

INTERNET A KÖNYVTÁRI RENDSZERBEN

Gyüre Péter, gyp@giant.lib.klte.hu
KLTE Egyetemi és Nemzeti Könyvtár, Informatikai Osztály

In the late 80's OPAC¹ was the only way library systems used the Internet. Today WWW² services are equal to the Internet in many people's mind, so telnet interfaces to database seems to be outdated. How could libraries preserve that domination in the Internet content services that they had achieved with OPAC services? One way is to catalog publicly available Internet services in library system, and provide library users with these information. These services can be considered as conventional information sources in many aspects; they all have a title, author or corporate author, issue date, place of publication and publisher. If we store the URL³ in addition to this, and enter subject headings, Internet resources in the library catalog can compete with popular Internet search services in the area of professional services. Libraries can and should provide leadership in the solution of the problems arised due to unprestistent behavior of URL's too, to give reliable, long-term access to information.

1. Bevezetés

A könyvtárak számítógépesítési története már elválaszthatatlanul összeforrt az OPAC-ok, azaz a mindenki számára szabadon használható számítógépes katalógusok megjelenésével. Ezáltal minden olyan könyvtár szolgáltatása elérhetővé vált az Internet felhasználói számára, amelyek gépe rendelkezett ilyen csatlakozással. Így a modern kor eszközeivel - számítógépekkel az Interneten keresztül - tudnak a könyvtárak tájékoztatást adni a hagyományos dokumentumok lelőhelyéről, azok elérhetőségéről. Ehhez felhasználják azokat a hosszú évtizedek alatt kialakult osztályozási és tárgyszavazási rendszereket, amit a hagyományos katalógusrendszerek igényeltek.

Tekintve a napjainkban és közeljövőben várhatóan robbanásszerűen szaporodó online módon elérhető szolgáltatások, elektronikus folyóiratok sőt komplett teljes szöveges dokumentumokból álló elektronikus könyvtárak létrejöttét, kérdéses, hogyan felelhetnek meg az egyre inkább elterjedő - de alapvetően csak papír alapú dokumentumok kezelésére felkészített - integrált könyvtári rendszerek az új médiák által felvetett követelményeknek.

2. Internet dokumentumok katalogizálásának technikai lehetőségei

A legtöbb Interneten elérhető dokumentum vagy szolgáltatás rendelkezik olyan jellemzőkkel amelyek alapján a könyvtárosok a könyveket vagy folyóiratokat azonosítják; van címük, szerzőjük vagy testületi szerzőjük, kiadási évszámuk, esetleg megjelenési (frissítési) gyakoriságuk, kiadási helyük és kiadójuk. Természetesen emellett az URL az, ami a szolgáltatást egyértelműen azonosítja, így ezt is tárolni kell ilyen dokumentumok esetén. Ez a legkevésbé kifinomult könyvtári rendszerek esetén is lehetséges egy további mező felvételével, illetve a rekord űrlapján való megjelenítésével. Az ilyen dokumentumok esetében is természetes az az igény, hogy az egyszer valahol már létrehozott rekordot a saját integrált könyvtári rendszerbe be lehessen tölteni. Igazság szerint még természetesebb kell legyen, hiszen ilyen esetben a rekordot nem is kell a helyi adatokkal kiegészíteni mint a hagyományos dokumentumok rekordjai esetén, mivel a lelőhelynek ilyen esetben

¹ Online Public Access Catalog = nyilvánosan használható számítógépes katalógus

² World Wide Web ≈ világméretű számítógépes hálózat

³ Uniform Resource Locator = egységes lelőhely megnevezés

is ugyanaz az URL felel meg. Az ilyen rekordletöltés tehát valamilyen nemzetközileg is elfogadott szabványon kell alapuljon, ami könyvtári számítógépes körökben egyértelműen a MARC formátumot jelenti. Az OCLC indított egy projektet Intercat néven, ami éppen az ilyen Internet dokumentumok katalogizálására készített egy szabványajánlást, és egyáltalán nem mellékesen jó néhány ezer dokumentum MARC rekordját is.

Az szabvány kulcsa egy speciális MARC mező, a 856-os tag használata. A 856-os mező megnevezése Electronic Location and Access, nevén túlmutatóan azonban számos egyéb információt is tárol. A teljesség igénye nélkül néhány érdekesebb almező és azok használati módja álljon itt, a részletes leírás megtalálható [2]-ben.

- első indikátor, a hozzáférés módja.
- a almező, a host neve. A szolgáltató szerver teljes neve. Több név is megadható, az 'a' almező ismétlésével.
- b almező, IP cím. A szolgáltató szerver IP címe. Az ajánlás szerint lehetőleg dinamikusan kell a rendszernek generálnia, mert ez a névvel szemben gyakrabban változhat.
- c almező, tömörítésre vonatkozó adatok. Ha az Internet dokumentum egy fájl, és tömörítve tárolják, akkor a tömörítő algoritmusra vagy programra vonatkozó információk helye. Ez lehetővé teszi letöltés után akár az automatikus kitömörítést is.
- d almező, az elérési útvonal (path)
- f almező, fájlnev
- g almező, fájlnev - intervallum vége. Akkor használható ha a dokumentum több fájlban van tárolva (pl. lista archívumok) és az 'f' almező nem ismételhető az adott rendszerben.
- h almező, az elektronikus kérések feldolgozójának neve (például listserv). Általában az 'a' almezőt megelőző @ előtti névként értelmezhető.
- k almező, jelszó. Egy ftp szerver vagy egy könyvtári katalógus kérhet a felhasználótól jelszót. Ha bármilyen karaktersorozat használható jelszóként, akkor ezt a mezőt nem kell kitölteni, elsősorban általános használatú (tehát nem titkos, vagy egyedi) jelszavak megadására szolgál. Például:
856 2_ \$aopac.lib.klte.hu\$b193.6.129.214\$kopac\$lopac\$tvvt220
- l almező, login név
- m almező, kapcsolattartó
- p almező, a port száma. Nem szabványos szerver szolgáltatás megadásához szükséges, ha a szolgáltatás neve nincs meg a services táblázatban.
- s almező, fájl méret. Igen hasznos információ lehet a fölösleges és lassú letöltés elkerüléséhez. Megoldható hogy bizonyos méret (és domain távolság) fölött a rendszer figyelmeztessen az esetlegesen órákig tartó letöltésre
- t almező, terminál emuláció. Telnet protokollal működő szolgáltatás esetén alapvető fontosságú információ. Ebben a mezőben specifikált tartalom alapján indítható a megfelelő terminálemulációt biztosító program.
- u almező, URL. Ez az almező használható a többi almezőben található információk egyszerűbb megadására. Ugyanakkor ha a megjelenítési követelmények olyanok, akkor a szabvány egyaránt javasolja az a, d, f, k, l, p és az u almezők megadását.
- 2 almező, elérési mód (protokoll). Ha az első indikátorban megadható protokolloktól eltér a hozzáférés módja, akkor azt a 2 almezőben kell megadni.

Mivel a 856-os MARC mező a szabvány szerint ismételhető, így egy dokumentumhoz több forrás is adható. Így például egy levelezési listához a feliratkozás címe és módja, mellette a FAQ ⁴ letölthető változata egy ftp szerveren, annak tükrözése, valamint az archívum ftp szervere.

```
000      00424nas  2200289  4504
001      bibKLT00075411
005      19960404105033.0
008      c19959999hu w p          0uuu 0hun d
```

⁴ Frequently Asked Questions = gyakran feltett kérdések, minden levelezőlista karbantart egy ilyen kérdés-válasz jellegű ismertetőt

```

020      $a1219-4263
040      $aHuDeKLEK
245 00  $aInternetto$h[számítógépfájl]
260      $aBudapest :$bIDG Lapkiadó KFT,$c[1995-]
300      $bill.
310      $ahetente
362      $aVol. 1. No. 1. (1995 július)
500      $aA cím a képernyőről
516      $aMultimédia elektronikus folyóirat
538      $aHozzáférés: World Wide Web
710 2    $aIDG Magyarországi Lapkiadó Kft. (Budapest)
856 7    $2http$uhttp://www.idg.hu/internetto/
949      $lIR
    
```

1. példa

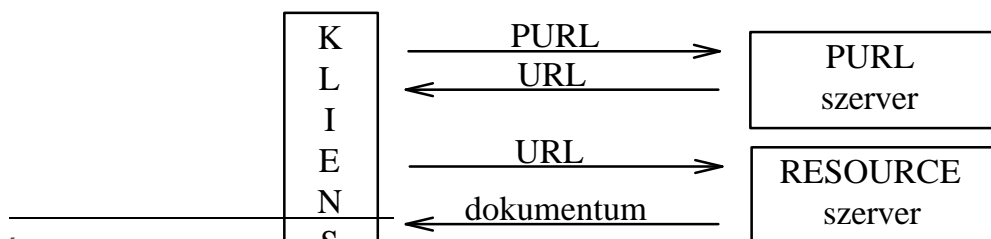
Az Internetto elektronikus folyóirat tagged MARC rekordja

A MARC egyéb mezői és almezői ilyen esetekben persze speciális tartalmat kapnak, mint az előző példában a 245 mező h almezője. A MARC szabvány használatával az Internet dokumentumok ‘bibliográfiai’ leírása mellett így azok leelőhelyének és az összes, letöltésükre vagy használatukra vonatkozó adatot megadhatjuk. A szabvány szerint elkészített rekordok minden a formátumot támogató rendszerbe betölthetők, és a saját számítógépes katalógus részévé válnak. Megjelenítésük formátumát a könyvtár határozza meg, a rendelkezésre álló rendszer keretein belül.

3. A ‘moving target’ problémaköre és annak lehetséges megoldása

Az Internet dokumentumok keresőrendszereinek felállítása során hamar szembesülünk azzal a ténnyel hogy az Internet rendkívüli gyorsasággal változik, új és még újabb szolgáltatások jönnek létre. Sajnos ugyanakkor egyes szolgáltatások helye megváltozhat, vagy akár szolgáltatások és dokumentum leelőhelyek meg is szűnhetnek. A könyvtárak számára ez veszélyes terepet jelent, hiszen a könyvtári szolgáltatások jellemzője az állandóság, a könyvtárak felelősséggel tartoznak olvasóik felé a tárolt dokumentumok mindenkori eléréséről. A katalógizálás szemszögéből pedig mérhetetlenül nagy munkát jelentene a már feldolgozott URL-ek naprakész voltának ellenőrzése, illetve módosítása a katalógus rekordjaiban.

Ez a problémakör persze érinti az összes Web lapokat indexelő rendszer üzemeltetőit is, így létezik egy szabványosítási törekvés az IETF ⁵ keretein belül, melynek célja az URN ⁶ szabvány létrehozása. Mivel azonban a szabványok kifejlesztése és elfogadása lassú folyamat, az URN technológia kifejlesztésének és elfogadásának gyorsításához az OCLC létrehozott egy név szolgáltatást általános Internet erőforrásokhoz. Ezek a nevek - melyeket PURL ⁷-nek neveznek - használhatók Web lapokon és katalógizáló rendszerekben egyaránt. Lényegét tekintve a PURL is egy URL, azzal különbséggel hogy nem közvetlenül egy Internet erőforrásra mutat, hanem egy közbülső név feloldó szolgáltatásra. A PURL szerver adja vissza kliensnek azt a PURL-nek megfelelő URL-t amely a szolgáltatás pillanatnyi leelőhelyét azonosítja. A kliens aztán az URL alapján ugyanúgy dolgozik tovább, hiszen nem történt más mint egy szabványos HTTP ⁸ átirányítás.



⁵ Internet Engineering Task Force = az Internet műszaki feladataival foglalkozó bizottság

⁶ Uniform Resource Name = egységes, állandó leelőhely megnevezés

⁷ Persistent URL = állandó URL

⁸ HyperText Transfer Protocol = hipertext továbbító hálózati szabvány

1. ábra A PURL működési elve

A PURL-t az teszi állandóvá, hogy csak a hozzá kapcsolódó URL-t változtathatjuk meg, magát a PURL-t nem. Természetesen valakinek üzemeltetni kell a PURL név feloldó szervereket, hogy ez az állandóság biztosítva legyen. Ez az üzemeltetés azonban szükségtelessé teszi az egyes rekordokban (utalókban) való adatmódosítást, és a PURL szerver a kért PURL alapján az éppen aktuálisan megfelelő URL címét adja vissza. Az OCLC-nél kidolgozott elképzelés szerint a PURL szerverek sokasága szolgáltatná az összes fontosabb Internet szolgáltatás URL-PURL megfeleltetését. Az ehhez szükséges szoftver az OCLC PURL szerveréről (purl.oclc.org) letölthető. A specifikáció egészen finoman határozza meg a technológiát, egészen a feloldást csak használó személyektől a karbantartó személyek és csoportok hozzáférési listájáig [2]. Meg kell azonban jegyezni hogy bár a technológia nagyban segít az állandóság problémájának megoldásában a lényeg továbbra is a szervezetek kezében lesz:

- a PURL szervereknek folyamatosan elérhetőnek kell lenniük
- PURL szervereket kell üzemeltetniük a hosszútávú, megbízható információszolgáltatásban érdekelt szervezeteknek, kiadóknak, könyvtáraknak és akadémiai intézményeknek.

A PURL tehát egy olyan megoldás mely nem teszi szükségessé a használt protokollok és az azokat implementáló Internet böngésző eszközök módosítását, hanem az azok által támogatott módon, köztes feloldó szerverek üzemeltetésével segíti az Internet folyton változó világában a navigációt. [3]

4. További lehetőségek, avagy amiről a szabvány nem szól...

A számítógépes katalógusba rögzített Internet dokumentumok a hagyományos dokumentumokkal együtt kereshetők, leggyakrabban cím, tárgyszavak vagy szerző alapján. Így egy keresés találatai részben tartalmaznak könyveket és folyóirat vagy cikk címeket (azzal együtt hogy melyik polcon vagy raktárban található az adott kötet) és Internet szolgáltatások és dokumentumok címeit, azok lelőhelyével együtt. Kézenfekvő hogy a könyvtári keresőrendszer képes legyen az egyébként már beírt és könnyen eltéveszhető URL-ek 'követésére', azaz egyetlen kattintással elkezdődhet a kívánt dokumentum letöltése. Erre a feladatra nyilvánvalóan egy professzionális Web böngésző (Netscape, hiszen az a könyvtárak számára ingyenesen használható) a legalkalmasabb. Ezzel a megoldással a könyvtári felhasználók számára gyakorlatilag megszűnik a hagyományos és az Internet dokumentumok közötti különbség, illetve az Interneten található dokumentum válik könnyebben hozzáférhetővé. Ez olyan további előnyökkel is jár, mint

- egy művet egyszerre többen is olvashatnak, nincs olyan konkurencia mint a hagyományos dokumentumok esetén
- a műveket, vagy azok részleteit tetszés szerint ki lehet nyomtatni, vagy levélben elküldeni;
- nincs szükség a dokumentum tárolásához helyre (lásd Encyclopedia Britannica Online ⇔ nyomtatott változat), drága raktárakra,
- a dokumentumok fizikailag nem használódhatnak el,
- nincs szükség raktárkezelő és kölcsönzői kiszolgáló személyzetre,
- nincs szükség a dokumentumok megrendelésére, szállítására és leltározására,
- a katalogizáló munka a rekordcsere folytán egyszerűbbé válik.

A MARC szabvány csak egy-egy Interneten található dokumentum teljes leírásáról és lelőhelyének megadásáról szól, azonban az URL kapcsolatokat követő technológia kis módosítással másra is használható. Ha egy hagyományos dokumentumhoz tartozó illusztrációkat, tartalomjegyzéket képi formában kívánjuk tárolni, akár egyanezt a módot is használhatjuk. A digitalizált képeket WWW szerveren tárolva, az URL-t a megfelelő helyre beírva, olyan multimédia képességgel vértelmezhetjük fel rendszerünket ami az Interneten keresztül is képes dokumentumokat szolgáltatni. Az így keletkezett rekordok szintén képezhetik rekordcsere alapját, hiszen a kapcsolódó médiák letöltése máshonnan is megtehető. A képek mellett természetesen bármilyen más, WWW böngészővel megjeleníthető médiafajtát is kapcsolhatunk a hagyományos

dokumentumok rekordjaihoz. A lényeg itt nem maga a multimédia képességgel rendelkező integrált könyvtári rendszer, hanem a szabványos rekordformátum. Ez - a szabvány megfelelő módosításával, például a második indikátor használatával - szintén megtehető a 856-os MARC mezőn keresztül. További lehetőség a hagyományos és elektronikus formában is létező dokumentumok egymáshoz rendelésére a hagyományos formához tartozó MARC rekordba illesztett 856-os mező. Bár rekordkapcsolatokat használva is hasonló eredmény érhető el, mégis imponáló például a Népszabadság bibliográfiai rekordjának és kardex adatainak megjelenítése mellett a WWW szerverről letöltött aznapi cikkek megjelenítése.

Az ilyen URL-t követő megoldások egyrészt alapulhatnak a már említett külső WWW böngészőt indító megoldáson, de ugyanilyen természetes lehet a könyvtár Web alapú OPAC szolgáltatásának használata is, alkalmassá téve a kapcsolatok követésére. Ilyen található a

http://www.lib.klte.hu/voyager/cgi-bin/complex_query.pl (pl. cím= internetto) címen.

Az Internet dokumentumok könyvtári rendszerből történő szolgáltatásának számos előre még nem látható problémát okozó eleme van. Ilyen például az, hogy egyre több a 'fizetős' Internet szolgáltatás, melyek gyakran egyetlen account megvásárlásán alapulnak és mint ilyenek nem felelnek meg a könyvtárak követelményeinek. Megoldás lehet, ha a szolgáltató egy teljes domain (például egy egyetem) számára engedélyezi a hozzáférést az előfizetési díj fejében - mint például az Encyclopedia Britannica. Bár sajnos refrénszerűen hangzik el újra és újra, mégis meg kell jegyezni a szűk vonalkapacitást is, mint a szolgáltatás színvonalát lényegesen befolyásoló tényezőt. Nincs igazán jó megoldás még az URL (vagy PURL) megadásának egyszerűbbé tételéhez sem, hiszen ezek többnyire hosszúak (URL) vagy számokat tartalmaznak (PURL) és emiatt könnyen elírhatók.

5. Az Internet dokumentumok katalogizálásának jövője

Az Internet töretlen fejlődését, az Internet erőforrások hihetetlen méretű gyarapodását tekintetbe véve nem állítható teljes bizonyossággal az, hogy az itt ismertetett módszerek használatán fog alapulni az Internet erőforrásai közötti tájékozódás. Azt azonban megállapíthatjuk hogy a fenti módszerek együttes használata

- lényegesen pontosabb keresési eredményeket adhat mint a jelenlegi indexelő rendszerek használata - persze a rekordok tartalmának részletességétől függően,
- lehetővé teszi azt hogy az információk helyének változásától függetlenül, akár évek múlva is megtaláljuk a szolgáltató szervert a PURL alapján,
- a szabványos formátum használatából adódó rekordcserén keresztül leegyszerűsíti az ilyen információk tárolását.

Kérdéses persze, hogyan 'gyarapítják' a könyvtárak állományukat az ilyen Internet rekordokkal, milyen 'beszerzési politikát' folytatnak. Nem valószínű hogy a többnyire csak néhány tíz- vagy százezer rekordra konfigurált és megvásárolt integrált könyvtári rendszereknek a teljes Web közösséget indexelő szuperszerverekkel kellene versenyezni, és az összes Internet erőforrást tartalmazó rekord betöltését megcéloznia. Ugyanakkor az olvasói igények indokolják és az erőforrások lehetővé teszik egy szűkebb tudományterület teljes, vagy egy tágabb terület válogatott Internet rekordjainak létrehozását/betöltését. Ez teljes mértékben megegyezik a hagyományos dokumentumtípusok beszerzési politikájával - azzal a különbséggel hogy itt még szinte minden ingyen van. Ilyen módon a hagyományos és nem hagyományos dokumentumtípusok címeinek és az elérhetőségére vonatkozó adatok szolgáltatása csaknem megegyezik, lehetővé téve a médiák között levő különbségek eltüntetését az olvasók számára. Mindez olyan módon érhető el, melyhez a szükséges technikai eszközök már teljes mértékben rendelkezésre állnak, legfeljebb a szoftvereken kell némi módosítást végezni az URL kapcsolatok követésére.

Irodalomjegyzék

- [1] Nancy B. Olson: Cataloging Internet Resources : A Manual and Practical Guide, 1995, OCLC Inc.
- [2] Keith Shafer, Stuart Weibel, Erik Jul, Jon Fausey: Introduction to Persistent Uniform Resource Locators, 1996, OCLC Inc.
- [3] PURL Frequently Asked Questions, PURL:<http://purl.oclc.org/OCLC/PURL/FAQ>, 1996