

A vállalati-üzleti szektorban már komoly veszélybe került a könyvtár fizikai léte. A világ egyik vezető gyógyszergyára 2004 elején bezárja szakkönyvtárait – ez mindenképpen precedenst teremthet, a következmények pedig beláthatatlanok.

A felsőoktatás és kutatás területén nagy sikerek születtek a konzorciumi licencszerződések területén, és hatalmas mennyiségű olyan tartalom lett elérhető, amely korábban egyáltalán nem állt rendelkezésre. Még nem dönt el, hogy a „Big Deal” valóban a legjobb megoldás-e a könyvtárak számára. A szerző meggyőződése, hogy a felsőoktatási könyvtárosoknak jelentős szerepük lesz a szabad hozzáférésű kiadási modellek lobbistáiként, az egyetemek kiadóiként, a tartalom gondozójaként és reklámozójaként, szolgáltatójaként stb. De ha nem ragadják meg ezt a lehetőséget, majd megragadja más.

Összefoglalva:

- A tudományos közlési folyamatban még nem lehet a közvetítők kiiktatásáról beszélni, de már a könyvtáraknak és a kiadóknak nyújtott szolgáltatásokban is sokkal több szereplő van jelen az új technológiai megoldásoknak és az új üzleti modelleknek köszönhetően. (A Content Complete cég is azért jött létre, mert az árakat egyre ritkábban diktálják a szolgáltatók, azokról inkább megállapodásokat kötnek. Ahogy egy könyvtáros fogalmazott: ismeretlenné válik az a fogalom, hogy „a rendelések automatikus megújítása”.)

- Egyre több nem alapvető tevékenységet kihe-lyeznek. A kiadóknál ez a tartalomszolgáltatás működtetését és az ügyfélszolgálatot érinti, a könyvtárakban pedig mind a nyomtatott, mind az elektronikus gyűjteménnyel kapcsolatos adminisztratív-üzviteli teendőket. Jelenleg jelentős az átfedés a kiadók, közvetítők és a könyvtárak ilyen tevékenységében; ha ezt sikerül csökkenteni, az a költségeket is mérsékelni fogja.
- Úgy tűnik, hogy folytatódni fog a cégek összeolvadása és felvásárlása. Lehetséges, hogy ennek a szabályozására is sor kerül, ha az Elsevier kezében még több hatalom összpontosul.
- Az alternatív kiadói modellek rövid és középtávon arra vezethetnek, hogy az üzleti kiadók ráébrednek, önmérsékletet kell tanúsítaniuk az ár-emelés terén.
- Az új, kísérleti árképzési és üzleti modellekről folytatódhatnak a szakmai viták a kiadók körében.
- A két vezető előfizetési ügynökség továbbfejleszti a kiadói és a könyvtári szakma számára nyújtott szolgáltatásait, de esetenként más-más hangsúllyal. Az EBSCO egyre inkább teljes jogú kiadói feladatokat vállal. Az is elképzelhető, hogy a két cég egy napon egyetlen intézményként működik tovább.

HARWOOD, Paul: Chain reaction: new roles and changing relationships in the scholarly communications process. = *Managing Information*, 10. köt. 9. sz. 2003. p. 4., 6–7./

(Hegyközi Ilona)

Keresés a világhálón: hogyan változtatta meg az internet az információkeresést?

Az információkeresés olyan összetett emberi tevékenység, amely az Alexandriai Könyvtártól a világhálón (World Wide Web, röviden web) terjed. Ez utóbbi az elmúlt években több millió embernek mutatta meg, hogy mi is az a keresés. Az információkereséssel foglalkozó szakemberek készek arra, hogy segítséget nyújtsanak a világhálón fellelhető adatok, információk megtalálásához. A klasszikus információkereső stratégiát, a honlapok tárgyszavazását, ezek metaadatokként való megjelenítését már kipróbálták, de az eredmények korántsem kielégítőek. Az emberi természet ellenállásába ütközött, hogy a weben publikált dokumentumokat szerzőik indexelhető tárgyszavakkal lássák el. Megfigyelték, hogy a webes dokumentumok HTML kódjában elemként szereplő „kulcs-

szó” metaadatok megbízhatatlannak bizonyulnak, és ez az oka annak, hogy a legtöbb internetes keresőgép ezekkel nem foglalkozik. A Dublin Core is tényként kezeli ezt a felfedezést.

Még most is folyik a vita arról, hogy ezek a metaadatok mennyire hasznosak. Az alapvető kérdés azonban az, hogy technikai szempontból mennyire fontos, hogy egy weblap tartalmazza ezeket az elemeket.

Ha úgy vesszük, hogy a web egy széles körben elterjedt dokumentum-adatbázis, és a HTML nyelven készülő honlapokat tartalmazza, akkor egyértelműen igen a válasz, azaz szükség van ezekre a metaadatokra. Mint ahogy más dokumentum-

adatbázisok is az indexelés révén válnak használhatóvá, úgy igaz lehet ez a világhálóra is. Ha viszont azon a véleményen vagyunk, hogy a világháló nem egy nagy dokumentum-adatbázis, hanem különböző megjelenések állandóan változó hálózata, akkor a válasz *nem*. Ebben az esetben ugyanis a HTML nyelv nem más, mint technológia, amely abban segít, hogy megjeleníthessünk valamit az interneten.

A legtöbb honlap illékony, tisztavirág-életű, ideiglenes tartalmú, amely a felhasználók, készítők széleye szerint változik. A negatív válasz igazából azt jelzi, hogy ezen tárgyhoz tartozó metaadatok értékéről vitatkozni addig felesleges, amíg nem válik világossá az, hogy technikailag fontos velejárói a honlapoknak.

E mögött a metaadat-vita mögött ott lappang annak a kérdése is, hogy mennyire megbízható a dokumentum koncepciójának alkalmazása a weben található tartalomra és megjelenésre. A dokumentum fogalmát még a klasszikus adatbázisokból örököltük, amikor a technikai környezet nem tette lehetővé a tartalom és a forma különválását. Ha ebből a hagyományos dokumentum-adatbázis szempontból közelítjük meg a kérdést, a honlapok indexelése csupán egy újabb adaléka a már meglévő gyakorlatnak, vagyis az új, digitális dokumentumformátummal egészítjük ki az eddigi információkeresést. Ha viszont a HTML oldaláról vizsgáljuk a kérdést, akkor a weblapok indexelése csupán összetéveszti a tartalmat és a formát. A metaadatok segítségével az információtartalom indexelése volt a cél, nem pedig az önkényes vagy személyes tartalmi feltárás. Márpedig a legtöbb weblap esetleges, és a böngészőktől, biztonsági beállításoktól, segédprogramoktól, stíluslapoktól stb. függ.

Az a felismerés, hogy a dokumentum metafora nem fedi le a világhálót, egyfajta paradigmaváltást jelez az információkeresésben. Lehet, hogy a webes keresők nem is információt keresnek vissza, hanem valami egészen mást tesznek. Inkább csak felfedezik, de nem keresik az információkat. A 20. század közepére a számítógépes adatbázisok megjelenése lehetővé tette, hogy ezeket a rendszereket a dokumentumok tárolására és visszakereshetőségére használjuk. A dokumentumokat ekkor információhordozókként kezeltük, s bár egyre elvontabbá vált a kifejezés, lassan eljutottak oda, hogy a szóhasználatban szinte szinonimaként kezelték az adatbázisrekord szókapcsolatot és a papírlapot.

A dokumentumok digitalizálásával az információkeresést, lehetőségeit, módszereit egyre rendszeresebben vizsgálták. A tartalomfeltárás nagy lépésekben haladt előre. A dokumentum-adatbázis keresési módszereit a világhálóra is alkalmazni lehetett, s ez a kezdetekben azt mutatta, hogy a web csupán az információkeresés egyik bővítője lesz, elvégre:

- rengeteg dokumentum megtalálható a világhálón (azaz a világháló igazából egy hatalmas adatbázis);
- ha egy hálólapp forráskódját megnézzük, akkor az egy szerkesztett dokumentumot mutat (pl. a dokumentumot a <html> </html> elemek zárják közre);
- úgy tűnik, hogy a Google indexeli a weblapokat (pl. a Google nem más, mint egy nagy index, és a weblapokon talált szavakból épül fel).

A kereséssel kapcsolatos elképzeléseink is társadalmi konvenciókon alapulnak, s ezt próbáljuk a webre is ráhúzni. A könyvtár még a technológiai fejlődés, az indexelési eljárások felbukkanása előtt is hagyományosan a keresés fellegvára volt (bibliográfiai leírások egységesítése, osztályozási rendszerek stb.). A számítógépes adatbázisok megjelenése a 20. század közepén arra ösztönözte az információs szervezeteket, hogy szabályokat hozzanak létre a dokumentumok feltárására. Társadalmi összefogás, technikai összhang jött létre egyes diszciplínákon belül annak érdekében, hogy irányításuk alatt tarthassák saját területükön belül az információkat.

A dokumentum-adatbázisok esetében létrejött összefogás feltételezte, hogy a kialakult, kidolgozott elméletek, gyakorlatok a világhálóra, mint társadalmilag összhangban álló információs környezetre is alkalmazhatóak lesznek:

- A világháló szerzői odaillő metaadatokkal látják el a honlapjaikat („Tárgyszavakkal és Dublin Core metaadatokkal fogom indexelni a hálólappomat, hogy az emberek megtalálhassák az interneten”).
- Mindenki a legmegfelelőbb, oda illő metaadatokot fogja használni.
- A keresőgépek, mint a Google is, a megfelelő metaadatokkal fognak operálni.

Csak mostanában derül ki azonban, hogy a világhálónak egészen más szociális dinamikája van. A web nem egy jóindulatú, társadalmi összefogáson alapuló környezet, hanem agresszív versenyszínpad uralkodik rajta, ahol is a szerzők arra törekednek, hogy minél jobban felhívják a hálólappjaikra a

figyelmet, akár annak árán is, hogy visszaélnek a metaadatokkal. Ennek eredményeképpen a keresőgépek önvédelemből az összes kulcsszót és metaadatot úgy kezelik, mintha reklámok és körlevelek lennének (azaz spamek). Fontos eldönteni, hogy a hálólapok mennyire felelnek meg az információkeresésben hagyományosan kezelt információtaroló dokumentumnak.

Az ERIC adatbázisban (Educational Resources Information Center = Oktatásügyi Források Információs Központja) akár harminc éve meglévő adatokhoz is biztosan hozzáférhetünk, s tudjuk, hogy ez így lesz a jövőben is. Ez jelenti körülbelül azt, hogy mit értünk az információkeresésen: ugyanazzal a keresőkérdelemmel újra megtalálhatjuk ugyanazt a dokumentumot. Ezzel szemben a világhálón meglévő források csak pillanatfelvételekkel látják el az embert. Elvégre a világhálón fellelhető eszközök alapvető jellemző vonása, hogy nem statikusak, állandóak, hanem fejlődnek, változnak. Nem is annyira a konstans adatbázisrekordokra, hanem inkább a szabadlapos könyvekre hasonlítanak. Ha a világhálón lévő oldalak csupán pillanatfelvételek, akkor fontos jellemzőként szerepelhet az aktualizálás kérdése. Míg jó pár ERIC rekord akár 30 éves is lehet, még a legidősebb HTML dokumentum is körülbelül tíz éve keletkezett, és a legtöbb tartalom a világhálón nagyon is illékony. Felmérések igazolják, hogy

- az összes weblap fele nem idősebb 100 napnál, s körülbelül a negyedük idősebb csak egyévesnél;
- a .com (üzleti szféra) területen a weblapok 40%-a naponta változik, míg az állami (.gov) és oktatási (.edu) szektorban az oldalak átlagos élettartama négy hónap;
- a világhálón megjelenő tartalom átlagos élettartama két év;
- az URL-ek átlagos élettartama négy év;
- a tudományos oktatásban használt URL-ek átlagosan ötvenöt hónapig élnek.

Az biztos, hogy az információkeresés alapvető törvényeinek megállapításakor senki sem sorolta a dokumentumok közé ezeket a pillanatfelvételeket.

A weblapok kulturális konvenciók alapján létrehozott termékek. A hálón fellelhető tartalom csak egy formai megjelenítő segítségével férhető hozzá, ezek a böngészők. A hálólapok megjelenése a biztonsági beállításoknak, monitoroknak, biztonságos színeknek, szkripteknek stb. is függvénye. Igazából a weblapok alkotói rengeteg időt és ener-

giát fordítanak arra, hogy konzisztensen jelenítsék meg a tartalmat a különböző platformokon. Ráadásul az, hogy a világháló tartalma úgy jelenik meg a böngészőnk ablakában, mint egy nyomtatott dokumentum, nem technikai szükségszerűség, hanem kulturális konvenció. A böngészők a felhasználók többsége társadalmi elvárásainak engedelmessé válnak, vagyis az információnak a nyomtatott oldalra kell hasonlítania, s ráadásul teljesen testre szabható is.

Az a honlap, amelyet az ember a böngészőjében lát, s melynek forrása a <HTML> és </HTML> elemek között található, azt mutatja, hogy a forrászerről hogyan kapja meg az adatot, arról viszont nincs tudomása, hogy az eredeti szerverten a tartalmat milyen formában tárolják. A világháló korai éveiben a legtöbb hálólapot HTML nyelven és kézzel írták. A weblapok nagy része még mindig így készül, az alkalmazhatóság aránya azonban arra kényszerítette a nagy mennyiségű tartalmat szolgáltatókat, hogy automatizálják a weblapok létrehozását. Ezért fordulhat elő, hogy a forrászerről megtalálható tartalmat adatbázisban, XML nyelven, vagy egyéb dokumentumban, szkriptekben tárolják. A weblapok a szervert, a programozás és az esetlegesség függvényei. Így történhetett meg a szakadás tartalom és forma között, amelyek a klasszikus információkeresés definíciói szerint összetartozó entitások. A böngészőben megjelenő dokumentum lehet, hogy külsőleg dokumentumra hasonlít, de valószínűleg semmilyen dokumentumra hasonlító eredete nincsen.

A Google nem indexeli a világhálót. Egy index abban segít a keresőknek, hogy megtalálják az információt egy bizonyos adatbázisban. Általában igaz: annak a sikeressége, hogy megfelelően releváns információt találunk egy adatbázisban, attól függ, hogy mennyire ismerjük az adott indexelési eljárást, az adatbázisban alkalmazott technikákat. A Google a legnépszerűbb keresőgép a weben, azonban senki sem tudja, hogy milyen algoritmus alapján dolgozik. Ha ez valaha is nyilvánosságra kerülne, akkor azonnal kihasználnák a weblapok szerkesztői, hogy előnyt szerezzenek, és láthatóbbá váljon az általuk prezentált tartalom. A Google gazdasági fennmaradása, piacvezető szerepe függ attól, hogy megtartsa a titkát. Ez az üzleti szemlélet teljesen különbözik az olyan adatbázis-készítőkétől, mint például a Dialog Corporation. A Google meg sem próbálja lefedni a teljes világhálót. Annak reményében teljesen felesleges tehát metaadatokat elhelyezni weblapunkban, hogy ez

alapján a Google megtalálja azokat. A Google ugyanis nem fogja elmondani, hogy felhasználja-e a kulcsszavainkat, a metaadatokat, ugyanis a vállalat túlélő stratégiája fekete doboznak maradni.

* * *

A világháló technológiáját felmérve azt láthatjuk, hogy a weblapok a megfelelő metaadatoknak nem jó hordozói. Nem a metaadatokkal mint koncepcióval van itt gond. Azt a következtetést vonhatjuk le, hogy ezek rosszul alkalmazhatók egy olyan technológiában, amelyet állandóan változó, önkényes tartalmi megjelenítések jellemeznek, s amelyeket ismeretlen algoritmusok alapján tárnak fel. A metaadatok alkalmazása csak akkor térül meg, ha olyan információs struktúrában alkalmazzák őket, amely időben állandó, felépítése, tartalma és identitása előre meghatározható. A metaadatok helyes alkalmazása a világhálón csak akkor lesz eredményes, ha a technikai és társadalmi körülmények támogatják az információkeresés hagyományos, információtároló dokumentum felépítését. Ezt úgy lehet elérni, ha lezárjuk a világhálót.

A hagyományos információkeresést támogató környezetet a világhálón is létre lehet hozni, jel-szavak mögött, intraneten, digitális könyvtárakban. Ezeket az alkalmazásokat olyan társadalmi csoportok hozhatnák létre, amelyek megállapodnak abban, hogy milyen információs struktúrát és metaadatokat használnak fel ahhoz, hogy az információ valóban visszakereshető legyen. Az ilyen elzárt hálózaton az ember hagyományos adatbázi-

sokat hozhatna létre, és klasszikus információkeresést hajthatna végre.

A mostani trend alapján, és ez várható a közeljövőben is, a webes keresők a nyitott hálózat több milliárd honlapját fogják átnézni. Az emberek arra használják a Google-t, hogy információt találjanak, és sokan információkeresésnek titulálják tevékenységüket, annak ellenére, hogy rengeteg felesleges adatot, rossz linket, eltűnő weblapokat fognak kapni eredményként. Inkább nevezhető ez a tevékenység keresésnek, állandó felfedezésnek. Igazából a világháló egy mesterkéletlen környezet, ahol nincs garancia arra, hogy a prezentációk megmaradnak, vagy hogy a szerverek sokáig működnek. Az egyetlen módja annak, hogy megőrizzük a világhálón fellelt tartalmat, ha pillanatfelvételt készítünk róla, ezáltal változatlan formában őrzünk meg egy állandóan fejlődő folyamatot.

Az információkereséssel hivatásosan foglalkozóknak át kell rágniuk magukat a meglévő módszereken, hogy olyan elképzelést találjunk, amelyet alkalmazni lehet a világhálóra. De még az is elképzelhető, hogy a világháló annyira új technikai találmány, hogy az eddig alkalmazott információkeresési hagyományok, fogalmak és módszerek a 20. század végi adatbáziskorszak óta túlhaladottká váltak.

/BROOKS, Terrence A.: Web search: how the Web has changed information retrieval. = Information Research, 8. köt. 3. sz. 2003. <http://informationr.net/ir/8-3/paper154.html> 12 p./

(Lepp Tünde)

A digitális formátumok globális regiszterének terve

A digitális formátumok problémája a digitális repozitóriumok architektúrájának és működésének minden szintjén érezteti hatását. Az anyagok befogadásával, tárolásával, a hozzáférés biztosításával és a megőréssel kapcsolatos döntések általában a formátumokkal függnek össze. A digitális objektumok hatékony cseréje ugyancsak megköveteli, hogy legyen közös megegyezés a formátumok szintaxisáról és szemantikájáról.

A formátumok egyetlen globális számbavételét a *MIME (Multipurpose Internet Mail Extensions)* Media Types regiszter biztosítja. Sok digitális repozitórium számára azonban a MIME kategóriák

nem eléggé részletesek. Ebben a regiszterben például az összes, egymástól különböző TIFF képfarmátumot *image/tiff* azonosítóval, a PDF teljes termékcsaládját pedig az *application/pdf* megjelöléssel rögzítik. Annak elkerülésére, hogy a különböző formátumokat azonos módon jelöljük, javaslat született egy, az egyértelmű azonosítást lehetővé tevő regiszter létrehozására. Fontos, hogy a regiszter részletes, megbízható megjelenítési információt nyújtson a formátumokról. A MIME regiszter nem egységesen kezeli a formátumok belső szerkezetének leírását. A javasolt regiszter célja egy megfelelő mechanizmus kifejlesztése a védett formátumok leírására, s ezen információk