különböző kategóriájú weboldalak között).
- Miként lehet biztosítani az összegyűjtött dokumentumok teljességét az összes működő ugróponttal, amikor az archívumból férünk hozzájuk (ez különösen nehéz azoknál az oldalaknál, amelyek JavaScript miniprogramokat, és egyéb beágyazott szoftvereket tartalmaznak. A NEDLIB gyűjtővel azt tapasztalták, hogy dinamikus oldalak nagy számát nem kezeli megfelelően, ami néha azzal jár, hogy kiválasztott oldalaknak csak kis részét vagy kezdőlapját gyűjti össze).
- A dinamikusan létrehozott weboldalak és a paraméterezett URL címek kérdése (veszélyes lehet úgy beállítani egy gyűjtőt, hogy paraméterezett URL-eket gyűjtsön, annak ellenére, hogy a webhelyek jelentős részét dinamikus generált tartalom foglalja el).

A fentiekből az a következtetés vonható le, hogy a mai információgyűjtő szoftverek még túl rugalmat-

lanok, és csak statikus online információ gyűjtésére alkalmasak.

Eredmények

A LIBIS ER alrendszer projektjében a NEDLIB gyűjtő mellett döntöttek. A gyűjtés az összes meglévő „.lt” domainre és a kapcsolódó weboldalakra vonatkozott. Egy kb. 300 tételes lista gyűjt össze a „.com”, „.net”, és egyéb végződésű litván oldalakról. A munkát a népszerű litván webcímtárral („Lithuania Online”) kezdték. A Litván Nemzeti Könyvtár rendelkezésre bocsátotta szerverét, amelyre a NEDLIB-et installálták. A gyűjtő a belső adatokat és URL listákat egy MySQL 3.23-as adatbázisban tartja. Úgy állították be, hogy hagyja ki a paraméterezett URL-eket, és csak a statikus dokumentumokat próbálja meg összegyűjteni. Az összegyűjtött dokumentumokat *gzip*-vel csomagolja a napi gyűjtési időszak befejeztével.

A kezdeti jó eredmények után 2002 októberében kezdődött az első gyűjtési ciklus. Három héten át, napi működés után az elérhető dokumentumok nagy részét lefedték.

A másik NEDLIB 2002 novemberében indult, azzal a céllal, hogy szelektíven gyűjtse össze az időszakos webkiadványokat. Kb. 60 hivatalosan bejegyzett URL-t vettek fel a jegyzékbe, e-folyóiratokat vagy folyóiratok elektronikus másolatait, megjelenési gyakoriságukkal. Egy saját fejlesztésű szoftver az időben történő gyűjtést biztosítja, mivel a NEDLIB nem alkalmas a változtatható időzítésre.

Mivel a legtöbb weboldal dinamikus készül, a litván webhelyek teljesebb reprezentációja érdekében az a döntés született, hogy a közvetlenül a kiadóktól beszerezhető elektronikus kiadványok élvezzenek elsőbbséget. A Litván Nemzeti Könyvtár által készített, az e-kiadványok összegyűjtésére és a szerzői jogokra vonatkozó szerződést a legfontosabb kiadóknak 2003 márciusában küldték el. A szerződések aláírása után megérkeztek az első dokumentumok is a metaadatokkal együtt. A korai eredmények egészen kielégítőek: a NEDLIB gyűjtő jó működési jellemzőket mutatott, nem omlott össze, és képes nagyon nagy adatmennyiséget kezelni, beleértve az 1 GB méretű MySQL adatbázist.

Az archívum mérete 2002 decemberében a következőképpen alakult:

Összes kinyert URL:	2 089 943	
Újragyűjtött:	612 600	
Új URL-ek:	1 477 343	
.lt-n kívüli domainek:	4 982	(0,3%)
Paraméterezett URL-ek:	96 501	(6,5%)
A dokumentumok teljes mérete:	75,5 GB	
Az URL-ek átlagos metaadatmérete:	142 bájt	

2002. október–2003. május között négy gyűjtési ciklus volt. Mivel ezek alkalmával az adatbázis mérete jelentősen nő, megnövelve az archívum hozzáférési idejét, az a döntés született, hogy a továbbiakban évente két gyűjtés legyen. A folyóiratok gyűjtése természetesen folyamatos lesz, mivel nem nagy a tárolási költségük.

A fejlesztés másik iránya az összegyűjtött információ indexelése és katalogizálása. Csak a válogatás kritériumainak megfelelő forrásokat dolgozzák fel, a bibliográfiai rekordok UNIMARC formátumban készülnek. Ezek a rekordok bekerülnek a Nemzeti Bibliográfiai Adatbázisba, és az OPAC-on keresztül érhetők el. Az adatbázisba nem kerülő dokumentumról egy automatikus konvertáló szoftverrel csak rövid bibliográfiai rekord készül, a dokumentumban található metaadat alapján. Mivel nagyon kevés HTML oldalnak van minőségi metaadata, a kiadóknak a *Dublin Core* szabvány alkalmazását ajánlották. Az összegyűjtött dokumentumok indexelésére még nem találtak megoldást.

Az ER Archívum továbbfejlesztése

Egy teljes digitális könyvtárhoz a következő lépések vezetnek:

1. A hozzáférés biztosítása a felhasználóknak; keresés a bibliográfiai rekordokban. Az archívumban szereplő legtöbb dokumentumról csak egy rövid rekord lesz hozzáférhető, amely az URL-ből és a metaadattól áll. Egy állandó egyedi azonosító szám biztosítja a következetes utalást.
2. A hozzáférés szervezése internetes interfészen keresztül is, böngészővel. Ezt már a Litván Nemzeti Könyvtár elkészítette a katalógusokra. A dinamikus weboldalakat ugyanaz a webszerver generálja, mint amely a gyűjtőket működteti. Az archivált tartalom az azonosító számból, vagy az URL-ből és gyűjtési időből nyerhető. A folyamatban lévő munkák az interfész javítására és az archívum kényelmesebb felhasználására irányulnak.
3. A szerzői jogok figyelembevételével az a döntés született, hogy az archivált dokumentumok másolása és nyomtatása nincs engedélyezve. Ezek a megszorítások a jövőben enyhülhetnek bizonyos weboldalakra, a kiadókkal történő megegyezések után.
4. A projekt következő szakasza az archívum tartalmának teljes szövegű indexelése. Nagyon növeli az értékét a kulcsszavas keresés megoldása.
5. Az archívum integrálása a Nemzeti Könyvtár honlapján.

JODELIS, Remigijus: Harvesting and archiving of electronic resources in Lithuania: towards virtual library. = INFORUM 2003. 8 p. <http://www.inforum.cz/inforum2003/english/sekce.asp?CisloSekce=9>

(Birkás Bence)

A német információpolitika fejlesztésének vezető eszméje a digitális könyvtár

A szakmai információban és kommunikációban mutatkozó nagy jelentőségű változások arra indították a Német Szövetségi Köztársaság kormányát, hogy újrafogalmazza információpolitikáját. Ennek középpontjában a használó és az általa támasztott követelmények állnak. A politika célja, hogy a tudományos információ mindenkinek egyenlő és igényeihez mérten álljon rendelkezésére.

Az Oktatási és Kutatási Szövetségi Minisztérium (*Bundesministerium für Bildung und Forschung = BMBF*) 2002-ben öntötte végső formába azt a

stratégiai tanulmányt, amely a tudományos és műszaki információ jövőjével foglalkozik. Erre alapozva rögzítette a fejlesztési-támogatási politikájának sarkalatos pontjait *Az információt hálózatosítani – a kutatást aktivizálni* címet viselő dokumentumban (Zukunft der wissenschaftlichen und technischen Information in Deutschland, illetve Information vernetzen – Wissen aktivieren).

Németország útját az információ, a tudás és a művelődés társadalmába az 1999-ben elfogadott kormányzati akcióprogram, az „Innovation und Ar-