

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 rágalmazásról 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 megörözt

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 oldalszámai és verziószámai 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 időkijelző, 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.html>

(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 rejtve

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://dlisr.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-

vert általában a tudományos szakirodalom adatbázisba szervezésére használják. Jól ismert példái a CogPrints (<http://cogprints.ecs.soton.ac.uk>) a kognitív tudományok, vagy az ArXiv (<http://arxiv.org>) a fizika területein.

A fent említett nehézségek miatt a könnyen elérhető programok ellenére a DLIST gyarapodása lassú. 2002 óta, noha a regisztrált használók száma (2004 közepén) 330, mindössze csak 105 egyéni feltöltés történt a gyűjteménybe. Ennek elsődleges oka a metaadat-készítés nehézsége.

A probléma ismert a szakirodalomban. A CiteSeer például maga keresi, gyűjti be és indexeli a kívánt dokumentumokat. A DLIST-ben más utat követtek: a gyűjteményeket más kollektívákkal együtt építették, és nem követelték meg azt, hogy a partnerek megfeleljenek a szabványoknak.

A nem OAI-konform források lehetnek nem strukturált személyes hálóoldalak, vagy valamely rendszerben strukturált hálózsek.

A technikai megközelítés alapelvei:

- A metaadatok automatikus kinyerésére szolgáló eszközök legyenek elérhetők nyílt forrásokból.
- A metaadatok begyűjtése legyen automatikus és rendszeres.
- A kifejlesztett eljárás legyen átvihető egyik forrásról egy másikra.

A szükségtelen kódmódosítások elkerülése és a hosszú távú stabilitás végett az OAI-PMH-t használták integráló rétegnek, és az Arc szoftvert a metaadatok begyűjtésére és szolgáltatására. A helyi metaadat-előállításra készült egyedi eljárás. A kísérlethez a Nemzeti Orvosi Könyvtár nagyméretű PMC gyűjteményét használták.

Az integrációs folyamat négy lépésből állt: az adatok begyűjtéséből, a metaadatok kivonásából, leképezéséből és importálásából.

A kívánt dokumentumok automatikus összegyűjtésére a WGET programot használták. Amikor ezek az új oldalak rendelkezésre álltak, egy nyílt forrású (open source) programozási nyelven, awk-nyelven írt elemzőszoftvert, parsert futtattak rajtuk a szükséges metaadatok kivonása végett. A program elemezte a HTML-szerkezetet, eltávolította a felesleges jelölőket, és az eredményt helyi fájlban tárolta. A program az állomány szerkezetétől függ. A parser az adatbegyűjtő programmal együtt periodikusan lefutó folyamatot (ún. cron job) képez. (Ez a gyűjteményspecifikus program látszik a folyamat leglényegesebb részének, és sajnos ennél többet nem tudunk meg róla a közleményből. – A ref.)

A következő lépés a metaadatok leképezése az OAI-PMH szokványos Dublin Core formátumába. Erre a feladatra az Old Dominion University által kifejlesztett Rapid Visual OAI Tool-t (RVOT) használták, amely egybe van építve egy egyszerű http szerverrel és egy OAI lekérdezőkezelő programmal. A leképezés definíciós folyamata grafikus interfész segítségével, kézzel történik. Ezután a helyi fájlokból az RVOT elkészíti a DC-metaadatokat tartalmazó és szolgáltató gép által begyűjthető fájlt.

Minden ilyen megközelítésnek vannak hátrányai. A parser helyspecifikus, ezért elég munkaigényes lehet a formálása, és mélyebb programozói ismereteket feltételez. Nagy dokumentumbázisok esetén kifizetődő, de kisebb erőforrások, vagy nem konzisztens szerkesztésű helyek esetén nem. Ráadásul a program érzékeny a távoli forrás bármely szerkezeti módosítására. Ennek ellenére és további fejlesztések hatására jól használható kiegészítő eszköznek látszik az OAI-folyamat gazdagítására.

/COLEMAN, Anita–BRACKE, Paul–KARTHIK, S.: Integration of non-OAI resources for federated searching in DLIST, an Eprints Repository. = D-Lib Magazine, 2004. július–augusztus, <http://www.dlib.org/dlib/july04/coleman/07coleman.html>

(Horváth Péter)

Az IMLS NLG program az USA-ban: könyvtárak, múzeumok és archívumok támogatása

Egy független, támogató állami ügynökség, a Múzeumi és Könyvtári Szolgáltatások Intézete

(Institute of Museum and Library Services = IMLS, USA) számos program révén segíti elő a múzeu-