

Egy webkereső és partnerei: a Google és a könyvtárak Hálóhelyzet

A keresőmotorok között éles ma a küzdelem. A számítástechnikai óriások, a behemót nagyágyúk mellett fürge, dinamikus „kiságyúk” is ringbe szállnak, hogy minél hangsúlyosabbá tegyék jelenlétüket az interneten. A hálózati kereséssel korábban nem igazán foglalkozó Microsoft erőteljes kampányba kezdett teljesen újraírt, *MSN-Search*-ként ismert programjának (<http://search.msn.com>) reklámozására.

A Yahoo! 2005 elején rukkolt ki *Y!Q* kódjelű új típusú, kontextus-érzékeny keresőjével, amely közvetlen kijelöléssel történő keresésnél bármely weben olvasott szövegből az adott szövegekörnyezet gyorslemezével hoz létre releváns tételekből álló találati halmazt (<http://yq.search.yahoo.com/search>). Az amazon.com környékéről friss hír, hogy saját keresőmotorján (<http://a9.com>) cégek böngészésekor az utcai üzletportálok fényképe is megjelenik, virtuálisan sétálni lehet előttük. Hasonló vállalat- és szolgáltatáskeresést kínál a Google *Local* (a magyar verzióban a neve: Helyi keresés) – itt az áruházak, üzletek, éttermek helymeghatározása nem fényképekkel, hanem rendkívül pontos földrajzi lokalizálással történik, térképen végzett, tízes skálájú közelítéssel.

A Vivisimo cég új, okos webkeresője, a *Clusty* (<http://www.clusty.com>) a mesterséges intelligencia eredményeit is fölhasználva, nevéhez híven klasztereket képez a visszakeresés során, és a találati halmaz képernyőjén szemantikai jellegű csoportokba rendezi a találatokat. (A *Clusty* tulajdonképpen metakereső, azaz más keresőmotorok eredményeivel végez finomításokat.) A *Jux2* szintén metakereső, a profiljában megadott keresőkérdést lefuttatja a Google-lal, a Yahoo!-val és

az AskJeeves-zel, és egybeveti a – tanulságos – eredményeket.

A hálót szövő pókok, a „bot”-oknak is nevezett robotok bejárják a world wide web, a világszéles világ(pók)háló szegleteit, és rendszeresen begyűjtik a honlapok adatait, amelyek aztán a cég adatbázisában várják a „kérőket”, a keresőket, hogy adott pillanatban találati halmazba tömörülve képernyőre kerüljenek a világ valamely zugában. A keresőkérdés szűkítése és szűrése, definíció, rövid magyarázat kérése (pl. define wikipedia), a nyelvi szűkítés (pl. language:hu) vagy az egy honlapon való keresés (pl. url:www.ifla.org), illetve az a megjelenítés, hogy a direkt link mellett a keresőmotor cache-ében eltárolt változatra is rá

lehet klikkelni, ma a legtöbb helyen alapszolgáltatásnak számít. A keresőgépek sémája, felépítése az első, 1994-ben megépített World Wide Web Worm (WWW) óta nemigen változott, ám az ördög a részletekben lakozik, és a ke-

resési folyamatban használt algoritmusok féltve őrzött titkok, ipari kémek valódi csemegéi.

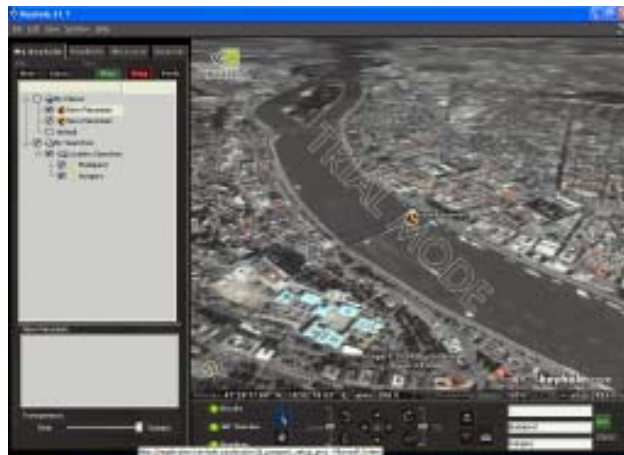
A keresőmotorok élmezőnyének az élén a kaliforniai Mountain View-ban székelő Google áll – s minthogy a két alapító egyike, *Sergey Brin* orosz származású, akár úgy is fogalmazhatunk, hogy a Google „minden keresőgépek cárja”. A kibernetet éjszakánként pásztázó csápjaival 8 milliárd weblapot gyűjt be, naponta 200 millió keresést kezel, ami a világ összkeresésének 57 százaléka. (A Google mögött a Yahoo! 21 százalékkal, az MSN pedig 9 százalékkal részesedik.) Bár egy friss szakmai felmérés (<http://brandchannel.com>) alapján a márkák elismertségében két éven át őrzött vezető helyét 2004-ben elvesztette az Apple ellenében, az 1998 októberében alapított Google ismertsége folyamatos, szolgáltatásainak rendszeres fejlesztése látványos és lenyűgöző. Megkülönböztetett figyelem övezte piaci vállallattá válását. A Google tőzsdére kerülése hosszú menetelés volt, 2004 őszére fejeződött be, ám bánkódásról szó sem volt, hiszen a két alapító egy csapásra bekerült a világ ötven leggazdagabb



embere közé. A már tekintélyesen hangzó Google, Inc. szolgáltatási palettájáról nézzünk néhány példát, hogyan is differenciálódik egyre – bár nem árt azért újra emlékeztetni, hogy a szolgáltatások jó részét más keresők kínálatában is fellelhetjük.

A Google palettája

A Google *Answers* a LibInfóhoz hasonló, de fizetős információszolgáltatás, ahol megajánlott összegért gondosan kiválasztott szakértők százai válaszolnak a kérdésekre. A Google *Directory*ban (magyar neve: Címtár) hierarchikusan témák szerint, emberi közreműködéssel előre rendezett menükben navigálhatunk. A Google *Groups* (Hírcsoportok) az évekkel korábban beintegrált, több évtizedre visszamenő Usenet szakmai levelezőlisták aktualizált mutatója, linkgyűjteménye. A Google *Images* (magyar neve: Képkereső) 800 millió képből válogathat. A *Froogle* (játékos és célzatos névválasztás, hiszen kiejtése egyezik a 'frugal', fukar szóval) az okos vásárlás weblapja. A Google *News* béta verziója hírportálnak indul, ahol a keresőkérdésre rímelő legújabb híreket is be lehet hívni. A Google *Scholar*, amely pár hónappal ezelőtti megjelenésével megriasztotta a tudományos adatbázisok forgalmazóit, kutatók és tudósok számára kínál specifikus keresést. Bár szűk háttéranyagból dolgozik, a megjelenítés a Google-nél megszokott világos és áttekinthető, a szócikkek hivatkozásai is azonnal átnézhető találati halmazként. 2004 áprilisában a Google elindította saját webalapú levelezőjét, az egy gigabájt tárolóhelyet biztosító ingyenes *Gmailt* (<http://gmail.google.com>). A Google *Suggest* prognosztizál, beírásakor már megpróbálja kitalálni a keresőkérést, és ahogy haladunk előre a kérdés leírásában, úgy módosulnak a javasolt találati halmazok a legördülő menüben. A Google ingyenes, némi képmanipulálást is megengedő fényképrendező rendszerének, a *Picasa Photonak* 2.0-s verzióját (<http://www.picasa.com>) nemrég hozták ki; a weben való közzétételt lehetővé tevő kapcsolódó alkalmazással (<http://www.hello.com>) karöltve azonnali siker, világsláger lett. 2005. január 25-én kezdte meg pályáját a Google *Video*, amely az amerikai tévéprogramok keresését teszi lehetővé (<http://www.google.com/video>). A videó azért ma még túlzás, hiszen a találatra klikkelve csupán képeket, képsorokat tudunk egyelőre megjeleníteni, de a következő logikus lépés a cég ré-



szerűl a valódi videoanyagokhoz való hozzáférés. Annál is inkább, mert például a Yahoo! hasonló szolgáltatása kapásból tartalmazza a videó letöltését, megjelenítését, lejátszását. A Google *Desktop* pótalkalmazás, plug-in a Windows-hoz, a Sajátgép fájljainak kitűnő keresőeszköze. A Google nemrég indította be a *Keyhole*-t, egy 2004 októberében megvásárolt kis helyi cég termékét. A *Keyhole* már ízig-vérig XXI. századi alkalmazás – csak nézünk, mint a moziban, vagy kukucskálhatunk a kulcslyukon. A szatellita eredetű képek, geográfiai szoftverek segítségével az egész Föld felszínének bármely pontjára fókuszálhatunk, és háromdimenziós képet kaphatunk. A drágább professzionális verzióban egészen méteres távolságig rázoomolhatunk bármely felszíni tárgyra, az olcsó, harminc dolláros verzióban 300 méter távolságból tudunk közelíteni akár saját házunkra. Mindezek után méltán kérdezhetik, miért e tolongás, hol ebben az üzlet, amikor például a Google szolgáltatásainak túlnyomó többsége ingyenes. Valóban, ami bevallottan a bevételt hozza, az – a reklám. A hirdetések ügyes, nem zavaró, releváns elhelyezését a Google két programja, az *AdSense* és az *AdWords* oldja meg. (Az *AdWords*-ben kínált reklámlehetőség esetén például a cégnek csak akkor kell fizetnie, ha rákattintanak a linkjére.)

A Google nagy durranásainak sorában a legnagyobbat az a 2004. december 14-ei bejelentés szolgált, amely szerint a cég az elkövetkező években öt nagy könyvtár több tízmilliós gyűjteményének digitalizálására vállalkozik. A hír bombaként hatott a könyvtárosokra is, a világon mindennél szalagcím volt, a New York Times cikkének címe szerint a „Google nagy könyvtárakkal bővíti saját adatbázisát”. A globális virtuális könyvtár megvalósulását fedezték fel a bejelentés mögött,

digitális kártyakatalógusról, a Google saját Dewey tizedes rendszeréről beszéltek. Az on-line szakmai hírlevél, a KIT másnapi, december 15-ei száma már szintén hozta a hírt, ilyen címmel: Könyvtárakat pakol fel a netre a Google (<http://www.gmconsulting.hu/inf/kit/index.html>).

A Kataliston nyomban élénk vita alakult ki a projektről, épp akkor, amikor a digitalizálásról, pontosabban a tankönyvek digitalizálásának jogi vonatkozásairól is már egy ideje zajlott a polémia.

Bacsa András 2005. január 8-ai levelében digitalizálási szakértőként aggályait fogalmazta meg – szerinte a projekt technikailag megoldhatatlan és minőségileg csak gyatra eredményt hozhat.

Drótos László szerint a Google „nyilván nem azt célozta meg, mint nálunk például a DIA (SGML/XML karakteres formátum, többszörös korrektúra, esetleg plusz szöveggondozás), hanem egy végletekig automatizált eljárást dolgozott ki a metaadatok katalógusból való átvételével”. *Moldován István* is az utóbbi véleménnyel ért egyet: „A Google nem fog »pepecselni« kritikai kiadásokkal, korrektúrával, mutatórendszerrel, hanem olyan technológiát használ, amellyel gyorsan, rövid idő alatt nagy tömeget lehet kielégítő minőségben digitalizálni.”

Király Péter így ír: „...amit a Google célul kitűzött, és amit esetleg vernegyulaizmusnak gondolunk, jelentős részben már meg is valósult, a nagy mennyiség nem jelent feltétlenül rossz minőséget...”. (A leveleket lásd a Katalist archívumában, a 2005. január hónap alatt: <http://listserv.iif.hu/Archives/katalist.html>.)

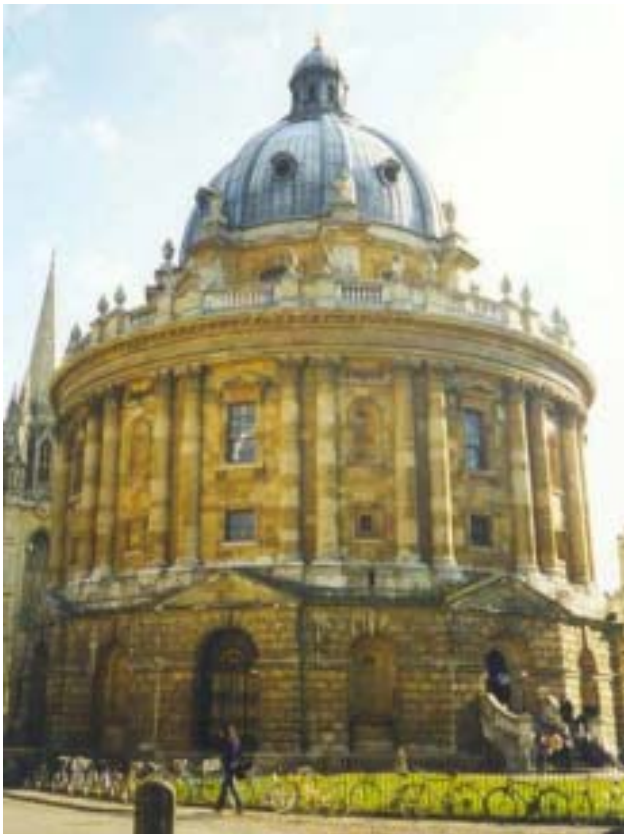
Az egyik hozzászóló sommás véleménye szerint a „»digitalizáljunk minden könyvet« a köznép kezébe adott, szép nagyra felfújt lufi!” Felfújt lufinak talán még viccből is erős nevezni egy tervet, amelyhez öt óriásgyűjtemény, egytől-egyig világszám, önként és dalolva csatlakozik – bár az is örök igazság, hogy még a nagyok is tévedhetnek, és nem baj, ha „van kéznél” némi egészséges szkepszis. Magából a tényből annyit azért bátran levonhatunk, hogy bizonyára profitálnak a végtermékből a könyvtárak is. Gondoljuk meg, a 2004. december 14-én világgá kürtölt projekt nem futó szeszélyből fakadt. Könyvtömegek digitalizálásának gondolata igazából még a Google gründolását is megelőzi, hiszen a két alapító a Stanfordon épp egy miniatűr könyvtári digitalizációs projekten dolgozott, amikor 1999 elején a

keresőmotor létrehozása vontta el attól *Larry Page* és *Sergey Brin* figyelmét. Idézzük pontosan, mit mondott a sajtótájékoztatón Page: „Még mielőtt elindítottuk volna a Google-t, arról ábrándoztunk, hogy a hihetetlen tömegű információt, amelyet a könyvtárosok oly szerető gondossággal szerveznek, on-line módon kereshetővé tegyük. Ma örömmünkre szolgál a program meghirdetése, amelynek keretében ezeknek a bámulatra méltó könyvtáraknak az állományát digitalizáljuk és szolgáltatjuk minden Google-használó javára. A könyvtárakkal való együttműködésünk kibővítése a meglévő Google *Print* programnak, amely lehetővé teszi, hogy a használó a könyvek teljes szövegében lokalizáljon egy adott szövegrészt. A Google küldetése a világinformáció organizálása, és izgalmasnak találjuk, hogy könyvtárakkal dolgozhatunk együtt ennek a küldetésnek a beteljesítésében.” Mindössze négymondatos nyilatkozat, s hemzseg benne a 'könyvtár' és a 'könyvtáros' pozitív értelmű kontextusban. Valljuk meg, nem így beszél egy könyvtárgyilkos technokrata, s nem is hízelgés mondatja vele e szavakat.

Az említett Google *Print* 2003 őszén indult, és bár máig béta verziónak van elkönyvelve, sikeres és jelentős program. A kiadókkal való széles körű együttműködés folytán – ez a program azóta is bővül újabb és újabb kiadók, többek között a HarperCollins, a Houghton Mifflin, a Scholastic és a Random House csatlakozásával – a Google a kiadók által felajánlott könyveket teljes terjedelemben beszkeneli, és az adott könyv szerzői jogi státusát figyelembe véve a halmazokban linket ad meg vagy a teljes szöveghez, vagy csak szövegrészekhez, kivonatokhoz és metaadatokhoz. A Google *Print* modellként szolgál arra, hogyan fog működni a keresőmotor a könyvdigitalizálási anyagokkal való gazdagodás után (<http://print.google.com/googleprint/about.html>). A Google digitalizálási projektje az öt könyvtár esetében a technológia azonossága mellett más és más kondíciókat tartalmaz, az egyes kétoldalú megállapodások eltérnek egymástól. Érdekes körülnézni a partnerek háza táján a sajátosságok felkutatására.

Oxford

Az Oxfordi Egyetem a Bodleiana gyűjteményével „nevezett be”. Sem az egyetem, sem a Bodleiana nem éppen parvenü intézmény. Oxfordban már 1096-ban folyt oktatás, de ugrássze-



rű fejlődésnek akkor indult, amikor II. Henrik 1167-ben megtiltotta az angol diákoknak, hogy a párizsi egyetemen tanuljanak. A Bodleiana Sir Thomas Bodley üdvös tevékenysége révén lett 1602-ben nyilvános könyvtár – bár igazából ez már az 1320-as években létrejött könyvtár újraalapítását jelentette I. Erzsébet uralkodásának utolsó évében. A Bodleiana gyűjteménye a British Library után a második legjelentősebb Albionban. Csaknem négy-száz éve (!) kötelezpéldány-könyvtár, a brit szigetek hat hivatalos letéti központjának egyike. Francis Bacon egy 1613-as mondása szerint a Bodleiana „Noé bárkája, amely megmenti a tudást az özönvíz-től”. Feltétlenül említsük meg nagyhírű tudós hazánkfiát, a nagykőrösi születésű Uri Jánost (1724–1796), aki a Bodleiana könyvtárosa és oxfordi tanár lett, 26 évig élt Oxfordban, ott is halt meg. Goldziher Ignác írt róla monográfiát 1908-ban (Értekezések a nyelv- és széptudományok köréből XX/7.). A Hollandiában bölcsész- és teológiai doktorátust szerzett Uri 1770-ben került Angliába; kifejezetten azért hívták meg, hogy leírja, rendszerezze és katalógusba foglalja a Bodleiana keleti kézíratait. Ezt a nyelvek tucatjának beható ismeretét kívánó munkát másfél évtizeden át végezte, 1787-ben nyomtatásban meg is jelent a katalógus. A *Bibliothecae Bodleiana Codicum manuscriptorum orientalium... Catalogus elkészítésével* Uri elnyerte a britek örök

háláját, hiszen megismertette őket a keleti kultúrákkal, s ez nem kis mértékben megkönnyítette világszerte terjeszkedésüket a következő évszázadban. Ilyen pazar, minket magyarokat is érintő gyűjteményről van tehát szó. A szóban forgó Google-projektben a Bodleiana sajátossága, hogy csak 1900 előtti könyveket digitalizáltak (<http://www.admin.ox.ac.uk/po/041214a.shtml>).

NYPL

Az 1895-ben három kisebb gyűjteményből alapított könyvtár 1911-ben nyitotta meg kapuját Manhattan szívében, az Ötödik sugárút és a 42. utca sarkán álló hatalmas márványépületben.



Ahogy a Fővárosi Szabó Ervin Könyvtár is szét-feszíti a 'közművelődési könyvtár' kereteit és tudományos könyvtárként is a főváros egyik büszkesége, a csaknem kétmillió beíratkozott látogatóval rendelkező New York Public Library hasonlóképpen jelentős kutatási központ, az USA vezető könyvtárai között van. *Paul LeClerc*, a NYPL elnöke szerint a Google-lal való együttműködés kitágítja a gyűjtemény horizontját és hozzáférhetőségét. A 89 fiókkönyvtárral rendelkező NYPL 50 milliós állományából 20 millió a könyv. A NYPL sajátossága, hogy kizárólag szabad hozzáférésű, copyright-mentes könyveket ad át digitalizálásra. A program befejezésével a könyveket nemcsak a Google-tól, hanem a NYPL-ből is el lehet majd érni. Ez módosítás a Google Printhez képest, s a könyvtári elérést hivatott biztosítani (<http://www.nypl.org/press/google.cfm>).

Harvard

A Harvard Egyetemet 1636-ban alapította a Massachusetts Bay Colony kormányzata, nevét John Harvard lelkészről kapta, aki első adományozóként könyvtárát és birtokát hagyta az új intézményre. Az Egyesült Államok legtekintélyesebb egyeteme, hét elnök (köztük J. F. Kennedy és George W. Bush) volt a Harvard diákja. A Harvard Egyetem és a Google partneri együttműködéséről, amely egyelőre kísérleti, ún. pilot-program, az egyetemi lapban olvashatunk. Az első fázisban mindössze 40 ezer könyvet szkennelnek – a külső raktárból (Harvard Depository – HD) szedik ki a műveket. *Sydney Verba* könyvtárigazgató szerint nagy a valószínűsége, hogy a teszt után a teljes állományra is sor kerül. A Harvardon a Widener Könyvtárban ülő használó a



HOLLIS-nevű OPAC-ban – <http://holliscatalog.harvard.edu> – kikeresi a könyv raktári jelzetét. Ha a könyv külső raktári, a HD-ban van tárolva, akkor nem kell behozatni, hanem a digitalizált változattal azonnali hozzáférést tudnak nyújtani (<http://www.thecrimson.com/article.aspx?ref=505139>).

UMich

A Michigan Egyetem Könyvtára az USA hatodik legnagyobb gyűjteménye. Larry Page egyébként a közelben született, és az egyetem Ann Arborban levő kampuszán diplomázott – ez is nyomhatott a latban a partnerek kiválasztásánál. (Az sem volt hátrány, hogy az egyetem élenjáró a digitalizálásban, és a Library of Congress American Memory című gyűjteménye mellett a UMich kilencezer kötetes Making of America című on-line archívuma a legjelentősebb történelmi forrásanyag.) A UMich a teljes anyagát digitalizálni fogja. A munka már szép csendben folyik, hónapok óta. A hétmilliós állományt hat év



alatt fogják digitalizálni. (Az egyetem jelenlegi saját digitalizálási tempója mellett ez a munka kb. 1600 évig tartana.) Teljes üzemben a napi mennyiség 3600 könyv lesz, ma még átlagban 2,25 perc alatt készül el egy könyv. A Google helybe hozta a technológiát, és a szkennelési műveleteket végző munkaerőről is gondoskodik.

A könyvtárosok készítik elő a könyveket és szabják meg a szkennelésre kerülő könyvek sorrendjét. Az egyetem tárolni fog egy példányt a szkennelt anyag minden egyes darabjából (<http://www.michigandaily.com/vnews/display.v/ART/2004/12/14/41becb60d0cab>).

Stanford

A Szilíciumvölgy kellős közepén fekvő Stanford Egyetemet 1891-ben alapította Leland Stanford, a transzkontinentális vasútépítés egyik emblematikus figurája, korábbi kaliforniai kormányzó. Az elnököt, Nobel-díjasok seregét, sőt két neves PhD-diákot, Page-t és Brint is adó intézmény, amely az Internet születésében is kulcsszerepet játszott, az USA egyik vezető oktatási-kutatási központja. A New York Times még 2004 februárjában beszámolt egy Stanford-Google közös, Project Ocean kódjelű tervről. Mint most kiderült, ez a terv már a digitalizálási megaprogram előjátéka volt, a részleteket sikerült csaknem egy évig titokban tartani. A Stanford saját ütemezésében – *Andrew C. Herkovic* osztályvezető könyvtáros szerint – először a public domain állományt, az 1923 előtti műveket, aztán az egész gyűjteményt fogják beszkenneálni. A Stanford sincs híján digitalizálási tapasztalatoknak, a kampuszon már jó ideje üzemel egy egymillió dollárt súroló árfekvésű szkennelő robotgép (<http://news-service.stanford.edu/news/2005/january12/google-0112.html>). A Stanford Report honlapjának jobb olda-



lán található az a link Stanford, Google collaboration will have far-reaching impact címmel, amely videofilmen mutatja be az olcsónak semmiképp nem mondható robotgép működését (<http://www.techan.ca/mt/paul/>).

A digitalizálás menete

A projektben alkalmazásra kerülő digitalizálási technológia bár számos, máshol kifejlesztett elemet tartalmaz, összességében a Google saját fejlesztése. A digitalizálás költségeiről annyi közvetett forrásból elhangzott, hogy tíz dollárba kerül egy könyv, ám más digitalizálási szakértők szerint ennél nagyobb fajlagos költséggel kell számolni. Hogyan zajlik a munka? Ennek mozzanatait a Michigan példáján mutatjuk be. A Google egy műhelyt állított fel az egyetemi kampuszon. A munka egyelőre manuális, de a nagyobb védelmet igénylő könyveket később is manuálisan fogják szkennelni. (Ritka könyveket – mint pl. a Gulliver első kiadása – egyáltalán nem szkennelnek.) A könyvtárosok a polcra a világszerte elterjedt könyvtári kocsira teszik a könyveket, vonalkódolvasóval leolvassák a vonalkódot, amivel úgymond „kikölcsönzik” a könyveket, és átszállítják az egyetemen létesített Google-műhelybe, ahol a szkennelés történik. Ez a kölcsönzés időtartamban olyan, mint az éjszakai vagy hétvégi, mert nagyon hamar visszakerülnek a polcra a könyvek. A vonalkódbeolvasás úgy van beállítva, hogy egyúttal a könyv OPAC-ban szereplő rekordja átkerül a Google-hoz, ahol a szkennelt szöveghez linkelik a bibliográfiai tételt. A folyamat további részében történik az optikai karakterfelismerés (OCR) és az indexelés.

Fontos körülmény, hogy a művek tulajdonjoga természetesen marad az egyetemé, és ez azt is jelenti, hogy ahogy az amerikai jogban a 'first sale' elv alapján a megvásárolt példányból a könyvtár szolgáltatathat, az 1923 utáni, tehát javarészt szerzői jogi védelem alatt álló műveket is teljes szöveggel, korlátozás nélkül rendelkezésre tudják bocsátani helyi használatra.

A digitalizált anyag megjelenítése

A könyvön belüli keresés nem új. Az Amazon például több százezer könyv digitalizálása és indexelése révén könyvek belsejében szavakra kereshetünk azoknál a találatoknál, ahol a kis ikon

megjelenik (Search Inside the Book). Előhívhatók mintaoldalok, a borító, a címlap, a verző, a hátsó borító és a tartalomjegyzék. Az amazon.com kereskedelmi webhely, ahol új vagy használt könyv vásárlását ösztönzik a bibliográfiai adatok és a mintaoldalok megmutatásával. A Google-ban ez a szempont ha nem is elhanyagolható, mégis számos alternatívát kínálnak fel. A keresési kifejezésből a halmaz tetején a Google „Book results” címmel külön halmazt készít, amennyiben a keresési kifejezés szerepelt az indexelt könyvben. Ha a könyv szerzői jogi védelem alatt áll, a keresési eredmény oldalán mintaoldal látható, elő lehet hívni néhány oldalt, a tartalomjegyzéket, kivonatokat, bibliográfiai adatokat. Itt is van vásárlási lehetőség: „Buy the book”-link mutat az amazon.com-hoz, a Barnes and Noble virtuális könyvesbolthoz és a Google saját áruházához, a Froogle-hoz. Ráklíkelhetünk még a használt és forgalomban nem kapható könyvek on-line áruházának, az Alibrisnek a honlapjára is.

Egészen eredeti a könyvtári lelőhelyhez mutató link. A „Find this in a library” keresőablakába be kell írni az irányítószámot, amelynek alapján a Google előhívja, melyik közelben lévő könyvtárban van meg a keresett könyv. A szabad, copyrighttal nem védett könyvek, kormánykiadványok stb. teljes szöveggel olvashatók lesznek. A még szerzői jogi védelem alá eső művek digitalizálásával kapcsolatban annyit meg kell jegyezni, hogy a könyvkiadóknak vannak kételyeik. Bár a Google csak korlátozott hozzáférést fog nyújtani az ilyen művek esetében, ám az tény, hogy teljes műveket kíván beszélni a jogtulajdonos, legtöbb esetben a kiadó engedélye nélkül, s ez egyes kiadói megnyilatkozások szerint összeütközésben áll a szerzői jogi törvényekkel.

A digitalizációs projekt járulékos haszna lehet – amire elsőre nem nagyon gondol az ember – a megőrzés kérdése. A digitális fájlok megőrzésének történelmi súlyú megállapodása alapján 2002 nyarától az Elsevier kiadó a Holland Királyi Könyvtárba helyezi el elektronikus anyagait megőrzésre. Ezzel egybecseng, hogy az öt könyvtár és a Google abban állapodott meg, hogy nem a kereskedelmi vállalat, jelen esetben a Google, Inc., hanem a nagy könyvtárak gondoskodnak a fájlok megőrzéséről. A Michigannél az eredeti fájlokat arany CD-ROM-ra mentik – ezek olyan, arannyal „felturbózt” eszközök, amelyek tűrése állítólag több száz év (erre szoktuk mondani, majd meglátjuk...). A Stanfordin mágnesszalagos mentést javasolnak, mindenből három példányt készítenek, és megszervezik a folyamatos frissítést és karbantartást. Az egész anyag nagyságrendje végül

is a petabájt-tartományba (1 petabájt = 1000 terabájt) esik.

Felmerült a könyvtári állományokban meglévő idegen nyelvű és idegen jelkészletű könyvek kérdése – a projekt nem fogja kikerülni a nem angol nyelvű könyveket sem. Mivel ismert, hogy a Google például kínai nyelven is szolgáltat, semmi gondot nem jelent a digitalizálási munka számára a kanji vagy az arab karakterek felismerése. Ezek szerint bízhatunk benne, hogy a magyar nyelvű könyveket sem fogják félretenni. Kovács Ilona áldozatos kutatása szerint a NYPL hungarica-anyaga mintegy 20 ezer kötetet tartalmaz. A többi részt vevő könyvtárban is jelentős magyar nyelvű anyag lehet, durván 100 ezerre becsülöm. Tiszta haszon, ha ezt az anyagot – ingyen – digitalizálni fogják. Megjegyzem: aránylag sok a hazai magyar könyv ezekben a könyvtárakban. Mindegyikük sok évtized óta az MTA Könyvtárának stabil cserepartnerre: ennek a viszonyoknak is köszönhető, hogy hungaricumok ezrei lapulnak az ottani polcokon.

Más projektek

A Google és az öt könyvtár digitalizálási gőzhengere tehát elindult. Voltak jelek arra, hogy a projektből kimaradt könyvtárak közül több zokon vette a mellőzést. Ám nem sokáig, mert egyre-másra kerülnek napvilágra partneri együttműködésben tervezett egyéb könyvtárdigitalizálási projektek. 2004. december 29-én a Brewster Kahle által 1996-ban alapított Internet Archive (<http://www.archive.org>) nemzetközi kezdeményezésében tíz nagy könyvtárral abban állapodott meg, hogy a nonprofit cég szabad hozzáférésű on-line archívumot hoz létre a könyvtárak gyűjteményéből szelektált könyvek digitalizálásával. A projektben részt vesz a pittsburgh-i Carnegie Mellon Egyetem Könyvtára és a Library of Congress – aporként saját, bámulatos digitális gyűjteményüket hozzák be: a CMU a már futó Million Book Projectjét, a LoC pedig a talán legismertebb digitális kincsestárat, az American Memoryt. További szereplők: a Torontói Egyetem, az Ottawai Egyetem és a McMaster University könyvtárai, egy kínai egyetemi könyvtár, az indiai Indian Institute of Science és az egyiptomi Alexandria Könyvtár. Az Internet Archive Text Archive elnevezésű projektje, amely nyílt hozzáférésű, Open Access archívumnak készül, a tagok gyűjteményéből millió fölötti könyvet fog digitalizálni, 50 ezret már 2005 első felében.

Bánhegyi Zsolt