

Tudományos és Műszaki Tájékoztatás

Könyvtár- és információtudományi
szakfolyóirat

57. évfolyam
HU ISSN 0041-3917

2010/08

Tartalom

Ajánló

DRÓTOS László: **Keresőgépesítés**

Cikkek

TÓTH Erzsébet: Az internetes keresők tárgyköri fogalomrendszere	319
TÓTH Erzsébet: Az internetes keresők működésének technikai háttere	326
WINKLER Bea: SEO a könyvtárban, avagy a SZIE Állatorvos-tudományi Könyvtár új portálja és a keresőoptimalizálás	335

Beszámolók • Szemlék • Referátumok

SALO, Dorothea: Névkezelés (authority control) az intézményi repositóriumoknál (Ref.: Dudás Anikó).....	345
HARINARAYANA, N. S. – VASANTHA RAJU, N.: Webkettes szolgáltatások az egyetemi könyvtárak honlapjain (Ref.: Drótos László)	350
BARRY, Eithene – BEDOYA, Jaclyn Kelli – GROOM, Carolyn – PATTERSON, Laurence: Virtuális referenz a brit felsőoktatási könyvtárakban (Ref.: Drótos László)	353
KAUR, Kiran: Egyetemi könyvtár marketingje a weben (Ref.: Naszádos Edit)	356
WAY, Doug: Olvasói könyvtárközi kérésre alapozó gyarapítás (Ref.: Burmeister Erzsébet).....	358



Scientific and Technical Information

Hungarian journal of library and information science

Vol. 57 • 2010/08

Summaries

TÓTH, E.: The system of subject concepts in internet search engines 319

Globalisation is affecting information search on the internet to a large extent, and its influence manifests itself in an increased degree in a sharp competition between search engines. Search effectiveness and its development are key issues in this competition, and also in the competition of users, as well as in the competitiveness in search marketing. The article analyses the comprehensive relationships of this complex theme, the impact of globalisation on searching, on internet-based information search, on its strategies, opportunities and models. It deals with the increasingly visual character of the internet (accessibility of images, videos etc.), and with the growing intensity of efforts to explore the invisible/deep web. The article discusses in detail the rich and complex network of related concepts. It presents the concept- and terminology-related issues of internet search engines; it highlights the relationships between concepts and synonyms. Further, it reviews related research areas, mentioning the invisible/deep web, data mining, and text mining etc., among others.

TÓTH, E.: Technical background of the operation of internet search engines 326

The essay provides an overview of the essential components of search engines and of the tasks they perform. It describes in detail the activities of search robots harvesting web pages, and discusses the importance of data files storing the web pages harvested. It summarises the tasks of indexers who analyse and extract relevant expressions from the web pages visited. Search engines primarily identify the indexed documents exactly matching the search questions. Ranking by various

aspects is another major characteristic of their operation. The essay lists the main principles of ranking hits, generally applied by search engines. Currently, the PageRank algorithm of Google enjoys special attention, and the essay presents a corrected version of the one spread in the library profession. The underlying idea of this algorithm is described, and how it models the behaviour of an „accidental surfer”. Finally, it tackles various problems related to internet-based searching, and the relevant solution attempts offered.

WINKLER, B.: SEO in the library, or the new portal of the University of Veterinary Science Library and search optimization335

Striving to meet user needs and requirements, most libraries are present on the World Wide Web. However, it is often not enough simply “to be present” if libraries want to capture the attention of people surfing on the internet. The Hungarian Veterinary Science Library, Archives and Museum wants to increase the number of visits to its new homepage, and uses Search Engine Optimization (SEO) as the most obvious tool to achieve this goal. The case study does not offer a general algorithm to SEO; its aim is rather to illustrate how SEO can be used in building a portal, and to call attention to the importance of SEO. The article analyses the present homepage of the library, and identifies the most interesting problems which arose during the development of the new portal, from the point of SEO. Some free tools used in SEO are also described. The major issues are the following ones: examining the sources of usage, keywords in search engines and in the homepage, trends of searching, keyword density, important tags, ranking of hit lists, using JavaScript, internal and external links, etc.

Tóth Erzsébet

Az internetes keresők tárgyköri fogalomrendszere

Az írás áttekinti a magyar nyelvű szakirodalom lényeges megállapításait az internetes keresők tárgykörében, rávilágítva annak gazdag és összetett fogalmi hálójára. Tárgyalja az internetes keresők fogalmi és terminológiai kérdéseit, rámutat az itt előforduló fogalmi kapcsolatokra. A szerző szerint egy ilyen jellegű szakirodalmi áttekintés hasznos kiindulópontként szolgál a tárgykör tanulmányozásához a felsőoktatási gyakorlatban.

Bevezetés

Korunk számos folyamata közül csak a legjelentősebbeket emelném ki, amelyek a következők: a politikai, katonai értelemben vett egyhatalmú világ kialakulása, a demokratizálódás, az európai integráció, a fokozódó ázsiai gazdasági együttműködés, a technológia forradalma, a globalizáció, az információs és a fogyasztói társadalom kialakulása. Az első és az utolsó három a világ minden részére közvetlenül, vagy közvetve ható, globális folyamatnak tekinthető. Napjainkban a „globális” jelzőt több, tartalmában eltérő jelenségre is rá lehet húzni, azonban helyesebb, ha csak a gazdaság, a tőke, az ipar, a kereskedelem, a szolgáltatás földrajzi és módszertani terjeszkedését nevezzük globalizációnak.

Az információs és kommunikációs technológia (IKT) rohamos fejlődésének és konvergenciájának eredményeként a társadalomban egy új életforma, újszerű működés és viselkedés alakult ki. Új értékrendek jöttek létre. Ezt a széles körben elterjedt új életmódot, magatartást, információs technológiára épülő gazdaságot nevezzük információs társadalomnak [2]. Az „információs társadalom” kifejezés az 1960-as évek második felében jelent meg, az '50-es '60-as évek fordulóján jöttek létre a „tudástársadalom”, „tudásgazdaság” stb. összetételek. Az „információs társadalom” a világban fellelhető információk általános gazdagságát tükrözi, míg a „tudástársadalom” arra a gazdagságra hívja fel a figyelmet, amit a tudás teremt, és arra a szegénységre, amelyet a tudástársadalom viszonyai között a tudás hiánya okoz [6]. Az információs társadalom kialakulása országonként különböző időszakban és ritmusban zajlik. A társadalom tág értelemben vett fejlettségi szintje és a kultúra nagymértékben befolyásolja azt. Lényeges, hogy erre a folyamatra

sokkal „könnyebben” lehet hatni, társadalmi szinten jó irányba terelni, mint a globalizációra. A fogyasztói társadalom megjelenése a globalizációhoz és az információs társadalomhoz kapcsolódó harmadik jelentős folyamat [2].

A könyvtárak viszonylag korán felismerték annak fontosságát, hogy meg kell felelniük az információs társadalom kihívásainak. Az *Európai Unió* könyvtárakkal kapcsolatos tevékenysége az *EU 3-4. Kutatás-Fejlesztési*, azon belül *Telematikai Keretprogramja* keretében ment végbe. 1991–1994 között a 3. keretprogramban három pályázati felhívás jelent meg, melyek eredményeképpen 81 ún. akcióterv kezdődött el. Ebből mintegy kétszáz intézmény részvételével 51 közösen finanszírozott projektre került sor. 1995–1998 között a 4. keretprogram nyitott volt a közép-kelet-európai országok számára is. Két pályázatot írtak ki, 15 kutatási projekt, 7 összehangolt közös nagy munka és 20 újabb, ún. horizontális támogatási program indult. A könyvtárakkal kapcsolatos európai uniós programok részletes ismertetését lásd [1]-ben.

A globalizáció gyorsuló és könnyörtelen versennyel jár együtt. Csak úgy lehetünk versenyképesek, ha az információs társadalom legfőbb értékét, magát az információt részesítjük előnyben. Egy adott szakmai kérdés megválaszolásának a leggyakoribb kiindulási pontja lehet számunkra az internet, amely a minket körülvevő globális társadalomnak egyik fontos eszköze [7]. Azonban az internetről nem tételezhető fel, hogy az minden feltett kérdésre kielégítő választ fog nyújtani. A világhálón történő információkereséskor egyre nagyobb gondot jelent számunkra a minőségi, releváns információk felkutatása és kiválogatása a ránk zúduló információáradatból. Ebben támogatnak minket a rendelkezésre álló keresőszolgáltatások, bár azok

sem minden esetben nyújtanak tökéletes megoldást.

Az interneten megjelenő keresőeszközök köré külön iparág szerveződött, amelybe kisebb-nagyobb méretű cégek, vállalatok nagy pénzeszegeket fektetnek be. Mindezt pedig saját versenyképességük, sikerességük és hatékonyságuk javítása érdekében teszik. A számadatok érzékeltetésére a *Search Engine Marketing Professionals Organization (SEMPO)* 2005-ös felmérésének lényeges megállapításaira utalnék: az Egyesült Államokban és Kanadában 5,75 milliárd dollárt költöttek 2005-ben keresőmarketingre (*SEM = Search Engine Marketing*). Ez az összeg 44%-kal haladta meg a 2004-es költségeket. Az előrejelzések szerint a keresőmarketingbe fektetett pénzeszeg 2010-re elérheti a 11 milliárd dollárt Észak-Amerikában [12]. A SEMPO 2009-es felmérése szerint a keresőmarketing-ipar Észak-Amerikában 16,6 milliárd dollárra növekszik 2010-re [13]. *Safa Rashtchy* internet média- és marketingelemző szerint a keresőpiac fizetős része 2005-ben megközelítőleg 10 milliárd dollár hasznot termelt globálisan, ami 41%-kal fog növekedni 2006-ban. Előrejelzése szerint a keresőpiacnak ez a része globálisan 37%-os éves növekedésre számíthat 2010-ig, ami több mint 33 milliárd dollárnak felel meg. A 2005-ös felmérésből kiderült, hogy a keresőmarketing-kampányok elsődleges célja a „branding” (egy márka ismertté tétele) és az értékesítés volt. A kisebb cégek inkább a termékek eladására helyezték a hangsúlyt, míg a nagyobbak (500 alkalmazott feletti) a weboldalukra érkező forgalom növelésére [12]. Mindkét felmérés adatai rávilágítottak arra, hogy a keresőszolgáltatások, valamint a különféle cégek, vállalatok abban érdekeltek, hogy minél több bevételre tegyenek szert. Ebben a kiélezett versenyhelyzetben a keresőeszközök folyamatosan törekszenek arra, hogy megújuljanak és minél több speciális, új szolgáltatással vonzzák a használókat maguk köré. Ezért a versenyben részt vevő szereplők számára rendkívül fontos, hogy az egyes keresőeszközök minőségét hogyan értékelik a kutatók.

Az információkeresésre irányuló kutatás több mint két évtizedes múltat tekint vissza. Ezen a területen a vizsgálódás egyik lehetséges iránya a kérdést feltevő felhasználók viselkedésének tanulmányozása, azaz milyen kérdést, hogyan, és milyen társadalmi rétegből, milyen tanultságúak tettek fel. A kutatók 1981-től számos modellt alkottak meg. A modellek kialakítását nagymértékben befolyásolta a kutatók világlátása, kutatási területe és jártassá-

ga. Ennek függvényében beszélhetünk kognitív perspektivikus, szociális, szociális-kognitív vagy szervezeti modellekről [7]. Mindez azt tükrözi számunkra, hogy a felhasználók weben történő keresése több szinten vizsgálható, beleértve a társadalmi és a szervezeti szintet, az információkeresés szintjét, az ember és a számítógép közötti kapcsolat szintjét, valamint a megfogalmazott keresőkérdés szintjét [15]. Csak jelzésszerűen hivatkoznék néhány jeles kutatóra, akik ezen a téren komoly eredményeket értek el: *Spink, Jansen, Saracevic, Ingwersen*. Az információkeresési viselkedésekkel, modellekkel kapcsolatos kutatási eredmények megtalálhatók [7]-ben.

A vizsgált témakör fogalmai

Fontosnak tartom, hogy meghatározzam ennek a komplex tárgykörnek az alapvető fogalmait és a közöttük lévő kapcsolatrendszerét. Először a „metaadat” fogalmának meghatározásával kezdeném, mert a hozzá tartozó információknak a megléte szükséges az internetes keresők működéséhez. Metaadat kifejezésen a weblapok intellektuálisan vagy automatikusan létrehozott másodlagos adatait értjük, amelyek magát a dokumentumot jellemzik [24]. A keresőrendszerek a saját adatbázisukat csupán olyan technikai metaadatokkal látják el, mint a begyűjtött dokumentum URL címe, fájlformátuma, mérete, begyűjtési dátuma stb. Egy másik meghatározás szerint metaadat alatt mindazokat a többletinformációkat értjük, amelyeket a weboldalak készítői a weboldalakhoz kapcsolnak a keresőkérdés pontosabb megválaszolása reményében [19]. Ezen adatok körébe tartoznak: a bibliográfiai leírás szabványosított adatelemei, a dokumentum tartalmát leíró kulcsszavak, tárgyszavak, deskriptorok és az osztályozási jelzetek. Elengedhetetlen követelmény volt a metaadatok egységes elektronikus kezelése, ami kiterjed ezeknek az adatoknak az elsődleges dokumentumokból való kinyerésére és a dokumentumok számítógépes leírására [24]. Számos metaadatrendszer jött létre a hálózati információk feldolgozására, például az *OCLC InterCat*, a *DublinCore*, a *WWW Semantic Header*, a *TEI (Text Encoding Initiative)* fejléc stb. Ezek közül a metaadatrendszerek közül a *DublinCore* jelentőségét hangsúlyoznám, mivel napjainkban ez az egyik legáltalánosabban elterjedt metaadat-alkalmazás. A *Dublin Core* formátum 15 leíró elemet tartalmaz, és ez áll a legközelebb a könyvtári katalogizáláshoz. Elterjedését elősegítette, hogy adatelemeit az európai szabványosítási szervezet, a *European Committee*

for Standardization (CEN) is elfogadta [8]. A Dublin Core-ra vonatkozó magyar nyelvű szabvány letölthető a mek.oszk.hu/dc oldalról.

A keresőszolgáltatásoknak két típusát különböztethetjük meg: az indexelőszolgáltatásokat és az internetkatalógusokat. Az előbbieken belül különleges változatként fordulnak elő a gyűjtő- és a metakeresők. A metakeresők (*meta search engines*, *Meta-Suchmaschinen*, *métamoteur*, *méta-chercheur*) segítségével több indexelőszolgáltatásban kereshetünk párhuzamosan anélkül, hogy az egyes szolgáltatásokkal külön foglalkoznunk kellene. A rendszer mindegyik keresőszolgáltatás adatbázisában végrehajtja a keresést, megjelenítve a találatoknál, hogy melyik szolgáltatás adatbázisában találta meg a rekordot, valamint a duplumszűrésre is törekszik. A metakeresők előnye, hogy rövid idő alatt valószínűsíthetően több releváns találatot juthatunk [23]. Továbbá, nehezebben csapják be őket azok az oldalak, amelyek mindenféle trükkös megoldásokkal a javukra befolyásolják a keresők találatrangsorolását, azonban ezeknek az oldalaknak nincs igazi, használható tartalmuk; „spam”-eknek hívjuk őket. A metakeresők azért képesek a „spam”-oldalak kiszűrésére, mert azok általában egy-egy keresőre szakosodnak és egyszerre több keresőt már nem tudnak becsapni [19]. A „spamdexing” kifejezés a „spamming” és az „indexing” szavak összeolvadásából született, amely a '90-es évek közepén jelent meg a keresőiparban. A *search spam*, *search engine spam*, illetve a *web spam* kifejezéseket szintén használjuk rá. Ez a folyamat számos módszert foglal magába, amelyeket azért alkalmaznak, hogy a kereső által indexelt oldalak relevanciáját vagy fontosságát növeljék. Használt módszerei azonban nincsenek összhangban a kereső indexelésének célkitűzésével. Néhányan úgy vélekednek, hogy a spamdexing a keresőoptimalizálás részét képezi. Több kereső ellenőrzi a spamdexing előfordulásait és eltávolítja a gyanús oldalakat indexéből [14]. Átmeneti típusnak tekinthető a gyűjtőszolgáltatás (*configurable unified search interface [CUSI]*, *all-in-one formular*, *sample service*, *Sammeldienst*), amely több keresőszolgáltatást ajánl fel, de mindig csak egyet választhatunk ki a lekérdezésre [23].

Az indexelőszolgáltatások („keresőgépek”-nek is hívjuk őket), (*search engines*, *Suchmaschinen*, *moteur de recherche*) emberi munka nélkül, számítógépes programok segítségével végzik a keresést a hálózaton. Ezek a szolgáltatások két fő részből állnak: a keresőrobotból (*crawler*, *web spider*, *web*

robot, *bot*) és az indexelőből (*indexer*). A robotok állandóan figyelemmel követik és begyűjtik a weboldalakat a világhálóról a keresőszolgáltatás adatbázisába. A webhelytulajdonosok adhatnak utasításokat a robotoknak begyűjtéskor, ekkor egy *robots.txt* állományt kell elhelyezniük a webhely gyökérkönyvtárában. A robotok úgy vannak kialakítva, hogy követniük kell az utasításokat, ezért megpróbálják megtalálni a *robots.txt* állományt és elolvasni az utasításokat belőle, mielőtt a webhelyről bármilyen állományt begyűjtenének. Ha ez az állomány nem található meg, akkor feltételezik, hogy a webtulajdonos nem kíván speciális utasításokat meghatározni számukra. A *robots.txt* állomány valójában egy olyan kérés a webhelyen, amely megszabja, hogy egyes robotok bizonyos állományokat vagy könyvtárakat figyelmen kívül hagyjanak begyűjtéskor. Ha a webhely több aldoménból áll, akkor azok mindegyikének rendelkeznie kell a saját *robots.txt*-jével [9]. Az indexelő elemzi a begyűjtött dokumentumokat, amelyekből előállítja az indexkifejezéseket. Létrehoz egy indexet, amely minden szóhoz – a stopword-öket kivéve – hozzárendeli az őt tartalmazó *Uniform Resource Locator*-ok (*URL*) listáját. A keresőszolgáltatás erre az indexre támaszkodik, amely révén elvégzi a keresést a felhasználó számára [19]. A keresőrobotot és az indexelőt integráló egységet „keresőgépnak”, „keresőmotornak”, „keresőműnek” (*search engine*), „keresőrendszernek” (*search system*) nevezik. Hibásan a teljes keresőszolgáltatást is „keresőgépnak”, „keresőmotornak”, „robotnak” hívják, ami a szolgáltató rendszernek csak az egyik részét jelenti. Ebbe beletartozik még a felhasználói felület és a szolgáltatott tartalom is [23]. Ezek a keresőszolgáltatások általában rendelkeznek egy egyszerű és egy összetett keresési lehetőséggel. Egyszerű kereséskor (*quick search*) rendkívül nagy lehet a visszakeresett, nem releváns dokumentumok száma, azaz a zaj. Ennek csökkentése érdekében tanácsos használnunk a részletes keresési lehetőséget (*advanced search*, *powered search*) [26].

Amikor egy vagy több releváns kulcsszót írunk be a keresőablakba, a kereső indexében megvizsgálja, hogy melyek a kérdésünkre legjobban illeszkedő találatok és azokat szolgáltatja számunkra. A találatlistában szereplő oldalakról általában egy rövid ismertetést kapunk, amely magába foglalja a forrás címét, valamint annak kiemelt szövegrészeit. Találati halmazaink mennyiségi viszonyait (a halmazok egymáshoz viszonyított terjedelmét, illetve helyzetét) logikai műveletekkel adhatjuk meg. Ezeket a műveleteket pedig logikai műveleti

jelekkel – ún. operátorokkal – fejezhetjük ki. A legtöbb kereső támogatja az ÉS, VAGY, NEM Boole-operátorok használatát, amelyekkel a keresés tovább finomítható. A keresők egy része megengedi a helyzeti operátorok (*proximity operators*) használatát is, amelyek lehetővé teszik számunkra, hogy meghatározzuk a kulcsszavak közötti távolságot (pl. NEAR, BETWEEN, WITH operátorok stb.). Kereséskor a találati halmaz terjedelmét úgy módosíthatjuk, hogy megengedjük, hogy a keresőszó elején, végén vagy meghatározott karakterpozícióin bármilyen karakter helyezkedjen el. Ehhez „jolly joker” jeleket (*wild card*) adhatunk meg a keresőszóban. Bővebb találati halmazokat nyerhetünk abban az esetben, ha a keresőszó elején („balról csonkolás”) és/vagy végén („jobbról csonkolás”) meghatározott karaktert használunk, amely minden megelőző és/vagy követő karaktert helyettesít. Ezt a műveletet csonkolásnak (*truncation*) nevezzük. A csonkoló jelek használata keresőrendszerenként eltérő [26]. A keresőknél létezik egy kifinomult keresési technika, a fogalomalapú keresés (*concept-based searching*). Ennél a technikánál statisztikai elemzéssel találjuk meg azokat az oldalakat is, amelyek nem tartalmazzák az általunk megadott kulcsszavakat. Ekkor azonban az oldalak olyan egyéb szavakat (pl. szinonimákat, tulajdonneveket, állandósult szókapcsolatokat) foglalnak magukba, amelyek ugyanabba a fogalomkörbe tartoznak, mint a beírt keresőszavak. Így a keresőrendszer akkor is relevánsnak minősíti az oldalakat, ha a megadott keresőszavak nem találhatóak meg bennük. Egy másik kereső funkció a fuzzy megfeleltetés/illesztés (*fuzzy matching*), amelynek az a lényege, hogy a keresőszót a szótőre redukálják és minden lehetséges szóalakot ráillesztenek különböző algoritmusokkal. Ez nagymértékben megnöveli a találati halmazunkat, mert minden kapcsolódó szót visszakeres, még a kevésbé relevánsakat is. Néhány keresőnél alapértelmezett funkció a *stemming*, ami a keresőkérdés összes toldalékolt alakjának a visszakeresésére alkalmas. Ha ezt a funkciót használjuk a keresőkérdésre, akkor még bővebb találati halmazt kapunk a csonkoláshoz képest.

Megállapítható, hogy egy kereső hasznossága valójában a szolgáltatott találatlistája relevanciájától függ. A legtöbb kereső rangsorolja a találatokat fontosságuk szerint arra törekedve, hogy a legjobb oldalakat jelenítse meg a találatlista elején. Keresőnként változó, hogy milyen rangsorolási módszert alkalmaznak erre a feladatra. A *Google Page Rank* algoritmus az egyik legismertebb rangsorolási módszer, amely az oldalak közötti linkstruktú-

rát veszi alapul és más egyéb tényezőket egyaránt figyelembe véve súlyozza a találatokat. Beszélhetünk olyan keresőkről is, amelyek nem egy egyszerű találatlistában jelenítik meg a találatokat, helyette inkább a keresőkérdéshez kapcsolódó tematikus kategóriákba rendezik azokat. Ezek a csoportok (klaszterek) abban segítenek bennünket, hogy könnyen áttekinthessük a keresett témát, és hogy kiválaszthassuk a megfelelő kategóriát. A találatok klaszterálása segítséget nyújt a keresés finomításában a korábbi keresés találati halmazára támaszkodva (pl. *clusty.com* kereső) [4]. Találkozunk olyan vizuális keresőeszközökkel is, amelyek a találatokat grafikusán jelenítik meg (*graphical visualization*) két- vagy háromdimenziós képekben (pl. *viewzi.com*, *eyexplorer.com*). Az internetkatalógusokat (*directories*, *annuaires internet*, *répertoires internet*) [26] „böngészőszolgáltatásnak” (*browsing service*, *browsing Dienste*) [24], „tárgyszótárnak”, „tématárnak” (*subject directory*, *Themenverzeichniss*, *annuaire thématique*) [21], valamint „webes katalógusnak” (*annuaire Web*, *répertoire Web*) is nevezik [19]. Továbbá a „linkgyűjtemény” és a „tematikus katalógus” megnevezések is ismertek. Ezek a katalógusok hierarchikus osztályozási rendszert használnak. Adatbázisaik többnyire intellektuálisan feldolgozott weboldalak rekordjait foglalják magukba, valamint kapcsolókat más adatbázisokhoz. Az osztályozást és a tartalmi kivonatok készítését szerkesztőségben végzik. Azonban sok linkgyűjtemény egyéni vagy közösségi munka eredménye és nincs mögötte szerkesztőség, lásd például a „Startlap” tematikus oldalait. Ezekben a katalógusokban osztályok alapján böngészhetünk, de lehetőségünk van arra is, hogy egy keresőkérdés megadásával, célzott kereséssel találjuk meg a kívánt osztályt. Vannak olyan katalógusok is, amelyek indexelőszolgáltatásként is működnek, ilyen például az *ok.hu/linktar*. Az internetkatalógusok adatbázisai sokkal kisebbek, mint az indexelőszolgáltatásokéi, azonban a keresés kevesebb zajt eredményez az intellektuális feldolgozásnak és a gondos osztályozásnak köszönhetően. A szakterületre specializálódott keresők nagy része internetkatalógusnak tekinthető. Egy részüket híres kutatóintézetek gondozzák, más részük kereskedelmi szolgáltatásnak minősül [26].

Kapcsolódó kutatási területek

Elsősorban a szemantikus webnek, mint perspektivikusan fejlődő területnek a jelentőségét hangsúlyozom, melynek feladata a jelentés megtalálása a webes tartalmakban. A szemantikus web kialakítá-

sára irányuló törekvések nyomán jelentek meg az ún. ontológiák. Gruber megfogalmazása szerint az „ontológia megegyezésen alapuló fogalmi rendszer formális, egyértelmű leírása” [3]. Ebben a meghatározásban a „megegyezésen alapuló” kitétel lényeges, hiszen azt a szemléletet tükrözi, hogy az ontológiák szemantikai szabályrendszerek, amelyek a dolgok rendezésére használhatók [25]. Az ontológiák lehetővé teszik, hogy tisztázzuk az alapvető fogalmakat és a közöttük lévő relációkat. Továbbá elősegítik, hogy az erre vonatkozó tudásunkat formálisan és gépi következtetésre alkalmasan fogalmazzuk meg [18].

Számos fejleménynek kellett ahhoz bekövetkeznie, hogy webes ontológiák jöhessenek létre. Ezek közül csak a legfontosabbakat emelném ki. 2000-ben közreadtak egy „tématérképnek” (*topic map*) nevezett hierarchikus fogalmi struktúrát kezelő szabványt. A weben jelenleg elérhető vizualizált fogalmi struktúrák többsége ezen vagy ehhez hasonló fejlesztéseken alapszik [27]. A W3C konzorcium irányítása alatt egy másik irányban kezdődött el a fejlesztés. Ennek egyik fontos eredménye, hogy 2000-ben a web metaadatainak leírására egy szabványt hoztak létre, az XML-en alapuló webforrás leíró nyelvet (*Resource Description Framework = RDF*). A weben található hierarchikus fogalmi struktúrák formális leírására is ezt a nyelvet használták fel. 2002-ben a W3C konzorcium kezdeményezésére hozzákezdték az ontológiák szabványának tekinthető webontológia-nyelv (*Ontology Web Language = OWL*) kidolgozásához [25]. Az OWL 2-re vonatkozó szabványajánlást 2009-ben adta közre a W3C konzorcium [5]. Jelenleg elérhető és már létező általános ontológiáknak tekinthetők például a *Dublin Core*, a *Magyar Egyesítés Ontológia*. Szakterületi ontológiaként megemlíthető a *Galen*, amely orvostudományi szakterületen használatos [18]. A „Museo24” projektben kifejlesztett ontológiának érdekes felhasználási területe a virtuális múzeum, amely gondolatvilágában közel áll a könyvtárákéhoz. (Lásd a projekt leírását [17]-ben.) Jelenleg egyfajta közeledés figyelhető meg hazánkban a könyvtári és az informatikai szakmai közösségek között az ontológiák terén, amit a W3C konzorcium magyar irodája szakmai előadások szervezésével egyaránt támogat [18, 22].

A szakirodalomban az „invisible web” (láthatatlan web), „hidden web” (rejtett web), vagy „deep web” (mély web) angol kifejezéseket használják mindazon dokumentumok és adatok körének az össze-

foglalására, amelyek számos oknál fogva nem érhetőek el a keresőszolgáltatások számára. A láthatatlan web csoportjába sorolhatók: a dinamikus weblapok (azaz pl. a kereshető adatbázisokból nyert oldalak), azok az oldalak, amelyek csak regisztráció után érhetőek el, a nem szöveges dokumentumok, valamint a keresőrobotok elől elzárt oldalak. Fontos hangsúlyoznunk, hogy a web csak egy szolgáltatás az interneten, tehát az nem azonos vele. Egy olyan hipertext-struktúrára épül, amelyben szabadon böngészhetünk a szöveges formában megjelenített információk közötti kapcsolatok (linkek) alapján. Ha egy weblapra nem mutat egyetlen link sem, akkor az nem kerül bele a kereső adatbázisába. Azoknak a weboldalaknak az összességét, amelyeket a keresők keresőmotorjai megtalálnak „felszíni webnek” (*surface web*) vagy „statikus webnek” nevezzük. Ennek nagysága a teljes web méretének a 0,18%-ára becsülhető. Ezzel szemben a láthatatlan web információmennyisége 550-szer nagyobb, mint a felszínié és növekedése, gyarapodása is sokkal gyorsabb ütemű [10, 7]. Sokféle törekvéssel igyekeztek a rejtett webet „láthatóvá tenni”, például bizonyos metakeresőkkel, intelligens keresőprogramokkal (ágensek), témakatalógusok kialakításával, egyéb speciális keresőkkel. Mindezeket a lehetséges megoldási kísérleteket, eszközöket bővebben kifejtve lásd [7]-ben.

Ehhez a tárgykörhöz kapcsolódóan hivatkoznék a szövegbányászat és az adatbányászat ígéretes lehetőségeire, amelyek a rejtett tudás kinyerésére törekednek a weben található, nagy mennyiségű strukturálatlan vagy félig strukturált HTML és egyéb formátumú dokumentumokból. Fiatal kutatási területnek számít a „web mining”, amely kiterjed az adatbányászatra, az internettechnológiákra, valamint a szemantikus webre [11].

A weben találkozhatunk speciális keresőszolgáltatásokkal is, például képek, videoanyagok visszakeresésére alkalmas keresőkkel, amelyek nagy népszerűségnek örvendenek a használók körében. Megjelenésük azt jelzi, hogy a használók rendkívül nagymértékben igénylik a nem szöveges dokumentumok eredményes megtalálását is. Ezen az új kutatási területen a megfelelő információkereső nyelvek létrehozása és azok további fejlesztése elengedhetetlenül fontos feladat mellett, hogy a tartalomgazdák metaadatokat helyeznek el a kép- és videofájlokba, továbbá, hogy egyre fejlettebb kép- és beszédfelismerő eszközöket használnak a keresőgépek.

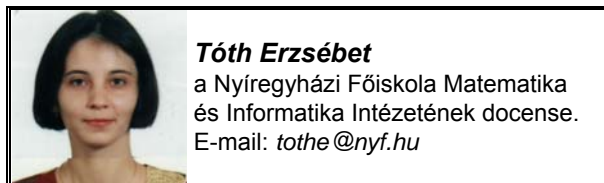
Egy másik, szerkezeti sajátosságaiból adódóan elkülönülő dokumentumcsoportot is meg kell említenünk a weben: a blogokat és a mikroblogokat (pl. Twitter). Számuk rohamosan növekszik, mert divatos véleménynyilvánítási forma a használok körében. Komoly kihívást jelent napjainkban a blogok hatékony visszakeresésének megoldása, amely a hagyományos információkereső rendszerek módszereitől eltérő mechanizmusokat követel meg. Ez abból is adódik, hogy a blogoknál rendkívül nagy szerepe van az aktualitásnak és a kapcsolódó linkeknek, azonban sokszor nehéz meghatározunkunk a témájukat. Megjegyzem, hogy az időtényező fokozottabb kiaknázása a keresésekben új vonásnak számít, például a blogok és a hírek esetében [4, 16]. 2006-ban a korábbi passzív internetezők önszerveződő közösségek aktív tagjaivá váltak, ami főként a web 2.0 térhódításának volt köszönhető. Ezáltal a web rendkívül nyitott közösségi színtérré alakult át az innováció, a kibontakozás, valamint az értékteremtés számára. A Google és más üzleti cégek fokozatosan teret engednek a tömeges együttműködés kultúrájának, még pedig olyan formában, hogy nyíltan hozzáférhetővé teszik alkalmazásaik programozófelületét (az API-t) platformjaikon. Az API-k megnyitása után a fejlesztők (akik közül néhányan korábban „hekkerek” voltak...) gyors tempóban kezdték el gyártani az új alkalmazásokat [20]. A Google számos szolgáltatásával támogatja az egyedi felhasználók kényelmét kereséskor. Ezek például a következők: *Google Suggest*, *Custom Search*, *Google Alert*, *Desktop Search*, *Google Toolbar*. Kifejezetten a közösségi tartalmak visszakeresésére alkalmas keresőként megjegyezhető például a *grub.org*.

Irodalom

- [1] Creating a European library space: Telematics for libraries programmes 1990–1998. = <http://cordis.europa.eu/libraries/en/intro.html> (2008.01.22.)
- [2] FODOR, I.: Merre megy a világ gazdasága, merre mehetünk mi? = Az információs társadalom. Összeáll. Demetrovics J., Keviczky L. Budapest, MTA, 2000. p. 95–113.
- [3] GRUBER, T.: A translation approach to portable ontology specifications. = Knowledge Acquisition, 5. köt. 2. sz. 1993. p. 199–220.
- [4] LANGVILLE, A. N. – MEYER, C. D.: Google's PageRank and beyond. The science of search engine rankings. Princeton-Oxford, Princeton University Press, 2006.
- [5] OWL 2 Web Ontology Language Document Review. (2009). = <http://www.w3.org/TR/2009/PR-owl2-overview-20090922/> (2010.04.09.)
- [6] NYÍRI, K.: Globális társadalom, helyi kultúra. = Az információs társadalom. Összeáll. Demetrovics J., Keviczky L. Budapest, MTA, 2000. p. 43–64.
- [7] PAJOR, E.: A láthatatlan/mély web felhasználása a könyvtári tájékoztatásban. [Doktori (PhD) értekezés]. Budapest, ELTE BTK, 2006. 214 p.
- [8] RÁCZ Ágnes.: A kiadványok bibliográfiai számbavétele; leíró katalogizálás. = Könyvtárosok kézikönyve. 2. köt. Feltárás és visszakeresés. Szerk. Horváth Tibor – Papp István. Budapest, 2003, Osiris. p. 187–295.
- [9] Robots Exclusion Standard (Robot Kizárási Szabvány) szócikk. = http://en.wikipedia.org/wiki/Robots_Exclusion_Standard (2010.04.10.)
- [10] RUTKOVSKY, E. – RUTKOVSKY, Á: A láthatatlan web keresése. (2003). [Előadásanyag]. = <https://nws.niif.hu/ncd2003/docs/ehu/EHU-61.htm> (2008.01.15.)
- [11] SCIME, A.: Web mining. Applications and techniques. 2005, Idea Group Inc.
- [12] SHERMAN, C.: The state of search engine marketing. (2006). = <http://searchenginewatch.com/showPage.html?page=3575926> (2008.01.20.)
- [13] SHERMAN, C.: The State Of Search Engine Marketing 2010. (2010). (A cím félrevezető lehet, mert a legfrissebb felmérés 2009-es, amit a SEMPO elvégzett. A SEMPO honlapján 2010. április közepén a 2009-es év felmérése érhető el csak tagoknak vö. http://www.sempo.org/learning_center/research/ <http://searchengineland.com/the-state-of-search-engine-marketing-2010-38826> (2010.04.10.)
- [14] Spamdexing szócikk. = <http://en.wikipedia.org/wiki/Spamdexing> (2010.04.10.)
- [15] SPINK, A. – JANSEN, B. J.: A study of web search trends. = Webology, 1. köt. 2. sz. 2004. <http://www.webology.ir/2004/v1n2/a4.html> (2008.01.27.)
- [16] SULLIVAN, D.: What is real time search? Definitions & players. (2009). = <http://searchengineland.com/what-is-real-time-search-definitions-players-22172> (2010. 04.10.)
- [17] SZÁSZ, B. – SARANIVA, A. – BOGNÁR, K. – UNZEITIG, M. – KARJALAINEN, M.: Cultural heritage on the semantic web – the Museum24 project. (2006). [Előadásanyag]. 10 p. <http://www.seco.tkk.fi/events/2006/2006-05-04-websemantique/presentations/articles/Szasz-museum24Paris.pdf> (2008.01.14.) <http://www.museo24.fi> („Museo24” portál honlapja) (2008.01.14.)

- [18] SZEREDI, P.: Ontológiák – egy matematikus-informatikus szemével. = Ontosz. Előadássorozat a formális ontológiákról. Az ontológia fogalmának, felépítésének, alkalmazási lehetőségeinek különböző megközelítései. Budapest, W3C, 2007. ápr. 25.
<http://www.w3c.hu/rendezvenyek/2007/ontologia/index.html> (2008.01.10.)
- [19] SZEREDI P. [et al.]: A szemantikus világháló. = A szemantikus világháló elmélete és gyakorlata. Szerz. Szeredi P. [et al.]: Budapest, 2005, Typotex. p. 17–59.
- [20] TAPSCOTT, D. – WILLIAMS, A. D.: Wikinómia. Hogyan változtat meg mindent a tömeges együttműködés. Szerk. Török Hilda; ford. Garamvölgyi Andrea. Budapest, HVG, 2007.
- [21] UNGVÁRY Rudolf: Az információkeresés értékelése. = Osztályozás és információkeresés: kommentált szöveggyűjtemény. 2. köt. Az információkeresés és elmélete. Szerk. Ungváry Rudolf, Orbán Éva. Budapest, OSZK, 2001.
<http://mek.oszk.hu/01600/01683/pdf/01683-2.pdf> (2007.11.17.)
- [22] UNGVÁRY Rudolf: Az ontológia fogalma, avagy az eltűnt teaurusz. = Ontosz. Előadássorozat a formális ontológiákról. Az ontológia fogalmának, felépítésének, alkalmazási lehetőségeinek különböző megközelítései. Budapest, W3C, 2007. ápr. 25.
<http://www.w3c.hu/rendezvenyek/2007/ontologia/index.html> (2008.01.10.)
- [23] UNGVÁRY Rudolf: A tartalom szerinti információkeresés az interneten: I. indexelőszolgáltatások. = TMT, 47. köt. 1. sz. 2000. p. 3–17.
http://tmt.omikk.bme.hu/show_news.html?id=1624&issue_id=15 (2008.01.27.)
- [24] UNGVÁRY Rudolf: A tartalom szerinti információkeresés az interneten: II. internetkatalógusok. = TMT, 47. köt. 2. sz. 2000. p. 55–67.
http://tmt.omikk.bme.hu/show_news.html?id=1625&issue_id=16 (2008.01.27.)
- [25] UNGVÁRY Rudolf: Tezaurusz és ontológia, avagy a fogalmi ismertetőjegyek generikus öröklődésének formalizálása. = TMT, 51. köt. 5. sz. 2004. p. 175–191.
http://tmt.omikk.bme.hu/show_news.html?id=3615&issue_id=450 (2008.01.27.)
- [26] UNGVÁRY Rudolf – VAJDA Erik: Könyvtári információkeresés. 2. jav. kiad. Budapest, Typotex, 2002.
- [27] XML-Topic-Map (XTM) Standard, ISO/IEC 13250:2000. XTM TopicMaps Org. = <http://www.topicmaps.org/xtm> (2008.01.14.)

Beérkezett: 2010. V. 25.-én.



Már 400 millióan használják a Firefoxot

A kontinensünkön egyetlen esztendő alatt 8,4 százalékkal csökkent az *Internet Explorer (IE)* piaci részesedése és jelenleg alig haladja meg az 50 százalékot. A legnagyobb konkurensének számító *Firefox* viszont köszöni szépen, jól van. Ugyanakkor *Tristan Nitot*, a *Mozilla Europe* elnöke tisztában van azzal, hogy nem ülhetnek a babérjaikon. Gőzerővel fejlődik ugyanis a *Google Chrome*, az IE 9-es verziója is sok tekintetben előrelépést jelent a korábbi változatokhoz képest, így a Firefoxnak is fejlődnie kell, ha lépést akar tartani a konkurensével.

„A Google rendkívül innovatív vállalat, amely tavaly felerősítette a böngészőpiacon zajló versenyt. Évek óta vágytunk erre a konkurenciaharcra, mert úgy véljük: sokat lendíthet a böngészők fejlesztésén. Ugyan a Mozillára nehezedő nyomás nem lett kisebb, de ez így van jól.” – jelentette ki *Tristan Nitot*. Az elmúlt hónapokban számos kritika érte a Mozillát, hogy túlzottan elkényelmesedett és nem figyelt eléggé a Firefox fejlesztésére. Ez különösen annak tükrében érthető, hogy a Firefoxot több mint 400 millióan használják világszerte.

„Néhány országban, mint Lengyelország vagy Németország már közel 50 százalékos piaci részesedésre tettünk szert, ami szenzációs teljesítmény. Minél nagyobb azonban a tortából általunk kihalított szelet, annál nehezebb tovább növekednünk. A hiányzó innovációs képességünkre vonatkozó kritikákat viszont visszautasítom, hiszen a kiadási stratégiánkat az igényekhez igazítottuk és gyorsítottunk a fejlesztéseken is. Ráadásul a Firefox 4-essel egy olyan verzióugrás következhet be, amely pont a sebesség és a dizájn tekintetében hoz magával jelentős javulást.” – hangsúlyozta a szakember.

Nitot hozzátette, hogy a jövőben szeretnék az online közösség kreatív potenciálját még jobban hasznosítani, ezért a céljuk az, hogy a Mozilla Labs kísérleti ötleteit közvetlenül integrálni lehessen a fejlesztési folyamatba. Emellett jelentősen leegyszerűsítene a kiegészítők fejlesztési lehetőségeit is és a modulok a böngésző újraindítása nélkül, azonnal használhatók lesznek.

/SG.hu Hírlevél, 2010. július 26., <http://www.sg.hu/>

(SzP)

Az internetes keresők működésének technikai háttere

Jelen írás azok számára lehet érdekes, akik mélyebben érdeklődnek egy komplex kereső működésének technikai háttere iránt. Összefoglalja számunkra egy keresőrendszer működési elveit és az alkotóelemeinek feladatait. A találatok rangsorolása a keresők egyik legmeghatározóbb sajátossága, ehhez kapcsolódóan bemutatja a Google kereső PageRank eljárását. Tárgyalja a PageRank algoritmus alap gondolatát és rávilágít arra, hogy hogyan modellezi a felhasználói viselkedést egy keresés során.

A világháló új helyzetet teremtett a hagyományos információkeresés területén, hiszen a rendszerezettség, a homogenitás és a rend helyett azt látjuk, hogy bárki létrehozhat rajta tartalmat, amelynek minőségét és megbízhatóságát senki nem vizsgálja. Tehát a világháló heterogén szintaktikájú és szemantikájú, továbbá többnyire nem ellenőrzött tartalmú dokumentumok halmazát reprezentálja. Ebből adódóan az internetes keresés alapvetően eltér egy lassan változó, kontrollált dokumentumgyűjteményben való kereséstől. Ez a különbség többek között abban is megnyilvánul, hogy a keresőknek meg kell találniuk a releváns webes tartalmaknak azokat a halmazait, amelyek jól használhatók a felhasználók számára, nem pedig egy hagyományos gyűjteményből kell kiválogatniuk a keresőkérésre pontosan illeszkedő dokumentumokat. Kereséskor a legjobb találatoknak egyéb jellemzőik is vannak (frissítési gyakoriság, minőség, hivatkozások száma, népszerűség stb.), amit a keresőknek szintén figyelembe kell venniük és nem elegendő csupán a keresésnek pontosan megfelelő dokumentumokat szolgáltatniuk. Egy-egy keresésre különböző válaszok adhatók, ezért nagyon lényeges, hogy mely találatok jelennek meg elsőként a felhasználóknak [10].

Még mielőtt rátérünk a téma részletes tárgyalására, pontosan meghatározzuk az internetes keresők fogalmát. Internetes keresők alatt a programoknak egy olyan általános csoportját értjük, amely lehetővé teszi a weben történő információkeresést a felhasználók számára. Ezek a programok dokumentumokat indexelnek és arra törekednek, hogy megtalálják a feltett keresőkérésre a releváns találatokat.

Keresőszolgáltatások előretörése az interneten

Lényeges változásnak tekinthetjük, hogy az ezredfordulóra az országok döntő többsége információs társadalomként jelent meg a világtérképen. Végbement egy technológiai forradalom, a digitális eszközök mindennapjaink részévé váltak, azokat nemcsak eszközöknek tekintjük, hanem érzelmileg is kötődünk hozzájuk. Mindez kihatott a médiafogyasztásra, megváltoztak a kulturális objektumok átvételi csatornái. Átrajzolódott a gazdaság, a kormányzás, a tartalomipar, átalakult a fogyasztó és az előállító viszonya. Legfontosabb jelenségként tapasztalhattuk az informatika, az internet és a számítógép hétköznapivá válását, bár ezek a folyamatok jóval korábban (10-20-30 éve) kezdődtek el. Az internet egy szűk kör számára hozzáférhető újdonságból a világ lakosságának nagyjából hatoda által használt eszközzé vált. 1998–2008 között az internetezők száma több mint tízszeresére növekedett világszerte [11].

Az 1990-es évek elején néhány webszerverről lista készült, amelyet *Tim Berners-Lee* állított össze és a *CERN (Conseil Européenne pour la Recherche Nucléaire)* szerverén helyezett el. Miután egyre több webszerver jött létre, ez a központi lista már nem volt elégséges. Később az *NCSA (National Center for Supercomputing Applications)* webhelyén jelentették be az új webszervereket „What’s new!” megnevezés alatt.

Az első keresőeszköz az *Archie* volt, amelyet 1990-ben *Alan Emtage* hozott létre. A program az anonim FTP szervereken lévő állományok könyv-

tár struktúráját töltötte le, azonban nem indexelte ezeknek a szervereknek a tartalmát. Ezáltal megszületett az állománynevek első kereshető adatbázisa.

1991-ben a *Gopher* megjelenése két új kereső-programhoz vezetett: a *Veronica*-hoz és a *Jughead*-hez. Mindkét program az *Archie*-hoz hasonlóan állománynevekre és címekre keresett a *Gopher* indexekben. A *Veronica* program (*Very Easy Rodent-Oriented Net-wide Index to Computerized Archives*) megoldotta a menücímkre történő kulcsszavas keresést a *Gopher* világban. A *Jughead* (*Jonzy's Universal Gopher Hierarchy Excavation And Display*) keresővel adott *Gopher* szerverekről szóló menüvel információkat lehetett visszakeresni, amelyekre feltelepítették.

A web első kezdetleges keresője a *W3Catalogus* volt, amit 1993. szeptember 2-án indítottak újtjára. 1993 nyarán *Matthew Gray* létrehozta az első keresőrobot-programot, amely *Perl* programozási nyelvre épült. Ezzel a keresőrobottal a „*Wandex*” nevű indexet állította elő. Keresőrobotját a web akkori méretének meghatározására használták 1995-ig. A web második keresője az *Aliweb* volt, amely 1993 novemberében jelent meg. Ez a keresőgép nem használt robotot az oldalak begyűjtésére, hanem kizárólag a webhely adminisztrátorok visszajelzésére épült azzal kapcsolatban, hogy létezik-e valamilyen indexállomány az adott webhelyen.

1993-ban kezdte el működését a *JumpStation*, amely már keresőrobotot alkalmazott az oldalak meglátogatására és az indexének létrehozására. Ez volt az első olyan kereső, amely egy keresőgép mindhárom alapvető jellemzőjét tartalmazta (a begyűjtést, az indexelést és a keresést). A szűkös erőforrások miatt a *JumpStation* indexelése és keresése a begyűjtött weboldalak címére korlátozódott.

Az első olyan kereső, amelynek a keresőrobotja már a begyűjtött weboldalak teljes szövegét vette figyelembe a *WebCrawler* volt (1994-ben jelent meg). Elődjeitől eltérően megengedte használóinak, hogy a weboldalakon lévő bármelyik szóra keressenek, ami ettől fogva alapvető elvárásként fogalmazódott meg a keresőknél. Szintén 1994-ben indult a *Lycos* a *Carnegie Mellon Egyetemen*, ami jelentős kereskedelmi sikerré vált. Hamarosan számos kereső jelent meg a piacon, amely egyre nagyobb lett. Ezek közé tartoztak a következők: *Magellan*, *Excite*, *Infoseek*, *Inktomi*, *Northern Light*,

AltaVista és a *Yahoo!* Azonban a *Yahoo* kereső-funkciói a webes katalógusra épültek, nem pedig az oldalak teljes szövegére. Használói böngészhetek benne a kulcsszavas keresések alkalmazása helyett [12]. Ezeket a keresőszolgáltatásokat főként vállalati tőkéből, reklámokból, illetve kutatási költségvetésekből finanszírozták.

1996-ra már a különböző folyóiratok, üzleti és napilapok is komoly figyelmet szenteltek a keresőknek. Megnövekedett a keresésre specializálódó szoftvertermékek száma, például webes katalógusok, metakeresők, szakterületi szolgáltatások, kereső ágensek és „push” szolgáltatások jelentek meg [8]. Ugyanebben az évben a *Netscape* öt nagyobb keresővel kötött megállapodást, mely szerint azok évenként felváltva kerültek fel a keresőoldalára, meghatározott pénzüsszeg fejében. Az öt kiválasztott kereső közé tartozott: a *Yahoo!*, a *Magellan*, a *Lycos*, az *Infoseek* és az *Excite*.

A '90-es évek végén a keresők fejlesztésébe jelentős pénzüsszegeket fektettek be. 1997-ben kezdett el növekedni a „dot-com” névre keresztelt gazdasági buborék. Az e-szektor részvényeinek árai gyorsan emelkedtek, a külső tőke is meghatározóvá vált. Számos cég bukkan fel a piacon, néhány közülük felhagyott a nyilvános kereső működtetésével, helyette pedig vállalatoknak szánt kiadást vitt a piacra (l. pl. *Northern Light*). Sok kereső belekerült ebbe a gazdasági buborékba, ami egy spekuláció-vezérelt piaci robbanásnak volt tekinthető. Ez a folyamat a tetőpontját 1999-ben érte el, és 2001-ben ért véget.

A *Google* 2000 környékén jelent meg a keresőpiacon és fokozatosan prominens keresővé vált. A cég felemelkedését annak köszönheti, hogy *PageRank* algoritmusával pontos találatokat szolgáltatott a használók kulcsszavas kereséseire. Ezenkívül a *Google* egyszerű keresőfelületet kínált. 2000-re a *Yahoo!* olyan szolgáltatásokat nyújtott, amelyek az *Inktomi* keresőre épültek. 2002-ben felvásárolta az *Inktomit*, valamint 2003-ban az *Overture*-t. 2004-ig áttért a *Google* kereső használatára, ekkor azonban újtjára indította a saját keresőjét, amely a felvásárolt cégeinek a technológiáin alapult.

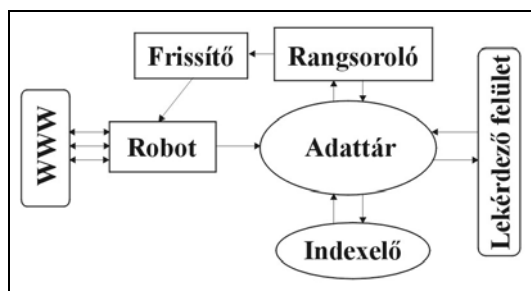
1998 őszén a *Microsoft* elindította az *MSN* keresőjét, amely az *Inktomi*-ból származó keresési találatokra támaszkodott. 1999 elején az *MSN* a *Looksmart* és az *Inktomi* keresési eredményeit jelenítette meg a találatlistájában, kivéve ugyanebben az évben azt a rövid időszakot, amikor az

Altavista találataira támaszkodott. 2004-ben a Microsoft elkezdett áttérni a saját keresőtechnológiájára, amelyet a keresőrobotjával, az ún. *msnbot*-tal támogatott. 2009 júniusában a Microsoft egy új, *Bing* nevű keresővel jelent meg a piacon. 2009 júliusában a Yahoo! és a Microsoft olyan egyezséget kötött egymással, amely szerint a Yahoo! is a Microsoft Bing keresőtechnológiájára épül.

2009 júliusában a keresők világméretű népszerűsége a következőket mutatta: Google (78,4%), *Baidu* (8,7%) és a Bing (3,17%). A Yahoo! 7,16%-os és az AOL 0,6%-os piaci részesedése szintén csökkent az előző évhez képest. 2009 májusában a Google használatának aránya 63,2% volt az Egyesült Államokban. 2009 júliusában a Baidunak 61,6%-os népszerűsége volt a Kínai Köztársaságban [12].

Weboldalak begyűjtése és indexelése

A továbbiakban részletesen ismertetjük egy keresőrendszer alkotóelemeit és a rájuk bízott elvégzendő feladatokat. Az 1. ábra vázlatos áttekintést nyújt egy kereső szerkezeti felépítéséről.



1. ábra **Egy keresőrendszer felépítése**

Forrás: FRIEDMAN E. – UHER M. – WINDHAGER E.:
Keresés a világhálón

A keresők első, lényeges feladata az oldalak meglátogatása és begyűjtése, amit speciális szoftverek, ún. keresőrobotok (*crawlers*, *web robots*, *web spiders*, *bots*) segítségével valósítanak meg. Ezek a programok folyamatosan és bizonyos időközönként újra és újra átfésülik a webet. Egy keresőrobot választhat egy népszerű, de megbízható oldalt kiindulópontjául, illetve dolgozhat egy korábbi, meglévő adatbázis alapján is. A robotnak le kell töltenie az általa meglátogatott oldalt, és át kell adnia azt az indexelőnek*. Ezután az oldalon lévő hiperhivatkozásokat nyomon követve ugyanígy kell eljárnia a hivatkozott oldalakkal is.

Számos esetben bizonyos időkülönbség jelentkezik a begyűjtés és az indexelés, valamint az eredmény keresőbe történő beépülése között. Ezért az oldalak begyűjtését és indexelését két, párhuzamosan zajló feladatnak kell tekintenünk. A keresőrobotok tehát nem végeznek semmilyen elemzést a meglátogatott dokumentumon, hanem csak nyomon követik a hivatkozásokat és letöltik a felfedezett oldalakat. Látszólag a robotok nagyon hasonló módon működnek, azonban jelentős különbségek figyelhetők meg a viselkedésükben. A robotok további feladata a meglévő, begyűjtött dokumentumok frissítése az adattárba. Ha például módosul egy oldal, akkor annak az újabb verzióját a megváltozott metaadataival együtt a robotnak el kell helyeznie az adatbázisban, a régit pedig törölnie kell.

Egy robot számára fontos szempont az, hogy mely hivatkozásokat kövessen nyomon, és mely oldalakat keressen fel, valamint lényeges kérdés, hogy milyen gyakran végezze el az oldalak begyűjtését. Egy keresőrendszer általában több robotot is alkalmaz a weblapok begyűjtésére. Emiatt a hálózati forgalom megnövekszik. A robotok igyekeznek nem folyamatosan terhelni egy szerveret különböző kérésekkel, hanem időben elosztva küldik neki a kéréseket [10].

A „crawler control” modul látja el a robotok irányítását és munkájuk összehangolását. Az oldalak begyűjtése közben egy prioritási sort használ, amelyben a még meg nem látogatott oldalak címei szerepelnek fontossági sorrendben. A sor elejéről kivesszi a címeket és a hozzájuk tartozó oldalakat, letölti és kigyűjti belőlük a hivatkozásokat. A felderített linkekről eldönti, hogy melyiket kell követniük a robotoknak, ezeket beteszi a prioritási sorba, a többit pedig elhagyja. A begyűjtés addig tart, amíg a helyi erőforrások, mint például a tárolókapacitás, vagy egyszerűen a meglátogatott oldalak el nem fogynak [3].

A webmestereknek vagy a tartalomgazdáknak módjukban áll a robotok számára megtiltani egyes oldalak begyűjtését, az oldalon lévő hivatkozások követését és az oldal archiválását. Ezt a *Robot Kizárási Szabványban* (*Robot Exclusion Standard*) megszabott módon tehetik meg [6]. Ha egy weblapra nem hivatkozik egy másik oldal, akkor a keresőrobot nem fogja azt megtalálni. Ezért az új

* A fenti ábra tükrözi, hogy a robot és az indexelő között csak közvetett kapcsolat valósulhat meg az adattáron keresztül.

honlapokat tanácsos manuálisan regisztrálnunk az egyes keresőknel, amelyek így indexelni tudják őket. A keresőket lekérdezhethetjük arról, hogy egy adott oldal indexelve van-e náluk. Azonban a különböző keresőknel eltérő lekérdezéseket kell alkalmaznunk erre a célra [9].

A keresőrobotok által begyűjtött oldalak az adattárba (*repository*) kerülnek, amelynek elsődleges feladata az oldalak egyenkénti tömörítése és szekvenciális tárolása. Ezenkívül a rendszer nyilvántartja egy állományban a dokumentumok pontos elhelyezkedését [3].

A keresők másik lényeges összetevője az indexelő (*indexer*), amelynek fő feladata az adatbázisban lévő begyűjtött oldalak elemzése és az indexelendő kifejezések begyűjtése [3, 10]. Az indexelő tulajdonképpen az adattárra támaszkodik. A feldolgozás elején két problémával találkozunk az indexelővel. A weben előforduló oldalak elemzése összetett feladat. Ezt nem csupán a dokumentumok heterogén kialakítása okozza, hanem az egy-egy adott formátum esetén előforduló hibák is, például szintaktikai hibák a HTML dokumentumokban. A másik probléma az, hogy az indexelőnek szét kell tudnia választani a fontos és a kevésbé fontos kifejezéseket egy dokumentumban.

Erre egy lehetséges megoldás az, hogy figyelembe vesszük a szavak gyakoriságát és eldobjuk a legkisebb, valamint a legnagyobb gyakoriságú szavakat. Az előbbieket azért, mert azok nem lehetnek fontosak, hogyha csak néhány alkalommal fordulnak elő, az utóbbiakról pedig nagy valószínűséggel állítható, hogy felesleges szavak a dokumentumban. Azt is feltételezzük, hogy a töltelés és egyéb szavak eloszlása eltérő egy dokumentumban. Tehát a szavak eloszlásának elemzésével a szavak gyakorisági kategóriákba sorolhatók.

A gyakorlatban azonban elterjedt egy másik megközelítés is, amelyben nyelvenként hoznak létre egy ún. tiltott szólistát (*stopwords*): ez foglalja magába a tartalmi szempontból feleslegesnek tekintett szavakat [10]. Az ilyen lista meggátolja a névelők, a kötőszavak és más, szinte minden dokumentumban előforduló szavak indexelését [3]. Tehát ez a módszer rendkívül gyors, egyszerű és könnyen használható.

A megmaradt releváns kifejezéseket bizonyos jellemzőikkel együtt gyűjti ki a dokumentumból az indexelő. Fontos jellemzőnek minősül a szó előfordulásának helye, mint például az oldal címe, a

metaelemek, az oldalon belüli pozíció [10]. Továbbá az indexelő létrehoz egy indexet, amely minden releváns kifejezéshez hozzákapcsolja az őt tartalmazó URL-ek listáját [3]. A begyűjtött indexelendő kifejezéseket és jellemzőiket a tényleges keresés és sorrendezés során veszik alapul a keresők.

A találatok sorrendezése, rangsorolása

A keresők működésének a leglényegesebb vonása a találatok megfelelő fontossági sorrendben történő megjelenítése a felhasználók számára. Ezért a keresőknek jelentős alkotóeleme a *Rangsoroló modul*, amely egy adott keresésre automatikusan sorrendezi a találatokat fontosság szerint [3]. Az indexelt adatmennyiség megnövekedésével vált egyre fontosabb feladattá a találatok pontos sorrendezése. Mivel a felhasználók csak az első 10-20 találatot szokták áttekinteni egy adott keresésnél, ezért rendkívül fontos, hogy a kereső mely találatokat jeleníti meg a találati lista elején [10]. Kereséskor célunk a témában íródott legszínvonalasabb weblapok felkutatása, melyhez az oldalakat rangsorolni kell [3]. Az egyes keresők által használt rangsorolási szempontokról általában keveset tudunk, de a fő elvek ismertek.

Az egyik legalapvetőbb sorrendezési szempontnak minősül a *keresőkifejezés helyének vizsgálata* a dokumentumban. A keresők nagyon gyakran előnyben részesítik azokat az oldalakat, amelyeknek a címében is megtalálható a keresendő kifejezés. A találatok sorrendezésénél azt is figyelembe vehetik, hogy a dokumentum mely részében jelenik meg először a keresőkifejezés. Itt az alapelv az, hogy a weblap szempontjából releváns kifejezések nagy valószínűséggel fordulnak elő már a bevezetésben is, vagy legalábbis a dokumentum elején. Egyes keresők az oldal fontosságának meghatározásához számításba veszik a fontméretet is, és következtetésekre jutnak a szavak közti távolságokból is, valamint elemzik a HTML metaelemeket. A metaadatok segítségével közölhetjük honlapunk tartalmának összefoglalóját, valamint az oldalunkra vonatkozó kulcsszavakat. Ezeket a háttér-információkat is hasznosíthatják a keresők a rangsorolás, valamint a keresés közben is.

Másik jelentős vizsgálati szempont a *keresőkifejezések előfordulási gyakorisága*. Itt azzal a feltételezéssel élhetünk, hogy ha egy dokumentumban egy bizonyos kifejezés gyakran fordul elő, akkor az fontos a téma szempontjából.

Ebben az esetben természetesen kivételt képeznek a tiltott szavak listáján lévő kifejezések. Továbbá lényeges, hogy ne csak az egyes szavak előfordulási gyakoriságát kövessük nyomon, hanem az adott szóösszetételekét is. A keresők sokszor tanulmányozzák felhasználóik reakcióit is. Ha például a felhasználók többsége nem az első találatra kattint a szolgáltatott találatlistában, akkor nagy a valószínűsége annak, hogy rossz a találatok rangsorolása, és nem az első helyen szereplő oldal a legrelevánsabb.

Ezek a felsorolt sorrendezési szempontok sajnos lehetővé teszik, hogy könnyedén befolyásoljuk a találatok rangsorolását. Megfigyelhető az a tendencia, hogy az egyszerűen manipulálható rangsorolási szempontok egyre inkább háttérbe kerülnek és csökken a súlyuk a végső sorrend kialakításában. Helyettük pedig olyan kritériumokra helyeződik a hangsúly, amelyeket nehezebb befolyásolni. Itt megemlíthetők például olyan módszerek, amelyek az oldalak közti linkstruktúrát veszik figyelembe [10].

A találatok rangsorolásánál kényes etikai kérdésként merülhet fel az, hogy a kereső jó pénzért nem árul-e kulcsszavakat a cégek számára. A megvásárolt kulcsszóért cserébe az adott cég webhelye az első 10 találat között szerepelhet. Ez nem jellemző a nagyobb keresőkre, azonban a felhasználói kulcsszavakhoz kapcsolódó reklámok eladása széles körben elterjedt gyakorlat. Ezekben az esetekben a szoftverfejlesztők úgy változtatják meg a keresők relevanciarangsorolási algoritmusát, hogy az eladott kulcsszó a felhasználót rögtön arra a webhelyre vezesse, amely korábban megvásárolta azt [4].

Egyes keresők a linkhez tartozó szöveget nem a linket tartalmazó, hanem a link által hivatkozott oldalhoz tartozónak veszik. Az ilyen típusú linket horgonynak hívjuk, amit bizonyos keresők a találatok rangsorolásakor használnak fel [10]. A Google együtt kezeli a linkek szövegét azokkal a weboldallal, amelyekre azok ténylegesen hivatkoznak. Ennek a módszernek számos előnye van: a linkek sokszor pontosabb leírást nyújtanak a hivatkozott oldalakról, mint maguk az oldalak; továbbá olyan oldalakat is megkaphatunk, amelyeket a keresőrobot nem gyűjtött be a webről [1, 10].

Sokan vélekednek úgy, hogy a Google kereső népszerűségét annak köszönheti, hogy a találatokat minőségileg jobban rangsorolja, mint a többi kereső. A Google alkalmazza a fentebb ismertetett

általános módszereket, továbbá kialakít egy olyan speciális algoritmust is, amely kizárólag a linkstruktúrát alapul véve határozza meg az egyes dokumentumok fontosságát. Ezt a fontosságot a kereső megfelelően súlyozza és a többi faktort egyaránt figyelembe véve, dönt a végső sorrend kialakításáról [10].

A Google kereső PageRank algoritmus

A PageRank (PR) valós szám, amely egy adott oldal fontosságát tükrözi. A Google kereső a PageRank algoritmust alkalmazza az általa indexelt oldalak fontosságának meghatározásához, amit figyelembe vesz a rangsorolás során. A Google más egyéb szempontokat is felhasznál a sorrend kialakításakor, amelyek közül csak egy a PageRank érték, azonban ez az egyik legfontosabb. A PageRank-kel kapcsolatos eredmények megtalálhatók [10]-ben. Fontosságuk miatt néhány alapvető megállapítást részletezünk a továbbiakban.

A PageRank algoritmus alap gondolata, hogy amikor egy oldal hivatkozik egy másik weblapra, akkor a forrásweboldal tulajdonképpen ajánlja a hivatkozott weblapot. Tehát az oldal létrehozója azért tüntette fel a linket az oldalán, mert a másik lapot valamilyen szempontból fontosnak tekintette. Emellett azt is figyelembe kell vennünk, hogy a hivatkozó oldal mennyire fontos, mert egy fontos oldalnak többet ér a hivatkozása. Eredményül egy rekurzív algoritmust kapunk, ami azt fejezi ki, hogy egy oldal fontos, ha mérvadó oldalak hivatkoznak rá. Ez a modell természetesen vitatható, hiszen lehetséges, hogy csak rossz példaként hozunk fel egyes weboldalakat, és nem arra szeretnénk velük célozni, hogy ők értékes oldalak. A gyakorlat azonban az eredeti alapötlet sikerességét igazolja, hiszen kevésbé meghatározók ez utóbbi linkek az interneten [10].

Az alapalgoritmust [7]-ben közzétették először. (Nagy valószínűséggel feltételezhetjük azt, hogy a Google most már egy másik változatát használja az itt tárgyaltaknak, amiről azonban nem tájékoztatják a nyilvánosságot [10]). Ez a rekurzív egyenlet a weboldal fontosságára egy megközelítőleges becslést nyújt [1]. Érdekes, hogy a szerzők egyik cikkükben pontatlanul adták meg az egyenlet első tagját és az így terjedt el a szakmában széles körben. Ez a változat megtekinthető az alábbiakban:

$$PR(A) = (1-d) + d \cdot \left(\frac{PR(t_1)}{C(t_1)} + \dots + \frac{PR(t_n)}{C(t_n)} \right)$$

Az egyenlet az A oldal PageRank értékét határozza meg. Az egyenletben $t_1 \dots t_n$ jelöli azokat az oldalakat, amelyek A oldalra mutatnak. $PR(t_i)$ fejezi ki az i . ilyen oldal PageRank értékét, azaz annak a fontosságát. A d paramétert egy skálázó faktornak tekintjük, aminek értéke 0 és 1 közé eshet. A d értékét a szerzők $0,85$ -nek határozták meg [1, 10]. C -vel jelöljük az egy oldalon lévő összes kimenő hivatkozás darabszámát. Például, ha $C(t_i)$ értékét 24 -nek vesszük, az azt jelenti, hogy az i . oldal összesen 24 darab kimenő hivatkozást tartalmaz, amelyek közül egy biztosan az A oldalra hivatkozik. Az eredeti algoritmus nem számol azzal az esettel, hogy mi történik akkor, ha egy oldalról több link is hivatkozik egy másik oldalra.

Az egyenlet tehát a következőt jelenti: az A oldal az első olyan oldalról, amely hivatkozik rá, $PR(t_i)/C(t_i)$ -nyi szavazatot kap, azaz a t_i -es oldal egyenletesen elosztja a saját fontosságát a kimenő hivatkozásai között. Ha t_1 oldalon egyetlen kimenő link található, akkor A megkapja a teljes $PR(t_1)$ értéket, ha három, akkor csak t_1 fontosságának a harmadát stb. Ugyanezt az elvet követjük az összes többi olyan oldal esetén, ahonnan találunk hivatkozást A -ra. Ezután ezeket a fontosságokat összeadjuk és megkapjuk A oldal fontosságát. Ebből tehát az következik, hogy kedvezőbb PR értéket kapunk, ha egy alacsonyabb PR értékű lap mutat ránk, mintha egy magasabb, ha az alacsonyabb fontosságú lapon nem sok kimenő link található. Egy dolgot azonban biztosan kijelenthetünk: ha oldalunkra több oldal hivatkozik, nem számít, hogy milyen rangos oldalak, valamilyen mértékben nőni fog a fontosságunk.

A d faktornak köszönhetően egy bizonyos oldal nem a teljes fontosságát osztja szét a kimenő linkjei között, hanem annak csak a 85% -át. Ahhoz, hogy megértsük ezt az összefüggést, szükségünk van egyrészt a javított PageRank egyenletre és a PageRank algoritmus egy újabb jelentésének bemutatására. A javított PageRank egyenletet tehát a következőképpen adhatjuk meg, ahol N az összes indexelt weblap számát jelenti.

$$PR(A) = \frac{(1-d)}{N} + d \cdot \left(\frac{PR(t_1)}{C(t_1)} + \dots + \frac{PR(t_n)}{C(t_n)} \right)$$

A PageRank algoritmus egy olyan modellnek is tekinthető, amely a „véletlen szörfölő” viselkedését

tükrözi. Egy ilyen felhasználó véletlenszerűen elindul egy weboldaltól és a hivatkozásokra véletlenszerűen kattintva folyamatosan előrehalad. Nem is figyeli meg, hogy hova kattint, hanem egyenesen eloszlás szerint választ a meglévő hivatkozások közül. Ezzel magyarázható az, hogy a PageRank algoritmus a kimenő linkek számával elosztja egy bizonyos oldal fontosságát. Mindez addig tart, amíg szörfölőnk meg nem unja a kattintgatást és egy másik véletlenszerűen kiválasztott weboldalon nem indul el. Ez az egyenlet egy valószínűségi eloszlást határoz meg, ahol egy-egy weboldal PageRank értéke egy valószínűségnek (0 és 1 közötti valós szám) felel meg. Ebben a modellben az összes weboldal PageRank értékeinek összege maximum 1 lehet. Ez a megállapítás csak abban az esetben igaz, ha a felhasználónk egy adott oldalon mindig talál legalább egy hivatkozást, amelyen továbbhaladhat [1,10].

Ha webszájtunk olyan oldalt tartalmaz, amelyre ugyan mutat link, de belőle nem indul kimenő hivatkozás, akkor a szájt nem veszi fel a maximális PageRank értéket. Lógó (*dangling*) oldalnak hívjuk az ilyen oldalakat. A Google figyelmen kívül hagyja a lógó oldalakat, mert azok ellentmondanak a PageRank algoritmus által használt „véletlen szörfölő” modellnek. A Google tehát szűri a lógó oldalakat (az elhagyások miatt esetlegesen újonnan keletkezett lógó oldalakat rekurzív módon szintén figyelmen kívül hagyja). A megmaradt linkstruktúrában kiszámolja a pontos PR értékeket. Ezután fokozatosan visszahelyezi a lógó oldalakat és meghatározza azok fontosságát is a már kiszámított PR értékek alapján.

A Google nem csupán a linkstruktúrát elemzi, hanem egyéb tényezőket is figyelembe vesz az oldalak rangsorolásakor. Például sokszor negatívan értékeli azt, ha bizonyos, megbélyegzett oldalakra mutató hivatkozásokat tüntetünk fel az oldalunkon. Nyomon követi azt is, hogy az oldalra történő hivatkozások ugyanabból a doménből, földrajzi területről származnak-e. Tehát a rangsorolás szempontjából többet ér az, ha egy „független” valaki hivatkozik ránk, mint ha egy „ismerős” szavaz nekünk bizalmat [10].

A PageRank módszer manipulálása sokkal nehezebb feladat, mint a szöveges dokumentumok tartalomalapú sorrendjének befolyásolása. Ennek az az oka, hogy a web nagyobb részét kell módosítanunk, valamint hivatkozások sűrű szövevényével kell azt ellátnunk. A Google által alkalmazott rangsorolási módszer nagyjából ismert a nagy

nyilvánosság számára, ezért a világban számos cég specializálódott különféle manipulatív megoldások használatára, amelyekkel a saját forgalmukat tudják indokolatlanul befolyásolni. A cégeknek ezt a törekvését finomabb változatban „kereső optimalizálásnak” hívjuk, erősebb változatban pedig „hivatkozás spam-nek”. A PageRank támadásának egyik közkezdvelt módszere a linkfarmok létrehozása. Ilyenkor nagyszámú és sok szerverre kiterjedő, részben értékes oldalak másolatát, részben számítógéppel generált weblapokat tartalmazó oldalcsoporthoz alakítanak ki. Itt az oldalak mindegyike a céloldalra hivatkozik, ezáltal magas fontosságot tulajdonítanak nekik [2].

Problémák az internetes kereséssel és a megoldási kísérletek

A kereséssel kapcsolatos problémákat öt fő csoportba soroljuk, amelyek a következők:

1. Általános problémának tekinthető az internet hatalmas mérete, ami nemcsak a keresést, hanem az oldalak begyűjtését is nagymértékben befolyásolja. A weblapok meglátogatása és feltérképezése időigényes feladatot jelent még a legjobb keresők számára is.
2. Az utolsó begyűjtés óta eltelt idő alatt az internet tartalma és szerkezete megváltozik, ami további nehézségeket eredményez [5].
3. A keresőrendszerek számára általában elérhetetlenek azok az interneten meglévő tartalmak, amelyek a mély web (*deep web*) körébe sorolhatók.
4. A keresőrobotok nem gyűjtik be a dinamikus weblapokra mutató hivatkozásokat.
5. Az internetes keresők nem a felkutatható dokumentumok és a keresőkérdés jelentésével foglalkoznak, hanem csupán a szöveges alakal.

A továbbiakban részletezem, hogy milyen módszerekkel próbálják megoldani ezeket a felmerülő problémákat.

Az óriási adattömeg kérdését oldják meg a metakeresők, amelyek párhuzamosan több rendszerrel kerestetnek. Így azok az internet nagyobb részét képesek átfésülni [10]. Növelik a találati esélyünket az ismeretlen témák esetében, valamint átfogóbb képet nyújtanak számunkra a weben fellelhető információkról egy adott témában.

A gyorsan változó tartalom kezelésére használt leglényegesebb módszer, hogy a változás mérté-

két és gyakoriságát is eltárolják a weblapok tartalmával együtt, majd a gyakran és jelentősen változó oldalakat sűrűbben látogatja újra a robot. Fontos továbbá az RSS csatornák indexelése is, mert így gyorsan értesül az új tartalmakról a keresőgép. További lehetséges módszer az oldalak begyűjtésének fókuszált módja (*focused crawling*). Ennek a módszernek lényege az, hogy nem követünk minden hivatkozást, hanem valamilyen szempontrendszer szerint egy bizonyos területhez kapcsolódó oldalakra szűkítjük a keresési teret, például nevezetes hírportálok meglátogatására. A fókuszált begyűjtést végző robotokkal kialakíthatunk egy-egy adott területre specializálódott keresőt is. Létrehozhatunk például egy olyan keresőt, amely orvosi tartalmak indexelésére és orvosi szakterületen feltett kérdések megválaszolására alkalmas.

A mély web csoportjába tartoznak a weben keresztül lekérdezhető adatbázisok, a nem szöveges formában található tartalmak, valamint a jelszavas vagy IP címalapú védelem mögé rejtett statikus oldalak. Ez az adatmennyiség azért nem elhanyagolható, mert becslések szerint a mély weben nagyságrendekkel több információt tárolnak, mint a hagyományos weboldalakon. A mély web és a sekély web (*surface web*) közötti lényeges különbség, hogy ez utóbbit az általános keresőgépek visszakeresik, a másikat azonban nem. A szerver mindkét típusú web esetében eljuttatja a kért weboldalt a klienshez – ha az jogosult rá. A kettő között húzódik az ún. szürke zóna (*gray zone*), amit egyes keresők látnak (pl. a torrent keresők, a Flickr képkeresője), mások viszont nem. A mély web kezelését úgy is támogathatjuk, ha a keresők számára is elérhető metainformációkat közlünk az adatbázisok tartalmáról, valamint különböző csatolóprogramokat hozunk létre a nem szöveges állományokhoz (PDF, Excel, JPG stb.).

Újabb nehézség, hogy a keresőrobotok nem követik a dinamikus weblapokra mutató hivatkozásokat, ezáltal sok információhoz nem férnek hozzá. Ennek az az oka, hogy a dinamikus linkek gyakran hoznak létre hatalmas vagy esetleg végtelen keresési tereket. Ezeket keresőcsapdának (*spider trap*) nevezzük, amelyeket a keresőrobotok megpróbálnak elkerülni. Előfordul az is, hogy bizonyos szerverek megkísérlik álcázni magukat, és egy keresőrobotnak más tartalmat nyújtanak, mint amit egy böngészőnek. Napjainkban számos technika terjedt el a dinamikus oldalak indexelésének támogatására; lényegük, hogy elhitetjük a keresőrobotokkal, hogy statikus hivatkozást követnek.

Az internetes keresők számára a legnagyobb probléma, hogy nem a fellelhető dokumentumok és a keresőkérdés jelentésével foglalkoznak, hanem csupán a szöveges alakkal. A nyelvi problémákat tulajdonképpen az okozza, hogy a mai eszközökkel történő információ-visszakeresés túlságosan a letárolt szöveges információ tényleges alakjára épül. Ennek egyik következménye, hogy a nem szöveges dokumentumok által hordozott információk nem kereshetők vissza automatikusan. További hiányosságként kiemeljük, hogy a keresőrendszerek nem ismerik a fogalmak jelentését és a fogalmak közötti kapcsolatokat, ezért nem képesek különféle következtetések levonására [10]. Ezt a problémát a szemantikus keresők orvosolják hatékonyan.

Az internetes keresőknek létezik egy másik fajtája, a webes katalógusok, amelyek emberek által összegyűjtött oldalakat tesznek visszakereshetővé. Ezek a katalógusok eredményesen oldják meg a jelentés, azaz a szemantika megragadását, ami az oldalak begyűjtését és indexelését végző emberek feladata. Előnyük, hogy oldalaik megbízhatók és minőségük garantált, mert emberek válogatják ki őket. Itt nemcsak szöveges keresésre van lehetőségünk, hanem témakategóriák között is böngészhetünk. Ezekben a gyűjteményekben nagy segítséget jelent, hogy az oldalakat emberek olvasták végig kategorizálásukkor. Legnagyobb hátrányuk viszont, hogy a létező weboldalnak csak kis hányadát tartalmazzák. Ezenkívül meg kell említenünk a kérdésátalakító keresőket is, amelyek szintén a jelentés megragadására törekednek. Feladatuk, hogy megpróbálják jobban értelmezni a feltett keresőkérdést és azt úgy átalakítani, hogy az új keresőkérdés már jobb találatokat eredményezzen. Egy ilyen átalakításhoz a keresőknek kell, hogy legyen valamilyen matematikai formalizmussal leírható háttértudásuk.

A szemantikus webírányzat hatékonyan oldja meg a jelentéssel kapcsolatos problémakört, amelynek fő célja, hogy jelentést vigyen a webre. Ez úgy válik lehetségessé, hogy a weboldalak előállítói a webes tartalmakhoz szabványos formában metaadatokat rendelnek, a szemantikus web pedig biztosítja számunkra, hogy ezen metainformációk alapján következtetéseket vonjunk le. Jelenleg a metainformációk ugyanolyan heterogén formában fordulnak elő, mint maguk a webes dokumentumok. Ezért a szemantikus web fejlesztőinek elsődlegesen a metainformációk és a következtetéshez szükséges háttértudás egységes és feldolgozható alakban történő leírására kell törekedniük [10].

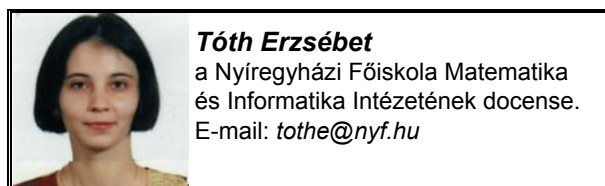
Irodalom

- BRIN, S. – PAGE, L.: The anatomy of a large-scale hypertextual web search engine. = *Computer Networks and ISDN Systems*, 30. köt. 1-7. 1998. p. 107–117.
<http://infolab.stanford.edu/pub/papers/google.pdf> (2007.03.02.)
- BENCZÚR A. – BÍRÓ I. – CSALOGÁNY K. – RÁCZ B. – SARLÓS T. – UHER M.: PageRank és azon túl: Hiperhivatkozások szerepe a keresésben. = *Magyar Tudomány*, 167. köt. 11. sz. 2006. p. 1325–1331.
<http://www.matud.iif.hu/06nov/07.html> (2007.07.17.)
- FRIEDMAN E. – UHER M. – WINDHAGER E.: Keresés a világhálón. = *Híradástechnika*, 58. köt. 3. sz. 2003. p. 20–24.
<http://www.cs.elte.hu/~hexapoda/kereses.pdf> (2010.04.10.)
- FROEHLICH, T. J.: Case study 5.1: Developing search engine evaluation criteria. = *Library evaluation. Libraries Unlimited*, 2001. p. 185–200.
- HAWKING, D. – CRASWELL, N.: Very large scale retrieval and web search. = *TREC: Experiment and evaluation in information retrieval / Ellen Voorhees, Donna Harman editors. MIT Press*, 2005.
http://es.csiro.au/pubs/trecbook_for_website.pdf (2007.07.10.)
- KOSTER, M.: A method for web robots control. Technical report, Internet Engineering Task Force (IETF), 1996.
<http://www.robotstxt.org/norobots-rfc.txt> (2010.04.22.)
- PAGE, L. – BRIN, S. – MOTWANI, R. – WINOGRAD, T.: The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.
<http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=&name=1999-66.pdf> (2007.07.17.)
- SCHWARTZ, C.: Web search engines. = *Journal of the American Society for Information Science*, 49. köt. 11. sz. 1998. p. 973–982.
- SULLIVAN, D.: Checking your listing in search engines, October 2001.
<http://searchenginewatch.com/webmasters/article.php/2167861> (2007.07.10.)
- SZEREDI P. [et al.]: A szemantikus világháló. = *A szemantikus világháló elmélete és gyakorlata. Budapest, Typotex*, 2005. p. 17–59.
- A világ előrehaladása az információs társadalom terén 1998–2008. *World Progress Report 2008*. Készít. a BME-UNESCO Információs Társadalom-

és Trendkutató Központjának (ITTK) kutatócsoportja. Budapest, 2007. március
http://www.ittk.hu/web/docs/ITTK_WPR1998-2008.pdf (2008.01.28).

12. Web search engine (Webes keresőgép) szócikk. = http://en.wikipedia.org/wiki/Web_search_engine#History (2010.04.22.)

Beérkezett: 2010. V. 25-én.



Hangtalan társalgás a számítógéppel

A *Silent Speech Interface* lényege, hogy az emberek beszédmozgásával kapcsolatos izommozdulatait figyeli, és ezek alapján értelmezi a hangtalanul közölt szavakat.

Tanja Schultz a *Karlsruhei Technológiai Intézet (KIT) Antropomatikai Intézetéhez* tartozó *Kognitív Rendszerek Tanszékén* dolgozik. Az intézményt februárban alapították, összesen 120 kutatója van. Schultz szakterületének az emberi biojeleken alapuló technológiák és alkalmazások számítanak, ide értve az izom- és agyi tevékenységen alapuló beszédfelismerést és -interpretációt. Az antropomatika egy mesterséges szó, amelyet a görög *anthropos* (ember) és az automatikára utaló *matik* szavakból hoztak létre. Az antropomatika alatt az ember és a gép szimbiózisát érti a tudomány. A célja az embereken alapuló rendszerek kutatása és fejlesztése informatikai eszközök segítségével.

„A Silent Speech Interface lehetővé teszi, hogy hangtalan beszédet továbbítsunk. Az alapötlet a következő: a beszéddel kiküldött akusztikus jeleket gyakran eltorzítják a háttérzajok, például egy vonaton ülve vagy egy zsúfolt csarnokban. Éppen ezért működik annyira rosszul a beszédfelismerés hangos környezetekben. A technológia az EMG-n, vagyis az izomtevékenység elektromos jeleinek felderítésén és rögzítésén alapul. Vagyis az akusztikai jelek helyett mi ezekre építünk. Ha valaki hangtalanul mozgatja az ajkait, akkor a szájmozgásával kifejtett EMG-jeleket továbbítjuk a számítógépnek. Egy szoftver képes e jelek alakja alapján megmondani, hogy az illető melyik izma mozgott, majd pedig kiszámolja a hangot. A jeleket később beszéddé alakítja át, majd ezt egy számítógépes hang segítségével közvetíti. Ez a technika nemcsak a beszédfelismerésben használható, hanem lehetőséget ad arra is, hogy akik egy betegség miatt nem tudnak beszélni, újra megtehessék ezt. Ilyenek lehetnek, mondjuk a gégerákban szenvedők, akiknek megsérültek a hangszálai. Ezenkívül lehetőség nyílik a hangtalan telefonálásra. Ez elsősorban olyan embereknek lehet érdekes, akiknek telefonálniuk kell, de nem szeretnének hangosak lenni. Érdekes lehet a szoftver az elektronikus banki és más beszédinterfészt használó programok esetében is. A Silent Speech Interface használatával még a bizalmas információkat (kódokat, jelszavakat) is biztonsággal lehetne kimondani.” – jelentette ki Tanja Schultz.

Az új interfész hibaaránya erősen függ a viselőtől és attól, hogy az illető mennyire jól artikulál. A jól artikulálóknál a hibaarány 5-10 százalék, az átlagos hibaarány pedig 100 szó esetében 10-20 százalékos. A rendszert tesztelték már kínai és német nyelvű embereknél is, valamint hamarosan kezdődnek a japán tesztek, de a technológia gyakorlatilag nyelvfüggetlen.

„A technológia még kiforratlan, ráadásul az embereket zavarja, ha kábelekkal az arcukon kell beszélniük. Ez a megoldás nem szép, de számos előnye van. A szenzortechnológia tovább fog fejlődni. Már most vannak kísérletek emberi szervezetbe ültethető EMG-elektrodákkal. Az elektrodák a jövőben még kisebbek lesznek; eljön az idő, amikor majd egyszerűen beinjekciózhatjuk őket.” – mondta Tanja Schultz.

A KIT-nél külön intézet foglalkozik azzal, hogy milyen következményekkel járhat a technika alkalmazása. Ha ugyanis valaki valóban adaptálható rendszereket fejleszt, akkor foglalkoznia kell azzal is, hogy ezeknek milyen morális, etikai és szociális következményei lehetnek a társadalomra nézve. Egy gép csak akkor hasznos, ha képes alkalmazkodni az emberi igényekhez. Sok technika vesz körül minket, de rengeteg időt és energiát fordítunk arra, hogy ezeket a saját igényeinkhez igazítsuk. Valójában ennek fordítva kellene lennie: a technikának kellene tudnia, hogy mit akarunk, és ahhoz kellene megfelelő szolgáltatásokat kínálnia.

/SG.hu Hírlevél, 2010. július 26., <http://www.sg.hu/>

(SzP)

Winkler Bea

SEO a könyvtárban, avagy a SZIE Állatorvos-tudományi Könyvtár új portálja és a keresőoptimalizálás

A felhasználói igényeknek megfelelően a könyvtárak többsége képviselteti magát a világhálón. Az interneten böngésző eléréséhez ez azonban ma néha már kevés. A SZIE Állatorvos-tudományi Könyvtára szeretné, ha most készülő új honlapját minél több látogató keresné fel és ennek eléréséhez a keresőoptimalizálás látszik a legkézenfekvőbb megoldásnak. Az esettanulmány célja nem a keresőoptimalizálás lépéseinek általános leírása, hanem egy konkrét könyvtár/könyvtáros induló lépéseinek szemléltetése, és remélhetőleg mások figyelmének felhívása a kérdés fontosságára. A cikk a meglévő könyvtári honlap feltérképezését, és az új portál építésének keresőoptimalizálási szempontból érdekes pontjait kívánja meghatározni, továbbá szeretne bemutatni néhány ingyenes és szabadon használható segítséget.

Bevezető

Egy unikális szakkönyvtár sokféle szempontot kell, hogy figyelembe vegyen az új könyvtári honlap tervezésekor, de talán a legfontosabb az alacsony költség és a nagy megtérülés. A könyvtárak többségének anyagi lehetőségei nem teszik lehetővé, hogy külső szakértők, cégek munkájának eredményeként szülessen meg honlapjuk. A következő cikk hasonló helyzetben egy könyvtáros lépéseit igyekszik összeszedni, aki a honlap elkészítésén túl szeretné minél több használó számára elérhetővé, megtalálhatóvá tenni a Szent István Egyetem (SZIE) Állatorvos-tudományi Könyvtár, Levéltár és Múzeum portálját.

Egy 2008-as felmérés szerint Magyarországon a leggyakoribb internetes tevékenységek között találjuk a keresőhasználatot (94%) [1]. Megvizsgálva a könyvtári honlap statisztikáját¹, megállapítottuk, hogy a látogatóknak 20%-a érkezik keresőn keresztül, 30% direkt (beírja az URL-t) és majdnem 50% más oldalakról linkeken keresztül talál meg minket.

A fenti adatok alapján céljaink megvalósításához a keresők találatában való minél jobb, gyakoribb megjelenésre kell koncentrálnunk, amelynek elősegítésével a keresőoptimalizálás (*Search Engine Optimization = SEO*) témaköre foglalkozik.

Keresőoptimalizálás

A keresőoptimalizálás az a folyamat, amely lehetővé teszi, hogy az információkeresők nagyobb valószínűséggel találják meg és nyissák meg a weboldalunkat. Kettős feladata van: a keresők által generált forgalom mennyiségének és minőségének javítása [2].

A keresőoptimalizálás elsődleges célja, hogy a keresőben oldalunk minél előrébb szerepeljen a találati sorban. Az ehhez elvégzendő feladatokat több szempontrendszer szerint lehet csoportosítani, melyek közül az egyik a *Malaga* [3] által leírt négyes felosztás:

1. kulcsszavak, szókapcsolatok felderítése – kiválasztása,
2. a keresők indexébe való bekerülés elérése,
3. az oldal optimalizálása (kód szinten),
4. az oldalunkra mutató linkek szerzése (külső linkek).

A fenti feladatokat illetően könyvtárunk kedvező helyzetben van, hiszen a meglévő honlap lehetőséget nyújt a keresők által hozzákapcsolt kulcsszavak felderítésére, és ezeket csak ki kell egészíteni a számunkra fontos témákban népszerű keresési kifejezésekkel.

A második lépés a keresők indexébe való bekerülés, amely a találati listákban való megjelenés

alapja. A működő weboldalt jelenleg is indexelik a nagy keresők, így ezzel nem foglalkozom.

Az optimalizálás (kód szinten) egy meglévő honlap esetén sok feladatot adhat, az új portálnál azonban, ha a lépéseket előre meghatározzuk, akkor ez nem jár külön munkával. Tervünk szerint így az elérendő célokat és a megvalósítás lehetőségeit még a portál tartalmi építése előtt kell feltérképezni, hogy az oldalak létrehozásakor már alkalmazhassuk őket.

Működő honlapunk esetében a látogatások 50%-a külső linkeken keresztül érkezik, ami jól mutatja a linkek fontosságát számunkra is. Külső linkek szerzésének több módszere létezik, például: *önkéntes* linkadás (jelenleg a könyvtár honlapjára elsősorban ilyenek mutatnak), linkcsere, linkkatalógusok (pl. a *lap.hu* oldalak) és térítéses szolgáltatások. Ezek közül véleményem szerint egy könyvtárnak mint kulturális, oktatási vagy tudományos tartalmak közvetítőjének az első három lehetőségre kell koncentrálnia.

A keresőoptimalizálás nem hasonlít a többi marketingtevékenységre, mivel költségei és megtérülése nem könnyen mérhető a megszokott módszerekkel. Hosszú távú befektetés ez, amely azonban segíthet az interneten található hatalmas információtömegben eligazodni vágyó közönség elérésében.

A megvalósítás során két jól elkülöníthető megoldási modell létezik, amelyek közül én csak az elsővel, az ún. „megengedett” vagy „fehér kalapos” technikákkal kívánok foglalkozni.

A másik a „fekete kalapos” (black hat), illetve tiltott módszerek csoportja, amelyek eredménye hosszú távon az oldal hátrasorolása lehet a találati listákon, de akár a keresőkben való teljes blokkolása is. Ide tartozik például a csak az indexelést végző robotok számára látható hamis tartalmak létrehozása (elrejtésük pozícionálással, színekkel, „hidden” tulajdonsággal stb.), az oldal tartalmával nem rokon, de keresésekkor nagyon népszerű kifejezések elhelyezése az oldalon, linkfarmok kiépítése, melyek oldalai mind egymásra mutatnak stb.

Google? Google!

Több adat szerint is a Google a legnépszerűbb, a legtöbbek által használt kereső, így az optimalizá-

lás során elsősorban erre a keresőre fogom helyezni a hangsúlyt.

Egy 2006-ban készült felmérés alapján a keresőket használók 62%-a csak az első találati lapon kattint, és csak 10%-uk teszi meg ugyanezt a 3. oldal után is [4]. Így az a cél, hogy honlapunk az első oldalon szerepeljen az általunk fontosnak gondolt kifejezések találati listájában, még inkább felértékelődik.

Az első oldalon való megjelenésnek két módja van: vagy az organikus találati listában² „odasorolódunk”, vagy szponzorált linkként (az organikus találatok előtt, vagy a jobb oldalon) kerülünk oda. Természetesen könyvtárunknál az első a megvalósítandó.

Az internet tele van a Google és más kereső rangsorolásra használt algoritmusainak leírásaival, én azonban a *SEOMoz.org* oldalán publikált felmérés eredményét választottam kiindulásnak. SEO szakértők válaszoltak több mint 100 kérdésre, amelyek a keresés rangsorolását befolyásoló tényezőket igyekeztek feltárni tapasztalataik alapján. Az általuk legfontosabbnak tartott öt jellemző [5]:

- az oldalra mutató linkek szövege (anchor text),
- az oldalra mutató linkeket tartalmazó oldal népszerűsége (mennyiségi és minőségi szempontból),
- a bejövő linkek sokfélesége (hány doménről vannak a linkek),
- a kulcsszavak elhelyezése a <title> tag-ben,
- az oldal biztonságossága (megbízhatósága) a doménre alapozva.

A fenti felsorolásból jól látszik, hogy a Google, illetve más nagy keresők igyekeznek rangsorolásuknál minél több, a honlapot készítő által nehezebben befolyásolható tényezőt is figyelembe venni. Itt természetesen felmerülhet a kérdés, hogy miért érdemes foglalkozni a keresőoptimalizálással, ha az igazán lényeges dolgokon nem tudunk közvetlenül változtatni. Ugyanakkor fontos tudni, hogy a felsorolt öt tétel nem a teljes lista, hiszen a keresők akár 200 szempontot is képesek egyszerre figyelembe venni a rangsor felállításához.

Az megadott listát kiegészíthetjük azokkal a SEO szakértők által lényegesnek tartott tényezőkkel, amelyek befolyásolására képesek vagyunk, és amelyeket a keresőoptimalizálás folyamán érdemes szem előtt tartani (Malaga 4. pontjához kapcsolódva, amely a kódszintű keresőoptimalizálásról szól) [5]:

- a <title> tag első szava legyen fontos kulcsszó;
- kulcsszó elhelyezése a doménnévben (esetleg aldoménben) például *konyvtar.univet.hu*;
- kulcsszó *h1* tagban, építsünk fel logikus és átlátható struktúrát a html által nyújtott jelölők (tag-ek) segítségével, vagyis használjuk a címek kiemelésére kitalált címkéket, amilyen a *h1*, *h2* stb.;
- anchor text – a belső linken és a kifelé mutató linkeken egyaránt;
- egyedi tartalom az oldalakon;
- országcód a doménben;
- nyelv megadása.

Véleményem szerint egy keresőkkel jól megtalálható oldal több látogatót vonz, és ez több, az oldalunkra mutató linket eredményezhet, mindez pedig tovább javíthatja a keresők találati listáiban elfoglalt pozícióinkat.

Tájékoztató

A keresőoptimalizálás és az akadálymentesség egymást erősítő tényezők, amire sokan rájöttek, hiszen a világháló leggazdagabb, legbefolyásosabbak felhasználói a Google, a Yahoo! és az MSN Live Search keresőrobotjai [6]. A könyvtári portál létrehozásakor tehát a szabványosság mellett az akadálymentességnek is fontos szempontnak kell lennie, kiegészítve a keresőoptimalizálással.

Az új honlap tervének kidolgozásakor megnéztük több magyar egyetemi könyvtár honlapját és néhány helyen személyesen is jártunk tapasztalatszerzés céljából (BME, BCE, Pannon Egyetem, SZTE). Végül a Joomla! tartalomkezelő rendszer (Content Management System = CMS) használata mellett döntöttünk.

A SZIE Állatorvos-tudományi Könyvtára 1997 óta tart fenn saját weboldalt. A 2010 márciusában működő honlap eredetileg 2005 közepén készült el. Annak ellenére, hogy az oldalaknak php végződésük van, a tartalom szempontjából statikus oldalról beszélhetünk, amelyek egy közös fejléccel használnak és a navigációt egy JavaScript menü adja. A következőkben sokféle vizsgálat alá vettem ezt a honlapot, hogy kiderítsem erősségeit és gyengeségeit, mert ezen tapasztalatok felhasználásával fog elkészülni az új portál.

Kulcsszavak

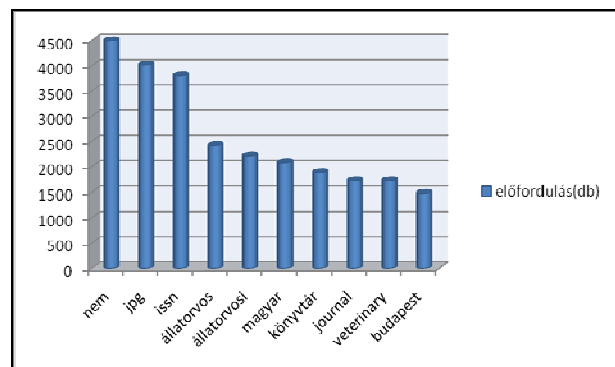
Első lépésként a kulcsszavak felderítését végeztem el több szempont szerint is.

A Google-ban keresett kifejezések

A Google Analytics adatai szerint a látogatók a következő kifejezések keresése esetén érkeztek könyvtárunk honlapjára (csak a legnépszerűbbeket említem, mivel összesen 1180 kulcsszót találunk a listán): *univet könyvtár*, *konyvtar.univet.hu*, *állatorvosi könyvtár*, *univet.hu*, *kazár*, *állatorvostudományi könyvtár*, *állatorvosi egyetem könyvtár*, *univet library*. A Google és a Yahoo keresőjében lekeresve ezeket, kiderült, hogy az előbbi kiemelt kifejezések esetén (az *univet.hu* és a *kazár* kivételével) a könyvtár honlapja az első a találati sorban, így ezekről a kulcsszavakról elmondhatjuk, hogy biztosan „erősek” az oldalak.

Kulcsszavak a honlaphoz (Google)

A Google Webmestereszközök szolgáltatás részeként megtudhatjuk a Google által a honlapunkhoz kapcsolt kulcsszavak listáját (1. ábra).



1. ábra A kulcsszavak megoszlása a honlapon való előfordulásuk szerint (a teljes honlap átlaga) [2010.01.12.]

Ezek az adatok azért érdekesek, mert az egész honlapot figyelembe véve készültek (ami több mint 400 oldalt tartalmaz) és a legjellemzőbbnek ítélt kulcsszavak semmitmondók és irrelevánsak (*nem*, *jpg*, *issn*) a honlap tartalmával kapcsolatban. A SEOmoz.org felmérésének eredményei közül, a doménnévben található kulcsszavak fontosságát

megerősítő tény, hogy az *univet* szó az előbb említett táblázatban nem is volt feltüntetve, mégis a legtöbb látogatást generáló keresőszavak között szerepel.

Google AdWords – Kulcsszóválasztó eszköz

A kulcsszavak keresését a saját meglévő oldalunkhoz köthetőkkkel kezdtem. Következő lépésként meg szeretném tudni, hogy két alapkifejezésnek számító szó esetén milyen szinonimákat, illetve hozzájuk kapcsolható kulcsszavakat ajánl a Google hirdetéssel foglalkozó oldalának ingyenes „Kulcsszóválasztó eszköz”-e (1. táblázat).

Az előző táblázatban megjelenített szavak a legnépszerűbb keresett kifejezéseket mutatják a keresések számával, az elmúlt 12 hónap átlagát figyelembe véve. Az értékek kerekítettek, azonban amikor a honlap különböző részeinek elnevezését, a linkek szövegeit szeretnénk megválasztani, akkor mégis érdemes lehet ezeket (is) figyelembe venni a döntéshez.

A *katalógus* vagy *opac* szó honlapon való használatának eldöntéséhez például jó támpontot ad, hiszen 120-szor több esetben kerestek a *katalógus* szóra, mint az *opac*-ra, továbbá még az angol *catalogue* szó is majdnem nyolcszoros eredményt ért el ugyanerre a mintára. Ha a használók érdekeit tartjuk szem előtt, az általuk ismert és kedvelt kifejezést alkalmazzuk elsődlegesként az oldalon.

1. táblázat

A „Kulcsszóválasztó” ajánlatai a könyvtár és a katalógus/opac szóra (csak a legnépszerűbbek)

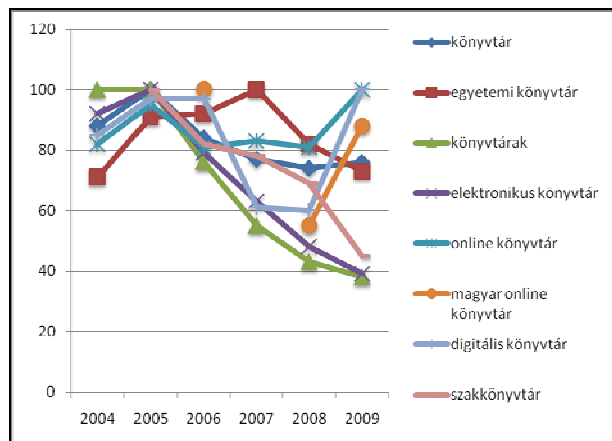
a könyvtár szóra		a katalógus, illetve az opac szavakra	
helyi keresés mennyisége, az elmúlt 12 hónap átlaga (2010. március)			
könyvtár	246 000	opac	1900
elektronikus könyvtár	18 100	online katalógus	4400
szechenyi	9900	katalógus	246 000
könyvtárak	8100	catalogus	22 000
katalógusok	3600	corvina	12 100
		katalógusok	3600

Kulcsszóválasztáshoz további segítséget jelenthetnek a *Microsoft adCenter Labs* szolgáltatásai (pl. a Search funnels), amelyeknél grafikusán ábrázolva tekinthetjük át az általunk keresett angol nyelvű szóhoz kapcsolódó kifejezéseket.

Keresési trendek

A Google „Keresési Trendek”-et bemutató oldalain lehetőséget adnak arra, hogy az általunk fontosnak tartott kifejezésekre történt kereséseket egyféle trendként, évekre visszabontva megtekinthessük. Lekerestem néhány érdekes szókapcsolatot, kifejezést (szűrők: internetes kereső, világszerte, 2004 – jelenleg, minden kategóriában). A túl hosszú kulcsszavak (*állatorvos-tudományi könyvtár* és *állatorvosi könyvtár*) nem mutattak értékelhető eredményt.

A 2. ábra néhány könyvtárral kapcsolatos kifejezés keresési trendjeit tartalmazza (az adott év legmagasabb értékét kiemelve a Google által alkalmazott normalizált eredményekből [7]). A keresési trendvizsgálat bonyolult kutatási terület és az ilyen egyszerű alkalmazással kapott eredményekből nem szabad hirtelen következtetéseket levonni és tényként kezelni őket. Azért tartottam fontosnak mégis megnézni ezeket, mert a *könyvtár* szóval kapcsolatos területen ezek voltak a legnépszerűbb keresési kifejezések az elmúlt években. Ha intézményünk nyitni szeretne a nagyközönség felé, akkor ezek keresése esetén kellene jobb eredményt elérnünk. Ráadásul néhány esetben a meglévő oldalaknak csak új, közérthetőbb nevet kell adni a jobb elérhetőség érdekében.



2. ábra Google „Keresési Trendek” eredményei [2010.01.23.]

A kulcsszavak használata

Metatag

A különböző keresők különböző módon értékelik az oldalakon elhelyezett metatag-eket, így például a Google nem foglalkozik velük, de a Yahoo!, a Bing vagy más kisebb keresők igen. A SEOmoz.org felmérésének eredményei szerint a leírásnak és a kulcsszónak is minimális a fontossága (19% illetve 5% a relevanciájuk).

A könyvtári webhely oldalai jelenleg közös fejléccet használnak, ami egyforma kulcsszavakat és leírást jelent (összesen 422 teljesen megegyező). A Joomla (illetve más tartalomkezelő rendszerek is, például a Drupal) lehetővé teszi, hogy minden cikkhez külön címet, leírást és kulcsszavakat adjunk meg. Ez az adott tartalom létrehozásakor nem okoz jelentős pluszmunkát, de hatása a keresők egy részénél mindenképpen fontosnak tekinthető.

Kulcsszavak a szövegben

Az előzőekben felderített és az adott laphoz kapcsolódó kulcsszavakat használhatjuk az oldal címében (<title>), a tartalom felépítésekor alkalmazott hierarchikus címrendszerben (*h1* a legfontosabb, de a többi is alkalmazható pl. *h2*), a szövegben (szinonimákat is) akár kiemelésekkel (), link-szöveggként (<a> anchor text) vagy képek leírásakor is gondolhatunk rájuk (). A kulcsszavak átgondolt használata esetén törekedni kell arra, hogy az oldalra látogatók számára továbbra is könnyen áttekinthető és jól olvasható szöveget adjunk, így felmerül a „kulcsszósűrűség” kérdése.

Kulcsszósűrűség

A kulcsszósűrűséget úgy is felfoghatjuk, mint az *oldal szövegeinek harmonizálására tett kísérletet*, így az itt kapott eredmény (egy százalékos arányszám) megmutatja, ha túl sokszor alkalmazzuk a számunkra elengedhetetlennek tűnő kifejezést. Az optimális arány 2-5% között lehet, bár vannak ennél magasabb értéket mutató eredmények.

Ha azonban az oldalunkat tudatosan építjük fel és a használt kifejezéseket átgondoltan alkalmazzuk, akkor mindenképpen érdemes kiszámolni ezt az értéket, nehogy igyekezetünk a megfelelő kifejezések használatára pont az ellenkezőjét hozza, és valamely kifejezés vagy egy szinonima aránytalan túlsúlyt kapjon.

A kulcsszósűrűség mérésére több ingyenes lehetőség létezik mind magyar, mind külföldi oldalakon. Én a Walter [6] által ajánlott RANKS.NL oldalt használtam a nyitó oldal áttekintéséhez (3. ábra).

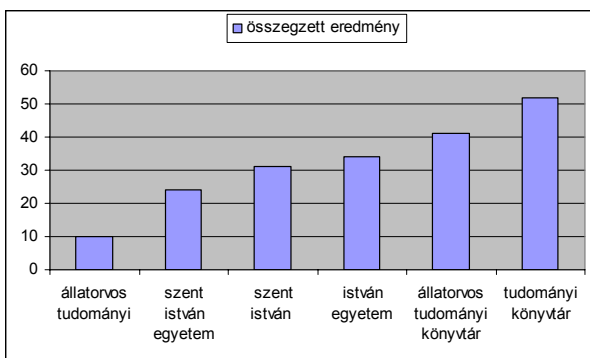
Single word repeats							
word	repeats	density	Prominence	word	repeats	density	Prominence
könyvtár	11 T,D,K,B,A	3.77%	66.98	állatorvos	7 T,D,K,L,B,A	2.40%	70.99
tudományi	6 T,D,K,L,B,A	2.05%	67.96	szie	4 L	1.37%	43.60
478	4	1.37%	64.53	nyitva	3 L,B	1.03%	46.12
istván	3 T,L,B	1.03%	61.11	állatorvosi	2 K	0.68%	86.24

Phrase repeats							
Total 2 word phrases: 12 - Total Repeats : 31				Total 3 word phrases: 6 - Total Repeats : 15			
Phrase	repeats	density	Prominence	Phrase	repeats	density	Prominence
állatorvos tudományi	6 T,D,K,L,B,A	4.11 %	76.16	állatorvos tudományi könyvtár	5 T,D,K,A	5.14 %	84.34
tudományi könyvtár	5 T,D,K,A	3.42 %	83.95	egyetem állatorvos tudományi	2	2.05 %	66.86

3. ábra Kulcsszavak és kulcsszósűrűség a <http://www.konyvtar.univet.hu/index.php> esetén [2010.01.20.]

A program eredményoldala részletes tájékoztatást ad arról, hogy az adott kifejezés hányszor és hol szerepelt az oldalon (T = <title>, D = meta description, K = meta keywords, L = linkszöveg (anchor text), B = bold [vastaggal kiemelt szöveg], A = alt tag). A szolgáltatás részeként megtudhatunk még számtalan érdekes információt: Google PageRank, külső és belső linkek, tagfelhő, hány stopszó szerepel és melyek azok.

A közös fejlécek használata miatt ugyanazon kifejezésekre látszik optimalizálnak az oldalak többsége. A következő ábrán 10 oldal (a könyvtár menüjében az első tíz, 2010. április 14-én) optimalizálási eredményének összegzése látható úgy, hogy a legjellemzőbb 1 pontot kapott, a második 2 pontot és így tovább, vagyis minél alacsonyabb az összpontszám, annál optimalizáltabb az adott kifejezésre a vizsgált tíz oldal. Az ábrán 10-es eredményt elérő *állatorvos tudományi* kifejezés tehát minden vizsgált oldalon első helyet ért el (4. ábra). Jól látszik, hogy a címben és a metaleírásban sokszor ismétlődő kifejezések milyen jó eredményt értek el, mégis miért nem szerencsés ezek ismétlése minden oldalon. Az egyedi oldaltervezés, főképp az egyedi tartalmú, nem általános információkat közlő oldalaknál elengedhetetlen. Az adatbázisok listáját tartalmazó oldal esetén például logikus volna, ha az *adatbázis* szóra is „erős” lenne az oldal.



4. ábra Kulcsszavak erőssége: 10 kiválasztott könyvtári oldal összesítése [2010.04.01.]

Keresőbarát URL-ek

A doménnévben (esetleg aldoménben) elhelyezett kulcsszavak fontosságára a *SEOMoz.org* listája már felhívta a figyelmet. Könyvtárunk egy aldoménen található, a *konyvtar.univet.hu* cím alatt, ami-

nek előnye az univet szóval kapcsolatos találatok száma is megerősítette. Az URL másik fele: a mappa, illetve az adott lap neve szintén jelentőséggel bír, mivel a találati listában az itt szereplő keresőszavak is ki lesznek emelve.

A jelenleg működő webhelyen általában az elérési út egy része angolul, míg az adott dokumentum neve magyarul szerepel, sőt ez sok esetben csak valamilyen rövidítés. Például *database/szakdbuj.php* annak a lapnak a címe, ahol az adatbázisok listáját találjuk, az ajánlott forma pedig – még ha hosszabb is – a következő lenne: *adatbázisok/szakteruleti_adatbázisok.php*.

Az új honlaponál alkalmazott Joomla lehetőséget teremt a keresőbarát URL-ek használatára, amit globális beállításként lehet engedélyezni (keresőbarát URL-ek, Apache *mod_rewrite* használata, utótag hozzáfűzése az URL-ekhez). A Drupal rendszernél szintén van lehetőség előre generált keresőbarát URL-ek létrehozására, illetve URL aliasként egyesével is megadhatjuk őket a tartalom létrehozásakor.

Linkek

Az internet állandóan változó világában az oldalak közötti kapcsolat mindennek az alapja, amelynek megtestesítője maga a link.

A linkekkel kapcsolatos legfontosabb fogalom az anchor text, vagyis a kattintható szövegrész. A keresőkben való megtalálhatóság egyik fontos tényezője ugyanis az oldalunkra mutató linket (backlink) tartalmazó oldal tulajdonságain felül a linkhez tartozó szöveg. Ez akkor a legkedvezőbb, ha az oldalunk tartalmával megegyező kulcsszavak fordulnak elő benne. Ha mi készítünk linket, ezt szem előtt is tarthatjuk és mind a belső, mind a külső linkelésnél kerüljük az *ide kattints, itt, erre, tovább* és ezekhez hasonló semmitmondó megfogalmazásokat.

A külső linkek vizsgálata, feltérképezése a keresőkkel és a keresőoptimalizálással kapcsolatos kérdésben fontos szerepet tölt be. Az egy adott oldalra mutató linkek mennyiségét és minőségét sokféle szolgáltatással megvizsgálhatjuk, a 2. táblázatban bemutatunk néhány példát a mennyiségek megjelölésével.

2. táblázat

Összehasonlító táblázat a Google és a SEO Toolbar adataival [2010.01.25.]

az oldal leírása az oldal linkje a <i>konyvtar.univet.hu</i> oldalon	Google speciális keresés „link alapján” az oldalra mutató linkek száma	SEO Toolbar az oldalra mutató linkek száma	Google Webmaster Eszközök az oldalra mutató linkek száma
nyitó oldal /index.php	83	448	280
Madárinfluenza /ugro/links/birdflu.htm	0	6858	1038
érmek, plakettek /portre/plak/tartplak.htm	2	6	23

Az eltérő eredmények nehezítik a tájékozódást, de nem pusztán pontatlanságból fakadnak. Minden kereső más és más honlapok feltérképezését végzi el, másokat helyez el az indexébe, és részben ez adja a közöttük lévő különbséget. A későbbiekben érdemes lehet a fenti eredmények részletesebb vizsgálata, hogy a sokak által linkelt oldalak elköltöztetése ne jelentsen egyet látszólagos megszűnésükkel. A kérdés gyakorlati megvalósításával az új honlap teendői kapcsán foglalkozom majd.

Oldaltérkép

A látogatók eligazodását segítő oldaltérkép régóta használt elem a weboldalakon, létezik azonban a keresők segítségére használt változat is. Az XML sitemap célja, hogy szabványos³ és mégis könnyen kezelhető formában hozzunk létre a keresők botjai által is értelmezhető oldaltérképeket. A fájlban megadhatjuk a portálon található összes oldal linkjét, az utolsó módosítás dátumával, a frissítés gyakoriságával (mindig, óránként, naponta, hetente, havonta) és a fontosságával (prioritásával).

Az xml oldaltérkép létrehozására lehetőségünk van saját kezűleg, vagy (ingyenes vagy fizetős) generátorok alkalmazásával. Joomla esetén például a *SEF Service Map 2* komponens teszi lehetővé a könnyedebb, ugyanakkor pontosabb generálást, Drupal esetén használhatjuk az XML sitemap modult. A Google Webmastereszközök oldalán lehetőség van saját oldalunk xml oldaltérképének feltöltésére a feltérképezést segítő, és a robots.txt-ben is hírt adhatunk létezéséről.

PageRank

A találati oldalakon történő rangsorolásnál számtalan szempontot figyelembe vesznek a keresők és

ezek együttes eredménye adja a listát. A mutatók közül az egyik legismertebb a *PageRank*. Maga a kifejezés a Google tulajdonában lévő védjegy és a technika leírása szabadalom⁴. Ennek lényege, hogy egy adott oldal értékét a rá hivatkozó oldalak számának és minőségének ismeretében egy algoritmus számítja ki.

Annak ellenére, hogy a rangsorolást befolyásoló tényezők közül ez csak egy (a több mint 200 mérőszámból [9]), sokan egyfajta bűvös számként tekintenek rá. A böngészőbe épülő kiegészítők (pl. Seo Status PageRank/Alexa Toolbar), vagy a már használt RANKS.NL egyaránt jó megoldás lehet ilyen irányú kíváncsiságunk kielégítésére. A PageRank maximális értéke 10, és jelenleg a könyvtár nyitó oldala 6-os értéket kap, a többi pedig 4 és 5. Ez az eredmény kielégítőnek mondható.

Alexa

Az *Alexa* egy internetes információs cég, amely lehetőséget nyújt arra, hogy egy, az oldalunk forgalmához köthető mérőszámhoz hozzáférjünk. Ez a szolgáltatás az érték kiszámításához a domént veszi alapul, így könyvtárunk (amely egy aldoménon található) honlapjának ellenőrzésekor az *univet.hu* oldal adatait adja vissza. Az eredményoldalon viszont az adott doménre néhány, már az előzőekben megismert információt is megkapunk, keresőkifejezéseket, bejövő linkeket stb.

mozRank

A *SEOMoz.org* oldal 2010 elején új szolgáltatást indított *Open Site Explorer* néven, az eddig is mű-

ködő *Linkscape* nyomdokain. Az itt elérhető *mozRank*-et a PageRank-tól több dolog különbözteti meg: például, hogy a szolgáltatás a SEOmoz saját adatbázisból dolgozik. Célja egy egyszerű, gyors és könnyen átlátható linkkapcsolatokat alapul vevő mérőszám biztosítása. Előnyének tekinthetjük, hogy nem csak egy egész számot ad meg, így árnyaltabb eredményt kapunk, mint az előző mutatóknál. Az 5. ábra adatai meglepőek voltak számunkra, ugyanis a *www* nélküli honlapváltozatnak jelentősen több linkje van és magasabb a MozRank értéke, mint a *www*-snek.

Page-Specific Metrics	URL
Page Authority	31
mozRank	4.82
mozTrust	Pro Only
Total Links	709
External Followed Links	482
Internal Followed Links	Pro Only
Linking Root Domains	Pro Only

<http://konyvtar.univet.hu>

Page-Specific Metrics	URL
Page Authority	19
mozRank	3.74
mozTrust	Pro Only
Total Links	6
External Followed Links	4
Internal Followed Links	Pro Only
Linking Root Domains	Pro Only

<http://www.konyvtar.univet.hu>

5. ábra Az Open Site Explorer eredmények a kétféle névváltozat esetén [2010.04.10.]

Ezeknek az eredményeknek valószínűleg az az oka, hogy néhány évvel ezelőttig az Állatorvostudományi Könyvtár honlapja csak a *www* nélküli változatban volt elérhető és a linkek egy része még akkor keletkezett. Jelenleg a két változat egyformán működik, de a jobb eredmények érdekében érdemes lehet kiválasztani az egyik változatot mint preferáltat és a másikat átirányítani oda.

JavaScript: átok vagy áldás

A *JavaScript* (*JS*) egy kliens oldali scriptnyelv, így működéséhez elengedhetetlen, hogy a látogató böngészőjében engedélyezve legyen futása. Általános ajánlás, hogy ne használjunk olyan JS funkciókat, amelyek nélkül az oldal működésében alapvető hibák keletkeznek, mert ezek károsan befolyásolhatják a keresőrobotok viselkedését is.

A 2010 tavaszán működő könyvtári webhely egyik legnagyobb problémája a JavaScript menü, amelyben hiba van, mert egyes böngészők (pl. a *Google Chrome* és az *Opera* bizonyos verziói) az oldal menüjét meg sem jelenítik. A menü hiánya az oldalra visszatérő látogatókat meglepheti, aki pedig először jár itt, csak jóval kevesebb funkciót érhet el – ha ilyen böngészőt használ.

A probléma nálunk is – mint sok más oldalon – abból fakad, hogy egy szabadon hozzáférhető kódtárból lett letöltve a menü 2004-ben (vagyis nem helyben készült), így javítása a könyvtári dolgozóknak szinte megoldhatatlan feladat.

Áttekintve az új, Joomla-alapú weboldalunkat, ha a böngészőben letiltjuk a JavaScript-et, nem észlelünk változást a használhatóságban.

Keresés a honlapon

Egy könyvtári honlap rengeteg információt továbbít a látogatóknak, ezért fontos kérdés, hogy hogyan találják meg a keresett információt? Az oldal szerkezete, a menü elhelyezése és felépítése segíti vagy hátráltatja a látogatót, de egy jól működő és pontos kereső mindenképpen fontos.

Kereső a jelenlegi könyvtári honlapon is van, egy *Google Egyéni Keresőmotor* formájában. A megoldás nagy előnye, hogy egyszerűen létrehozható és mégis profi eredményeket kapunk. Hátránya, hogy a találati lista is az ismert Google oldalon jelent meg, ami kilépést jelentett a könyvtári webhely

keretéből. További gond volt, hogy a menüpontok sokasága miatt az oldal már a kereső nélkül is zsúfolt képet mutatott, így egy külön menüpontot kellett neki létrehozni (nem lehetett keresőmezőt elhelyezni), emiatt viszont a kereséshez ki kellett lépni az adott oldalról, esetleg elfelejtve merre is járt, mit keresett a látogató.

A Joomla portálrendszer rendelkezésünkre bocsát egy beépített keresőt, amelynek előnye, hogy találati listája illeszkedik az oldalba, hátránya viszont, hogy nem minden esetben ad olyan jó eredményeket, mint a Google. Az előző honlaphoz képest több hely lehetővé teszi viszont, hogy minden oldalon megjelenjen a keresőmező.

Új honlap: első lépések

A változtatások célja, hogy a használók jobb, modernebb szolgáltatásokhoz jussanak, de a keresők szempontjából ez kockázatos dolog is lehet. Idővel az új portál indexelve lesz, de a megváltozott linkek, menüpontok és szolgáltatások miatt értékük is meg fog változni a keresőgépek szemében. Az eddig folytatott vizsgálatok célja volt, hogy ez az átállás ne hátrányként, hanem előnyként jelentkezzen. Minden pozitívum és negatívum, amit felderítettünk és kijavítottunk vagy átvettünk az új oldalra, elősegítheti, hogy a későbbiekben még többször, akár általánosabb keresőkifejezésekre is megtalálják honlapunkat az érdeklődők.

Az egyik legjelentősebb kérdés mégis az, hogy mi lesz a jelenlegi oldalakra mutató linkek sorsa, mert a látogatók fele érkezik ilyen módon. Ha csak egyszerűen kikapcsoljuk a régi szervert, akkor elvesznek az oda mutató hivatkozások és az általuk generált eredmények, találatok.

Azon oldalaknál, amelyeket fontosnak gondolunk, érdemes a teljes megszűnés helyett átirányítást megadni. Ennek lényege, hogy tudatjuk a látogatóval (illetve böngészőjével) vagy a bottal, hogy a keresett oldal végleg elköltözött (301-es státusz-kód) és tovább irányítjuk az új helyre.

A megvalósításra több lehetőség is van: például apache esetén (a `mod_rewrite` engedélyezésével), htaccess fájlban, vagy a php átirányítás (lehetséges egy adott domént átirányítani teljes egészében, de lehetőség van mintaillesztéssel sokféle oldal, vagy akár konkrét oldalak átirányítására is).

Fontos azonban, hogy nem szabad kliensoldali (például JS) megoldást alkalmazni, mert az csak a valódi látogatóknak szól (és csak ha engedélyezték a JS futását), de a robotok nem tudnak tovább lépni. Ha viszont a bot rájön, hogy a látogatók nem ugyanazt az oldalt „látják” mint ő, ezt már a fekete-kalapos SEO technikák közé sorolja és büntethet is miatta – akár annak ellenére, hogy célunk nem volt szabályszegő.

A keresőoptimalizálás minden?

A SEO igazából azért érdekes terület, mert bár látszólag léteznek szabályok, ajánlások és figyelembe vehető mérőszámok, csak az eredményesség dönti el a változtatások értékét. Az interneten számtalan fórumon lehet a keresőoptimalizálásról olvasni, segítséget kérni és kísérletek sora folyik ebben a percben is, hogy az eddig leírtak egy-egy elemét bizonyítsák vagy megcáfolják.

Mindig akadnak meglepő eredmények és ezek ismeretében én azt tartom fontosnak, hogy végezzük el a keresőoptimalizálást, de mindig figyeljünk arra is, hogy értékes, egyedi és hasznos tartalmakat tegyünk közzé, mert igazi eredményt csak azal érhetünk el (pl. látogatók visszatérése).

A korszerűbb és jobban használható honlap szolgáltatása magasan felülírja az esetleges kezdeti visszaesés veszélyét a keresőkben illetve a látogatók számában, mivel hosszú távon az akadálymentes, szabványos és a keresőoptimalizálás szabályait is figyelembe vevő oldal hozhatja meg könyvtárunk számára a szélesebb érdeklődést. A cikk tartalmazza az új honlap készítéséhez elvégzett előzetes keresőoptimalizálási vizsgálatok eredményeit, de mivel a megvalósítás folyamatban van, a tényleges eredményekről egy későbbi időpontban tudok majd beszámolni.

Jegyzetek

- ¹ A <http://konyvtar.univet.hu> oldalain a Google Analytics segítségével készítjük a forgalmi statisztikákat 2006 júliusa óta. Ahol Google Analytics eredményekre hivatkozom, ott a 2009. január 1. és december 31. közötti időszak forgalmi adatait használtam. Az adatok pontosabbá tétele érdekében a forgalom figyelésénél a könyvtári dolgozók gépeiről érkező kérések nem szerepelnek az eredmények között, a könyvtárban található publikus gépeken pedig nem a könyvtár honlapja a kezdőoldal.

- ² Az „organikus találat” jelentése, hogy „természetes módon” a keresőgép eredeti rangsorolási szabályai szerint előre sorolt találatokról van szó, ellentétben azokkal, amelyeket a keresőgép egyéb okok miatt tesz a lista elejére vagy mellé (pl. a kulcsszóhoz kapcsolt fizetett linkek, frissességük miatt kiemelt hírek, ízelítőnek berakott képek és videók, helyi üzleti információk, az adott felhasználó által preferált webhelyek stb.) Vannak témák, amelyeknél az első 10 tételben már alig van normál módon odakerült találat.
- ³ Az XML sitemap-ről bővebben a <http://www.sitemaps.org> oldalon lehet tájékozódni.
- ⁴ A szabadalmat 1998-ban *Larry Page* adta be, aki a Google egyik alapítója.

Irodalom

- [1] Internet használati szokások, 2008. = <http://www.nhh.hu/dokumentum.php?cid=16023> [2010.05.17.]
- [2] WALTER, A.: Keresők és webhelyek. Budapest, Kiskapu, 2009. p. 255.
- [3] MALAGA, R. A.: Worst practices in search engine optimization. = Communications of the ACM, 51. köt. 12. sz. 2008. p. 147–150. = <http://www.huomah.com/dojo/Download-document/31-Worst-practises-in-SEO.html> [2010.05.17.]
- [4] iProspect Search Engine User Behaviour Study, 2006. április = http://www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf [2010.05.17.]

- [5] Search Engine Ranking Factors 2009. = <http://www.seomoz.org/article/search-ranking-factors> [2010.05.17.]
- [6] WALTER, A.: Keresők és webhelyek. Budapest, Kiskapu, 2009. p. 255. ISBN 9789639637597
- [7] Gogole Keresési Trendek = <http://www.google.com/support/insights/bin/answer.py?hl=hu&answer=87285> [2010.05.17.]
- [8] My site's PageRank has gone up / gone down / not changed in months! = <http://sites.google.com/site/webmasterhelpforum/en/faq--crawling--indexing--ranking> [2010.05.17.]

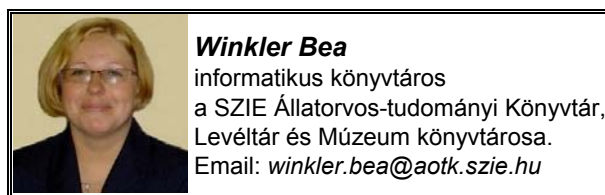
Hivatkozások

ENGE, E. – SPENCER, S. – FICHKIN, R – C. STRICCHIOLA, J: The Art of SEO. O'Rely, 2009. 608 p. ISBN 0596518862 = <http://books.google.com/books?id=4VvOLL4KlesC> [2010.05.17.]

KENT, P.: Search Engine Optimization for Dummies (3rd Ed.). Hoboken, N.J., Wiley, 2008. 432 p. ISBN 978-0-470-26270-2 = <http://books.google.hu/books?id=xmFLSuDcN2wC> [2010.05.17.]

Keresőoptimalizálás.lap.hu = <http://keresooptimalizalas.lap.hu/> [2010.05.17.]

Beérkezett: 2010. V. 16-án.



A felhasználók vakon megbíznak a Google-ben

A *Northwestern University* kutatói kiderítették, hogy az internetezők a legjobban a népszerű kereső hivatkozási rangsorában bíznak, a fiatal felhasználók vakon megbíznak a keresési eredményekben és nem kérdőjelezik meg, hogy miért az egyik vagy a másik került a találatok közül az első helyre.

Hargittai Eszter kutatásvezető szerint sok hallgató úgy gondolja, hogy azáltal válnak az oldalak megbízhatóvá, hogy a Google a keresésekkor a találatok között az első helyen jeleníti meg a címüket. Ez komoly problémává válhat, mivel a keresők az eredményeiket különböző szempontok alapján rangsorolják. S hogy mi lehetne a megoldás? Hargittai szerint az oktatási intézmények feladata lenne, hogy a hallgatókban tudatosítsák, kifejlesszék a portálok megítélésének képességét. A fiataloknak meg kellene tanulniuk, hogy miként mérhetik le egy oldal hitelességét.

A szakember és kollégái azzal a feladattal bízták meg a hallgatókat, hogy különböző információkat keressenek az interneten. A tesztben résztvevők a leggyakrabban a *Google*, illetve a *Bing*, a *Wikipedia*, az *AOL* és a *Facebook* segítségét választották. A diákok szinte az összes esetben az első találatra kattintottak és a rangsorpozíciónál sokkal kevésbé tartották fontosnak a tartalommal kapcsolatos információkat.

„A statisztikák azt mutatják, hogy a felhasználók legfeljebb az első kettő, esetleg a harmadik találati oldalt nézik meg. Ha nem találják meg a számukra szimpatikus találatot, akkor inkább módosítják a keresési kulcsszót. Éppen emiatt fontossá válik, hogy melyik linket hányan ajánlják és mekkora a szociális elismertsége” – közölte *Christoph Wendl*, az *Iphos IT Solutions* cég ügyvezetője.

Stefan Keuchel, a Google szóvivője kiemelte: a keresőoptimalizálás napjainkra egész üzletgá fejlődött. Ha valakinek megígérik, hogy egy honlap két héten belül a találatok között az első helyen szerepel majd, akkor a háttérben gyakran tisztességtelen módszereket (linkfarmtechnikákat, vagy az úgynevezett Cloaking-eljárást) vetnek be. E módszerek használatát a Google szabályzata tiltja és igyekszik is ellenük fellépni. A társaságnak ugyanis saját érdeke, hogy a keresőindexét tisztán tartsa. Keuchel hozzátette, hogy egyrészt szoftverekkel, másrészt munkatársak bevonásával igyekeznek ezeket a csapdákat felderíteni. Ha kiderül, hogy valamelyik oldal ilyen módszerekhez folyamodott, akkor akár a keresőindexből is kikerülhet.

/SG.hu Hírlap, 2010. július 28., <http://www.sg.hu/>

(SzP)



Névkezelés (authority control) az intézményi repozitóriumoknál

Az intézményi repozitóriumoknál a felhasználók saját közreműködésükkel hozzák létre a tartalom-szolgáltatást. A munkafolyamatokba viszont nincs közbeiktatva semmilyen minőségellenőrzés, nem került sor a könyvtári gyakorlatban jól bevált konzisztens adatkezelési módszerek átültetésére, így a szerzői nevek egységesítésére sem fordítottak figyelmet. A szerző szerinti keresések összehasonlító vizsgálatai azt bizonyítják, hogy a könyvtárosok közreműködésével épített repozitóriumok sokkal jobban megállják a helyüket, mint azok, ahol a tartalomfeltöltés szabadjára engedve, kizárólag a szerzőkre hárul.

A repozitórium-szoftverekre (DSpace, EPrints, Fedora) és a metaadat-szüretelő rendszerekre nem jellemző a következetes névkezelés: a szerzői nevek bevitelére, a bibliográfiai tételekhez való hozzárendelés módjára vonatkozóan semmiféle szabályozás nincs. A letölthető szoftverekhez nincs kész, kidolgozott modul a szerzőségi adatok kezelésére, a szerző szerinti keresés könnyűszerrel esetlegessé válik. Javíthatna a helyzetet, ha csatolhatók volnának a helyi, intézményi személyzeti nyilvántartó rendszerek, és ha külső forrásokat is be lehetne vonni a feldolgozás menetébe. A legtöbb repozitóriumkezelő eszköz azonban csak a saját névlistáját tudja némileg egyeztetni a feltöltések után lefuttatott ellenőrzéssel. Az összemosódó adatok az aggregáló szolgáltatásokhoz is átkerülnek, ahol még inkább összekuszálódnak a felhalmozott adatok tömkelegében. A nevekkal így módon nem lehet megkülönböztetni az egyes szerzőket, személyeket.

A problémakör

Egy 2006-os díjnyertes cikk (Vannevar Bush Díj) a nem könyvtárosok, azaz az amatőrök által létrehozott gyér minőségű metaadatokat vizsgálja. Szóvá teszi, hogy a *Nemzeti Tudományos Digitális Könyvtár (National Science Digital Library)* az automatizált metaadat-aratás technológiáján (*Nyílt Archívumok Kezdeményezés Metaadatgyűjtési Protokollja = Open Archives Initiative Protocol for*

Metadata Harvesting, OAI-PMH) dolgozva nem törődött az authority controllal, a nevek minősített kezelési környezetének kialakításával. A díjazott cikk e jelenséget szembeállítja a WordCat paradigmával, a hivatásos könyvtárosok által épített közös katalógusépítési kultúrával. Azt is megállapítja, hogy kevés repozitórium hajlandó arra – pontosabban kevés ilyen gyűjtemény engedheti meg magának –, hogy megfelelő humán erőforrás alkalmazásával katalogizálják a bekerülő dokumentumokat. [11]

A szakirodalomban kevés beszámolót olvashatunk a repozitóriumi feldolgozók munkájáról, napi gondjairól; mi több, maguk a katalogizálók is úgy gondolják, a repozitóriumoknál egyáltalán nem éri meg a fáradságot a nevek cizellált kezelése. [14] Az intézményi e-dokumentum-lerakatokat a könyvtárosok ugyanis olyan alacsonyabb rendű képződményeknek tekintik [17], amelyekre nem érdemes energiát pazarolni. A tudományos közvélemény sem tartja túl sokra az efféle dokumentumtárakat. [6] Ezután nem meglepő, ha a feldolgozó könyvtárosok csak vonakodva vállalják a repozitóriumi anyag katalogizálását – ha egyáltalán vállalnak is valamit ebből a hálátlan munkából.

Az egységesítési, egyeztetési eljárások híján viszont megnehezül a szerző szerinti keresés. A DSpace repozitórium szoftvercsomag alapkészletében nincs kidolgozva az azonos személyt jelölő nevek összekapcsolása, a betűrendes szerzői listába ezért szétszóródva kerülhetnek be az ugyanazt a személyt jelölő különféle névformák; a betűrendes névjegyzék hamar teli lehet hibás beírásokkal, elütésekkel.

Ugyanezzel a problémával lehet számolni az OAIsternél, az OCLC digitális gyűjteményénél. [15] Itt a szerzői keresés, a közties szerzői felsorolást kihagyva, közvetlenül a tételek listázását eredményezi. Ha valaki a szerző keresztnévét például teljes formájában adja meg, akkor a listából hiányozni fognak azok a tételek, amelyeknél az adott auktor neve csak a kezdőbetűkkel van megadva, és fordítva. Az OAIsternél csonkolásra is lehetőség

van, így teljesebb találatot lehet kapni, ám egyáltalán nem biztos, hogy az átlagos felhasználó erre magától is rájön. Nincs gyorsan átfutható szerzői betűrendes névlista, amit átfutva feltűnhetnének a furcsaságok, következetlenségek, ha pedig névváltozásról van szó, össze lehetne gyűjteni ugyanannak a személynek a korábbi vagy későbbi nevével a teljes munkásságát.

Az intézményi repozitóriumok több szakterület anyagát tartalmazzák, a szerzők felől indítva a keresést, a több területen működő azonos szerzői név miatt, sok olyan információ felbukkanhat, ami eleve irreleváns a kereső előtt. [5] Az azonos nevű személyek megkülönböztetése nem csak az emberi felhasználónak jelent nehézséget: az OAIster aratásakor is összemosódnak ezek az adatok. A gondot részben csökkenti a forrásgyűjtemény szerinti tartalom-megkülönböztetés, teljességgel viszont nem küszöbölhető ki, minthogy a kutatók több intézmény életében is részt vesznek, az intézmények közötti kutatóprogramok során egy-egy cikkhez pedig bármennyi szerző és intézményi adat kapcsolódhat. A tartalom diszciplína vagy téma szerinti megkülönböztetése lehetséges ugyan, ám, hogy a keresés hatékonysága e szempont bevonásával javulhasson, komplex szöveg-bányászati eljárások beépítésére volna szükség. A feltöltött tartalmakhoz rendelt tárgyszavakra nem lehet bizton támaszkodni: ezek esetlegesek és rendkívül változatosak.

A probléma forrásai

A könyvtárosok a feltöltött dokumentumok leírásait rendszerint egyesével készítik el. Kötegelt feltöltést is alkalmazhatnak például az intézeti vagy tanszéki weboldalakon közzétett, általában hiányos, beazonosítatlan metaadatokkal érkező közlemények betöltésével. Csoportos áttöltéssel kerülhetnek be a tudományos kiadóknál megjelentetett közlemények adatai is, változó részletességgel és minőséggel. Egy-egy szakfolyóirat kötegelt befogadásakor megfigyelhető, hogy ugyanannak a szerzőnek a neve még ugyanazon folyóiraton belül is más-más formában szerepel az egyes füzetekben. Nehézség, hogy a lapokban sok szerző neve a családnévre és a további kezdőbetűkre korlátozódik. A metaadatok rendkívül változatos formában érkeznek az aggregálóhoz. Ha a metainformációk nem mezőkre bontva, hanem sorfolytonos bibliográfiai formában töltődnek be, az adatsort szét kell bontani részekre, külön ki kell hámozni a neveket,

ami nem túl egyszerű programozási feladatot jelent.

Egyes tudományágazatok repozitóriumai iránt, mint amilyen például az *arXiv*, már régóta nagy az érdeklődés. [3] Az intézményi lerakatok iránti igény is nő: segítségükkel az intézetek kutatási tevékenysége láthatóbbá, a feltöltött tartalom pedig megőrizhetővé válik. A laikus adatfelvivőktől viszont nem lehet elvárni a szabályok és konvenciók ismeretét, így, ha volnának is repozitóriumai besorolási adatokra vonatkozó szabályzatok, ezeket akkor sem alkalmaznák. A szakszerűtlen metaadat-bevitel következményei mind a kétfajta archívumot egyaránt sújtják. Kevés helyen van mód kontrollált névkezelésre, ez a funkció többnyire hiányzik a beviteli űrlapokról. Megbízható utalórendszer kifejlesztésével több tematikus repozitóriumnál is kísérleteznek. [13, 18]

Szabványok

Az intézményi repozitóriumoknál alkalmazott metaadat-szabályzatok nem tesznek említést az azonosító–egységesítő eljárásokról. Az OAI-PMH metaadatgyűjtő szabvány a szolgáltatóktól *Dublin Core* tagolást vár el, de más metaadat-formátumot, köztük MARC-ot is megenged. Problémát jelent, hogy a Dublin Core nem fordít gondot a nevek személyek szerinti megkülönböztetésére, sem a variánsok egyeztetésére. Elkelnének a szerzők intézményi hovatartozását jelölő megkülönböztetett mezők, e nélkül akár a névtípusok is – például egy személy- és egy intézménynév – feltűnés nélkül összekeveredhetnek.

Az OAI-PMH egyirányú adatáramlás: a szüretelők az adatszolgáltatóktól gyűjtik össze az adatokat. A protokoll a szüretelőknek minden adatot úgy ad át, mintha minden rendben volna: súlyos technikai hibák esetén, például rosszul formált XML-nél is figyelmeztetés nélkül juthatnak át az adatok. A metaadatok korrekciója visszatérő nehézség. A hibát ott kellene javítani, ahol keletkezett, az eredeti adatbázisban. A metaadatok minőségének jobbítását a repozitóriumai forrásoldalon a metaadatok igényesebb előállításával kellene elérni; ennek kezelése még nem oldódott meg kellőképpen.

További gond, hogy az OAI-PMH-nál nincs jó visszajelzési mechanizmus a forrásbázisnál módosított adatokról, amelynek alapján a módosítást a szüretelt adatok közé is át lehetne vezetni. Az aggregátor annyit segít, hogy átirányítja a felhasználót a forrásgyűjteményhez, ott viszont előfordul,

hogy megtévesztő módon már más bejegyzések szerepelnek. Az *Open Archives Initiative's Object Reuse and Exchange (Nyílt Hozzáférés Kezdeményezés Objektum Újrafelhasználása és Cseréje = OAI-ORE)* specifikációja sem oldja meg a metaadat-minőség problémáját. Az OAI-ORE célja az összetett objektumok alkotóelemeinek összerendezése. Alkalmazható a szakterületet átlapoló folyóirat (overly journal) „szerkesztésére” is. Ennek a folyamatnak a során a rendszer több különböző intézményi lerakathoz fordul, és az ott tárolt cikkek közül valamint mellékleteiből vagy egyéb kellekiből periodikus, folyóiratszerű összeállítást képez le. Az eljárás azonban inkább adatmodellnek és sorozatgeneráló szabványnak tekinthető semmint kommunikációs protokollnak. Az OAI-ORE nem tartalmaz specifikációt arról, hogy az adatcserében hogyan játszódik le kommunikációs folyamat. Nem foglalkozik sem a módosításokkal, sem a metaadat-frissítésekkel, s ez a hiány ugyanazon objektumot leíró rekordvariánshoz vezethet az újrafelhasználási láncolatban. Ha például az átlapoló folyóirat OAI-ORE eszközt használ arra, hogy a repozitóriumban tárolt cikk adatait duplikálja, és ha a forráshely később a cikk szerzőjének nevét kijavítja a leíró adatok között, akkor az ugyanazon cikkhez tartozó két rekord közül a folyóiratnál lévőhöz nem juthat el a későbbi névvariáns.

Szoftverek

Sem a DSpace sem a Fedora nem rendelkezik automatikus névkiegészítő funkcióval, utalórendszerrel vagy más segítségül hívható menüponttal, amivel a névvariánsok köre szűkíthető volna: a nevek rögzítése egyedül a feltöltőkre van bízva. A nyílt forráskódú repozitóriumszoftverek közül az EPrintsnél a begépeléskor felbukkanó szövegkiegészítő aktivizálódik, az adott betűhelyen a felkínálkozó listából át lehet emelni a megfelelő nevet. Az EPrintsnél a kiadók és a folyóiratok nevét is ezzel a könnyítő eljárással lehet beírni, a rendszert ugyanis hozzá lehet kapcsolni a kiadók nyílt hozzáférési politikáját számon tartó SHERPA/ROMEÓ adatbázishoz. Ezzel a segédalkalmazással az adatbevitelkor csökkenthető a névváltozatok sora. A személyenkénti névegységesítést viszont már nem támogatja a rendszer, ezért nem lehet megelőzni a félregépelésből, betűtévesztésből adódó szóródást. Az automatikus karakterkiegészítő viszont nem is oldja meg a névvariánsok összekapcsolásának problémáját: bármennyi új nevet a már bent lévő állomány egyeztetése nélkül szabadon fel lehet vinni, ebben az esetben tehát pusztán névvariánsok jegyzékéről beszélhetünk.

A DSpace a „létrehozó” vagy „közreműködő” mezőkben szereplő elemeket egyetlen listába sorolva, szerzői abc-ben mutatja meg, attól függetlenül, hogy valódi szerzőről vagy valamilyen más közreműködőről van-e szó. Olykor azt sem lehet megkülönböztetni, ki egy dolgozat szerzője és ki annak mentora.

Sem a DSpace, sem a Fedora nincs felkészítve a kötegelt metaadat-szerkesztésre: ha ugyanaz a hiba több rekordban jelentkezik, nem lehetséges a csoportos módosítás, valamennyi rekordot egyenként kell elővenni. A DSpace-nek még egy szépséghibája van: a tartalom feltöltése után csak kétféle metaadat-szerkesztési jogosultsági szerepkört engedélyez, mégpedig a gyűjtemény referensének és a rendszeradminisztrátoroknak. A katalogizáló a névegységesítést a teljes adatbázisra kiterjesztve csak úgy végezheti, ha adminisztratori jogosultsága is van. Ez a feljogosítás azonban a rendszer biztonságát veszélyezteti.

Az EPrints újabb fejlesztései közé tartozik a metaadatok kötegelt módosítása, de a névkezelés még így sem ideális. A rekordok kiválasztása ugyanis a több szempontú keresőfelületről indított globális keres-és-felülír művelettel történik. Egyes helyesírási és átírási változatok, valamint a személy házasságával járó vagy egyéb körülmény folytán megváltoztatott, ugyanazon személyt jelölő név ezzel a módszerrel nem keríthető elő, így az összes módosításra szoruló névváltozat sem jeleníthető meg. Ennek ellenére, a funkció jelentős előrelépés a rekordok egyesével történő javítgatásához képest.

A nevek normálásához használható források

Nem megkerülhető lehetőség a külső névjegyzékek bevonása a repozitóriumok névterébe, a módszer azonban még nem terjedt el. Csatlakoztatható például az OCLC online felülete vagy a *Nemzetközi Virtuális Azonosító Adattár (Virtual International Authority File = VIAF)*. A két névtár csak a könyvek szerzőire vonatkozóan tartalmaz névváltozatokat és azonosító adatokat, ezzel szemben sok kutatási ágazatban, különösen a „kemény” tudományok körében, a cikkeken van a hangsúly, következésképp, a két említett névállomány a szerzői névadatoknak csak töredékét tartalmazza.

Felhasználhatók a helyi névállományok, amelyeket az egyetemi oktatói nyilvántartásokból lehet kinyerni. Teljes körű megoldásra itt sem számíthatunk, ugyanis az intézményközi együttműködések során született cikkeknek alkalmasint külső szerzői is lehetnek, a külső szerzőkre viszont nem terjed ki

az intézményi azonosítás és regisztráció. Egyszerű és járható útnak tűnik a repozitóriumok szerzői jegyzékeinek megosztása, önmagában azonban ez sem elegendő, minthogy a DC elemei között nem szerepel az intézményi hovatartozást rögzítő mező.

További lehetőség a személyi portfólió-szolgáltatásokban rejlik. A RePEc gazdasági preprint-szolgáltatásnál auktorokat számon tartó szolgáltatás is van. [2] A kiadóknál hasonló kezdeményezés a cikkeket linkelő CrossRef kapcsán merült fel. [8] Profitorientált vállalkozásokat is foglalkoztat a kérdés, ilyen például *Research Crossroads*, [19] ahol minden betagosodott kutatónak egyedi azonosítót (ResearchCrossroads ID) nyitnak. Néhány intézmény hasonló szolgáltatással kísérletezik. [7] Könyvtárakon kívül a *Nemzetközi Szabványügyi Szervezet (International Standards Organization = ISO)* dolgozik a „felet” (személyeket, testületeket) azonosító szabványon (*International Standard Party Identifier* [10]). A *Joint Information Systems Committee (JISC)* a névazonosítás problémáját a Names Project keretében kutatja; a holland SURF alapítvány pedig saját hatáskörében bevezette a Digitális Szerzői Azonosítót. [20]

A gomba módra szaporodó azonosítók nem kis fejtörést okoznak. A nehézség abban van, hogy bonyolult az egyes repozitóriumokat összekapcsolni az összes, autorizálási információt nyújtó potenciális forrással. Egy másik fontos kérdés a megbízhatóság és fenntarthatóság: felmerül ugyanis a kérdés, ellenőrizik-e vajon ezek a szolgáltatások az adatok valóságát és minőségét. Hány hasonló szolgáltatást lehet nyílt hozzáféréssel, illetve bármilyen hozzáférhetőséggel fenntartani, tekintettel a fenntartásra költhető szűkülő pénzforrásokra?

Egy következő lehetséges megoldás: OpenID identitáskezelő séma. [16, 9] A kutató az OpenID-vel az intézményi repozitóriumba is bejelentkezhet, ami őt a publikációival is összekapcsolja. Ha az illető egy másik intézményhez szegődik, akkor az új repozitórium elfogadja ugyanazt az azonosítót, és kezkesedik a két tárnál archivált tartalmak, valamint a személyt reprezentáló név és azonosító adatlapja helyes összekapcsolásáért. Az elképzelés a társszerzőség problémájának hálóján akad fenn: a deponálók vajon meg tudják-e adni szerzőtársaik pontos azonosítóját? Hajlandók lesznek-e az összes azonosítót kikeresni és hozzárendelni a megfelelő tartalomhoz?

Megoldási törekvések

Ha az intézményi repozitóriumban korrekt névkezelést akarunk látni, nem marad más hátra, mint a változatosan bevitt adatok kézi tisztogatása, amihez az EPrints-nél nyílnak kedvezőbb körülmények. Az intézmények hajlanak arra, hogy az ellenőrzést akár ki is hagyják a folyamatból, a nevek megfeleltetésére, egységesítésére ugyanis megfelelő munkaerőt és munkaidőt kellene felszabadítaniuk. A felhasználókra bízott és hirtelen benépesülő lerakatok metaadatainak javíthatására sokkal több időt kell fordítaniuk, mint a szokásos, évi egy-két alkalommal végzett utólagos felülvizsgáló munkálatokra.

A repozitóriumok közötti technikai és üzemeltetési együttműködésre érdemes nagyobb figyelmet szentelni. A JISC és az Open Archive Initiative a leginkább felkészült társulások, amelyek e kérdésekkel foglalkoznak; a leggyorsabb fejlődést az OAI-PMH fejlesztésével, a metaadatok minőségére, a visszajelzések kezelésére kiterjedő bővítéssel kell elérni.

A névkezeléshez használható külső források szempontjából sem biztató a helyzet: az azonosító adatlapsémák változatai visszafordíthatatlanul elburjánzottak. Manapság API-kon keresztül már jó lehetőség nyílik a személyek azonosításához szükséges információk egyeztetésére a szolgáltatások és a repozitóriumok között, a különböző névállományokban könnyebbé vált az azonos egyéneket vagy testületeket reprezentáló rekordok közötti összefüggések számítógéppel támogatott feltárására. Ezek az alkalmazások várhatóan a repozitóriumok gyakorlatába is be fognak épülni. A teljesebb körű névállományok létrejöttéig a legkézenfekvőbb az automatizált alkalmazások előtt is elérhetővé tenni a meglévő nemzeti és más névtárakat. Az EPrintshez hasonlóan a repozitórium-szoftverek fejlesztőinek a külső névállományok hasznosítására is érdemes figyelniük. Az azonosítási eljárások technológiája erőteljesen fejlődik, az elérhető szerzői névtér hamarosan oly méretűvé válhat, amit érdemes lesz az új eszközökkel bevonni a szolgáltatásokba.

Ez azt is jelenti, hogy az authority control során, egyes autorizálási (összehasonlítási, egységesítési, azonosítási) műveletek kívül kerülhetnek az adott helyi repozitórium körén. A tematikus cikkgyűjteményekből szüretelők és az online cikkgyűjtőregátorok a SWORD API [1] segítségével egyik repozitóriumról a másikba továbbíthatnak egy-egy

cikket. Példaként szolgálhat a BibApp egyetemi kutatási kísérleti projekt, [4] amely begyűjti a bibliográfiai leírásokat és átcsomagolja a repozitóriumok számára. Az egyéni publikációs jegyzékek mellett a kutató érdeklődési köréről, társszerzőiről és tevékenységének egyéb jellemzőiről is tartalmaz leírást. Mivel sok forrásból szerzi a bibliográfiai tételeket, a szerzőkre vonatkozó bejövő információk meglehetősen maszatosságok. A bejövő kusza névhalmoz rendezéséhez a BibApp úgy fog hozzá, hogy a szerző első bejelentkezésénél a szerzőt megkéri, válassza ki a már meglévő listából, melyik névalakok tartoznak hozzá, valamint jelölje ki, melyik névalakot szeretné elsődlegesen használni. Ezzel lényegében egy belső autorizálást, egységesítést, névnormálást végez, és létrehozza a saját azonosító besorolási rekordját. [12]

Az eljárás kézi és automatizált egyeztetésre épít, gyors és elnagyolt automatikus heurisztikával. Az eredmény nem egészen tökéletes: az eljárás ugyanis megengedi, hogy a tételek homályban hagyott szerzőinél annak alapján lehessen egy szerzőnek valójában nem hozzá tartozó művet tulajdonítani, hogy a BibAppnak kiről van több információja. Ettől függetlenül érvényes: mennél több információt sikerül a BibAppnak összegyűjtenie, a jelöléstechnikája annál pontosabb lesz. A program fejlesztői a szerzőségi adatok föderált egyeztetésére és a különféle programokkal kezelt autorizáló információkra is számítanak. A protokollok és az API-k jó eséllyel segítik a névföderálást, a kuszaság csökkentését. A névállományok bővülő forrásai és az azonosítási–autorizálási eljárások körül folyó párhuzamos fejlesztések várhatóan meghozzák gyümölcsüket, a repozitóriumok karbantartását egyszerűsítik, a keresést pedig megbízható névbázissal támogatják. Amíg azonban ezek az automatizmusok nem hoznak látványos eredményeket, addig az intézményi repozitóriumokkal járó munkára jelentős ráfordítást kell tervezni.

Irodalom

A webhelyek és internetről elérhető cikkek legutóbbi megtekintésének dátuma: 2010. május 20.

- [1] ALLINSON, Julie – FRANÇOIS, Sebastien – STUART, Lewis: *WORD: Simple Web-service Offering Repository Deposit*. = *Ariadne*, 54. köt. 2008. január, <http://www.ariadne.ac.uk/issue54/allinson-et-al/>
- [2] RePEc Author Service, <http://authors.repec.org/>

- [3] arXiv, <http://arxiv.org/>
- [4] BibApp projekt, University of Wisconsin at Madison és University of Illinois at Urbana-Champaign, <http://code.google.com/p/bibapp/>
- [5] BARTON, Jane – CURRIER, Sarah – HEY, Jessie M. N.: *Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice*. = SUTTON, Stuart [et al. eds.] *Proceedings 2003 Dublin Core Conference: Supporting communities of discourse and practice – metadata research and applications*. Seattle, Washington, 2003. Nyílt hozzáférésű változat (a továbbiakban: Ny.h.): *Erpaeprints*, <http://eprints.erpanet.org/83/>
- [6] BROWN, Laura – GRIFFITHS, Rebecca – RASCOFF, Matthew: *University publishing in a digital age*. = Ithaka report, 2007. 69 p. <http://www.ithaka.org/strategic-services/Ithaka%20University%20Publishing%20Report.pdf>
- [7] Cornell University, VIVO, <http://vivo.library.cornell.edu/>
- [8] CrossRef, http://www.crossref.org/CrossTech/2007/02/crossref_author_id_meeting.html
- [9] HABIB, Michael: *Managing your identity online*. *Library Journal netConnect* 10/15/2007, <http://www.libraryjournal.com/article/CA6486511.html>
- [10] International Standard Party Identifier, <http://collectionscanada.ca/iso/tc46sc9/27729.htm>
- [11] LAGOZE, Carl – KRAFFT, Dean – CORNWELL, Tim – DUSHAY, Naomi – ECKSTORM, Dean – SAYLOR, John: *Metadata aggregation and “automated digital libraries”: a retrospective on the NSDL experience*. = *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*. New York: Association for Computing Machinery, 2006. p. 230–239. Ny.h.: arXiv:cs/0601125v1, <http://arxiv.org/abs/cs.DL/0601125>
- [12] LARSON, Eric – SHREEVES, Sarah: *BibApp : information architected : presentation for the Digital Library Federation Spring Forum, 2008*. Kézirat, Ideals, <http://hdl.handle.net/2142/5441>
- [13] McRAE-SPENCER, Duncan M. – SHADBOLT, Nigel: *Also by the same author : AKTiveAuthor, a citation graph approach to name disambiguation*. = *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*. New York, Association for Computing Machinery, 2006. p. 53–54. ISBN 1-59593-354-9. Ny.h.: University of Southampton, <http://eprints.ecs.soton.ac.uk/12704/>
- [14] MEDEIROS, Norm: *The catalog's last stand*. = *OCLC Systems and Services*, 23. köt. 3. sz. 2007. p. 235–237. Ny.h.: E-LIS, <http://eprints.rclis.org/11466/>

- [15] OAster, OCLC, <http://www.oaister.org/>
- [16] OpenID, <http://openid.net/>
- [17] SALO, Dorothea: Innkeeper at the Roach Motel. = Library Trends, 57. köt. 2. sz. 2008. p. 98–123. Ny.h.: Minds@uw, <http://minds.wisconsin.edu/handle/1793/22088>
- [18] TAN, Yee Fan – KAN, Min-Yen – LEE, Dongwon: Search engine driven author disambiguation. = Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries. New York, Association for Computing Machinery, 2006. p. 314–315. ACM Digital Library, <http://portal.acm.org/citation.cfm?id=1141753.1141826>
- [19] Research Crossroads, <http://www.researchcrossroads.org/>
- [20] SURF, Digitális Szerzői Azonosító, <http://www.surffoundation.nl/smartsite.dws?ch=eng&id=13480>
- /SALO, Dorothea: Name authority control in institutional repositories. = Cataloging & Classification Quarterly, 47. köt. 3–4. sz. 2009. p. 249–261./

(Dudás Anikó)

Webkettes szolgáltatások az egyetemi könyvtárak honlapjain

„A web 1.0 eljuttatta az embereket az információhoz; a web 2.0 az információt viszi el az emberekhez” – mondta *Ian Davis*. Az olyan internetes vállalkozások, mint az *amazon.com* vagy az *ebay.com*, meghatározó szerepet játszottak abban, hogy a korábbi statikus webhelyek dinamikus és interaktívokká váltak, méghozzá úgy, hogy a felhasználók maguk is alakíthatják azok tartalmát (pl. címkézéssel, vélemények írásával, vagy pusztán azzal, hogy statisztikai adatokat szolgáltatnak a „mások ezt vették még”-típusú automatikus ajánlásokhoz). Ez a fajta „közösségi internetezés” rendkívül népszerű manapság: például a 2007-es *Pew Internet and American Life Project* felmérése szerint a 12-17 év közötti korosztály 64%-a használt már akkor is napi rendszerességgel közösségi oldalakat. Más szektorokhoz hasonlóan a könyvtári szféra is elkezdte beépíteni virtuális szolgáltatásai közé az újgenerációs webes funkciókat; nem sokkal a „web 2.0” elnevezés elterjedése után megszületett a „library 2.0” is – kifejezve ezt a változást. Az indiai szerzők által írt jelen tanulmány, a téma szakirodalmának rövid áttekintése után, egy olyan kutatást ismertet, amelyben a világ vezető egyetemeinek könyvtári honlapjait vizsgálták meg abból a szempontból, hogy hányféle webkettes funkciót tartalmaznak.

Előzmények

A „web 2.0” kifejezést 2005-ben *Tim O’Reilly* alkotta meg és kezdte népszerűsíteni. Az azóta született sokféle értelmezés és definíció leginkább a következő jellemzőket tulajdonítja ennek a technológiai és webhasználati korszakváltásnak: olyan alkalmazásokról van szó, amelyek annál jobbak válnak, minél többen használják őket; ebben a

dinamikus és rugalmas információs térben bárki létrehozhat és megoszthat tartalmakat; a közös információ-előállítás és -megosztás a kollektív intelligencia elvére épít. A webkettes eszközök közé sorolják a közösségi oldalakat (*Facebook*, *Bebo*, *MySpace*, *Friendster* stb.), a wikiket és a blogokat, a linkgyűjtő oldalakat (*Delicious*, *Furl*, *Digg* stb.), a képmegosztó helyeket (*Flickr*, *Photobucket* stb.), és még sok más alkalmazást is. *Breeding* [1] egyik cikkében meg is jegyezte, hogy a „web 2.0” egy marketinges divatszó lett, mindenre ráaggatják, amit „menő”-nek akarnak beállítani, függetlenül attól, hogy mennyire valósítja meg az eredeti értelmezés alapelveit. Mindenesetre arra jó apropó, hogy a könyvtárosok ennek hatására felülvizsgálják a webről alkotott elavult nézeteiket, és ők is alkalmazni kezdik az új technikákat, szolgáltatásokat. Ezért is jelent meg a „library 2.0” kifejezés, melyet elsőként *Michael Casey* írt le „librarycrunch” nevű blogjában, 2006-ban. *Maness* [2] meghatározása szerint a 2.0-s könyvtár „az interaktív, az együttműködésen alapuló, illetve a multimédia technológiák alkalmazása a webes könyvtári szolgáltatásokban és gyűjteményekben”. A könyvtárak is csatlakoztak az új generációs web által kiváltott társadalmi mozgalomhoz, és a „library 2.0” lényegében egy – szintén divatos szóval élve – *mash-up* a hagyományos könyvtári és az újszerű webes szolgáltatások között.

Miller [3] röviddel a fogalom első megjelenése után már elkezdte vizsgálni, hogy hogyan lehetne a webkettőt könyvtári célokra használni, és felhívta a könyvtárakat és szervezeteiket, hogy aktívan kapcsolódjanak be ezekbe a folyamatokba. *Mannes* [2] gyakran hivatkozott cikkében a „kettőpontnullás könyvtár” négy alappillére is definiálta: felhasznál-

ló-központú, multimédia-jellegű, szociálisan nyitott és újító közösségi szemléletű. *Habib* [4] dolgozatában pedig egy módszertani keretet javasolt a felsőoktatási könyvtárak számára ahhoz, hogy hogyan építsék be a web 2.0 eszközrendszerét a szolgáltatásaikba. Az átfogó, elméleti cikkek mellett egy sor olyan publikáció is megjelent már, amelyek ennek az eszközrendszernek egy-egy elemét vizsgálják. Sokan írtak például a wikikönyvtári tudásbázisként való alkalmazásáról. Mások az RSS, a podcast, az azonnali üzenetküldés (IM), vagy éppen a blogok és a közösségi oldalak (SNS) használatát kutatták. Az eredmények nemcsak az idővel, hanem könyvtártípusonként és földrajzilag is változnak: például az ausztrálzásiai egyetemi könyvtárak honlapjainak 2008-as elemzése szerint ott az RSS volt a leggyakrabban és az IM a legritkábban felkínált funkció. Ugyanakkor az amerikai intézményekben a blogok és az IM is az elterjedt eszközök közé tartoznak az RSS mellett, míg az SNS és a podcast használata jóval kevésbé gyakori.

A kutatás ismertetése

E tanulmány szerzői két kérdésre szerettek volna választ kapni: milyen fajta webkettes technikákat használnak az egyetemi könyvtárak, és milyen célra? A megvizsgálandó intézményeket a *timeshighereducation.co.uk* oldalon közzétett 2007-es toplistából választották ki – a világ száz legjobb egyeteméből végül 57-nek a könyvtári honlapját nézték végig. Csak olyan webhelyeket vettek bele az elemzésbe, amelyeknek angol felülete is van, és található rajtuk legalább egy web 2.0-s alkalmazás (max. két kattintásra a nyitólaptól). Ennek alapján több amerikai, brit, kanadai, ausztrál, továbbá néhány ázsiai, valamint egy-egy svéd és finn felsőoktatási intézmény könyvtára került bele a 2008 júliusában elvégzett kutatásba.

RSS

Az RSS egy egyszerű XML formátum egy webhely tartalmának, illetőleg változásainak közzétételére. Hasonlít a könyvtárak hagyományos témafigyelési szolgáltatására vagy újdonságlistáira, és kiváló eszköz hírek, aktualitások gyors elterjesztésére. Az 57 vizsgált könyvtári honlap közül 37-nek voltak RSS csatornáik. A hírek/események/közlemények mellett tanfolyami információkról, új beszerzésekről, adatbázisokról, friss blogbejegyzésekről és podcastokról tájékozódhatnak ezek segítségével a felhasználók.

Blog

A webes napló a legelterjedtebb webkettes műfaj. A 2007-es statisztika szerint az amerikai tinédzserrek 28%-a csinált már saját blogot; és a blogokban alakult ki az a fajta „állampolgári újságírás” is, amely egyre fontosabb hírforrásként funkcionál a hivatalos média mellett. A könyvtárak információterjesztésre, szolgáltatásaik ismertetésére és népszerűsítésére használják ezt az eszközt. A blogbejegyzésekhez érkező kommentek segítségével pedig megismerhetők az amúgy rejtve maradó olvasói vélemények és igények, s ezek alapján javítható a szolgáltatások minősége. A kutatásba bevont könyvtári weblapoknak csak mindössze 26%-án volt blog, ami meglepő eredmény, tekintve hogy ez máshol mennyire népszerű. A kézenfekvő, hírlevélszerű könyvtári blogok mellett egyes helyeken tematikus és személyes hangvételű weblapok is találhatóak. Mivel egyes blogok – a wikik mellett – idővel fontos információforrásokká nőnek ki magukat, a könyvtáraknak el kell gondolkodniuk azon, hogy mit és hogyan archiváljanak közülük a digitális gyűjteményükben.

Wiki

A wiki olyan oldal, amelyet bárki szerkeszthet, aki hozzáfér. A *Wikipedia* sikere önmagában is bizonyítja, hogy mennyire jó tudás- vagy információgyűjtő eszköz ez. Ennek ellenére az 57 könyvtár közül csupán egynél találtak ilyet: az *Emory University Libraries* egy „ReservesDirect” nevű wikit hozott létre az egyetemi kurzusok számára fenntartott, elektronikus, illetve fizikai hordozókon levő dokumentumok nyilvántartására.

Podcast/vidcast

Ez a két fogalom lényegében az interneten át hozzáférhető audio- és videotartalmakat jelenti, melyek online hallgathatók/nézhetőek és esetleg le is tölthetőek. Az iPod készülékek elterjedése nagy hatással volt erre a szegmensre: a már említett Pew Internet and American Life Project felmérése szerint az amerikai internetező fiatalok 19%-a hallgat podcastokat. A könyvtárak sokféle hang- és filmfelvételt szolgáltatathatnának ilyen módon: például felolvasott meséket (gyermekkönyvtárakban), helyi lakosok (pl. veteránok, ismert emberek) visszaemlékezéseit, könyvtári események (kiállítások, közönségtalálkozók, előadások, tanfolyamok) felvételeit, oktató és népszerűsítő anyagokat, könyv/zene/DVD/játék-ismertetőket és -bemutatókat, valamint más forrásokból ingyenesen átvehető podcastok válogatásait. A kutatók azonban csak

három olyan könyvtárat találtak, ahol voltak ilyen weblapba ágyazott hangfelvételek – főleg könyvtárismertető és aktuális hírek. Videókat hat helyen tettek fel. Például a *University of Pennsylvania* könyvtárában rövid *Libclip*-eket készítettek, amelyek a különféle információforrások és szolgáltatások használatát oktatják. A *University of Auckland* könyvtárosa pedig egy „Ta Punga” című animációsorozat segítségével mutatják be a *Voyager* katalógusban való kutatás menetét.

IM

Az azonnali üzenetküldő és csevegő szolgáltatások ma már gyakran nemcsak szöveg, hanem hang és videokép átvitelére is alkalmasak, és nagyon népszerűek a magánemberek közt és a vállalati szférában egyaránt. A könyvtárakban leginkább a ma már hagyományosnak mondható, e-mailre vagy webes úrlapra alapozott virtuális tájékoztató szolgáltatások kiegészítésére, esetleg kiváltására használják. Az „élő” online kapcsolatnak köszönhetően a könyvtárhasználók távolról is kérhetnek azonnali segítséget. A tanulmányozott egyetemi könyvtárak közül 37, vagyis kb. 65% kínált ilyen valós idejű „kérdezd a könyvtárost!” szolgáltatást, legtöbbször az ingyenes *Meebo* alkalmazáson keresztül.

SNS

Az *Alexa.com* népszerűségi listája szerint a *Wikipedia*, a *MySpace*, a *YouTube*, az *Orkut*, a *Facebook*, a *Friendster*, a *Flickr* és társaik, vagyis az úgynevezett *social networking sites* a leglátogatottabb helyek a weben. Közös jellemzőjük, hogy használói információkat osztanak meg egymás között, amelyek a személyestől a tudományosig terjedhetnek. A vizsgált 57 könyvtár közül 5 volt jelen ilyen közösségi oldalakon. Három intézmény a Facebook-on alakított ki magának profilt. A *University of Adelaide* könyvtárának 481 „ismerőse” volt itt a kutatás időpontjában. A *University of Texas* még egy keresődobozt is feltett a Facebook oldalára az online katalógusához. Illinois állam egyik egyetemi könyvtára a Flickr-en osztott meg egy képgyűjteményt; a *Rice University* könyvtárosa pedig a *goodreads.com* oldalon hoztak létre egy könyvklubot.

Bookmarking/tagging

Az olyan közösségi könyvjelzőgyűjtő oldalak, mint amilyen a *Delicious*, a *Furl*, a *Digg*, a *CiteUlike* stb., szintén jól ismertek az internetezők között. A koráb-

ban már többször hivatkozott 2007-es amerikai felmérés szerint a felhasználók 28%-a szokott online tartalmakat címkézni. A *tag*-ek és *bookmark*-ok elhelyezése nemcsak a weben való eligazodást segíti, hanem összehozza az azonos érdeklődésű embereket. E tanulmány készítői öt olyan egyetemi könyvtárat találtak, ahol megengedik ezt a fajta közösségi tevékenységet. A *University of Michigan Library* például saját eszközt kínál, amellyel felhasználói címkéket rendelhetnek egyes online tartalmakhoz. A *University of Pennsylvania* „PennTags” nevű alkalmazásával pedig a könyvtár katalógusában keresgélők is meg tudják címkézni a számukra érdekes tételeket.

Értékelés

A világ 100 nagyhírű egyetemének könyvtárát megvizsgálva 57-nek a honlapján sikerült webkettes funkciókat találni, leginkább RSS és IM szolgáltatásokat, népszerűségben ezt követik a blogok, míg a wiki a legkevésbé használt műfaj. Meglepő volt az is, hogy az akkortájt legismertebb közösségi oldalon, a *MySpace*-en egyik sem volt jelen közülük.

A világ változásával a könyvtárak is változnak. Jelen kell lenniük ott, ahol a használói vannak, összhangba kell hozniuk a szolgáltatásaikat a 2.0-s web lehetőségeivel. Okosan és gyorsan használatba kell venniük az új eszközöket, és segítségükkel még inkább a felhasználói igények kiszolgálását állítani a középpontba. A részvétel és az együttműködés a webkettő világának két legvonzóbb jellemzője, ezen a téren érdemes erősíteniük a könyvtáraknak is. Az itt bemutatott kutatás csak az újgenerációs technológiákra alapozott szolgáltatások meglétét vizsgálta, nem tért ki ezek minőségi elemzésére és fogadtatására a felhasználók oldaláról. De az e nélkül is bátran kijelenthető a világban végbemenő trendeket látva, hogy utóbbiak különösebb vonakodás nélkül, örömmel fogadnak minden ilyen újszerű lehetőséget.

Hivatkozások

- [1] BREEDING, M.: We need to go beyond Web 2.0. = *Computers in Libraries*, 27. köt. 5. sz. 2007. p. 22–26.
- [2] MANESS, J.: Library 2.0 theory – Web 2.0 and its implications for libraries. = *Webology*, 3. köt. 2. sz. 2006.
<http://www.webology.ir/2006/v3n2/a25.html> (letöltve: 2008.06.25)

[3] MILLER, P.: Web 2.0: building the new library. = Ariadne, 45. köt. 2005. október
<http://www.ariadne.ac.uk/issue45/miller/>

[4] HABIB, M. C.: Towards academic library 2.0: development and application of a library 2.0 methodology. Master's paper. University of North Carolina, Chapel Hill, NC., 2006.

/HARINARAYANA, N. S. – VASANTHA RAJU, N.: Web 2.0 features in university library web sites. = The Electronic Library, 28. köt. 1. sz. 2010. p. 69–88./

(Drótos László)

Virtuális referenz a brit felsőoktatási könyvtárakban

A *Virtual Enquiry Project* egy 2008-ban és 2009-ben lefolytatott kutatási program volt az Egyesült Királyságban, amelyet kettős céllal indítottak. Egyrészt fel kívánták mérni, hogy a felsőoktatási szférában mennyire terjedt már el a virtuális könyvári tájékoztatás. (Ez alatt jelen esetben csak az azonnali üzenetküldő vagy csevegő szoftverekkel megvalósított „élő” szolgáltatásokat értették, tehát az e-mail vagy webes űrlap formájában működőket kizárták a vizsgálódásból.) Másrészt igyekeztek hasznos információkat szolgáltatni olyan könyvtáraknak, ahol még csak most terveznek ilyen referenzszolgáltatást.

Előzmények

Az üzenetküldő és a csevegő programok nagyon népszerűek bizonyos felhasználói típusoknál. Egy 2007-es JISC felmérés például kimutatta, hogy az alsó- és középfokú oktatásban tanulók 90%-a használ ilyeneket rendszeresen vagy időnként. Az *Oxford Internet Institute* ugyancsak ebben az évben publikált jelentése szerint pedig a brit internetezők mintegy 60%-a szokta azonnali üzenetküldésre is használni a hálózatot. Természetes tehát, hogy az egyetemi hallgatók szívesen vennék igénybe a könyvtárak tájékoztató szolgálatát ilyen formában is, különösen azok, akik távoktatási vagy levelező képzés keretében tanulnak, és a tandíjuk fejében jogosan elvárhatják, hogy a nappali tagozatos diáktársaikkal egyenlő szolgáltatásokat kapjanak.

Az angol szakirodalom viszonylag szegényes ebben a témában, különösen az amerikaival összehasonlítva, ahol úgy tűnik, hogy a virtuális referenz sokkal elterjedtebb és jobban beépült már a könyvtári szolgáltatások közé. Néhány olyan kivételtől eltekintve, mint amilyen az *Open University* (ahol, az intézmény távoktatási jellegéből fakadóan elég korán, még 2001-ben elkezdtek

kísérletezni a *LivePerson* szoftverrel – a hallgatók nagy öröme), más brit egyetemi könyvtárakban ez a fajta szolgáltatás még kezdeti stádiumban van. Az eddig megjelent publikációk többsége vagy az aszinkron kommunikációs eszközökkel (mint amilyen az elektronikus levelezés) megvalósított tájékoztatásról szól, vagy esettanulmányokat közöl. Például *Hvass* és *Myer* [1] 2008-as cikkükben a *University of Teeside* könyvtárában bevezetett, *MSN Messenger*- és *Meebo*-alapú szolgáltatást ismertetik, amely induláskor munkanapokon négy órán át volt elérhető. A diákok nagyon jól fogadták ezt a lehetőséget, és kezdetben a könyvtárosok is pozitívan álltak hozzá, de az új kommunikációs formához való alkalmazkodás időnként nehézségeket okozott nekik. És bár ezen a csatornán viszonylag kevés kérdés érkezett, a szerzők hangsúlyozzák, hogy minden eszközt érdemes megragadni, ami elősegíti a kommunikációt az olvasókkal. A *University of Wolverhampton* tájékoztató szolgálata 2006 végén indított „valós idejű” referenzst. Ezt *Wendy Haynes* [2] mutatta be egy 2008-as *webinar*-on (online szemináriumon), és beszámolt egyrészt az igen alaposan kidolgozott marketingstratégiájukról (pl. szóróajándékokat is osztogattak a diákoknak), másrészt a használati statisztikákról (heti átlagban 30–40 kérdés jött, ez a szám persze erősen fluktuált a tanév különböző szakaszaiban). Azt is megemlítette, hogy hallássérültek számára különösen hasznos a forma.

Arról is vannak már információk, hogy egyes könyvtárakban miért döntöttek úgy, hogy megszüntetik ezt a fajta tájékoztatást. A legismertebb ilyen eset a *Strathclyde University* és két ausztrál könyvtár együttműködése. A partnerintézmények egy *eGain* nevű, eredetileg nem könyvtáraknak szánt ügyfélszolgálati szoftverre alapozva működtettek egy non-stop szolgáltatást, de két év után leállították, mert nehézkes volt a rendszer használata és nem volt rá annyi igény, hogy érdemes legyen fenntartani.* Egy másik tanulságos példát

Bains [3] ismertet egy 2003-ban megjelent hírlevélben. A brit egyetemi könyvtárak konzorciuma (*CURL*) tesztelte az *OCLC QuestionPoint* elnevezésű virtuális referenzrendszerét, ahol is a beérkező kérdéseket automatikusan a megfelelő szakkönyvtárnak továbbították. A bétateszt végén az volt az általános vélemény, hogy további kutatás-fejlesztés szükséges. A chat funkciót instabilnak és nehézkesnek találták akkor (ezt később az *OCLC* ki is vette a rendszerből). Végül csak az *Edinburgh University* használta tovább ezt a szoftvert.

Az amerikai *Bicknell-Holmes* [4] 2008-ban állította össze a téma válogatott bibliográfiáját, és ahogy kommentárjában megjegyzi: Bár a publikációk többsége azt sugallja, hogy egy csevegő/üzenetküldő referenzszolgálat elindítása „nem nagy ügy” ..., valójában még nem eldöntött kérdés, hogy érdemes-e ezeket az eszközöket ilyen célra használni. Van ugyanis néhány nehezen megoldható, bosszantó probléma velük: mint például az alacsonyabb forgalom, a hosszabb tranzakciók és az ügyeletet ellátó személyzet költsége.

A felmérés módja és eredményei

A *SurveyMonkey* nevű szoftverrel készített kérdőívet a *Virtual Enquiry Project* keretében 2008 augusztusa és októbere között küldték ki az Egyesült Királyság felsőoktatási könyvtáraiba. Öt részből állt, mert külön kérdések vonatkoztak azokra a könyvtárakra, ahol már működik virtuális referenz, továbbá azokra, ahol még csak most kezdtek próbálkozni vele, illetve ahol korábban volt, de jelenleg nincs, valamint azokra, amelyek szeretnének, és amelyek nem akarnak ilyet indítani. A válaszadási kedv ösztönzésére díjakat is kítűztek: egy *iPod Touch* készüléket és az *Amazon.co.uk* oldalon beváltható utalványokat lehetett nyerni. A felhívást megjelentették a különböző könyvtári levelezőlistákon, hírlevelekben és weblapokon. Összesen 82 felsőoktatási intézmény és 42 továbbképző intézet válaszolt (plusz még hat „egyéb”, ahol többnyire mindkét fajta képzés folyik). Az országos statisztikával összevetve ez azt jelenti, hogy az első típusból 49%-ról, a másodiktól pedig 10%-ról sikerült információt kapni. Mivel egyes helyeken több könyvtáros is kitöltötte a kérdőívet, így összesen 193 érvényes válasz jött a 130 intézményből. Ezek negyedében volt akkortájt működő online tájékoztatás (hét esetben még csak teszt üzemmódban), három helyen pedig létezett korábban ilyen, de megszüntették. 25 könyvtárban egyálta-

lán nem tervezték bevezetni, viszont 70 helyen igen.

A kutatók arra is rákérdeztek, hogy a könyvtárosok amúgy használnak-e üzenetküldő vagy csevegő programokat. A többség azt válaszolta, hogy igen, bár sokan nem rendszeresen. Van némi korreláció a között, hogy mely könyvtárakban vannak csetelő munkatársak, és hogy melyekben van ilyen típusú referenzszolgáltatás. Ezen intézmények 77%-ában legfeljebb két éve működtetik, és egyik helyen sem tervezik a leállítását, sőt inkább növelni és továbbfejleszteni szeretnék nagyobb hírveréssel, az ügyeleti idő meghosszabbításával, funkciógazdagabb szoftverekkel.

A szolgáltatáshoz használt program kiválasztásánál a válaszadók az egyszerű telepíthetőséget és használhatóságot értékelték a legmagasabbra, ezt követte a megbízható működés, a felhasználói felület és az akadálymentesség. Arra a kérdésre, hogy jelenleg milyen rendszert alkalmaznak, és hogy mennyire vannak vele megelégedve, sajnos csak kevesebb mint 30 helyről jött válasz. A legnépszerűbb a *QuestionPoint* és a *Meebo*, egy-két könyvtárban pedig a *LivePerson*, illetve a *LiveHelp* mellett döntöttek. (Ezek közül csak a *Meebo* ingyenes.)

Egyes kérdésekkel azt próbálták felderíteni a kutatók, hogy mennyire és mikor tekintik sikeresnek a könyvtárosok ezt a fajta szolgáltatást. Erről is elég kevés információt sikerült begyűjteni, de a reakciók többnyire pozitívak voltak, bár az is kiderült az átlagolásnál, hogy a könyvtárosok úgy gondolják, hogy a felhasználók jobban szeretik ezt a fajta tájékoztatást, mint ők maguk, és azt is lehet tudni, hogy az olvasószolgálatokhoz ezen a csatornán érkező kérdések aránya nemigen szokta elérni a 10%-ot. A kérdőív egyik pontja így szólt: „Mit tanácsolna egy olyan egyetemi könyvtárnak, ahol a virtuális referenz bevezetését tervezik?” A válaszadók ezzel kapcsolatban a szolgáltatás népszerűsítésének fontosságát emelték ki, illetve azt javasolták, hogy „bátran bele kell vágni”, de előtte azért nem árt megismerni a felhasználói igényt, végiggondolni a munkaerő-szükségletet, és kellően rugalmassá tervezni a szolgáltatást. Arra is figyelmeztettek néhányan, hogy nagyon fontos a részt vevő könyvtárosok elkötelezettsége, és hogy az elején csak kicsiben érdemes kezdeni: néhány órás ügyeket vállalva.

* Lásd: JOINT, Nicholas: A könyvtárak és a web 2.0 kihívás. = TMT, 57. köt. 1. sz. 2010. (A ref.)

Ahol nincs és a tervek között sem szerepel ilyen szolgáltatás, ott az okokról faggatták a kutatók a könyvtári dolgozókat. A felsőoktatási könyvtárakban a leggyakoribb érv az volt, hogy nincs idő tanulmányozni a dolgot, a továbbképző intézetek pedig inkább a munkaerőhiányra panaszkodtak; a harmadik fő ok pedig a forráshiány volt. De például olyan problémát is jeleztek, hogy egyes helyeken le van tiltva a csetprogramok használata, illetve hogy a diákoknak olyan időszakokban lenne leginkább szüksége az azonnali könyvtárosi segítségre, amikor a könyvtár zárva van (mert a hallgatók többnyire éjszaka és hétvégén szoktak tanulni). Van ahol pedig annyira elégedettek az e-mailre alapozott tájékoztató szolgáltatással, hogy nem látják értelmét, hasznát egy csetalapú referenzbevezetésének is.

Szoftverkinálat

Az aktuális helyzet feltérképezése mellett a Virtual Enquiry Project keretében a jelenleg kapható vagy ingyen letölthető virtuális referenzszoftverek összegyűjtését, tesztelését, értékelését és összehasonlítását is elvégezték. Ezzel azoknak a könyvtáraknak kívántak segítséget nyújtani, amelyek szeretnének ilyen szolgáltatást indítani, de még nem döntötték el, hogy milyen rendszert válasszanak hozzá.

A kutatók letöltöttek minden fontosabb ingyenes szoftvert, illetve tesztpéldányokat kértek a kereskedelemben kapható rendszerek gyártóitól. Mindegyik programot két ember értékelte a könyvtárosok és az ügyfelek szempontjából egyaránt, és időnként más munkatársakat is bevontak a tesztelésbe, hogy „kezdő felhasználói szemmel” is megnézzék az egyes szoftvereket. Az eredményeket egy mátrixban összesítették, így könnyebben összevethetők az azonos funkciók, az előnyök és a hiányosságok. A szempontok között voltak csak szubjektíven értékelhetők (pl. általános benyomás, a telepítés nehézsége, használhatóság, megbízható működés) és objektíven összehasonlíthatók (pl. ár, szükséges technikai háttér, hang- és videokapcsolat-üzemmód, fájlküldés, anonimitás megőrzésének lehetősége, naplózási és statisztikai funkciók, többfelhasználós működés).

Mindösszesen 11 programcsomagot teszteltek (*PIDGIN, Trillian, Meebo, AIM, Yahoo Messenger, MSN Messenger, Elluminate Live!, Library3lp, QuestionPoint, VRLPlus és LiveHelp*), de a hasonlókat azután összevonták egy kategória alá. (A

jelen cikk függelékében ezek közül is csak néhánynak az értékelését közlik a szerzők.) A Library3lp az egyetlen, amelyet kifejezetten könyvtári tájékoztatásra fejlesztettek, a többieknél hiányozhatnak egyes funkciók, amelyek pedig ilyen környezetben fontosak lehetnek (pl. az automatikus statisztikakészítés). Ahogy várható volt, az olyan professzionális, fizetős rendszereknek, mint amilyen a LivePerson és a QuestionPoint van a legjobb ügyféltámogatásuk és a legtöbb szolgáltatásuk: például az egyszerre érkező felhasználói kérések sorba állítása vagy átirányítása másik válaszadóhoz, több könyvtáros egyidejű bejelentkezésének lehetősége. Ugyanakkor egyesek már túl bonyolultnak találhatják ezeket a szoftvereket, ami pedig vonakodást válthat ki a könyvtárosokban, ez pedig komoly kockázati tényező. Úgyhogy ezeknél szükség lehet előzetes oktatásra – egyes esetekben maga a gyártó cég is nyújt erre lehetőséget. Kezdesnek, a felhasználói érdeklődés felméréséhez és a valós idejű online tájékoztatás előnyeinek és hátrányainak megtapasztalásához jó választás lehet akár egy olyan egyszerűen beüzemeltető és gyorsan megtanulható ingyenes eszköz is, mint amilyen a Meebo.

A teljes mátrix, valamint a kérdőíves felmérés részletes eredményei, továbbá különböző egyetemi könyvtárakból gyűjtött esettanulmányok megtalálhatók a projekt honlapján, a www.virtualenquiry.net címen.

Hivatkozások

- [1] HVASS, A. – MYER, S.: Can I help you? – Implementing an IM service. = *Electronic Library*, 26. köt. 4. sz. 2008. p. 530–544.
- [2] HAYNES, W.: Promoting virtual reference in a UK university. 2008. = <http://www.wlv.ac.uk/Default.aspx?page=17551> (letöltve: 2009.03.01.)
- [3] BAINS, S.: The OCLC QuestionPoint electronic reference service: experiences of a group of CURL libraries. = *SCONUL Newsletter*, 29. köt. 2003. p. 21–26. <http://www.sconul.ac.uk/publications/newsletter/27/ARTICLE6.RTF> (letöltve: 2008.06.16.)
- [4] BICKNELL-HOLMES, T.: Chat and instant messaging for reference services: a selected bibliography. Faculty Publications, UNL Libraries. 2008. = <http://digitalcommons.unl.edu/librarianscience/151> (letöltve: 2009.02.28.)

/BARRY, Eithne – BEDOYA, Jaclyn Kelli – GROOM, Carolyn – PATTERSON, Laurence: Virtual reference in UK academic libraries – The virtual enquiry project 2008–2009. = *Library Review*, 59. köt. 1. sz. 2010. p. 40–55./

(Drótos László)

Egyetemi könyvtár marketingje a weben

A könyvtáraknak szembe kell nézni azzal a tényvel, hogy nő a világhálón elérhető információ mennyisége. Hosszú ideig a könyvtár volt a nyomtatott információk egyedüli forrása, az információk közvetítője – ez a szerepe most már jelentősen csökken. A használók többsége információkeresésnél szívesebben használ keresőmotorokat ahelyett, hogy a könyvtárba menne. Mivel a hálózatot használók száma folyamatosan nő a könyvtárba látogatókkal szemben, az egyetlen járható út az olvasók megtartására, hogy a honlapot marketing és közönségkapcsolati (PR) eszközként is használni kell.

A marketing és a könyvtári honlap

A könyvtár – marketingstratégiákat alkalmazva – láthatóbb lesz, ezáltal hatékonyabban tudja elérni céljait, a könyvtárhasználók is eredményesebben tudják használni egyetemi és kutatási céljaik érdekében a kínált szolgáltatásokat. A marketing kétoldalú kommunikációs folyamat: az egyik oldalon a kínálat, ami a könyvtárban rendelkezésre áll, és amit a könyvtár tehet a kutatók érdekében, a másik oldalon pedig a kutatók egyre növekvő elvárásai, szükségletei és igényei. A gyűjtemények és az online szolgáltatások fejlesztése nem lehet hatékony, ha a könyvtárnak nincs kapcsolata a használókkal, nem kezeli a használói visszajelzéseket. Egy 2008-as publikációban 69 egyetemi könyvtár honlapját elemezve [1] az alapvető szolgáltatások és gyűjtemények népszerűsítése tekintetében megállapították, hogy a könyvtárak fontosnak tartották ugyan a kari honlapokon való megjelenést, de a digitális gyűjteményeket, a repositóriumokat, a nyílt hozzáférést nem reklámozták. Egy másik kutatásban [2] a „chat-alapú” referenzszolgáltatás marketingjét nézték, és azt találták, hogy a szolgáltatás használatának növekedése attól is függ, hogy hol van a honlapon ez az ikon. Az USA 106 egyetemi könyvtárában végzett felmérésben [3] a könyvtári honlapon elhelyezett marketing- és PR-linket vizsgálták a *Kutatókönyvtárak Egyesülete (Association of Research Libraries = ARL)* 1999-es tanulmányában kidolgozott négy fő szempont szerint. A jelen tanulmány is e négy szempont mentén elemzi a honlapokat:

- Láthatóság – direkt link az intézményi honlapról a könyvtári honlapra.
- Adománygyűjtés – direkt link baráti csoportok és egyesületek honlapjára, valamint adományokhoz.

- Tájékoztatás – hírekre, rendezvényekre, új gyűjteményekre mutató direkt linkek.
- Kérdések, megjegyzések – direkt link a „Kérdezd a könyvtárost!” oldalra, az e-mail címhez vagy az üzenetküldéshez.

A vizsgált honlapokon általánosan jellemző volt, hogy pénzszerzés és tájékoztatás tekintetében fejlesztésre szorultak. Nem volt jellemző, hogy adományok gyűjtéséhez közvetlen link lett volna, annak ellenére, hogy egyes könyvtárak a pénzügyi segítség növekedését tapasztalták, ha a könyvtári honlap felső részében megjelenítették ennek a lehetőségét. A könyvtári honlap alkalmas a könyvtári források, szolgáltatások elérésére, különböző információk közvetítésére, és egyben kommunikációs csatorna a könyvtár és használói között.

Jelen tanulmány a malajziai egyetemi könyvtárakban végzett kutatásról, az alkalmazott módszerről és az eredményekről számol be. Azt vizsgálták, hogy milyen módon reklámozzák és teszik nyilvánossá forrásaikat és szolgáltatásaikat honlapjukon a könyvtárak.

Módszer

Malajzia 18 állami és 4 magánegyetemének könyvtári honlapját vizsgálták. A könyvtárak mindegyike a PERPUN (együttműködési fórum az egyetemi könyvtárak vezetői és a Maláj Nemzeti Könyvtár főigazgatója részvételével) tagja. Első lépésben összeállították azon tulajdonságok listáját, amelyek alapján a könyvtárak honlapján azonosítani lehet a marketingstratégiákat. Az értékelést kiképzett hallgatók végezték a kutatók segítségével, majd az eredményeket véletlenszerűen ellenőrizték a kutatók. Először az egyetem honlapján nézték meg a könyvtári honlapra mutató direkt linket, majd a könyvtár weboldalait vizsgálták. Ha egy könyvtárnak több honlapja is volt, akkor csak a központi könyvtárét elemezték. Második lépésben interjúkat készítettek – hat személyes és három telefonos – kiválasztott könyvtárigazgatókkal, hogy kiderítsék, mit gondolnak arról, hogy a könyvtári honlapot marketingeszközként használják.

Eredmények

2008 júniusa és augusztusa között 22 egyetemi és könyvtári honlapot ellenőriztek; az ez utáni változtatásokat már nem tartalmazza a tanulmány. Az

eredményeket a korábban említett négy fő szempont szerint csoportosították.

Láthatóság

A láthatóságot az egyetemi honlapról a könyvtári honlapra mutató közvetlen hivatkozás meglétével határozták meg. Direkt link az a link, amely azonnal látható, máskülönben indirekt (ha csak akkor látszik, ha az egeret fölé visszük, vagy ha felugró ablakból vagy lenyíló menüből választható). Csak 13 esetben (59%) mutatott direkt link az intézményi honlapról a könyvtári honlapra. Két könyvtárnál egyáltalán nem volt, hét könyvtár esetében (31%) volt ugyan link, de csak a második oldalon, vagy csak az egérrel vált láthatóvá. Fontos, hogy az egyetemi menedzsment értékelje, és értékes vagyonnak tekintse a könyvtárat. A láthatóság jelen esetben olvashatóságot is jelent, mivel Malajziában az angol a második nyelv. Az egyetemi könyvtáraknak helyi és külföldi használói is vannak, ezért fontos a többnyelvűség. A négy magánegyetemi könyvtár honlapja angol volt, 16 állami egyetemi könyvtár (88%) szintén használta az angolt, de közülük csak három volt kétnyelvű (angol és maláj). Két könyvtári honlap csak maláj nyelvű volt, és bár tisztában vannak az angol honlap szükségességével, nincs anyagi forrásuk az elkészítéséhez.

Hálózat

A hálózat fontos a könyvtárak számára, hiszen tekintettel az állandó forráshiányra, a honlapot tevékenységük reklámozására használhatják. Direkt linkek egyesületekhez és adományozásra buzdító oldalakra – ez az egyik lehetséges módja annak, hogy a könyvtár kapcsolatot tartson olyanokkal, akik tehetnek valamit az érdekében. Malajziában csak egy könyvtári szervezet van: a *Maláj Könyvtárosok Egyesülete (Malaysian Librarians Association = Persatuan Pustakawan Malaysia, PPM)*, ennek ellenére csak hét könyvtárnál (31%) volt közvetlen hivatkozás az egyesület honlapjára. Szintén csak három könyvtári webhelynek volt direkt linkkel az IFLA honlapjára, a többi könyvtárnak egyáltalán nem volt könyvtári szervezetekre mutató linkje. A vizsgált könyvtárak mindegyike a PERPUN tagja, ezért elvárható volt a PERPUN oldalára mutató link – a többségüknél, 18 könyvtár (81%) esetében volt is. Másik egyetemi könyvtárra csak 12 könyvtárból (54%) volt továbblépési lehetőség, holott fontos lenne egyrészt az egyetemi könyvtárak együttműködése, másrészt célszerű a használók teljesebb kiszolgálása érdekében, hogy forrásaikat és szolgáltatásaikat megosszák, és ennek megfelelően a könyvtári honlapot információs portálként használják.

Tájékoztatás

A PR fontos része a tájékoztatás. Ebből a szempontból a könyvtárról, a könyvtári programokról és rendezvényekről, az új gyűjteményekről szóló hírek, az eseményekről készült fotók, valamint az online könyvtári hírlevél elérhetőségét vizsgálták. A könyvtárak többsége: 14 (63%), adott hírt a honlapján könyvtári programokról. Csak hat (33%), – mindegyikük állami egyetemi könyvtár – közölt információt küszöbön álló rendezvényekről, 15 (68%) tájékoztatta a használókat újdonságokról, és csak öt (22%) tett fel képeket könyvtári eseményekről. Három állami egyetemi könyvtárnál (16%) találtak online könyvtári hírlevelet, egy könyvtárnál a *LISA (Library and Information Science Abstracts)* által is indexelt kiadványt. 11 könyvtár (50%) tájékoztattott a honlapján új gyűjteményeiről (új könyvek, folyóiratok, kipróbálható adatbázisok), 13 könyvtárnál (59%) volt tudományágak szerint csoportosított tartalomszolgáltatásokra mutató linkgyűjtemény. A szakkönyvtáros által válogatott és rendszerezett online források a keresőmotorok használata helyett a könyvtári weblapok felé terelik a használókat. Kurrens újságok, hírforrások elérését csak 12 könyvtár (54%) tette lehetővé a honlapjáról.

Kérdések, megjegyzések

A honlapokon a használóval való kapcsolattartásra e-mail cím, vagy a referenz-könyvtároshoz vezető link, népszerű nevén: „Kérdezd a könyvtárost!” szolgál. A vizsgált könyvtárak közül 15 (68%) adott e-mail elérési lehetőséget a használóknak, és csak 9 (40%) kínálta fel az online tájékoztató szolgáltatást. A könyvtár missziója az állami egyetemi könyvtárak 77%-ánál, míg a magánegyetemi könyvtárak mindegyikénél megtalálható volt a weblapon. Szerzői jogokra (copyright) vonatkozó információt csak 5 (22%) honlap tartalmazott. Tíz könyvtár (45%) kínálta linket keresőmotorokhoz. A könyvtárvezetők többnyire nem akarják a hallgatókat az internethasználatra ösztönözni, jobban bíznak az online adatbázisokban.

Fontos, hogy minden hallgató és egyetemi alkalmazott megértse a szellemi tulajdon védelmét, és a plagizálás veszélyeit, különösen a digitalizált környezetben, ahol az útmutatók és a törvények folyamatosan változnak. Csak öt egyetemi könyvtár (22%) fordít arra energiát, hogy használóit a copyrighttal és a plagizálással kapcsolatban oktassa.

A marketingstratégia fontos eleme a visszacsatolás. A használókat folyamatosan emlékeztetik, hogy

véleményük fontos a szervezet számára, és segíti a könyvtár és használói közötti kapcsolat kiépítését is. Kilenc könyvtár (68%) tette lehetővé, hogy a használó megírja elektronikus levélben vagy egy webes úrlapon panaszát, javaslatát. Az állami könyvtárak közül csak három (16%), a magánegyetemi könyvtárak közül pedig kettő (50%) alkalmazott online kitölthető kérdőívet a visszacsatolásra.

Konklúzió és javaslatok

A könyvtári marketing célja, hogy több használót vonzzon a könyvtárba. A honlapot ne csak a nyitva tartásról, kölcsönzési lehetőségekről stb. való tájékozódásra használják, hanem kiindulási pontja legyen az információkeresésnek. Az üzleti világgal ellentétben, a könyvtárakban a marketing nem hoz kézzel fogható profitot. Az egyetemi könyvtárak honlapjukon keresztül láthatóvá válnak az egyetemi közösség számára, ugyanakkor figyelmüket használók igényeinek kielégítésére fordítják. A láthatóságon kívül fontos a honlap használhatósága is, ami ebben az esetben a többnyelvűséget jelenti, tekintettel a Malajziában beszélt/használt nyelvekre, valamint a külföldi diákokra. Fontos továbbá a honlapon a könyvtári szervezetekkel és más könyvtárakkal való kapcsolat az együttműködések érdekében, hiszen az internet korában már nem az önellátás a cél.

A könyvtáraknak szembe kell nézniük azzal a ténnyel, hogy a használók, ha információhoz akarnak jutni, már számos lehetőség közül választhatnak.

Energiát kell fordítaniuk a hatékony reklámozásra, tudatni a használókkal, hogy mi történik a könyvtárban, milyen új gyűjtemények, új elektronikus szolgáltatások érhetők el. Ki kell építeni a kapcsolatot a használókkal, igényeik minél teljesebb kielégítése érdekében, és el kell érni, hogy újra a könyvtárhoz forduljanak segítségért. Ezért a honlapon digitális referenzszolgálatot kell nyitni, vagy chat-szolgálatot indítani, és jól látható helyen elhelyezni az észrevételek írására szolgáló ikont. E tanulmány keretében vizsgált könyvtáraknak csak nagyon kis százaléka használta ezeket az eszközöket. A felhasználói visszajelzések és igények fokozottabb figyelembevételére erősíti a lojalitás érzését a könyvtárak virtuális látogatóiban, így valószínűbb, hogy gyakran visszatérnek a könyvtár honlapjára.

Irodalom

1. GARDNER, S. J. – JURICEK, J. E. – XU, F. G.: An analysis of academic library web pages for faculty. = *Journal of Academic Librarianship*, 34. köt. 1. sz. 2008. p. 16–24.
2. VILELLE, L.: Marketing virtual reference: what academic libraries have done. = *College & Undergraduate Libraries*, 12. köt. 2005. p. 65–79.
3. WELCH, J. M.: The electronic welcome mat: the academic library web site as a marketing and public relation tool. = *The Journal of Academic Librarianship*, 31. köt. 3. sz. 2005. p. 225–228.

/KAUR, Kiran: *Marketing the academic library on the web.* = *Library Management*, 30. köt. 6–7. sz. 2009. p. 454–468./

(Naszados Edit)

Olvasói könyvtárközi kérésre alapozó gyarapítás

A könyvtárak manapság gyakran használják az olvasói kérésre alapozó gyarapítási technikákat, vagyis akkor szerzik be a dokumentumot, amikor azt egy olvasó könyvtárközi kölcsönzéssel kéri. A szakirodalom gyakran tárgyalja ezt a fajta gyarapítási eljárást, de főleg közkönyvtárakat vizsgáltak eddig. *Doug Way* cikkében egy általános egyetem, a *Grand Valley Állami Egyetem (Grand Valley State University = GVSU)* programját mutatja be. Általános egyetemen (comprehensive university) az oktatásra, s kevésbé a kutatásra hangsúlyt helyező egyetemeket értik az Egyesült Államokban. A GVSU az USA, Michigan államában található, 24 000 hallgatót képez, legmagasabb fokozatként master diplomát ad.

A közkönyvtárakban különböző szempontokat vesznek tekintetbe ennél a gyarapítási módnál: van,

ahol minden kért könyvet megvásárolnak, van, ahol különbséget tesznek: az árat, a könyv témáját, korát, beszerezhetőségét mérlegelve. A módszer eredményességét nem értékelték eddig részletesen, általában azt vizsgálták, milyen hatással volt a költségtakarékosságra, a dokumentumok forgási idejére, a beszerzést követő kölcsönzések számára.

A kutatóegyetemen a legfontosabb értékelési tényezőt a beszerzést követő használat jelentette. Az egyetemen mindenütt korlátozták az így beszerzett dokumentumok körét, csak megadott kritériumoknak megfelelő könyveket vásároltak meg: kikérték a művet, mennyire könnyen elégíthető ki a kérés könyvtárközi kölcsönzéssel stb. Az egyetemek nemcsak a gyarapításra vonatkozóan határoztak meg szempontrendszert, hanem a technika eredményességének értékelésére is.

A GVSU könyvtárában a bevált gyűjteményszervezési politika szerint végzik a gyarapítást: kiadói listák, kari ajánlások. Ennek ellenére az elmúlt hat évben a könyvtár állományának csak 31%-át kölcsönözték. Ugyanakkor mintegy 33%-kal nőtt a könyvtárközi kölcsönzések száma. Ezért fontolóra vették a könyvtárközi kölcsönzéssel kért művekre is alapozó gyarapítás bevezetését. Ehhez igénybe vették az OCLC Worldcat (*Worldcat Collection Analysis = WCA*) gyűjteményelemzés szolgáltatását, melynek segítségével az elmúlt öt év kölcsönzési listáját kinyerték és állományukat összehasonlították a WorldCat katalógussal, illetve más könyvtárak állományával. Az elemzések lefolytatása után úgy döntöttek, hogy belefognak a könyvtárközi kölcsönzési kérésekre alapuló gyarapításba. Erre 5000 dollárt különítettek el, korlátozó szempontként az utóbbi három évben megjelent és még kapható, 75 dollárnál olcsóbb monográfiákat határozták meg. Az árkorlátozással a tankönyvek vásárlását zárták ki. A programot egy egyetemi félévben át tesztelték, aztán teljes mértékben bevezették a következő tanév elején. A kezdeti időkből aránylag alacsony volt az így beszerzett dokumentumok száma, de rájöttek az okára: az integrált rendszerük automatikusan továbbküldte a kéréseket. Ezt egy beállítással korrigálni tudták, így már a könyvtárközi kölcsönzést végző munkatársak tudomására jutottak a kérések.

Az ILLiad elnevezésű könyvtári integrált rendszerük és a már említett WCA adatait használták fel a teszt értékeléséhez. A beszerzett dokumentumok 51%-át a karok oktatói, 41%-át egyetemi hallgatók, 11%-át végzős hallgatók, 2%-át dolgozók és 1%-át egyéb besorolású olvasók kérték.

A program első teljes évében 114 tétellel gyarapították állományukat ezen a módon. A megoszlás: 6 db videofelvétel és 108 könyv volt. 41 dokumentumot több mint egyszer kölcsönöztek ki. Az átlagos kölcsönzési szám 1,5 volt, egy tételt 9-szer is kivettek, míg 12-t egyszer sem.

Az értékeléskor megnézték azt is, hogy a beszerzett könyvek tipikusak-e egy általános egyetem esetében; e célból hasonló adottságú és nagyságú általános egyetemek könyvtáraival hasonlították össze magukat. Ezért néhány más könyvtár munkatársai is áttekintették a vásárlásokat szakterületi megfelelőség szempontjából. De ennek az eljárásnak meg lehetőséget nagy a szubjektivitása, így az elemzés nem sokat segített. Mivel a széles gyűjtőkörű általános egyetemek könyvtáraiban, ahol sokféle témájú dokumentumot vásárolnak, de egy-egy témában nem teljességre és mélységre törekvően, nem lehet tényleges összehasonlításokat tenni. Ráadásul eltér az állománygyarapításra fordítható pénzügyi keret és az egyes szakterületek lefedettsége. Ezért szem-

pontként inkább a megvásárolt művek tényleges fellelhetőségét vizsgálták kiválasztott könyvtárakban. Nemzeti központi nyilvántartások statisztikai adataira támaszkodva választottak ki öt hasonló állománynagyságú könyvtárral rendelkező egyetemet. A WorldCat katalógusát használva állapították meg, hogy az egyes beszerzett művek megvannak-e a többi könyvtárban. A felmérés kimutatta, hogy a könyvek 59%-a megvolt legalább egy másik könyvtárban, 36%-a pedig több mint egyben. Ugyanezt megnézték csak a megvásárolt könyvek (videók nélkül) esetében: a számok 60% és 22%. Azt is megvizsgálták, hogy ezeket a legtöbbször kölcsönzött műveket mely olvasói csoportok kérték leginkább. Egy másik elemzés azt derítette fel, hogy a más intézményekben is megtalálható könyveket hányszor kölcsönözték. A számok azt mutatták, hogy több intézményben is meglevő dokumentumok többszöri kölcsönzése magasabb értéket mutat.

Az elemzések egyértelműen bizonyították, hogy a könyvtárközi kölcsönzési kérésekre alapozó gyarapítás hatékonyan egészíti ki a szakreferensek által használt hagyományos módszereket. A beszerzett művek általában illenek a gyűjteménybe és a könyvtárhasználók igénylik őket. Ezeknek a könyveknek magasabb a forgása, mint a hagyományos módszerekkel megvásároltaké.

Más hasonló adottságú könyvtárak állományának elemzése is nagy segítséget jelenthet, hiszen azokat a könyveket kölcsönözték gyakrabban, amelyek több más könyvtárban is megvoltak. További vizsgálatok kimutathatják, hogy ez a megállapítás általánosítható lesz.

A program eredményei alapján a GVSU könyvtára úgy döntött, hogy a gyarapításnak ezt a hatékonynak bizonyuló módját továbbra is fenntartják és ki is terjesztik a korlátozások szűkítésével. Bár a gyűjteményszervezésnek ez a módja sohasem fogja teljesen felváltani a hagyományos módszereket, arra utal, hogy a könyvtárak egyre inkább bevonják használóikat a beszerzési döntésekbe. Ezáltal elősegítik a költségvetési keretek hatékonyabb felhasználását, a munkatársak és a könyvtárosok idejének és erőfeszítéseinek jobb kihasználását.

WAY, Doug: The Assessment of Patron-Initiated Collection Development via Interlibrary Loan at a Comprehensive University. = Journal of Interlibrary Loan, Document Delivery & Electronic Reserve, 19. köt. 4. sz. 2009. p. 299–308./

(Burmeister Erzsébet)

Science of Synthesis

Houben–Weyl Methods of Molecular Transformations®



Pharmaceutical Substances

Syntheses | Patents | Applications

- The most effective and reliable methods for group transformations
- Covers all fields of organic chemistry
- Continuously updated (4 releases per year)

- All approved active pharmaceutical ingredients (APIs)
- Continuously updated
- Structure searchable

Science of Synthesis 3.8

Table of Contents

- Organometallics
 - Vol. 1: Compounds with Transition Metal-Carbon Bonds and Compounds
 - Vol. 2: Compounds of Groups 7-9 (Mn, Co, V, Ti, Sc, La, Ce, Ac)
 - Vol. 3: Compounds of Groups 10 and 11 (Ni, Pd, Pt, Ag, Au)
 - Vol. 4: Compounds of Groups 12 (Zn, Cd, Hg, Cu, Ag, Au)
 - Vol. 5: Compounds of Groups 13 (Al, Ga, In, Sn, Pb)
 - Vol. 6: Compounds of Groups 14 (Si, Ge, Sn, Pb)
 - Vol. 7: Compounds of Groups 15 and 16 (As, Sb, Bi, N, P, As)
 - Vol. 8: Compounds of Group 17 (Cl, Br, I, At)
- Heterocycles
 - Compounds with Four and Three Carbon-Heteroatom Bonds
 - Compounds with Two Carbon-Heteroatom Bonds
 - Compounds with One Carbon-Heteroatom Bond
 - Compounds with All-Carbon Functions

Method 2: Diene-Diene Cycloadditions

Montgomery, J., in: Science of Synthesis, 1 (2001), p.13

Information on Science of Synthesis Volume 1

Feedback

The most widely used application of $\text{Ni}(\text{DPPF})_2\text{Cl}_2$ -diene complexes is the dimerization of 1,3-dienes. Pioneering studies by Wiles demonstrated many different modes of coupling, including dimerization, trimerization, and oligomerization of 1,3-dienes.^[1-7] An overview of the product classes that may be obtained from 1,3-dienes is provided in Scheme 4 (see also Houben-Weyl, Vol. 4, 16, pp. 35 and 302-327). The initially formed $\text{Ni}(\text{DPPF})_2$ complexes 5 and 6 have not been isolated. However, the complexes may be stabilized by the addition of phosphines, and *m*-allyl complexes 7-9 have been prepared and characterized.

Scheme 4 Products of $\text{Ni}(\text{DPPF})_2$ -Catalyzed Diene-Diene Dimerization and Trimerization^[8,9]

Science of Synthesis: Organic and organometallic synthetic methods critically evaluated by experts – an authoritative information resource

Pharmaceutical Substances:
Your reliable source for the syntheses of active pharmaceutical ingredients.

Thieme Pharmaceutical Substances

Atorvastatin calcium

Synonyms: C1-981; Y18-548

ATC: C10AA05

Use: Hypocholesteremic, HMG-CoA-reductase inhibitor

Chem. Name: (S)-[2-(4-chlorophenyl)-5-(1-methylethyl)-3-phenyl-4-(phenylamino)oxazolyl]-1H-pyridin-1-one calcium salt (2:1)

Name: Heparinic acid calcium salt (2:1)

Formula: $\text{C}_{27}\text{H}_{35}\text{ClN}_2\text{O}_5\text{Ca}_2$

EW: 1155.36 g/mol

CAS-RN: 134523-03-8

Free acid

Formula: $\text{C}_{27}\text{H}_{35}\text{N}_2\text{O}_5$

EW: 508.60 g/mol

CAS-RN: 134523-00-5

Substance Classes

Antilipids
Pharmaceutical acids
Heparinic and Heparinic acids
Hydroxy acids
Pyrimidines

Synthesis Path

very synthetic methods

For more information and a free trial please contact:

Thieme Institutional Sales
E-mail: eproducts@thieme.de
Tel.: +49-7 11-89 31-407

