



A Google keresőjén túl

Az „Oxford English Dictionary” szócikke szerint a „google” szót először egyfajta speciálisan csavart ütésre használták a krikett és a tollaslabda sportban. De az 1998 szeptemberében indult Google kereső nem innen, hanem egy matematikai szlengszóából kapta a nevét: a „googol” a 10-nek a századik hatványát jelenti. Ezzel a gigászi névvel az új keresőgép azt sugallta, hogy képes lesz megbirkózni a web elképzelhetetlen méreteivel. A gyors sikert az is jelzi, hogy egy év múlva már igeeként is használták az új kereső nevét a *Usenet* hírcsoportokban, és az akkori vezető keresőgépek: a Lycos, az AltaVista és a Yahoo! gyorsan háttérbe szorultak. A 21. század első évtizedét egyértelműen a Google uralja és már nemcsak a webkeresőjével, hanem az „istállójába” tartozó egyéb szolgáltatásokkal is, mint például a Google Images, a Google Books, a Google Scholar, a Google Maps stb.

Az amerikai *Rutgers University* tájékoztató könyvtárosa cikkében sorra veszi azokat az alternatívákat, amelyeket a Google helyett használhatunk, ha másfajta módon szeretnénk keresni, vagy valamilyen speciális témában kutakodunk az interneten, vagy amikor weblapok helyett más típusú dokumentumokat szeretnénk megtalálni.

Eltérő megoldások

A Google általános és szabadszavas keresőjének egyik alternatíváját az emberi közreműködéssel készülő válogatások jelentik, ahol is szakértők által összegyűjtött, minőségi információforrásokban tudunk keresni, melyek relevánsabb találatokat adnak. Ezek egyik jó példája a brit *Intute* (www.intute.ac.uk) adatbázisa, ami már több mint 100 ezer, oktatási, illetve tudományos jellegű forrást tartalmaz – jellemzően kormányhivatalok, kutatóintézetek és szervezetek honlapjait. A keresést leszűkíthetjük nagyobb témákra: művészetek és humán területek, egészségügy és élettudományok, mérnöktudományok és technika, valamint

társadalomtudományok. Egy másik érdekes oldal az önmagát „tematikus indítóállvány”-ként definiáló *Pinakes* (www.hw.ac.uk/libwww/irn/pinakes/pinakes.html). Az *Alexandriai Könyvtár* katalógusáról elnevezett oldal nem tartalmaz saját keresőt, hanem csak ugrópontokat a legfontosabb tematikus webkatalógusokhoz, olyan változatos témákban, mint például a jog, az építészet, az űrhajózás, a történelem, a zene, a várostervezés.

Egy másfajta megközelítést jelent a természetes nyelvű keresés. Az *Ask.com* (www.ask.com) – korábbi nevén: *Ask Jeeves* – oldalán a szabad szavas keresés mellett kérdő mondatokat is beírhatunk, és azután a rendszer további opciókat kínál a találatok pontosításához. Kevésbé megbízható válaszokat ad az *Answers.com* (www.answers.com), ahol szintén lehet angol mondatokkal keresni, de a szótárak és lexikonok szócikkei mellett az internetező közönség által összeállított, kétes minőségű WikiAnswers oldalai is megjelennek a találatok közt.

Egyes keresők a vizuális megközelítést választották: A *Quintura* (www.quintura.com) például címkével készíti a beírt keresőszavakhoz, így további releváns szavakra kattintgatva tudjuk finomítani a keresésünket. Ennél is látványosabb megoldást használ a *KartOO* (www.kartoo.com), amely Flash formátumú tématerképet jelenít meg találati lista helyett, és az eredmények közötti kapcsolatot is mutatja. Bár a cikk szerzője nem igazán látja a hasznát és értelmét, de a vizuális beállítottságú embereknek ez lehet, hogy áttekinthetőbb.

Egy további eredményprezentálási technika a klaszterezés. Erre a *Clusty* (clusty.com) a legjobb példa. Ha ezt a keresőt használjuk, akkor a találatok altémakörökbe csoportosítva jelennek meg. Például a globális felmelegedésre keresve egy oldalsávban ilyen témacsoportok tűnnek fel: „ég-hajlatváltozás”, „hogyan állítsuk meg a felmelegedést”, „széndioxid-kibocsátás” stb., és ezekre kattintva finomítható a találati lista. A *Grokker*

(www.grokker.com) szintén ezt a kézre álló megoldást használja, de ez csak a Yahoo! és a Wikipédia keresőjére épül, míg a Clusty több keresőgépet is lekérdez.

Végül ott vannak az általános keresők számára láthatatlan „mélyweb” elérését biztosító eszközök. A *Complete Planet* (aip.completeplanet.com) saját állítása szerint 70 ezer adatbázist és speciális keresőt kérdez le. Az *OAIster*-hez (www.oaister.org) hasonlóan mélyen belenyúl a nagy tudományos és kormányzati adattárakba és webhelyekbe. Az *OAIster* közel ezer szolgáltatótól gyűjti be a metaadatokat és így több mint 16 millió, másként nehezen hozzáférhető adatrekord között kereshetünk vele.

Speciális témák

A legjobb válaszokat általában az adott témakörre specializált források nyújtják. Gazdasági témákhoz például ilyen a *BizSeer* (bizseer.ist.psu.edu), amivel oktatási, tudományos, céges és kormányzati webhelyeken tudunk cikkeket, műhelytanulmányokat, jelentéseket és más hasonlókat keresni. Ráadásul a *Penn State University* által fejlesztett *BizSeer* idézetségi keresésre is alkalmas. Ide kapcsolhatók még az álláskereső oldalak is: ezek közül a két legnagyobb a *Monster* (www.monster.com) és a *Career Builder* (www.careerbuilder.com). A tudományos és műszaki témák fontos forrása az *Elsevier* által fenntartott *Scirus* nevű kereső (www.scirus.com). Ez már több mint 450 millió folyóirat-cikket, preprintet, szabadalmat, tudományos weboldalt és más hasonlókat indexel le. Az eredmények három kategóriába osztva jelennek meg: előfizetéses folyóiratok, kiemelten fontos webhelyek, és a web többi része. A *ChemSpider* (www.chemspider.com) egy ennél is specializáltabb kereső, amely molekulaszervezetek adatait és kísérő információit gyűjti egy hatalmas adatbázisba. Orvosi témákban kitűnő keresőgép a *Healia* (www.healia.com): a weben, a Pub-Med adatbázisban, valamint a klinikai vizsgálatok között kereshetünk vele. A nagyközönség számára ennél is érdekesebb a nagyon népszerű *WebMD* (www.webmd.com), amivel mindenféle egészségügyi kérdésre kaphatunk választ.

A kormányzati weboldalak leghatékonyabban az erre a célra létrehozott szolgáltatásokkal kereshetők. Az Egyesült Államokban ilyen a szövetségi és a tagállami hivatalos honlapokat és dokumentumokat egyaránt leindexelő *USA.gov* (www.usa.gov). Nagy Britanniában a *Direct.Gov* (www.direct.gov.uk) szolgál hasonló célra. A kormányzatok egyik erőssége a statisztikagyártás, de a megfelelő

adatsorokat nem mindig könnyű megtalálni. Ebben segít a *Census Bureau* által fejlesztett *DataFerret* (dataferrett.census.gov). Itt egy Java alkalmazás segítségével kereshetünk az USA kormányzata által fenntartott demográfiai, gazdasági, környezeti, egészségügyi és más jellegű adattárakban.

Különleges típusok

A YouTube a legnagyobb és legnépszerűbb videogyűjtőhely a weben. A *Truveo* (www.truveo.com) keresővel nemcsak a YouTube, hanem a Google Images, az AOL és a MySpace videói is megtalálhatók. Az Internet Archive *Moving Image Archive* nevű gyűjteménye (www.archive.org/details/movies) nem annyira a „csinálj saját műsord” mozgalmat támogatja, hanem inkább az ismeretterjesztő és a kulturális filmekre helyezi a hangsúlyt. Ugyanez a helyzet a hangfelvételek esetében is: míg a *Podcast Alley* (www.podcastalley.com) elsősorban az önkifejezési célból létrehozott podcast-ok között keres, addig a *LearnOutLoud* (www.learnoutloud.com) inkább az információs és az oktatási jellegű felvételekre összpontosít, mint például a beszédközpontú rádiók és a hangoskönyvek. A blogok is az önkifejezés egyik új formáját jelentik, és természetesen ezeknek is megvannak a speciális keresők; közülük a legjobb a *Technorati* (technorati.com).

Képek esetében a Google és a Yahoo! képkeresői mellett ott van a *Ditto* (www.ditto.com), amely magát a legnagyobb vizuális tartalomkeresőnek definiálja. Hangokat és effekteket a *FindSounds* (www.findsounds.com) és a *Soundsnap* (www.soundsnap.com) segítségével találhatunk legkönnyebben: állathangok, repülő nyíl surrogása, nevetés, almaharapás hangja, szerszámok és járművek zaja, és még milliónyi egyéb hang között kereshetünk itt. És végül, ha világtérképre van szükségünk, akkor az egyik jó hely erre a *National Geographic* honlapján a *Map Machine* (maps.nationalgeographic.com/map-machine).

Bár most a Google dominálja az internetes keresést, de mindig jönnek új versenytársak újfajta megoldásokkal, ami kikényszeríti a folyamatos fejlesztést. Ez az állandó versengés különösen hasznos azoknak, akik az információkeresésből élnek.

/MAXYMUK, John: Searching beyond Google. = The Bottom Line: Managing Library Finances, 21. köt. 3. sz. 2008. p. 97–100./

(Drótos László)

Van-e optimális kategóriaszám a könyvtári honlapok forrásgyűjteményeinél?

A felsőoktatási könyvtáraknak komoly kihívást jelent, hogy hogyan vonzzák a honlapjukra a felhasználókat. Egy 2007-es amerikai felmérés [1] kimutatta, hogy a hallgatók közel háromnegyede a könyvtár helyett már inkább az internetet használja információkeresésre; nem mindegy tehát, hogy a könyvtári weboldalak mennyire felhasználóbarát felületet és tartalmat kínálnak. Ezeknek a honlapoknak az egyik tipikus szolgáltatása az internetes források tematikusan rendezett címgyűjteménye. A könyvtárosok által összeállított forráslisták eléggé hasonló témakörneveket használnak, ugyanakkor a témák száma jelentősen változik: néhol csupán csak öt kategória van, máshol viszont akár száznál is több. A cikkben ismertetett kísérlet során a szerzők arra próbálták választ kapni, hogy a tematikus kategóriák száma mennyiben befolyásolja a keresés hatékonyságát?

Szakirodalmi áttekintés

McGillis és Toms 2001-ben publikált kutatásuk [2] során azt találták, hogy a felhasználóknak gondot okozott megkeresni egy adatbázist a könyvtári honlapon való navigálással: nehezen tudták eldönteni, hogy a honlapon felsorolt kategóriákból melyiket válasszák, hogy eljussanak hozzá. Jeng 2006-os disszertációjában [3] egy sor tesztet elemzett a digitális könyvtárak használhatóságával kapcsolatban. Többek között az alábbi megállapításokat tette:

- A felhasználók „elveszetteknek” érzik magukat: nem látják át a digitális könyvtár struktúráját.
- Félreérthető terminológia: a szakszavak és a fogalmak helytelen megválasztása volt a felelős a használhatósági problémák 36 százalékáért.
- A folyóiratcikkek és a tematikus adatbázisok megtalálása komoly nehézséget okozott a felhasználóknak.

Más, nem könyvtári témájú publikációk is érdekesek a jelen kutatás szempontjából, mindenekellett Miller híres tanulmánya [4] a rövidtávú emberi memória 7 ± 2 tételes kapacitásáról. Ezt a felfedezést néha tévesen úgy interpretálják, hogy nem ajánlott 9 elemnél hosszabb menüket használni a weblapokon. Valójában a későbbi kutatások kimutatták, hogy a „tömbösítés” (*chunking*) módszerét használva – vagyis amikor az elemeket csoportokba foglaljuk úgy, hogy az ugyanazon csoporton

belüli elemek szorosabb kapcsolatban legyenek egymással, mint más csoportok elemeivel – a rövidtávú memóriával is sokkal nagyobb halmazt lehet kezelni annál, mint amit Miller kimutatott.

A menük hosszára és mélységére vonatkozó vizsgálatok némileg ellentmondásos eredményeket adtak. Larson és Czerwinski [5] például arra a következtetésre jutottak, hogy akár az egyes menüszintek kiterjedését, akár a szintek számát növeljük, a felhasználó reakcióideje, hibázási gyakorisága és komplexitásérzése egyaránt növekedni fog, ezért közepesen hosszú (8-9 elemű) és lehetőleg nem túl mély menük használatát ajánlották. Ugyanakkor Snowberry, Parkinson és Sisson [6] azt találták, hogy a tesztelt 2^6 , 4^3 , 8^2 és 64^1 struktúrák közül (ahol pl. a 2^6 egy kételemű és hatszintű menürendszert jelent) a kategóriákba csoportosított 64^1 bizonyult a legelőnyösebbnek mind a keresési sebesség, mind pedig a pontosság szempontjából. Zaphiris is megerősítette azt, hogy a weben minél mélyebb egy menürendszer, annál lassabban találja meg a felhasználó, amit keres. [7] Azt is megmérte, hogy a kinyitható menüknél a felhasználók válaszideje 50 százalékkal hosszabb, mint az egymást követő menük esetében. [8]

A következőkben ismertetett vizsgálat csak az egyszintes menükre terjedt ki, és azt próbálta meghatározni, hogy az ezekben felsorolt témakörök száma és jellemzői hogyan befolyásolják a keresés hatékonyságát.

A tesztelés ismertetése

A kutatáshoz 30 angolszász könyvtári weblapról választottak ki tematikusan rendezett forráslistákat, majd ezek számát 12-re szűkítették, melyek jól reprezentálták a többit is, egymáshoz eléggé hasonló témakör-megnevezéseket tartalmaztak, és méretük a kicsitől (5 téma) az egész nagyig (72 téma) terjedt. A tesztalanyokból véletlenszerűen ugyancsak egy tucatnyi csoportot alakítottak ki és minden csoport egy-egy ilyen tematikus listát kapott – kinyomtatva, ábécérendben és egységes külalakkal. A résztvevők 90%-a diák volt, a többiek tanszéki alkalmazottak. Mindegyiküknek ugyanazt a 11 kérdést tették fel, és az illetőnek ki kellett választania a kapott menülistából azt az egy menüpontot, amely megítélése szerint legvalószínűbb

ben elvezetné a válaszhoz. A kérdések az aztékokra, a gördeszkázásra, az ismétlődő mikrotraumára, a francia forradalomra, a farkasok visszatelepítésére, egy üzletlánc marketingstratégiájára, az atomenergiára, Emmanuel Kantra stb. vonatkoztak – vagyis a legkülönbözőbb témákra, amelyek egy könyvtári kutatás során előfordulhatnak. A tesztelést ketten felügyelték: egyikük feltette a kérdést és feljegyezte a megjelölt témakört, a másikuk pedig mérte azt az időt, amire a tesztalanynak szüksége volt a témakör kiválasztásához.

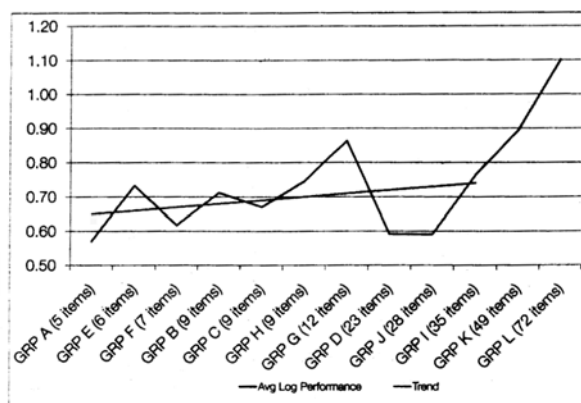
Előzőleg minden lehetséges kérdés-válasz kombinációhoz két könyvtáros relevanciaértékeket rendelt, vagyis eldöntötték, hogy egy adott tematikus listában egy-egy témakör mennyire releváns az egyes kérdések esetében. A vizsgálat során azt szerették volna megtudni, hogy a hasonlóan magas relevanciájú válaszokat összehasonlítva van-e különbség a válaszidőben a különböző tematikus listák között?

A vizsgálat eredményei

A mérési adatokat egyváltozós általános lineáris modellanalízis segítségével értékelték ki, külön-külön mindegyik kérdésre. A modellben a függő változó a válaszidő volt, a magyarázó változó (kovariáns) pedig a válasz relevanciája, vagyis hogy mennyire illeszkedett egymáshoz a kérdés és a válaszként megjelölt téma. (A „Gazdaság” témakör pl. igen magas relevanciaértéket kapott az aktuális munkanélküliségi rátára vonatkozó kérdéssel kapcsolatban, ugyanakkor egyáltalán nem volt releváns a Szaturnusz négy holdjának nevét firtató kérdés esetében.) Az analízis főfaktora az a témalista volt, amit a tesztalany használt. A statisztikai modell így foglalható össze: válaszidő = lista + relevancia + (lista x relevancia) + hiba.

Az általános lineáris modell használhatóságához két feltételnek kell teljesülnie: egyrészt az adatokat véletlenszerűen kell kiválasztani egy normális sokaságból, másrészt a csoportok varianciájának azonosnak kell lennie (homoszkedaszticitás). A tesztelés során mindkét feltétellel voltak problémák, ezeket a szokásos logaritmikus transzformációval mérsékeltek. A 11 kérdésből így már csak három nem felelt meg a követelményeknek, ezért az ezekre adott válaszokat a további elemzésből kizárták. A maradék 8 kérdés adatait feldolgozva és grafikonon ábrázolva – ahol a vízszintes tengelyen az egyes, változó elemszámú listák vannak, a függőleges tengelyen pedig a válaszidők logarit-

musa –, nem látható semmilyen trend, ami azt mutatná, hogy a listák elemszámának növekedésével a válaszidő is változik. Úgy tűnik, hogy inkább a kérdés és maguk a kategóriák befolyásolják a válaszidőt, és nem annyira a kategóriák száma. Egy 28 témakörös lista hasonló válaszidőt eredményezett ugyanazon a relevanciaszinten, mint egy 6 témakörből álló. Ha statisztikailag hasonlítjuk össze az egyes tesztcsoportok válaszidejének középértékét ugyanazon kérdéseknél, akkor is ezt kapjuk: nincs semmi jelentős eltérés közöttük. Ha viszont a 8 kérdés válaszidejét átlagoljuk, és ezt ábrázoljuk az egyes listák elemszámának függvényében, akkor némileg tisztább képet kapunk (1. ábra). A trendvonalat is megrajzolva azt látjuk, hogy a válaszidő fokozatosan növekszik az elemszám növekedésével nagyjából az 50 tételes listáig, majd ezután hirtelen megugrik.



1. ábra Az egyes tesztcsoportok 8 kérdésre adott átlagos válaszideje és a trendvonal (zárójelben a csoport által tesztelt témalista elemszáma)

A válaszidők és a relevancia között viszont semmilyen kapcsolatot nem sikerült kimutatni: a gyors válasz nem feltétlenül jelentett releváns választást, és ugyanígy: a lassú válaszidő sem hozott irrelevánsabb eredményeket. Kétféle viselkedésmintát lehetett megfigyelni. Voltak, akik gyorsan és könnyen döntöttek, miután átfutották a választható témaköröket, de ez nem mindig volt releváns választás. Más esetekben viszont a tesztalanyok elbizonytalanodtak, ha elsőre nem sikerült dönteniük, ilyenkor próbálták átértelmezni a kérdést, hogy egy megfelelő kategóriát tudjanak hozzá találni. Bár a szemmozgásukat nem regisztrálták műszerrel, de úgy tűnt, hogy elkezdik újra átfutni oda-vissza a listát, hogy találjanak egy illeszkedő kategóriát. Ez a fajta késlekedés két esetben fordult elő: amikor egyik témakör sem tűnt megfelelőnek, illetve amikor a listák kezdtek hosszúvá válni.

Feltételezhető, hogy ha a lista elemszáma túl nagyra nő, akkor ez az átfutás egyre jobban ellassítja a választási folyamatot.

A kategórianemek érzékelhetően nem befolyásolták a tesztalanyok teljesítményét. Három olyan témakörlista is volt, amelyek mindegyike 9 elemből állott, és ezek különböző elnevezéseket használtak ugyanazon témakörökhöz, de a válaszdők nagyon hasonlóak voltak mindhárom esetében. Ám elképzelhető, hogy egy későbbi vizsgálat, amely több, azonos elemszámú, de különböző témanemeket használó listát hasonlít össze, kimutathat majd eltéréseket a reakcióidőkben. Ugyancsak másfajta eredményeket hozhatnának más, esetleg szerencsésebben megválasztott tesztkérdések, mert bár az ismertetett vizsgálat során a teszt tervezői igyekeztek minél változatosabb témájú és eltérő nehézségi fokú kérdéseket feltenni, de végül a tizenegyből hármat el kellett vetniük, mert nem feleltek meg a statisztikai modell követelményeinek.

Irodalom

- [1] JONES, Steve: The Internet Goes to College. Ed. Mary Madden. Washington, D.C.: Pew Internet and American Life Project, 2002. p. 3.
- [2] MCGILLIS, Louise – TOMS, Elaine G.: Usability of the Academic Library Web Site: Implications for Design. = *College & Research Libraries*, 62. köt. 4 sz. 2001. p. 361.

- [3] JENG, Judy H.: Usability of the Digital Library: An Evaluation Model. PhD dissz., Rutgers University, New Brunswick, New Jersey, 2006. január, p. 38–42.
- [4] MILLER, George A.: The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information. = *Psychological Review*, 63. köt. 2. sz. 1956. p. 81–97.
- [5] LARSON, Kevin – CZERWINSKI, Mary: Web Page Design: Implications of Memory, Structure and Scent for Information Retrieval. Los Angeles, ACM/Addison-Wesley, 1998. p. 25
- [6] SNOWBERRY, Kathleen – PARKINSON, Mary – SISSON, Norwood: Computer Display Menus. = *Ergonomics*, 26. köt. 7. sz. 1983. p. 705.
- [7] ZAPHIRIS, Panayiotis G.: Depth vs. Breath in the Arrangement of Web Links.
<http://www.soi.city.ac.uk/~zaphiri/Papers/hfes.pdf> (letöltve: 2007. nov. 1.).
- [8] ZAPHIRIS, Panayiotis G. – SHNEIDERMAN, Ben – NORMAN, Kent L.: Expandable Indexes Versus Sequential Menus for Searching Hierarchies on the World Wide Web
<http://agrino.org/pzaphiri/Papers/expandableindexes.pdf> (letöltve: 2007. nov. 1.).

/MILES, Mathew J. – BERGSTROM, Scott J.: Classification of library resources by subject on the library website: Is there an optimal number of subject labels? = *Information Technology and Libraries*, 28. köt. 1. sz. 2009. p. 16–20./

(Drótos László)

Digitális gyűjtemények és adatok felügyelete: „kurátorság” és más alternatívák a tudományos könyvtárakban

A digitális anyagok sérülékenyek. Bár ez a nyomtatott dokumentumokra is elmondható, de azok helyes kezelésének módszereit kidolgoztuk az elmúlt évszázadok alatt, míg a digitális információ megőrzéséről sok mindent meg kell még tanulnunk. A „digitális kurátorság” egyre ismertebb fogalom: azoknak a tevékenységeknek a gyűjtőneve, amelyek a digitális adatok és tudományos eredmények karbantartására és használhatóságának megoldására irányulnak, azok teljes életciklusa alatt, a jelenben és a jövőben egyaránt. Magában foglalja a digitális archiválás és megőrzés folyamatát, de az adatok létrehozásának és menedzselésének helyes gyakorlatára is kiterjed, valamint azokra az értéknövelő eljárásokra is, melyekkel új információ és új tudás állítható elő. A hollandiai *Digitális Megőrzés Koalíció* koordinátora cikkében

azt vizsgálja, hogy a tudományos könyvtárak milyen típusú digitális adatokat kezelnek, és hogy ezeket hogyan igyekeznek megőrizni:

- átadják-e ezt a terhet az erre szakosodott intézményeknek;
- vagy maguk vállalják fel a digitális megőrzés felelősségét, de a tényleges munkát kiszervezik és mással végeztetik el;
- vagy pedig teljesen magukra vállalják a feladatot és saját digitális repozitóriumot építenek ki az intézményen belül?

Elektronikus publikációk: a raktározástól a hozzáférésig

A digitális kor egy sor új kihívást hozott a szakkönyvtárakba, ezek egyike az áttérés a nyomtatott

könyvekről és folyóiratokról az elektronikus változatokra. Ez a váltás másfajta beszerzési gyakorlatot eredményez, például licenctárgyalásokat – jellemzően nagybani – előfizetésekről. És változik a hozzáférés módja is, mert többnyire az interneten át, online érhető el a tartalom. Ezáltal szinte észrevétlenül eltűnt a könyvtárak egyik alapfeladata: az információ raktározása. A könyvtárak számára meglehetősen nyugtalanító érzés, hogy nem birtokolják többé a tudományos publikációkat, hanem meg kell bízniuk a kiadókban, illetve a közvetítőkből, hogy betartják a szerződés „folyamatos hozzáférés”-re vonatkozó pontját. A kiadók pedig egyre inkább a szaporodó digitális őrzőhelyekkel szerződnek, mint amilyen például a Holland Nemzeti Könyvtár e-Depot szolgáltatása, vagy a Portico és a LOCKSS; gyakran akár többel is, hogy az eltérő megőrzési technikákkal minél jobban csökkentsék az adatvesztés kockázatát. Más esetekben törvényi szabályozás alapján történik a digitális dokumentumok biztonságba helyezése – többnyire a nemzeti könyvtárakban.

Leginkább jól szerkesztett, PDF formátumban tárolt publikációkról van szó, melyek archiválása és a jövőbeli hozzáférés biztosítása nem különösebben nehéz műszaki feladat. Emellett az ilyen elektronikus kiadványok raktározása nem is elsősorban a tudományos könyvtárak felelőssége, hanem a kiadóké és a digitális megőrzéssel foglalkozó nemzeti vagy nemzetközi szervezeteké.

Helyi digitalizált gyűjtemények

A megőrzendő anyagok másik csoportját a digitalizált gyűjteményrészek jelentik. Mivel a könyvtárak elkezdtek digitalizálni állományuk egyes darabjait, hogy így könnyebben hozzáférhetővé tegyék őket az olvasóknak, logikusnak látszik, hogy egyúttal azt a felelősséget is felvállalták, amit ezeknek a – sokszor jelentős költséggel létrehozott – számítógépes állományoknak a hosszú távú megőrzése jelent, hogy ezek a jövő generációi számára is használhatók maradjanak. Ám valójában ez nincs mindig így. A cikk szerzője számos olyan előadást hallott az éves LIBER-konferenciákon, amikor az előadó ezzel zárta a mondandóját: „Digitalizáltuk a gyűjteményt és ezáltal megőriztük a jövőnek. Köszönöm a figyelmet!” Az élet azonban nem ilyen egyszerű. Nem lehet eleget hangsúlyozni, hogy a digitalizálás nem ugyanaz, mint a digitális megőrzés. Ha valamit számítógépes formára alakítunk, az csak az első lépés – keletkezik egy digitális

objektum, ami ki van téve a jól ismert veszélyeknek:

- a hordozóeszköz tönkremenetele,
- a hardver vagy a szoftver elavulása,
- a szervezet megszűnése, és végül, de nem utolsósorban:
- emberi hibák.

Senki nem ismeri a végleges válaszokat ezekre a problémákra és valószínűleg nem lesz tökéletes megoldás később sem. Viszont a kurátorság már most is képes minimalizálni a kockázatokat egy olyan mozgó környezetben, ahol a digitális adatok folyamatosan özönlenek és a számítástechnika pedig állandóan változik. A gyakorlatban egy biztonsági mentés házon kívüli tárolása megfelelő kockázatcsökkentő megoldásnak látszik, de valójában csak egy óvintézkedés, ami nem nyújt védelmet az itt felsorolt veszélyek mindegyike ellen. Ha azt akarjuk, hogy a digitalizált anyag fennmaradjon, akkor folyamatosan gondozni kell. Itt nem létezik a „jótékony hanyagság”, mint a nyomtatott művek világában, amikor is örökre eltűntnek hitt antik könyvek csodálatos módon egyszer csak újra felbukkannak poros padlásokon. A digitális információ életben maradása teljesen a megfelelően működő hardver- és szoftverkörnyezetre függvénye. Az *Ausztrál Nemzeti Archivum* ezért is vezette be a „teljesítőképesség-mutató” (performance model) fogalmát arra az esetre, amikor egy digitális objektumot elő kell hívni: objektum+(hardver+szoftver)=teljesítőképesség. Ennek a „teljesítőképességnek” a megvalósítása többet jelent annál, mint hogy biteket tárolunk egy szerveren. Olyan szervezetet igényel, amely nemcsak arra ügyel, hogy ezek a bitsorozatok ne sérüljenek, hanem folyamatosan figyeli a világban zajló technológiai változásokat, amelyek hatással lehetnek a megőrzendő anyagra és olyan stratégiákat dolgoz ki, amelyekkel megoldható, hogy ezek a digitális objektumok a következő generációs számítógépeken is megjeleníthetők legyenek. Ilyen jól ismert stratégiák például a migrálás, az emuláció/szimuláció és a normalizálás – ezek mind meglehetősen bonyolult műszaki eljárások. Kulcsrakész és mindenre kiterjedő termékek még nem léteznek a piacon ezekre a feladatokra. Vagyis: minden olyan könyvtárnak, amely saját maga akarja felügyelni és megőrizni a digitalizált gyűjteményét, bizonyos fokig késznek kell lennie arra, hogy részt vegyen az alkalmazott műszaki megoldások továbbfejlesztésében is, különösen, ha egy már meglévő hozzáférési rendszert szeretne fenntartani.

Jó hír, hogy a saját digitalizált gyűjteményrészeknél rendszerint csak néhány fájlformátum fordul elő (PDF, JPG és TIFF), melyek kurátorsági szempontból viszonylag könnyen megőrizhetők hosszabb távon is. Egy másik fontos tényező, amely megkülönbözteti ezeket a digitális anyagokat a következő fejezetben tárgyalt tudományos kutatási adatoktól, hogy ezeket rendszerint maguk a könyvtárak hozzák létre, így feltehetően megfelelően strukturált és részletes metaadatok is tartoznak hozzájuk, ami megkönnyíti a velük való munkát és a megőrzést. De ezen könnyebbségek ellenére a hosszú távú megőrzés ebben az esetben is komoly vállalkozás egyrészt a költség miatt (ami nem annyira a tárolóeszközök árát jelenti, mert azok egyre olcsóbbak, hanem a szervezeti háttér fenntartását), másrészt a szükséges szakértelem miatt (mert bár már egyre több üzleti vállalkozás is megjelenik ezen a területen, az archívum üzemeltetése és összekapcsolása a könyvtárban működő információs rendszerekkel, sok helyi és számítástechnikai ismeretet igényel). Kérdés ezek után, hogy érdemes-e a tudományos könyvtáraknak felvállalni ezt a feladatot és felelősséget, különösen most, hogy a költségvetés egyre szűkebb és a helyzet a belátható jövőben sem lesz jobb? A nemzeti könyvtáraknak nem nagyon van más választási lehetőségük: többnyire törvény írja elő, hogy a nyomtatott és a digitális anyagok őrzőhelyként kell működniük. De a szakkönyvtáraknak több alternatívájuk is van:

- Egyszerűen házon belül tárolják a digitális gyűjteményüket és remélik a legjobbakat. Talán ez a leggyakoribb megoldás, de kockázatos. Viszont ha az eredeti állomány épségben marad, akkor lehet azzal érvelni, hogy hosszú távon olcsóbb lesz újradigitalizálni egyes tönkrement fájlokat, mint gondosan őrizni a teljes digitális gyűjteményt. Ám ez csak akkor fogadható el, ha valóban a veszélyek alapos megfontolása után döntöttek így; vagyis például garantálni tudják, hogy a fizikai hordozókon levő eredeti dokumentumok sértetlenek maradnak a jövőben.
- Keresnek egy külső intézményt, akár egy nemzeti szervezetet, akár egy fizetős céget, és ott tárolják a digitális állományt. A tudományos könyvtárak egy részének ez járható út lehet, különösen amelyek túl kicsinek tartják magukat ahhoz, hogy saját digitális raktárt üzemeltessenek. Fontos viszont, hogy valóban megbízható őrzőhelyet találjanak, olyat, amely tényleg képes a szükséges felügyeletet ellátni, és amely gondoskodni tud arról, hogy a tárolt anyag integrálható legyen a könyvtári rendszerbe. Az adattárolásra szakosodott vállalkozások valójában gyakran csak biz-

tonsági mentést képesek nyújtani, miközben hosszú távú őrzőhelyként hirdetik magukat. (Hasonló ez ahhoz, mint amikor úgy reklámoznak egy optikai tárolóeszközt, hogy ezer évig is megmaradnak rajta az adatok – ami jól hangzik, de ez semmi védelmet nem jelent a hardver és a szoftver elavulása ellen. Hiába maradnak rajta a bitek ezer évig, ha nem lesz olyan számítógép, amely fel tudná dolgozni őket.)

- A harmadik lehetőség egy saját digitális archívum kiépítése. Ez a leginkább nagyratörő és mélyreható választás a lehetséges alternatívák közül.

Digitális kutatási adatok

Miközben a szakkönyvtárak megszabadultak a kiadói e-folyóiratok őrzésének felelősségétől, egy másik feladat jelent meg náluk: a kutatási adatok tárolása és gondozása. Ez igazi kihívás, tekintettel arra, hogy a tudományos adatok a legkomplikáltabb formáját jelentik a megőrzendő digitális információknak. Mind az adatok, mind az előállítók, rendkívül változatosak, és olyan komplex digitális objektumokat is tudni kell kezelni, mint az „élő” adatbázisok. Ugyanakkor a Google-szerű vállalkozások korában, amikor úgy tűnik, hogy a szakkönyvtárak lassan elveszítik az összes tradicionális életcéljukat, itt egy lehetőség arra, hogy új, mással nem pótolható helyet vívjanak ki maguknak – egyenesen az oktatási vagy kutatási anyaintézményük információs rendszerének közepében.

Jelenleg elég kevés egyetemi könyvtárnak van olyan digitális repozitóriuma, amely kellően feltöltött ahhoz, hogy valódi segítséget jelentsen a tudományos kutatásban és az oktatásban. Mindazonáltal a szakkönyvtáraknak legalább három olyan alapvető sajátosságuk van, amelyek különösen alkalmassá tehetik őket a tudományos kutatási eredmények kurátorságára:

- a küldetésükben benne van a hosszú távú megőrzés,
- strukturális támogatást kapnak,
- részei a tudományos közösségnek.

Igaz, hogy a második érv megkérdőjelezhető, mert a könyvtárak finanszírozása ritkán elégséges, viszont ők legalább költségvetésbe beépített támogatást kapnak. Mióta a tudományos kutatás projektrendszerben és határozott időre szóló pénzügyi támogatásokkal folyik, a folyamatos finanszírozás hiánya miatt a kutatás során összegyűjtött adatok gyakran odavesznek, miután a projektek lezárul-

nak. Ennél még a könyvtárak szerény, de legalább biztos költségvetése is nagyobb garanciát jelent a hosszú távú megőrzéshez.

Ami azt az előnyt illeti, hogy a könyvtárak részei a tudományos közösség hálózatának: ez nem vehető biztosra. Ehhez a könyvtári és adatkezelési szolgáltatásokat speciálisan az adott kutatói kör igényeire kell(ene) szabni, és elsődleges prioritást adni az ő szükségleteiknek és kéréseiknek. Nem véletlen, hogy a digitális megőrzés területén leginkább használt Open Archival Information System keretrendszer kidolgozói kiemelt helyet és külön szakkifejezést rendeltek a „célközönséghez”.

Ezek a magas elvárások arra a döntésre is vezethetnek, hogy a tudományos könyvtár végül is nem száll be ebbe a folyamatba. Főleg, hogy a könyvtári szféra még inkább csak küzd a digitális világ különböző kihívásaival, miközben egyes tudományágak kutatói idejekorán felismerték a digitális jövőt és ügyesen megszervezték a maguk információs hálózatát, akár európai, sőt akár világméretben is. Ezek a nagy hálózatok általában jól megalapozottak és az adott tudományág területén dolgozók igényeire vannak optimalizálva. Így a tudományos könyvtáraknak nem sok szerep jut bennük, főleg mert a legnagyobb problémájuk a stabil és jelentős méretű pénzügyi háttér megszerzése, amiben a könyvtárak nem tudnak segíteni.

Tanácsok

Befejezésül néhány gyakorlati, „tedd” és „ne tedd” típusú jó tanács azoknak a könyvtáraknak, amelyek azon gondolkodnak, hogy hogyan kezeljék a digitális gyűjteményüket és az általuk szolgált kutatók digitális információkra irányuló igényeit:

- Találjuk meg a „célközönségünket”.
- Üljünk le, írjunk egy tervet, és alakítsuk ki a célközönségünkkel kapcsolatos politikánkat. Sokszor már ennyi is elég, hogy felismerjük az erősségeinket és gyengeségeinket, és helyes sorrendbe tegyük a prioritásokat.
- A tervben vegyük számba azokat a digitális gyűjteményeket, amelyeket gondoznunk kell, és próbáljuk felbecsülni, hogy milyen problémákat okozhatnak a belátható jövőben.
- Beszéljünk a célközönségünk tagjaival az információs igényeikről és építsük be őket a tervünkbe.
- Selejtezzünk, és azután selejtezzünk újra. Bár néhány archiváló szakember még mindig az „őrizzük meg mindent” elvet hirdeti, mert „kik vagyunk mi, hogy eldöntsük, mire lesz szükség a jövőben?”, de a legtöbben már egyetértenek abban,

hogy „mindent” megőrizni nem lehetséges és nem is célszerű.

- A méret számít. Az első digitális objektum, amit a kurátorságunk alá veszünk, rendkívül költséges lesz, de a milliomodik már csak fillérekbe kerül majd. Ha nem vagyunk elég nagyok ehhez, akkor válasszuk a következő pontot.
- Keressünk partnereket, lehetőleg hozzánk hasonló területen és mérettartományban. A gyakorlat azt mutatja, hogy az együttműködés leginkább a hasonló partnerek között jön létre, és egyre nehezebbé válik, minél távolabb vannak a szervezetek egymástól a hagyományaikat és a céljaikat tekintve.
- Válasszunk egy szakértői hálózattal rendelkező ernyőszerületet. Ez lehet regionális, országos vagy nemzetközi.
- Kényeztessük el a célközönségünket. Engedjük, hogy a kutatók kutatók maradhassanak, például ne erőltessük rájuk a mi metaadat-sémáinkat, hanem inkább igyekezzünk nekik olyan adatkezelő eszközöket és megoldásokat kínálni, amiket könnyen be tudnak építeni a munkafolyamataikba.
- Döntsük el, hogy miben vagyunk jók, és ennek alapján keressük meg a helyünket az adatkuratori szervezetek hálózatában.

De ha ez nem megy, akkor legalább ennyit tegyünk meg:

- Soha ne tároljunk fontos digitális információt floppy lemezen, CD-n, vagy lokális PC-n, hanem töltsük fel megbízhatóbb hardverre.
- Rendszeresen mentsük az adatainkat és állapotunk meg egy kollégánkkal, hogy vigyázunk egymás back-up állományaira.
- Készítsünk listát a felügyeletünk alá tartozó digitális anyagokról és tartsuk naprakészen.
- Keressünk megbízható őrzőhelyet az adatainknak. Ez lehet a nemzeti könyvtár, vagy egy nemzeti archiváló szervezet, vagy akár egy olyan kolléga, aki digitális repozitóriumot üzemeltet. Ha nem áll módunkban hivatalosan meggyőződni arról, hogy egy archiváló hely mennyire megbízható, nézzük meg az intézmény egészét és tegyük fel magunknak a kérdést: valószínű, hogy ez a szervezet még ötven vagy száz év múlva is létezni fog?
- Az erőforrásainkat a hozzáférés kérdésére összpontosítsuk, mert hosszú távon csak az fog számítani. Minden más csak eszköz ahhoz, hogy a hozzáférés lehetőségét megőrizzük.

/ANGEVAARE, Inge: Taking care of digital collections and data: 'Curation' and organisational choices for research libraries. = LIBER Quarterly, 19. köt. 1. sz. 2009. p. 1–12./

(Drótos László)