

- élethosszig tartó tanulás, a képzés folytatása a kommunikáció, az információ és az informatika területén;
- helyi információ előállítás, a helyi ismeretekhez való hozzáférés javítása az általános műveltség és az IKT használatának oktatásával;
- nemzetközi szabványok és az élenjáró módszerek használata a kommunikáció és az informatika területén az UNESCO hatáskörének megfelelően;
- hálózati együttműködés a helyi, nemzeti, regionális és nemzetközi szinten az információ és a tudás szférájában.

A tevékenység fő irányai

A program öt fő irányt, ezeken belül számos feladatot tartalmaz:

1. Az információs politika kidolgozása nemzetközi, regionális és nemzeti szinten.
2. Az emberi erőforrás és az információs korszak követelményeinek megfelelő lehetőségek fejlesztése.
3. Az intézmények szerepének erősítése az információhoz való hozzáférés biztosításában.
4. Információfeldolgozó és -kezelő rendszerek fejlesztése.
5. Az információs technológia az oktatás, a tudomány, a kultúra és a kommunikáció szolgálatában.

Együttműködés és partnerség

A program sikere attól függ, hogy egyre több ország csatlakozzon hozzá, alakítson ki különféle kapcsolatokat, és folytasson eredményes együttműködést. Ennek érdekében az UNESCO-n belül

is szorgalmazzák a partnerkapcsolatok létesítését, és a különböző nemzetközi programokba való bekapcsolódást az ENSZ-en belül és azon kívül is.

Értékelés

A program menetét folyamatosan figyelemmel kísérik, befejezésekor olyan értékelést állítanak össze, hogy annak alapján meg lehessen tenni a szükséges módosításokat. A program és a várható eredmények általános értékelését a hetedik évre tervezik, amely egybeesik az UNESCO középtávú stratégiájának utolsó évével, azaz 2007-re.

A program megvalósításának elvei

Gyakorlati szinten a program keretében minden megvalósítandó projekt:

- ösztönző hatású,
- kísérleti jellegű,
- tartalmazza a rendszeres ellenőrzést,
- értékelési kritériumokat tartalmaz,
- biztosítja a projektek kölcsönös kiegészíthetőségének elvét.

A program irányítása

A programot nemzetközi tanács tervezi és irányítja. A 26 UNESCO-tagországból delegált tagokat a Közgyűlés választja az igazságos földrajzi megoszlás és a rotáció elvének betartásával.

/KUZ'MIN, Evgenij Ivanovič: Programma UNESCO „Informaciá dlá vseh”. = Bibliotekovedenie, 2005. sz. p. 82–85./

(Viszocsek Eszter)

A Paradigma projekt

A Norvég Nemzeti Könyvtár 2001 augusztusában indította a *Paradigma projektet*, amelynek célja a norvég digitális dokumentumok köteles példánykénti megőrzése volt. Ez magában foglalta a technológia fejlesztését, az összes digitális dokumentumtípus kiválasztásának, gyűjtésének, leírásának, azonosításának módszereit és gyakorlatát. A projekt feladata az volt, hogy hozzáférést biztosítson internetarchívumához a felhasználók számára. Négy munkatárs teljes munkaidőben foglal-

kozott a projekttel, és további harminc dolgozó vett részt különféle feladatokban. A projekt 2004. december 31-én zárult le.

Gyűjtési és kiválasztási stratégiák

A projekt általánosan gyűjtött minden digitálisan elérhető dokumentumot a norvég webtérből (.no), valamint ezzel egy időben a „.com”, „.org” és a

„net” doméneken találhatóakat is. Az általános gyűjtésnek több oka volt:

- nehezen jósolható meg, mi lesz értékes egy jövőbeli kutatás számára;
- a digitális tárolás napról napra olcsóbb lesz;
- a korlátok nélküli gyűjtés megőrzi a forrásfelhasználás kiválasztását;
- az internetarchívumban szabad szöveges kereséssel található meg a dokumentumok, köztük a kézi katalogizálásból kimaradtak is.

A köteles példány osztály 2001-től kezdve egy HTTrack szoftver segítségével félautomata módon gyűjti a kiválasztott webdokumentumokat, amelyeket a nemzeti könyvtár katalógusa számára katalogizálnak. Ez a részleg eseményhez kapcsolódó dokumentumokat is gyűjt, mint például a politikai pártok weblapjait választások előtt és után. Más osztályok részt vesznek egyéb gyűjtési tevékenységekben, például a könyvtár hang- és képparchívuma azon dolgozik, hogy megoldást találjon a digitálisan készített rádió- és televízió-programok megőrzésére.

Nagy kihívást jelent a „deep web” megőrzése, amely internetes folyóiratokból, webkamerával készített dokumentumokból, interaktív médiumokból és mindenféle elektronikus anyagból áll, amelyeket különféle adatbázisokban őriznek. A Paradigma projekt körülbelül 65 internetes folyóirat gyűjtését kezdte el, valamint néhány teljes folyóirat-adatbázis letöltését tervezi a közeljövőben, hogy kiegészítsék a meglévőket a napi eseményekkel.

A Norvég Nemzeti Könyvtár több csatornán keresztül jut digitális dokumentumokhoz:

- automatikus dokumentumgyűjtés az internetről,
- kötegelte formában érkező adatbázis-frissítések,
- előfizetett folyóiratok,
- levelezőlisták,
- NetNews vitacsoportok,
- fizikai adathordozók, például CD-ROM-ok.

Kiválasztás

A projekt résztvevői kidolgozták azoknak a dokumentumoknak a kiválasztási kritériumait, amelyekről érdemes bibliográfiai leírást készíteni. Egy olyan rendszerarchitektúrát alkalmaztak, amely háromfázisú kiválasztást tesz lehetővé. Az első lépésben megkeresi és összegyűjti a norvég és számi dokumentumokat az internetről. A második fázis lehetőséget ad a könyvtárosnak, hogy automatikusan rangsorolt listákat készítsenek speciális keresőkérdések alapján. A listák metaadatokat

tartalmazó vektorokon alapulnak, amelyeket automatikusan az összegyűjtött dokumentumokból állítanak össze. A harmadik lépésben a könyvtárosok a listákból kiválasztják a megfelelő dokumentumokat nyilvántartásra a fenti kritériumoknak megfelelően.

A norvég internet domén

A Paradigma projekt – a 2004-es eredmények alapján – két fázisban 7,2 millió URL-t gyűjtött össze a NEDLIB-harvester segítségével, melynek gyűjtési paraméterei a HTTP protokollra és a .no-ra korlátozódtak. Norvég oldalak 45-55%-ban a .no hatókörön kívül találhatóak, például finnországi és svédországi doméneken.

Hozzáférési stratégia

Az első kérdés, amit fel kellett tenni, hogy *kiknek* készítik az archívumot, és *mit* fognak benne keresni. A jövőbeli kutatók keresőkérdéseit nehéz megjósolni, de néhány felhasználócsoporthoz, illetve típust meg lehet határozni:

- Az internetet és a digitális dokumentumot mint *médiumot* tanulmányozók kíváncsiak lehetnek a nyelvhasználatra, a nyomtatott és digitális adathordozók vagy a technológiai fejlődési trendek és a tartalom közti összefüggésekre, a web-dizájnról.
- A digitális dokumentumokat *forrásanyagként* vizsgálók különféle tudományterületek képviselőiből állhatnak. Fontosak-e ehhez a dinamikus tartalmak, animációk, interaktív megjelenítések, beépített hang- és videoanyagok? Szabad szöveges keresést kell-e biztosítani, vagy a különböző források információit hozzuk-e összefüggésbe?

Keresőrendszer

Az internetarchívumhoz olyan eszközt kellett biztosítani, amely lehetővé teszi a keresést. Ehhez figyelembe kellett venni, hogy a könyvtárosok csak nagyon kevés dokumentumot fognak katalogizálni. A Paradigma projekt a *Skandináv Webarchívum Keresőrendszerén* (NWA) teszi lehetővé internet-archívuma kereshetőségét.

Metaadat-megoldások keresése

A visszakeresést szolgáló metaadat megtalálása a technikai metaadatok automatikus kinyerésének megoldása mellett az elmúlt év egyik fő tevékenysége volt.

Nancy Olsen három érvet sorakoztat fel az internetes dokumentumok katalogizálására:

- sok értékes információt tartalmaznak,
- a kereshetőséghez rendszerezni kell őket,
- a visszakeresésre a könyvtári eljárások és online katalógusok a legalkalmasabbak.

A projekt résztvevői szerint jó, ha az archivált oldalak 1%-a lesz bibliográfiai leírás tárgya. A hagyományos könyvtári dokumentumok leírásában is vannak eltérések: a könyveket és folyóiratokat részletesebben írják le, mint a rövid érvényességű dokumentumokat. Az internetes dokumentumokat 100%-ban a FAST indexelő szoftverrel dolgozzák fel, ami lehetővé teszi a könyvtárosoknak a szabad szöveges vagy más index alapján történő keresést. Csak kevés érhető majd el közülük teljes szövegű formában és a könyvtár katalógusán keresztül.

Mi az internetes dokumentum?

Amikor kiválasztunk egy internetes dokumentumot megőrzésre, a dokumentum fogalma nagyon félrevezető lehet. Melyek azok az elemek, amelyeket a dokumentum részeként kell archiválnunk? A projekt dolgozói szerint minden olyan részt, amely befolyásolja a weboldal kinézetét, meg kell őriznünk: a háttérképeket, a kerettartalmakat, a gombok képeit. Azok a dokumentumok, amelyeket csatolókon keresztül tudunk elérni, különböznek, de mégis kapcsolódnak a hivatkozó dokumentumhoz. Vagyis gyakran egy dokumentumcsoportot kezelünk egy nagy egységként.

Bibliográfiai szempontból

Sosem bízhatunk annyira a számítógépben, hogy meg tudja állapítani, hol végződik egy internetes dokumentum. Szerencsére a könyvtárosok elég jók ennek a megállapításában. *Bibliográfiai szempontból az internetes dokumentum egy olyan információegység, amely bibliográfiai adatokkal leírható.* A definíció megengedi, hogy a könyvtáros eldöntse: az egész weboldalt egy rekordban írja-e le, vagy egy adott részlethez külön leírást készít-e. A leírásban szerepelhetnek vagy kihagyhatók a háttérképek és hangok, vagy egy dokumentumba kerülhetnek szorosan kapcsolódó weboldalak, mint például egy mű fejezetei. A jövőben automatizált folyamatok miatt fontos a dokumentumok tartalmi alapú meghatározása, mert ennek hiányában a beágyazott képek, a közvetlenül hivatkozott hang- és videoklipek, stíluslapok a dokumentum részei lesznek.

Metaadatok

Az internetarchívumban többféle metaadat használható:

- *adminisztratív*, amely nyilvántartja a rekordok létrehozását és módosítását;
- a *jogi és hozzáférés-kezelő* a szerzői jogi információkat, valamint a használatra, illetve archiválásra jogosult felhasználó csoportokat tartalmazza;
- a *szerkezeti* a logikai kapcsolatokat mutatja;
- a *hosszú távú megőrzésre vonatkozó* a fájl típust, a megtekintéshez szükséges szoftvert és a dokumentum átalakításának vagy költöztetésének lépéseit rögzíti;
- a *technikai* a dokumentumra vonatkozó technikai információkat tartalmazza;
- a *leíró és analitikus* a visszakeresést segíti.

A leíró modell kiválasztása

Több megoldás létezik arra, hogy milyen szintű leírást kell készíteni a digitális dokumentumokról. A kétszintű és háromszintű leíró modellek közül az utóbbit találtuk hasznosnak.

Háromszintű leírás:

1. Katalogizálás a Nemzeti Bibliográfia, illetve a Nemzeti Könyvtár katalógusa számára.
2. Leírás egyszerűsített formátumban.
3. A metaadatok automatikus kiválasztása a dokumentumból és a kommunikációs protokollból.

Néhány értékes digitális dokumentumot érdemes MARC formátumban leírni, ilyenek például az audiovizuális dokumentumok, melyeket hosszú távú megőrzésre szánnak. Az egyszerűsített leírás a Nemzeti Könyvtár azon szolgáltatásához kapcsolódik, amely a jövőben a kiadók számára lehetővé teszi, hogy metaadataikat a beküldéssel egy időben megadhassák. Ebből később egy részletesebb bibliográfiai rekordot lehet készíteni. A metaadatok automatikus kiválasztását a nagy mennyiségű dokumentum indokolja. Az ily módon kinyert metaadatot az eredeti dokumentummal együtt tárolják, ami összetett keresést tesz lehetővé az internetarchívumban. A technológia még nem elég fejlett ahhoz, hogy a dokumentumtípust automatikusan megállapítsa, de segít csökkenteni a dokumentummennyiséget. Jelenleg a következő tulajdonságokat ismeri fel:

- nyelv, szókincs, nyelvtan;
- a dokumentum mérete és szerkezete;
- forrás, kiadó, webszerver;
- cookie-k;
- a dokumentumtól elvárható élettartam;

- hangok, képek, animációk, videók és más fejlett információ típusok;
- a felhasználói interakciókat segítő eszközök, gombok;
- a csatolók száma és forrása;
- URL-összetevők, mint például speciális szavak és karakterek;
- a technikai kommunikáció részletei.

Az FRBR szerepe az internetarchívumban

A Paradigma projekt az archivált digitális dokumentumokat és metaadataikat rendszerezett formában akarja szolgáltatni, hogy segítse a felhasználók tájékozódását. Az IFLA FRBR modellje alapul szolgál az internetarchívum megtervezéséhez. Különösen fontos lehet dinamikus adathordozók esetén, például az internetes, multimédiás, vagy más folyamatos forrásoknál. Az FRBR modell alkalmazásakor a kinyilatkoztató jellegű, valamint a tételszerű fogalmak átértékelésére van szükség.

Az FRBR modell alkalmazása dinamikus dokumentumok esetén

Dinamikus dokumentumok

Az internetes dokumentumok gyakran *dinamikusak*, például az online folyóiratok, melyeket naponta frissítenek. A felhasználó hivatkozhat erre a dinamikus dokumentumra mint fórumra vagy információs csatornára: „A Daily News jelenti, ...” Azt mondhatjuk, hogy a dinamikus dokumentum nagyjából megegyezik egy URL-lel. Formális szempontból egy weboldal frissítése hasonlít egy könyv új kiadásához. De az olvasók szempontjából egy internetes folyóirat folyamatosan változó nyitóoldala csak egy változó entitás, nem önálló kiadás. Kibővítve az FRBR modellt az általános összetevőkkel, a dinamikus dokumentum: „egy folyamatosan változó weboldal – vagy ehhez hasonló internetes dokumentum – teljes életciklusa”. Egy folyamatosan változó dokumentum tartalmának feldolgozása olyan, mintha egy rádióadásból rögzítenénk egy részletet. Minden ilyen mintát speciális dokumentumnak nevezünk. Amikor egy dinamikus dokumentumot keresünk, a megkapott *találat* teljesen különbözhet az összes többi találattól ugyanabból a dokumentumból. Ez különböző tényezőktől függhet: a felhasználótól, a böngészőtől, a korábbi belépésektől ugyanabba a dokumentumba (megőrzött cookie-k), a felhasználó által megadott paraméterektől, valamint az adatbázis aktuális állapotától. Ezért egy adatbázis a dinamikus dokumentumok félig állandó fizikai megjelenítésének

tekinthető, amelyből a speciális dokumentumok származnak. A tételeknek nincs saját megjelenítésük, vagyis ideiglenesek, kivéve, ha internet-archívumban őrzik őket.

Speciális dokumentumok

A *speciális dokumentum* egy fontos szempontból különbözik a hagyományos dokumentumtól: egy dokumentumcsoport tagja, melyek ugyanazt a dinamikus dokumentumot jelenítik meg. Teljes szövegű keresés esetén a dinamikus dokumentumokba legfeljebb egy belépést fogunk kapni a találati halmazban. Ha a felhasználó a találat megjelenítését kéri, a dinamikus dokumentum egy egységként jelenik meg, amiből a felhasználó kiválaszthat egy adott tételt egy időrend szerinti menüből, ami a dokumentum élettartamát mutatja. Minden megőrzött változat – azaz speciális dokumentum – egy jelzéssel jelenik meg az időskálán. A felhasználó hozzáférhet az összes speciális dokumentumhoz az adott dátum vagy időpont jelölésére kattintva.

A kiadó vagy felhasználó által meghatározott dokumentum- és metaadat-definíciók

A fenti elképzelésekből kiindulva javasolták a Nemzeti Könyvtár azonosító-kiosztó szolgáltatásának felülvizsgálatát. Napjainkban ez a webes szolgáltatás adja ki az URN:NBN-eket az egyetemeknek, és más intézményeknek a norvég URN:NBN névtérből. De látnak arra esélyt, hogy önálló ISBN-eket utaljanak ki a szolgáltatáson keresztül.

Jövőbeli működés – a forgatókönyv

Az első azonosító sorozatok, amelyeket a szolgáltatás rendel hozzá a dokumentumokhoz, megkövetelik a felhasználótól, hogy megadják a metaadatokat egy minimális csoportját és az azonosított dokumentum pontos definícióját. Egy ismert kiadó vagy egyetem meghatározását sokkal inkább elfogadják, mint egy egyéni felhasználóét, majd az azonosítót archiválják.

Metaadatmezők

A kötelező és nem kötelező metaadatmezők elérhetők lesznek minden FRBR szinten a dokumentumleírás számára, ahol a szintek URN:NBN-nel azonosíthatók. A metaadatokat azonosítókkal tárolják, hogy az internet alapú szolgáltatás felhasználói ezek szerint találják meg a dokumentumokat. Miután a jövőbeli metaadat-meghatározó eszköz mezőit kitöltötték, a kiadóknak lehetőségük lesz például egy <HTML Dublin Core> gombra kattintva, hogy egy külön ablakban megnézhessék a

metaadatokat HTML formátumban. Így a felhasználó a metaadatokat bemásolhatja a webes dokumentum <HEAD> fejlécébe, mielőtt folytatná az azonosító helyének meghatározását. Minthogy elmenthetik a digitális dokumentumot a beágyazott metaadatokkal, a felhasználó könnyedén tárolhatja az ily módon kibővített másolatot a könyvtár archívumában a böngésző frissítés gombjára kattintva.

Lehetséges felülvizsgáló és hitelesítő szolgáltatás

Mindannyian hallottunk olyan hatóságról, amely visszavonta hivatalos közleményét az interneten, majd később megtagadta a korábbi változatok elismerését. Ehhez hasonlóan hallhattunk olyan cégekről, amelyek termékeiket bizonyos áron reklámozták, majd sokkal többet számláztak a vásárlóknak. Az ilyen esetek megelőzésére a Paradigma projekt egy felülvizsgáló és hitelesítő szolgáltatást ajánl, amely lehetővé teszi a felhasználóknak, hogy kérjék egy adott dokumentum letöltését, azaz egy pillanatfelvételt a kereskedelmi ajánlatról, a jogi felelősségről szóló nyilatkozatról vagy rágalmozásról stb. Ha további kételyek merülnének fel ezekkel a dokumentumokkal kapcsolatban egy adott időre vonatkozóan, a könyvtár hitelesíteni (vagy cáfolni) tudna minden követelést. Akkor is, ha jogi szempontok nem vetődnek fel, a megőrzött

speciális dokumentumtétel jól meghatározott képet ad a dinamikus dokumentumról egy adott időben idézés és hivatkozás céljából. Ez különösen fontos akkor, ha felismerjük, hogy a legtöbb internetes dokumentumnak nincsenek oldalszámok és verziószámok stb. Az internetarchívumban a speciális dokumentumot abban a formátumban határozzák meg, ahogyan a webszerverről beérkezett. A dokumentum minden összetevőjének jól meghatározott bitfolyama van (szöveg, kép stb.). A speciális dokumentum úgy határozható meg, mint a dinamikus dokumentum adott összetevői és metaadatai:

- minden egyes összetevő forrása (például egy URL);
- visszakereséskor egy kliens meghatározza az összes paramétert;
- egy időkieljelző, amely a visszakeresés idejét mutatja;
- a dokumentumba foglalt összetevők csoportja.

NUYS, Carol van-ALBERTSEN, Ketil-PEDERSEN, Linda et al.: The Paradigma Project and its quest for metadata solutions and user services. [IFLA paper] Code number: 009-E, World Library and Information Congress: 70th IFLA General Conference and Council, 22-27 August 2004, Buenos Aires, 89. Cataloguing. 15 p. <http://www.ifla.org/IV/ifla70/prog04.htm/>

(Szalóki Gabriella)

Nem OAI-szabványos dokumentumok integrálása a DLIST nevű, Eprints szoftverrel épített repozitóriumba

A világhálón nagy mennyiségű tudományos és kulturális dokumentum található a legkülönbözőbb intézményi és személyes forrásból. A függetlenül épülő archívumok, digitális könyvtárak kölcsönös elérését teszi lehetővé műszaki szempontból az Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH) kezdeményezés, illetve szabvány. Ennek lényege, hogy a szövetkező tartalomszolgáltatók egységes formátumban készítik el metaadataikat, ezeket a közös szolgáltató rendszeresen begyűjti, és közös szerkezetben szolgáltatja. Az elképzelés sikerét felgyorsította a protokollon alapuló programok kifejlesztése.

Továbbra is probléma maradt ugyanakkor, hogy az OAI-PMH nem tudja kezelni a szabvánnyal nem harmonizáló gyűjteményeket, amelyek tulajdonosai nem tudnak vagy nem akarnak részt venni a közös munkában, ezért sok értékes dokumentum rejte

marad egy-egy témakörben, hacsak nem fordulunk témakatalógusokhoz vagy portálokhoz. A szerzők az adatbegyűjtés ezen hiányosságait próbálták kiküszöbölni a saját, *DLIST (Digital Library of Information Science and Technology, <http://dlist.sir.arizona.edu>)* elnevezésű elektronikus gyűjteményükben.

A DLIST a könyvtár- és információtudomány, valamint az informatika elektronikus dokumentumgyűjteményének, repozitóriumának készült az Arizonai Egyetem Információs Erőforrások és Könyvtártudományi Intézete, illetve az Arizonai Egészségtudományi Könyvtár kezdeményezésére, és egy nemzetközi konzorcium létrehozására törekszik.

A DLIST szabad szoftvereket (Linux, Apache, MySQL, Paracite stb.) használ, és az Eprints archiváló szoftver képezi a magját. Az Eprints szoft-