



Web-archívum made in Slovakia: Kísérleti projekt az elektronikus információforrások gyűjtésére és archiválására

A web mint kulturális örökség

A weben található elektronikus információforrások a kulturális örökség fontos részei. Valamennyi, a webes források archiválását célul kitűző állam számára nagy kihívást jelent a gyűjtés, tárolás és megőrzés komplex feladatának megoldása. Míg a svédek és a finnek országos szinten vállalkoznak a – .se és .fi végződésű doménnevekhez tartozó – webtartalmak archiválására, addig az amerikai *Internet Archive* (www.archive.org) az összes webes tartalom gyűjtésére törekszik.

Máshol szelektív gyűjtésre vállalkoznak (ausztrál *Pandora* projekt, amerikai *Minerva* projekt). 2004-ben Amszterdamban létrejött a *Web Archive* non-profit szervezet, amely az európai kulturális intézményeknek nyújt lehetőséget webes tartalmaik nyilvános hozzáférésű, hosszú távú archiválására (<http://europaarchive.org>).

A *Cseh Nemzeti Könyvtár* és a *Pozsonyi Egyetemi Könyvtár (PEK)* a *CULTURE 2000* európai program keretében vállalta a web archiválási módszereinek, szempontjainak kidolgozását. Ezzel egyidejűleg Szlovákiában is megkezdtek a web minőségi és mennyiségi felmérését, a szlovák nemzeti doménnel rendelkező weboldalak feltérképezését. 2006. májusi adatok szerint a szlovák nemzeti domén – .sk – keretében összesen 92 ezer doménnevet regisztrált mintegy 46 961 felhasználó.

A kísérleti projekt

2006 áprilisában a PEK-ben kísérleti projekt indult azzal a céllal, hogy összegyűjtsék az ISSN-nel rendelkező webforrásokat. (A szlovák weben 260 olyan forrás található, amelynek van ISSN-je, ezek közül 164 csak elektronikus formában létezik.) A kiválasztott webforrások archiválására viszonylag szerény hardverfeltételek szolgálnak (Intel Pentium 4-es számítógép, 3 GHz-ces processzorral és 2 GB

RAM-mal, 100 Mbit/s-os internetes kapcsolat). A használt szoftverek nyílt forráskódúak (a rendszerháttér a Debian GNU/Linux szolgáltatója). Az archivált webes dokumentumok indexelését és rekonstrukcióját a NutchWAX

(<http://archive-access.sourceforge.net/projects/nutch>) és a Wera (<http://archive-access.sourceforge.net/projects/wera>) végzi. A projekt keretében 34,5 GB-nyi forrást archiváltak, a legkisebb 45 KB, a legnagyobb 5,3 GB. A források formátumai között hatalmas előnnyel vezet a HTML (68,5%), a JPEG formátum aránya 12,2 %, az összes többi megoszlik a további formátumok között (GIF, PDF, TEXT, DOC stb.). A HTML előnye valószínűleg annak tudható be, hogy a források többsége saját webcímmel rendelkező cikkekből épül fel, valamint a stíluslapok (CSS-ek) használatával egységes dizájn alakítható ki.

A képi formátumok között egyértelműen a JPEG a legnépszerűbb. A csekély számú (0,04%) multimédiás forrás között legnagyobb részben video/x-ms-asf (54,8 %) és audio/mpeg formátumúak (14,4 %) fordulnak elő. A folyóiratok online publikálásának kedvelt formátuma a PDF. A tömörítést főképp ZIP-formátumban végzik. Meglepő, hogy egyetlen OpenDocument formátumú forrást sem találtak, jóllehet egy most már a Microsoft által is támogatott ISO-szabványról van szó. Ugyanakkor viszonylag gyakori a javascriptes alkalmazások használata, kevesen (nyolcan) élnek viszont az RSS nyújtotta lehetőségekkel.

Stratégiai célok

Mindent összevetve a *Web Cultural Heritage* projekt keretében végzett felmérések eredményeként megfogalmazhatók azok a stratégiai célok, amelyek a webes források gyűjtéséhez és őrzéséhez nélkülözhetetlenek:

- az online publikált eredeti elektronikus források köteles példányainak őrzésére, feldolgozására és hozzáférhetővé tételére önálló rendszert kell alakítani a Szlovák Nemzeti Könyvtár és a PEK együttműködésével;
- meg kell oldani az elektronikus források gondozásának törvényi háttérét: szükség van a köteles példányokról szóló, a kiadással, a nyomtatással és a szerzői jogokkal kapcsolatos törvényi szabályozás módosítására;
- meg kell oldani az elektronikus források hosszú távú őrzési módszereinek a kidolgozását, az időszakosan megjelenő elektronikus források azonosítására, a nemzeti ISSN Iroda segítségével;
- szervezeti és technikai feltételeket kell kidolgozni a nemzeti domén keretében publikált elektronikus források rendszeres gyűjtésére, archiválására, webarchívum megvalósítására;

- szakmai irányító központ kialakítása a PEK-ben az online hozzáférésű elektronikus források archiválására vonatkozó módszerek és kísérletek technikai feltételeinek biztosítására;
- szervezeti és adminisztrációs feltételeket kell teremteni a közgyűjtemények, kiadók és szerzői jogok tulajdonosai közötti kooperáció összehangolására.

Ezeknek a céloknak a megvalósítása megoldhatja a weben hozzáférhető elektronikus források gyűjtésének és archiválásának törvényi, szervezeti és technikai feltételeit.

/ANDROVIČ, Alojz: Web-archive made in Slovakia: Pilotní projekt zberu a archivácie elektronických informačných prameňov. = ITlib, 3. sz. 2006. p. 38–41./

(Prókai Margit)