

Burai István

Az internetes információkeresés jövője és a könyvtárak

E rövid írásommal az volt a célom: bepillantani az információkeresés közeli és kicsit távolabbi jövőjébe, és annak könyvtárakra vonatkozatható hatásaiba. Nem alkalmas a terjedelem miatt áttekintést nyújtani minden, a szakterületet érintő hatásról, azaz még véletlenül sem merészelném hiánytalannak vélni (erre talán a terület szaktekintélyei sem vállalkoznának). Feladatom a megfelelő kérdések megfogalmazása, illetve azok előkészítéséhez némi információ megosztása az Olvasóval. Határozottan gondolatébresztő, vagy ha úgy tetszik, vitaindító céllal született tehát ez a cikk.

A felvetés indoklásaként talán elegendő annyi, hogy a XXI. század világában az információ mint erőforrás értékesebbé vált, mint valaha. Jelentősége elérte, vagy talán meg is haladta a nyersanyagok vagy a közlekedés fontosságát, az üzleti, politikai, katonai, társadalmi, kulturális stb. élet szinte minden területén a legmeghatározóbb tényezővé vált. Ez az írás – talán elfogadható módon – a mai világban fellelhető legfontosabb ismerethalmaz, az internet és az információkeresés jelenét és közeljövőjét kutatja, és szerzője ki kíván térni mindezek könyvtárakra gyakorolt hatására, különös tekintettel a XXI. századi könyvtárhasználók igényeire.

Az információkeresés fontosságát jól mutatja, hogy a legjelentősebb innovációk az internetet érintően pontosan ezen a területen zajlottak az utóbbi években. Talán az sem véletlen, hogy az internettel foglalkozó IT-cégek közül a legjelentősebbek is konkrétan erre a területre specializálódtak. Az is egyértelműnek látszik, hogy az internetes innovációk közül kettő emelkedik ki, és válik meghatározóvá: a keresés és a közösségi szolgáltatások. (Meglátásom szerint ez utóbbi is jelentős részben visszavezethető az információkeresés területére, azaz a két terület nem válik el igazán egymástól. Sőt!). A terület maga pedig olyan sebességgel változik, fejlődik, hogy igazán pontos előrejelzést ma senki nem képes adni.

Nem kétséges, hogy globálisan tekintve az elektronikus dokumentumok aránya a nyomtatottakéhoz képest folyamatosan és megállíthatatlan módon nő, még akkor is, ha egyébként a nyomtatott dokumentumok kiadásának száma a XXI. század elején elérte csúcspontját. Tehát a relatív elmozdulás már érzékelhető. Mint ahogy beszédes az a tény is, hogy a világ egyik legnagyobb kiadóvállalatának, az Amazonnak az elmúlt időszak forgalmát tekintve a digitális tartalmak (e-book) eladásban elhagyták a nyomtatott termékekét. Az eleve az interneten létrejövő tartalmak, valamint a digitalizálás útján oda felkerülő (a céges, pl. Google-books¹ – illetve a

különbféle nemzeti, pl. a francia Gallica² – és közösségi, pl. az Europeana³ – projekteknek köszönhetően egyre nagyobb számban) tartalmak mellett ezeknek is mind jelentősebb szerepük lesz. Ez az exponenciálisan gyarapodó adat- és információhalmaz állítja kihívás elé az ebben keresni, eligazodni vágyókat.

Miért olyan nagy kihívás ez a terület, ha csak a keresés problematikáját tekintjük? Már pusztán a méret, a mennyiség megsejtése – mert valódi méreteit felfogni, sőt megbecsülni is több mint merész, lásd a deep-webet érintő kutatásokat; egyes eredmények szerint ugyanis a webkeresők által nem látható web mérete 2–5000 szerese az azok által (és így általunk) közvetlen elérhető tartalomnak!⁴ – is elegendő ahhoz, hogy elhiggyük, a keresés két legfontosabb minősítő jelzője, a teljesség és a relevancia körében a létező legnagyobb kihívások elé néz bármifajta fejlesztés ezen a téren.

A ma legelterjedtebb keresők a tankönyvek definíciója szerinti teljességre törekvő keresési metódusokat támogatják, de ez megtévesztő megközelítés. Nem nevezhetünk csak azért egy találathalmazt a kérdést tekintve teljesnek, mert több tízezer potenciális találatot adott. A keresők által el nem ért információk magas, illetve a releváns információk – a teljes találati számhoz viszonyítva – meglepően alacsony száma mutatja, hogy a jelenleg elterjedt és népszerű rendszerek gondokkal küzdenek. (Nekem néha úgy tűnik, hogy jelenleg még inkább – egy hasonlattal élve – az „eredeti információfelhalmozás” korát éljük, mint a hatékony keresés korát... Igazolja ezt az, hogy minden, ezen a területen tevékenykedő cég a legváltozatosabb információkhoz kíván hozzájutni, maga alá gyűjteni, lásd pl. a Street-View, kapcsolati-közösségi hálók, e-mail címek stb.)

Mindezek után nem véletlen, hogy a jelenleg elterjedt indexeléses és algoritmusokkal rendezett keresési eljárások mellett komoly kutatások folynak olyan (pl. szemantikus) keresők kifejlesztésének az irányába, amelyek már nem az információ vélhetően helyes forrásához irányítanak, hanem magát az információt, az adatot adják válaszul a feltett kérdésre. Nem feledhetjük, hogy a jelenlegi keresők a relevancia biztosításának kérdésében többségében statisztikai eszközöket alkalmaznak. Ezen megoldások „jósági foka” természetesen javul a feldolgozott információ mennyiségének az exponenciális növekedésével, de nem érheti el az adatszolgáltató rendszerek hasonló adatát. A relevancia és a teljesség mellett másik fontos tényező a hitelesség, amelynek biztosítása fontos, ám jelenleg komoly hiányosságok mutatkoznak körülötte. A webet alkotó adatmennyiség tekintetében talán ez biztosítható a legnehezebben. De vannak biztató fejlesztések: egyszerre mindhárom feltételt kívánja teljesíteni egy innovatív fejlesztés, a WolframAlpha⁵. A készítői által „computational knowledge engine”-nek nevezett-kategorizált tudásrendszer valóban nem a klasszikus keresők képét mutatja, sokkal inkább hasonlatos egy adatszolgáltató rendszerhez, mint klasszikus internetes keresőkhöz. Ugyanis nem a kognitívumokhoz vezeti el a keresőt (jelen esetben nem is pontos a „kereső” kifejezés, talán helyesebb a „kérdező” szó használata), hanem a feltett, természetes nyelven beírt kérdést, stringet matematikai algoritmusokkal közvetlen értelmezve, magát a választ próbálja megadni. Ilyenformán talán jelen pillanatban a szemantikus keresők egyik legjelentősebb képviselője, még akkor is, ha a felépített (és ellenőrzött tartalmú!) tudáshalmaz a teljességet még csak meg sem közelíti.

Hogyan lehetne ezeket a képességeket kiterjeszteni az internet egészére? Vagy másképpen feltéve a kérdést: hogyan lehetne létrehozni a szemantikus webet⁶? Az

bizonyosnak látszik, hogy egy lépésben ez a feladat nem megoldható, semmilyen módszer vagy eszköz nem áll rendelkezésre ehhez.

De érdekes módon bizonyos lehetőségek kihasználatlanul állnak, és már régóta alkalmazhatóak lennének. Ilyen pl. a W3C által már 1998-ban elfogadott XML specifikáció, amelyben lehetőség van a weboldal szövegének leíró adatokkal, címkékkel való ellátására. Továbbá az említett WolframAlpha is igazolja, hogy a szemantikus rendszerek másik két kulcsterületén, a „mesterséges intelligencia” kutatásban és az ontológiák alkalmazásában is sikerült jelentős fejlődést elérni. Továbbá az eddig el nem érhető tartalmak hozzáférhetővé tétele szempontjából komoly lehetőségeket rejt még pl. a P2P⁷ technológia és a keresők ötvözése. Szintén a hatékonyságnövelő megoldások közé sorolható a Google elképzelése, egyfajta „statisztikai intelligencia” alkalmazásba vétele. A közel végtelen mennyiségű információ statisztikus elemzésével valóban el lehet érni egy olyan tudásbázis-minőséget, amely már mutatja a szemantikus web bizonyos jellegzetességeit.

Itt szükséges megjegyezni, hogy a közeljövő keresőrendszereinek egy igen fontos képességet kell biztosítaniuk, ez pedig a géppel kérdezhetőség feltétele. Problematikus terület ez, és nem is technológiai akadályokkal tűzdelt, hiszen a különféle keresési szolgáltatók üzleti érdekei az együttműködés ellen szólnak. Jól példázza ezt a közelmúlt egyik eseménye, miszerint a Google kitiltotta adatai lekérdezéséből a Facebookot, a kölcsönösség hiányára hivatkozva, bár korábban is csak egyoldalú volt a lehetőség, a Google soha nem jutott hozzá a Facebook adataihoz.

Összefoglalva: az internetes keresés jövője szükségszerűen a szemantikus, tudásalapú, továbbá hagyományos indexelős (hiszen a webet alkotó oldalak nagy része nem fog megváltozni azonnal) eljárások rugalmas, egymást kiegészítő, előnyöket, felületet ötvöző alkalmazásában keresendő.

Mégis, hol van ebben a rendszerben a könyvtárak szerepe? Ki az, aki a közeljövő hálózatalapú társadalmában megcélózható, aktív felhasználója lehet a jövő könyvtárainak?

A munkahelyemen, a Méliusz Juhász Péter Megyei Könyvtárban, Debrecenben végzett statisztikai vizsgálatok azt mutatják, hogy a könyvtárba látogatók immár nagyobb hányada nem konkrét dokumentumért, hanem információért jön a könyvtárba.

A tendenciákat figyelve (amelyek az IT-infrastruktúra magyarországi otthonokban való elterjedtségét mutatja) nem kétséges, hogy a közeljövő felhasználója a számára szükséges információk döntő többségét az internetről fogja beszerezni. Becslésem szerint ennek bekövetkeztéig pedig ténylegesen nem több, mint háromöt év van hátra.

Hogyan őrizhetők meg ezek az olvasók a könyvtárak számára? Csak olyan módon, hogy a könyvtárak mennek el hozzájuk. Az olvasói igények kielégítése csak a könyvtárak előzőekben vázolt rendszerekbe való integrálásával képzelhető el. Az XML technológiák alkalmazására már ma is van lehetőség. További képességeket biztosítanak a metakeresők alkalmazásba vétele, helyi metaadatbázisok építése, a lokális rendszereken túl kistérségi, regionális metaadatbázisokba való integrálás, egységes keresési portálok létrehozása. Kiemelt szerepet kell kapjon a könyvtárak különgyűjteményeinek elérhetővé tétele ugyanezen rendszereken keresztül. A nyugat-európai Dublin Core, OAI (Open Archives Initiative) alapon

működő információs rendszerek jó példák lehetnek ennek kiépítésére. De ezek is továbbfejlesztésre szorulnak! El kell érni, hogy a „Google-Facebook-generáció” is elérhesse a könyvtárakban felhalmozott tudás nagy részét. És mindezt úgy kell kidolgoznunk, hogy közben a világ legnagyobb nyereségérdekelte cégeivel kell versenyeznünk a felhasználók figyelméért. Ebben a versenyben pedig sem az adatmennyiségben, sem az frissességben nem lehetünk nyertesek; egyedül a tartalomban. A releváns, máshonnan be nem szerezhető tartalomban. De abban is csak akkor, ha a jelenlétünk az adott felületen egységes, teljes és folyamatos. Mindez pedig nem biztosítható más módon, csak a szemantikus rendszerek által támasztott igényeket kielégítve, a meglévő rendszereink metakeresőkbe való integrálásával, tömegében is jelenlős szolgáltatások működtetésével.

Úgy gondolom, a könyvtárak jövőjének érdekében ezekre a kérdésekre mihamarabb és eredményes válaszokat kell találnunk.

JEGYZETEK

- 1 Lásd: <http://books.google.com/>
- 2 Lásd: <http://gallica.bnf.fr/>
- 3 Lásd: <http://www.europeana.eu/portal/>
- 4 Lásd még: http://www.technet.hu/hir/20101026/a_web_sotet_oldal/
- 5 Lásd: <http://www.wolframalpha.com/>
- 6 Lásd: <http://www.w3.org/2001/sw/>
- 7 Lásd: <http://hu.wikipedia.org/wiki/Peer-to-peer>

FELHASZNÁLT IRODALOM

- Bánhegyi Zsolt: Egy webkereső és partnerei. A Google és a könyvtárak. = Könyvtári Levelező/lap, 2005. 2. sz. 23–29. p.
- Benediktsson Dániel: Az információ értékelése mint ökológiai jellegű folyamat. 2. rész. = Könyvtári Figyelő, 2003. 1. sz. 89–101. p.
- Gottdank Tibor: Szemantikus web. Bevezetés a tudásalapú internet világába. Budapest, ComuterBooks, 2005.
- The Horizon Report 2010 edition. / The New Media Consortium. Internetes dokumentum. Lásd: <http://www.nmc.org/pdf/2010-Horizon-Report.pdf>
- Koprowski Gene J.: The Future of Human Knowledge: The Semantic Web. Internetes dokumentum. Lásd: <http://www.technewsworld.com/story/31199.html>
- THESEUS – New technologies for the Internet of Services. Internetes dokumentum. Lásd: <http://www.theseus-programm.de/en-us/about-theseus/default.aspx>
- Tikk Domonkos (szerk.): Szövegbányászat. Budapest, Typotex, 2007.
- Ungváry Rudolf–Orbán Éva: Osztályozás és információkeresés: Kommentált szöveggyűjtemény. 2. köt. Az információkeresés elmélete. Budapest, OSZK, 2001. 239–529. p.
- Ungváry Rudolf–Vajda Erik: Könyvtári információkeresés. Budapest, Typotex, 2002. 102–115. p.