

URN:NBN – elektronikus dokumentumok hosszútávú azonosítása

Aradi Bálintⁱ, Ittész Péter, Horváth Ádám

Országos Széchényi Könyvtár, 1827 Budapest, Budavári Palota F. épület

Kivonat

Az URN (Universal Resource Name) egy olyan azonosító, amely az interneten található elektronikus dokumentumok hosszútávú azonosítására szolgál. Ennek egy altípusa az NBN (National Bibliography Number). Az URN azonosító típust a céllal hozták létre, hogy kiküszöböljék az URL-ek használatából fakadó hátrányokat. Mivel az URL-ek gyakran változnak, a hivatkozások folyamatos karbantartása nagy erőfeszítést igényelne. A problémát még tovább növeli, hogy a változást közölnünk kell mindazokkal, akik az általunk közzétett forrásra hivatkoznak. Az URN azonosítók (természetesen az URL-el együtt) ezekre a nehézségekre próbálnak megoldást nyújtani. Az Országos Széchényi Könyvtárban is éppen most zajlik az URN-NBN rendszer bevezetése egy nyílt projekt keretében, amely révén lehetőség nyílik a magyarországi könyvtárak és internetfelhasználók számára, hogy elektronikus dokumentumaikat egyedi hosszútávú azonosítóval láthassák el.

1. Bevezetés

A világháló dinamikus fejlődésének köszönhetően rohamosan növekszik az elektronikus (is) elérhető információk mennyisége. A papíralapú információközléssel szemben az elektronikus tartalomszolgáltatás egyik nagy előnye az információk összekapcsolásában rejlik, vagyis hogy nem csak hivatkozni tudunk más (szintén elektronikus) forrásokra, de azok egy kattintással meg is tekinthetők. Mindez azonban csak akkor működik hosszú távon, ha az azonosítók, amely alapján a dokumentumokra a hivatkozás történik, nem változnak az időben, és mindig ugyanazt az objektumot jelölik.

Az elektronikus tartalom azonosítására az interneten egyelőre szinte kizárólag az URL (Universal Resource Locator) azonosítókat használják, amely minden egyes dokumentum esetében annak leőhelyét adja meg. Elvileg ez az azonosítási rendszer is lehetővé teszi a dokumentumok hosszútávú azonosítását, ám az elterjedt gyakorlat alapján ezek tartósságára nem lehet számítani, mert igen sokszor előfordul, hogy egy-egy dokumentum leőhelye megváltozik. Sajnos, ezek nem csak olyankor fordulnak elő, amikor valamilyen honlapot más szerverre (vagy más szolgáltató által üzemeltetett szerverre) telepítenek át, hanem igen gyakoriak az olyan változások is, amikor csupán a szerveren kialakított könyvtárstruktúrát alakítják át, ám ezáltal az érintett dokumentumok eredeti URL-je érvénytelenné válik [1]. Az átrendezés során érintett dokumentumok közötti koherenciát (a belső hivatkozásokat) az üzemeltetők általában helyreállítják, ám a mások által kezelt (külső) hivatkozások frissítése a legtöbb esetben nem történik meg, így olyan hivatkozások maradnak hátra, amelyek nem létező dokumentumra mutatnak.

Erre a problémára jelentenek megoldást az elérési helytől független egyedi azonosítók. Ezek az azonosítók nem tartalmazzak információt a dokumentum tényleges leőhelyéről, így alapját képezhetik egy hosszú távon is koherens hivatkozásrendszernek. Természetesen a dokumentum eléréséhez annak leőhelyére (URL-jére) is szükség van, ám az adott azonosítóhoz tartozó tényleges leőhelyet elegendő csupán egy központi helyen megadni és frissíteni, ahonnan aztán az azonosító alapján történő kereséskor ez kiolvasható. Amennyiben a dokumentum gazdája a leőhely megváltoztatásakor az azonosító és a leőhely közti hozzárendelést ezen a központi helyen frissíti, az azonosítón keresztül történő külső és belső hivatkozások koherenciája adminisztrációs munka nélkül is automatikusan megmarad.

2. Az URN

Az hosszútávú azonosítás előző pontban leírt követelmények többek között az Országos Széchényi Könyvtár (OSZK) által is bevezetni kívánt URN (Universal Resource Name) is

ⁱ Email: aradi@oszk.hu

megfelel. Az URN egy olyan azonosító, amely kiadása szabályozott módon, erre felhatalmazott intézmények által történik. Szintaxisát az RFC 2141-es számú ajánlása rögzíti [2], amely alapján egy URN

urn:<nid>:<nss>

alakú, ahol a <nid> (name space identifier) az URN-ek típusát (névtérét) jelöli, az <nss> pedig az adott típusú (adott névtérbe tartozó) URN-nek közötti egyértelmű azonosítót ad meg.

A névtér szerinti megkülönböztetés (<nid>) szükségessége abból fakad, hogy több fajta rendszer is létezik, amely alapján egyértelműen be lehet sorolni az elektronikus dokumentumokat. Ilyen jól ismert rendszer lehet például könyvek illetve folyóiratok esetén az ISBN ill. az ISSN szám, de képezheti egy, a nemzeti könyvtárak által adott szám (NBN -- national bibliography number) is az azonosítás alapját. Ez utóbbi előnye, hogy tetszőleges könyvtári kategóriába eső elektronikus dokumentumnak adható.

3. Az URN:NBN

Az URN:NBN azonosítók az URN azonosítók egy olyan névtérbe tartoznak, amely felett minden országban a nemzeti könyvtár rendelkezik. Maga az URN:NBN névtérület a Finn Nemzeti Könyvtár kezdeményezésére került bejegyzésre, felügyeletét és adminisztrációját az Egyesült Államokbeli Library of Congress végzi. Minden nemzeti könyvtár igényelhet ezen a névtéren belül egy tartományt magának, amelyen belül azonosítókat adhat ki. Az URN:NBN szintaxisa

urn:nbn:<országkód>-<egyedi azonosító>

alakú, ahol az <országkód> a kérelmező ország kétbetűs (ISO 3166 szabvány szerinti) rövidítését, az <egyedi azonosító> pedig tetszőleges, az URN szintaxisán belül megengedett egyértelmű azonosítót jelent. Érvényes URN:NBN azonosítók lehetnek például az

urn:nbn:fi-fe20010312 és urn:nbn:hu-11234

azonosítók, ahol az előbbi a finn, az utóbbi pedig a magyar nemzeti könyvtár hatáskörébe tartozik.

4. URN:NBN kezdeményezés az OSZK-ban

Az Országos Széchényi Könyvtár a közeljövőben a bevezetés végleges fázisába érő URN:NBN rendszer segítségével lehetőséget kíván nyújtani a magyarországi könyvtárak és a hazai internetfelhasználók számára, hogy dokumentumaikat egyedi hosszútávú azonosítókkal láthassák el. Az azonosítók igénylése teljesen nyilvános, bárki számára elérhető szolgáltatás. A bevezetendő rendszer kialakítása olyan, hogy az azonosítók igénylése, mind HTML-űrlapokon, mind pedig automatikusan, program segítségével is történhet. URN:NBN azonosító csak nyilvános, mindenki számára hozzáférhető dokumentumnak igényelhető. A bevezetés első fázisában az igénylés csak HTML (text/html) típusú dokumentumokra történhet, ez azonban később egyéb dokumentumformátumokra (pl. pdf) is bővíthet. Szintén csupán az első időszakot jellemző korlátozás, hogy egy dokumentumhoz csak egy URN:NBN azonosítót lehet igényelni, valamint, hogy egy URN:NBN azonosítóhoz csak egy URL rendelhető. (A később bevezetendő többes hozzárendelés azt a célt szolgálja, hogy ugyanazon dokumentum különböző helyeken megtalálható példányainak mindegyike nyilvántartható legyen a rendszeren belül.) A projekt során készülő szerver, illetve az URN:NBN igénylést és karbantartást segítő programok forráskódja nyílt, így az bárki számára hozzáférhető lesz.

4.1. URN:NBN igénylése

Az URN:NBN azonosítók igénylése http protokollon keresztül történik, amelyben az azonosítót előállító (és HTTP GET metódussal meghívott) szervernek paraméterként át kell adni azt az URL-t, amelyen az azonosítóval ellátni kívánt dokumentum található, például

<http://nbn.oszk.hu/GetNBN?url=http://igenylo.hu/dokumentum>

alakban. (A példákban szereplő szerver- és paraméternevek még nem véglegesek, ezért a nyilvános verzióban eltérhetnek az itt feltüntetettektől.) A szerver ezután elvégzi az adott URL-lel rendelkező dokumentum ellenőrzését (html-formátumú dokumentum-e, nincs-e még azonosítója, stb.), majd sikeres ellenőrzés után egy URN:NBN azonosítót ad vissza válaszként. Ezt az azonosítót el kell helyezni a html formátumú dokumentum fejlécében (a <head> és </head> elemek között) egy <meta> elemben a következő formában:

```
<meta name='dc.identifier' scheme='urn' content='urn:nbn:hu-1234'>
```

(Feltételezve, hogy a szerver által szolgáltatott azonosító az "urn:nbn:hu-1234" volt.) Ezekután megerősítésként másodszor is meg kell hívni a szerveret az első hívással teljesen megegyező módon. Ekkor a rendszer ellenőrzi, hogy a dokumentumba ténylegesen belekerült-e <meta> elem formájában az urn:nbn azonosító, és ha igen, akkor véglegesíti a bejegyzést. (Ellenkező esetben a bejegyzési kérelem érvénytelennek minősül.) A bejegyzett URN:NBN - URL hozzárendelés később természetesen módosítható, ennek részletei azonban itt hely hiányában nem kerülnek ismertetésre.

4.2. Az URN:NBN azonosítók feloldása

Az URN-nek automatikus feloldását, vagyis hogy egy kliensprogram (pl. egy böngésző) csupán az URN alapján megtalálja egy dokumentumot, a DDDS (Dynamic Delegation Discovery System) szolgálná, ám ez a rendszer egyelőre gyakorlatilag csak papíron létezik, és nem valószínű, hogy a közeljövőben kiépül az általános használatához szükséges infrastruktúra. Ugyanakkor az URN azonosítók értelmüket veszítik, amennyiben nem lehet automatikusan megtalálni az adott azonosítóhoz tartozó dokumentumot. Az OSZK ezért egy, több nemzeti könyvtár által is használt, meglehetősen népszerű hibridmegoldást alkalmaz erre a célra, amelyben az URN:NBN feloldása egy fix URL-lel rendelkező feloldószerveren keresztül történik. A folyamat az ún. THTTP (Trivial convention for using HTTP in URN resolution) protokolon keresztül történik [3], amely egy erősen leszűkített részhalmaza a HTTP-nek. A feloldás úgy jön létre, hogy az URN-t URL-lé átalakító, ún. N2L szolgáltatást nyújtó szervernek (egy GET hívás során) paraméterként át kell adni az elérni kívánt dokumentum URN azonosítóját

<http://nbn.oszk.hu/N2L?urn:nbn:hu-1234>

amire válaszul a szerver egy HTTP átirányítás ad az adott URN-hez tartozó URL-re, vagy hibaüzenetet amennyiben ilyen nem létezik. A felhasználónak tehát általában (pl. amennyiben egy böngészővel kérdezi le a szerveret) nem kell azzal törődnie, hogy a visszakapott URL-t még külön letöltse, ez ugyanis az átirányítás révén automatikusan megtörténik.

A THTTP protokollal történő feloldás hátránya, hogy egy fix URL (az N2L szolgáltatást végző szerver URL-je) szükségeltetik hozzá. Ugyanakkor nagyon nagy előnye, hogy ezen keresztül az URN-ek feloldása már a jelenlegi internetes infrastruktúrával is megvalósítható, így a DDDS rendszer bizonytalan jövője nem hátráltatja az URN azonosítók bevezetését.

4.3. Biztonsági kérdések

A bevezetendő URN:NBN rendszer megtervezésénél fontos szempont volt, hogy használata a lehetőleg egyszerű legyen, de ugyanakkor biztosítsa azt, hogy csupán a dokumentum tulajdonosa igényelhesen azonosítót dokumentuma számára, valamint hogy csak ő módosíthassa a dokumentuma URN:NBN azonosítója és URL-je közötti leképezést. További szempont volt, hogy a folyamat nagymértékben automatizálható lehessen, hogy nagy mennyiségű dokumentum feldolgozásakor programból is történhessen az igénylés vagy a módosítás. Ez utóbbi kritérium gyakorlatilag kizárja a jelszavas azonosítást, ezért az URN:NBN-nel kapcsolatos műveletek során mindig azt tekinti a rendszer a dokumentum gazdájának, aki bizonyítani tudja, hogy írással rendelkezik az adott dokumentum felett. Ezt szolgálja az igénylésnél leírt kétlépéses módszer, amely során a rendszer ellenőrzi, hogy

belekerült-e (egy <meta> elem formájában) az URN:NBN azonosító a dokumentum fejlécébe, vagyis az igénylő módosítani tudta-e azt a dokumentumot, amelyre az azonosítót igényelte. Amennyiben ez megtörtént, a rendszer elfogadja, hogy ő a dokumentum tulajdonosa, és véglegesíti a bejegyzést, míg ellenkező esetben a bejegyzés érvénytelen lesz.

4.4. Kitekintés

Az URN:NBN szolgáltatást nyújtó infrastruktúra (programok) fejlesztése jelenleg is folyik az OSZK-ban. A nyilvános szerver üzembe helyezése, valamint a bejegyzést és karbantartást segítő oldalak elkészítése 2003. május közepére várhatóak.

[1] Természetesen a webserver segítségével megoldható lenne, hogy a kívülálló semmit se érzékeljen a könyvtárstruktúra megváltozásából, de a gyakorlat azt mutatja, hogy az oldalak üzemeltetői sok esetben nincsenek erre figyelemmel.

[2] <http://www.ietf.org/rfc/rfc2141.txt>

[3] <http://www.ietf.org/rfc/rfc2169.txt>