

IOEHLSCHLÄGER, Susanne-ALTENHÖNER, Reinhard-GÖMPEL, Renate-HENGEL, Christel-JAHNS, Yvonne-JUNGER, Ulrike-MAHNKE, Christel-WERNER, Claudia: Weltkongress Bibliothek und Information, 72. IFLA-Generalkonferenz in Seoul, Korea. Aus den Veranstaltungen der Division IV

Bibliographic Control, der Core Activities ICABS und UNIMARC sowie der Information Technology Section. = Bibliotheksdienst, 40. köt. 10. sz. 2006. p. 1127-1148./

(Papp István)

A speciális keresés segédeszközei

Egy szűkebb szakterületen belül végzett keresés (*niche search*) egyre inkább előtérbe kerül, ahogy a keresőmotorok túl nagy merítésű találati halmazai nem felelnek meg a használó speciális igényeinek. Célzott keresési lehetőséget kell ma már neki nyújtani. Ez része az ún. „*Long Tail effektusnak*”, a hosszú távú felülkerekedésnek, amelyet *Chris Anderson a Wired magazin* rovatában és könyvében fejtett ki: a web által inspirált gazdasági modellben a kis volumenű, élesen fókuszált üzlet maga alá gyűrheti a nagykutyaikat. Az információkeresésben ugyancsak ez a trend kezd kibontakozni. A három behemót keresőmotor (Yahoo, Google, MSN) – gyakran a szakértők által végzett Boole-i keresés ellenére – sokszor nem elég hatékony találati listát hoz képernyőre. Speciális, élesen körülírt keresést végző kis cégek már eddig is léteztek – méregdrágán. Szolgáltatásuk fokozatosan feltárul és elérhető a széles nagyközönség számára, korábban zárt adatbázisok anyagához rajtuk keresztül hovatovább ingyen is hozzá lehet jutni.

ACCOONA

<http://www.accoona.eu>

Az ingyenesen elérhető speciális keresőmotor neve a szuahéli *Hakuna Matatától* származik, a maszajok egyik kedvenc mondása „accoona”, magyarul: „rá se ránts, ne aggaszd magad” – a használónak küldött pozitív üzenetként is felfogható. Az *Accoona* felhőtlen előnyei közé tartozik, hogy kifejezetten európai irányultságú keresőgép, a keresési folyamatban úttörő módon alkalmazza a mesterséges intelligencia technológiáját, s van egy dedikált *Dunn & Bradstreet (D&B)* üzletjelentési funkciója is.

A 2004 februárjában alapított *Accoona* a *China Daily Information Company*val közösen került piacra, európai megjelenése előtt már jelen volt a kínai

és az amerikai piacon. Az egyetlen kereső, amely a helyi eredményeket hét nyelven (angol, holland, francia, német, olasz, portugál, spanyol) képes megadni, s már a keresőkérdésből érzékeli, milyen nyelvet használunk. Ezt egy szabadalmaztatott felismerőprogrammal végzi, amely aztán az eredményeket is a nyelvhasználat alapján rendezi. Mindez a mesterséges intelligencia markáns szoftverbéli alkalmazásáról árulkodik. Szinte megérti a keresőkérdést, hiszen az Oscar-nyertesekre való kereséskor rákeres az összes szinonimára – például „*Academy Award winners*” – is. A keresés finomítására saját eljárást, egy *SuperTarget Your Search* nevű programot használ, amely mintegy „hab a tortán”. A finomítás a publikálás dátuma, kiadó, illetve az egyes keresőszavak súlyozásával is végezhető.

Az üzleti világban az *Accoona*t professzionális keresőgépnek tekintik, amelyben szabad hozzáférés kínálkozik több mint 25 millió európai vállalat adataihoz a *D&B* üzleti adatbázisán keresztül. Ahol a *D&B* ikon megjelenik, ott egyetlen kattintással képernyőre hozhatjuk az adott cég gyorsprofilját. Egyéb üzleti adatok azonban már csak pénzért kaphatók: a szabványos vállalati profil elérése 4, az üzleti háttérjelentés 39, a hitelezési jelentés 59, a részletes átfogó jelentés pedig 129,5 USD-be kerül.

GOOGLE SCHOLAR

<http://www.scholar.google.com>

A *Google Scholar* hamar népszerűvé vált azzal, hogy szabad hozzáférést nyújt szakmai bírálatban részesült tudományos cikkek, könyvek, műszaki jelentések tömegeihez. Érzékeny hiány azonban, hogy a találati halmazban nem szerepelnek az *Elsevier Kiadó* által publikált anyagok, kétségtelesen azért, mert a kiadónak megvan a saját ingyenes tudományos keresőszolgáltatása, a *Scirus*

(l. később). A Google Scholar jó szolgálatot tesz a könyvtáraknak is, hiszen a halmazban szereplő könyvrekordoktól – az *OCLC Open Worldcat* adatbázisban szereplő állományrekordok segítségével – közvetlen ugrópont mutat az adott régió belül a könyvet birtokló könyvtárhoz. A keresőmotor nagy aduja mégiscsak az, hogy használatával esetenként hozzájuthatunk az előfizetéses adatbázisokban pénzért kapható cikkek ingyenes változataihoz, preprintgyűjteményekben rejtőző szabad elérésű változataihoz. Az összetett keresés (*Advanced Scholar Search*) lehetőséget nyújt nemcsak tematikai szűkítésre, hanem szerzői és megjelenési idő szerinti keresőoperátor használatára is. A szerzői keresőminta akkor fontos igazán, amikor a szerzői név megegyezik egy köznévvvel, és így a rengeteg hibás találatot eleve ki lehet zárni.

A pontatlanság vádja azért merülhet fel a Google Scholar esetében, mivel a Google nem hajlandó közétetni azoknak a tudományos periodikumoknak a listáját, amelyek pártázásával az adatbázisát építi – s a frissítési gyakoriságról sem lehet tudni semmit. A „cikket idézte” funkció azonban, amely a citációs indexek mintájára az eredeti cikket idéző dokumentumokat jeleníti meg, ezt a speciális keresőgépet a gyors és hatékony kutatás remek eszközévé avatja.

ROLLYO

<http://www.rollyo.com>

Az ingyenes *Rollyo* a *Roll Your Own* szavak játékos rövidítése, magyarul talán így adható vissza: „indítsd a saját keresődet”. Ám itt nem a meta-keresők gyakorlatáról van szó, amelyek szimultán körkeresést végeznek egy tucat kiválasztott keresőmotorral, hanem a speciális webhelyeknek a használó által kijelölt csoportjára leszűkített keresésről. A személyre szabott keresés száma elvben végtelen lehet, sőt, a „privát keresőgépet” közzé is lehet tenni, és megosztani másokkal: a *Rollyo* erre a keresődarabkára is saját terminust alkalmaz – *Searchroll* (keresőhenger). A keresőhenger szerkeszthető és módosítható, mások keresőhengerét is saját beállításra alakíthatjuk. A keresés tartományát, a speciális webhelyeket az adott keresőhengerben nem tudjuk megváltoztatni, de a szerkesztéssel új, saját keresőnként menthetjük el. A keresőhenger egyik problémája azonban a tartományi korlát, a maximum 25 webhely. Bár ez elégnek is tűnhet, hiszen a *Rollyo* a hatékonyság maximális

kihasználásához 5–15 webhely kiválasztását javasolja, a tesztelés során több keresőhengert kellett létrehozni még egyetlen tárgyon belül is, majd az eredményeket összesíteni – ami az eredeti célt, az egyszeres szűkített mélykeresést végül is nem tudta teljesíteni.

Másik probléma, hogy a keresett oldalakat lehet ugyan definiálni, ám hogy melyik keresőmotorral végezzük a keresést, azt nem – a valódi személyre szabásnak erre is ki kellene terjednie. Doménszintű mélységi korlát is van, hiszen a *www.cnn.com* meghatározható keresési tartománynak, de a *www.cnn.com/politics* már nem, mivel a perjel utáni részt a rendszer nem veszi figyelembe. A *Rollyo* mögött a *Yahoo* keresőmotor működik, a blog-eredményeket pedig a *Feedster* szolgáltatja.

SCIRUS

<http://www.scirus.com>

A 2001 áprilisában indított *Scirus*, a tudományos információterjesztésre koncentráló ingyenes keresőmotor ma a *Google Scholar* legkomolyabb riválisa. Az Elsevier Kiadó tulajdonában lévő *Scirus* hatalmas előnye a konkurenciával szemben, hogy lefedi a csaknem kétezer Elsevier-folyóirat teljes tartományát – amelyből a *Google Scholar* mindössze csipegetni tud. A 250 millió tudományos weboldal mellett hozzáférése van 15 millió szabadalomhoz Európából, az USA-ból és Japánból, 212 ezer teljes szövegű cikkhez az *Institute of Physics Publishing* állományából, 6,6 millió teljes cikkhez a *Science Direct*-ből, 240 ezer disszertáció teljes szövegéhez a *Networked Digital Library of Theses & Dissertations (NDLTD)* adatbázisból stb. Nem egyszerű szépséghiba viszont, hogy a keresés ingyenessége gyakran nem eredményez ingyenes hozzáférést, hiszen Elsevier-folyóiratból csak az előfizetőknek nyújtanak teljes szöveget. Mindazonáltal a *Google Scholar*-nál fejlettebb keresési interfészt ad: találati szűrést lehet alkalmazni ISSN, szerzői munkahely, webforrás, illetve információtípus (cikk, disszertáció, kivonat, szabadalom stb.) szerint is.

TECHNORATI

<http://www.technorati.com>

A *Technorati* az „élő web” keresésének eszköze, amely nemcsak a *blogoszférát* – az elmúlt évek kirobbanóan sikeres kommunikációs formáját, a

vállalatok és magánszemélyek által egyaránt alkalmazott webnaplókat –, hanem a civil médiumok egyéb újszerű formáit (podcast, videoklip) is végigpásztázza. Nem elhanyagolható mennyiségekről van itt szó: a Technorati 54 millió blogot figyel; naponta 75 ezer új blog születik, a bloggerek vagy bloggerek 1,2 millió új bejegyzést készítenek naponta. A mennyiségen túl a hagyományos keresőmotorok még egy problémával szembesülnek. A Google és a hasonló szolgáltatások szoftverpókokat – webfűrész robotokat – küldenek ki a kiberterre, amelyek végigjárják a weblapokat, és visszahozzák a tartalmat – ezt a „hozott anyagot” indexelik aztán a keresési adatbázisban. A mennyiségi növekedéssel a webhelylátogatások közötti időszak óhatatlanul megnő. Ilyen technológiával képtelenség a blog valós idejű közegét, ezt az *instans* médiumot indexelni. A Technorati megoldása egy *pingelési* eljárás, amelyet még korábban a *Radio Userland* blogoló-fejlesztői az új blogbejegyzés körtájékoztatására dolgoztak ki. Ha egy blogger valahol megszólalt, egy pingertesítés érkezett a szerverhez, amelyet aztán a Userland robotja felszedett, és az új bejegyzéseket indexelésre küldte. A Technorati és a nagyobb blogszolgáltatók közötti megegyezés szerint a pingelés üzemelésével az új blogbeírásokat már a közzététel utáni másodpercekben indexelik.

A Technorati eléggé hagyományos relevancia-rangsorolást alkalmaz, amelyben a fő kulcsok: a blogüzenethez mutató csatolók száma, és a tartalom belüli kulcsszó elhelyezkedése. Támaszkodik továbbá a bloggerek saját címkézésére, privát tárgyszavazására is – nem szabad persze feledni, hogy a blogkeresés messze nem egzakt tudomány. A Technorati élen járt ugyan az új műfaj menedzselésében, de a konkurencia napjainkban már erősödik. A Google például amellet, hogy megszerezte a *Blogger* szolgáltató céget, saját blogkeresőt (Blog Search) bocsátott ki, amely gyorsabbnak tűnik a Technoratinál – ez utóbbi viszont

sikeresebben birkózik meg a hatalmas mennyiségű bloghulladék (blog spam) kiszűrésével.

ZOOMINFO

<http://www.zoominfo.com>

A *Zoominfo* erőssége a társadalmi jellegű keresés – célja az emberek, cégek, és a közöttük lévő kapcsolatrendszerek feltérképezése és megjelenítése. (Ezzel a keresőmotorral üzemel az *Amazon A9* elnevezésű keresője és a *Business Week* magazin is.) A magát összesítő keresőgépnek nevező ingyenes *Zoominfo* online forrásokból, weblapokból, hírszolgálatokból, üzleti-kereskedelmi adatbázisokból, tőzsdei állományokból szűri ki az adatokat: egy adott személynév vagy vállalati név keresőkérdésére 32 millió üzletember és 2,5 millió vállalat adatait kutatja át. A kihámozott és összesített adatsorból egy tömör és könnyen olvasható formátumú rekordot állít elő. A weben egyébként is szabadon hozzáférhető adattömegből a *Zoominfo* ingyenesen készíti a profilt: az elsősorban üzletemberek keresésére szolgáló eszköz 1 584 171 igazgatósági tagot, 323 336 vállalatvezetőt, 2 862 886 igazgatót és 128 960 jogászt tart nyilván indexében. A természetes nyelvi elemzéssel és mesterséges intelligenciával támogatott rendszer a tesztelés tanúsága szerint nem hibátlan. Tévesztés a gyakoribb neveknél érhetően többször fordul elő. A személyi profilok minősége egyenetlen, az összesítés sokszor avult. Azonosítás után egyének is kezdeményezhetik, és el is végezhetik saját profiljuk frissítését a *Zoominfo* adatbázisában. A tévesztéseket minimálisra szűkítő összetett keresés lehetősége azonban csak az előfizető használóknak jár.

WINDER, Davey: *Research tools: Specialist searches.* = *Information World Review*, 228. sz. 2006. október, p. 22–25. <http://www.iwr.co.uk/>

(Bánhegyi Zsolt)

RETROKON technológia – a katalógusdigitalizálás eszköze és az internetes elérés Csehországban

A RETROKON technológia alapjait tíz éve rakták le, amikor a *Cseh Nemzeti Könyvtárban* elkezdtek a retrospektív konverziót, a papíralapú cédulakatalógus és a bibliográfiai leírások digitalizálását. Ez

az eredeti cseh technológia a nemzeti könyvtár és a *Comdat* cég együttműködésének az eredménye. Az elmúlt tíz évben ennek a technológiának a segítségével durván tízmillió, különböző cseh könyv-