



Szakértői keresés

Találati listák megjelenítésének módozatai: szokásos megoldások

Jacsó Péter folytatja „Szakértői keresés” (*Savvy searching*) című sorozatát, az online keresést, az adatbázisok használatának évek óta tartó feltérképezését az *Online Information Review* legutóbbi számaiban is. (A sorozat korábbi darabjáról Bánhegyi Zsolt készített referátumod, I. TMT 51. köt. 12. sz. p. 558–559.

http://tmt.omikk.bme.hu/show_news.html?id=3780&issue_id=457 – A ref. megjegyzése.)

Jacsó itt bemutatott közleményei, amelyek a neves periodika két külön számában jelentek meg, ám a sorozati témán belül azonos altémát boncolnak, a különféle információs rendszerek találati listáinak megjelenítését vetik össze. Míg az online adatbázisok keresési eszközeit, segédleteit és lehetőségeit örökös figyelem kíséri, a kimeneti oldal, a találati listák megjelenítése kevésbé foglalkoztatja az adatbázis-tervezőket. Igaz ugyan, hogy az eseti felhasználó számára elegendő néhány releváns tétel és egy sallangmentes megjelenítési forma, a szakavatott kereső-kutató azonban a hosszadalmas és veszélyes keresési eljárás, a keresőkérdések kidolgozása, a közbülső listák finomítása nyomán igényli az időt megtakarító professzionális találati módot. Az adatbázis-szolgáltatók töreksenek arra, hogy a keresés eredményeképpen tiszta, jól szerkesztett találati lista kerüljön képernyőre a felhasználó által beállított formátumban, a lehető legtöbb információval. Ehhez sokféle dolgot kell rendelkezésre bocsátani: előre meghatározott kimeneti formátumok variációit, köztük a rekordtartalom és a megjelenítési formátumok több változatát, csoportosítási és exportálási lehetőségeket; ez utóbbi a kijelölt rekordok bibliográfia-kezelő programokba való töltését teszi lehetővé.

Rövid találati lista

Talán ez a legkritikusabb kimeneti opció, mivel minden felhasználó szembesül vele. Általános

gond, hogy a tételek túlságosan hosszúak, ezért a találati listát nehéz gyorsan áttekinteni.

A *DialogWeb* irányított keresés módjában elég jó ez a formátum: csak a címet és a megjelenés évét tartalmazza (méltán „picklist” a neve). Főként jeleníti meg azonban a cím után az adatbázis-azonosítót. A szerző tapasztalatai szerint a két legfontosabb adat a cím és a megjelenési év. Ezért a két információért a fizetős helyek sem kérnek külön díjat. Fontosságban a harmadik a folyóirat címe, a szerző neve csak ezután következik. Ha a cikk címéből csak az első 30-40 karakter jelenne meg, a folyóirat címéből 30, a kiadási év 4 karakterével együtt ezek az adatok együttesen is csak egyetlen sort tennének ki a rövid találati listán.

A *Dialog* adatbázis *REPORT* parancsával a felhasználók maguk állíthatják be, hány karakterig akarják megjeleníteni az általuk fontosnak ítélt adatelemek sorát. Ez a parancs azonban csak címtár/névtár jellegű adatbázisokban érhető el, azok közül sem mindben, pedig föltöbb hasznos lenne.

A kiváló *MetaLib* metakeresőgép tömör formátumot kínál fel, ami a kiadási évet, a szerző nevét és a dokumentum címét tartalmazza a forrásadatbázissal együtt.

Az *EBSCO* rövid találati listáján csak a dokumentum címe és teljes szövegű elérhetősége van feltüntetve, a kiadási év hiányzik. Praktikus vonás, hogy szűkíthetünk teljes szövegű elérhetőségre

(sajnos csak az összetett keresési módban), s azután egyetlen ikonnal kijelölhetjük az elérhető találatokat. A *Cambridge Scientific Abstracts* (CSA), a *Gale*, az *OCLC* és a *ProQuest* adatbázisaiban nincs ilyen opció; némelyikük tartalmazza a kiadás helyét. A folyóirat megjelenési helye és hasonló információk valójában még a teljes formátumban sem igazán lényegesek, csak kevés kutató

tartja ezeket igazán fontosnak. A legjobb megoldás az lenne, ha formátumot lehetne választani: év-cím-folyóirat vagy év-cím-szerző szerint, ilyen alternatíva viszont egyik rendszerben sincs.

Töredékek a találati listán

A „snippet” – nevezzük töredéknek – a találati lista tételeiben megjelenő szövegdarab, amely a keresőkérdest szűkebb környezetében mutatja meg. Az összes webes keresőmotor találati listája tartalmaz ilyen szöveggörnyezet-mintát (néhány helyen, zavaró módon más-más helyről halászott szövegszilánkokat ragasztanak össze). Meglepő viszont, hogy a legtöbb kereskedelmi online közvetítő-szolgáltatónál, aggregátornál hiányzik ez a lehetőség. A *Dialogban* az ún. *KWIC-formátum* segítségével (KeyWord In Context, egyfajta előre definiált formátumváltozat) különböző hosszúságúra állítható be. 30–50 szavas ablakban jelenik meg, így 80–100 eredmény is gyorsan átfutható. Sok *Dialog*-adatbázisban ez a megjelenítés ingyenes, ahol nem, ott is olcsó: a rövid vagy közepes formátum 10%-ába kerül.

Előre definiált rekordtartalom-opciók

A *DialogWeb* az imént említett *KWIC*-formátummal együtt öt előre definiált tartalommegjelenítést kínál fel. Az *OVID*-ban négy ilyen van, a *CSA*-ban és az *EBSCO*-ban három-három. A többi szolgáltató csak rövid, vagy teljes verziójú rekordmegjelenítést tesz lehetővé.

Testre szabható rekordtartalom-opciók

Az előre definiált tartalommegjelenítési formátumok csekély számát ellensúlyozza, hogy a *DialogWeb*, az *OVID* és a *CSA* adatbázisaiban a felhasználó beállíthatja az általa kedvelt rekordtartalom-formátumot, és azt el is lehet menteni. (A többi aggregátor viszont ezt sem kínálja fel, pedig az előre definiálás hiánya miatt az egyéni beállítás lehetőségére éppen ezeknél lenne a legnagyobb szükség.)

Előre definiált és testre szabható adatelem-megjelenítés

A rekordtartalom formáját tehát néhány online szolgáltatónál állíthatjuk, de az adatelemek megjelenítésének sorrendje, formája, térrendezése egyiknél sem alakítható. Az *OVID*-ban van néhány előre definiált megjelenítési formátum, ám nincs köztük olyan, amit a felhasználó maga alakíthatna

ki. A keresőkifejezés kiemelésére viszont kétféle – vastag, illetve dőlt betűs – karakterformázási lehetőséget tesz lehetővé, és a kiemelés színe is beállítható. Az *OCLC*-ben egyáltalán nem emelik ki a keresőkifejezést a találatokban, a *DialogWeb*-en pedig a felhasználó aktiválhatja vagy kapcsolhatja ki ezt az opciót.

Némelyik rendszerben a találati listák mentéséhez, nyomtatásához, e-mailben való küldéséhez beállítható a kívánt hivatkozás-stíluslap (tartalom, megjelenítés, tipográfia és központosítás kombinációja). Ebből a szempontból a *CSA QuikBib felülete* (bibliográfia-készítő felület) a leggazdagabb.

Rendezési lehetőségek

A találatok szerző, cím, folyóiratcím, megjelenés dátuma vagy relevancia szerint csoportosíthatók. A relevancia szerinti rendezés algoritmusai általában kereskedelmi titoknak minősülnek, ezért sokszor megfejtethetetlen egy csökkenő relevancia szerinti találati lista. Az *OVID* tartalmazza a legtöbb keresési kulcsot (ezek adatbázisonként különbözők). A felhasználó elsődleges vagy másodlagos csoportosítási kulcs közül választhat, emelkedő vagy csökkenő sorrendben. Például a szerzőt jelöli be elsődlegesként, a megjelenés évét másodlagosként (szerző: növekvő sorrendben, a szerző publikációi időrendi sorrendben visszafelé haladva). Némelyik csoportosítási kulcs értelmetlen, vagy nem igazán szükséges, például a kivonatok szerinti csoportosítás. A *Dialogban* szintén sokféleképpen rendezhetők a találatok (többnyire praktikusán). Irányított keresési módban egyszerre csak egy rendezési feltétel használható, parancsmódban viszont még három feltétel (pl. szerző, cím, megjelenés éve) szabható hierarchikusan. Az *EBSCO*-ban rendezési elemnek állítható be a dátum, a forrás, a szerző és a relevancia. Az *OCLC*-ben cím szerint is csoportosíthatunk, ám hiányzik a relevancia szerinti csoportosítás lehetősége.

A legtöbb rendszer mérethatárt szab. Ezt általában a helyi rendszerkönyvtáros állítja be (pl. az *EBSCO*-nál 50 rekord), vagy maga a szolgáltató (pl. az *OVID*-ban 1000, az *OCLC*-ben 200 rekord).

A küldés formátumai

Az *ISI* tulajdonában lévő bibliográfia-menedzselő programok (*ProCite*, *EndNote*, *Reference Manager*), illetve egyéb programok (*Biblioscope*, *DB/Textworks*, *RefWorks* stb.) lehetővé teszik a tárolást, és a kijelölt rekordok leválogatását is. Bár

ezek a segédeszközök némely eddig tárgyalt beállítási hiányosságot pótolnak, áruk azonban meglehetősen magas. A rekordok címkézett ASCII-formátumban exportálhatók, de előtte fárasztó kézi beállításokat kell megtennie a felhasználónak. A legjobb exportopciókkal a ProQuest, az EBSCO és az OCLC rendelkezik, ahol standard RIS formátumban menthetők a rekordok: a rövidítés a *Reference Manager* import stíluslapjára utal. Ezt a legtöbb bibliográfia-kezelő szoftvercsomag támogatja.

Az output formátumok – mint láttuk – befolyásolhatják a keresés hatékonyságát. A fenti, széles körben alkalmazott megoldások mellett a hivatkozásokat feldolgozó, citációs adatbázisokban egyéb lehetőségek is vannak – ezeket tárgyalja a szerző cikke második részében.

Találati listák megjelenítésének módozatai: hivatkozáskeresési opciók

A hivatkozott és hivatkozó rekordoknak is megvannak a maguk speciális megjelenítési követelményei, amelyek segítik a felhasználót abban, hogy szerző, folyóirat vagy téma szerint megtalálja a legtöbbször idézett cikkeket. Jelenleg a citációs találati halmazok megjelenítése, tartalma, rendezése rendkívül szűk keretek közé van szorítva, s a legjobb hivatkozáskeresést kínáló szolgáltatók is másodrangúként kezelik a hivatkozásra vonatkozó információkat.

A hivatkozáskeresés jelentősége vitathatatlan, ebből következően egy cikk idézettségi adataira növekvő igény mutatkozik, mint ahogyan az is fontossá válik, hogy a rendszer az idézettség rangsorolása szerint csoportosítsa a találatokat.

Hosszú ideig csak az *ISI* adatbázisai: a *Science Citation Index (SCI)*, a *Social Science Citation Index (SSCI)*, és az *Arts and Humanities Citation Index (A&HCI)* tartalmazták az idézett hivatkozásokat. Ebből az unikális adatbázisrendszerből aztán a néhány éve beindított *Web of Science (WoS)* jóval több lehetőséget aknázott ki, mint a csekély számú továbbszolgáltató (*DataStar*, *DIALOG*, *DIMDI*). 2004-ben az *Elsevier* piacra lépett a legnagyobb indexelő és kivonatoló adatbázisnak kikiáltott *Scopus*-szal, amelyben csaknem 8 millió rekord 180 millió hivatkozást tartalmaz. Tételei többnyire 1995 után publikált tudományos cikkek, konferencia-anyagok. Más adatbázisok új kiadásai szintén bővítik állományukat hivatkozásokkal. A *PsycINFO*-ban 2004 végén már több mint 30 000

rekord 12,5 millió hivatkozást tartalmazott. A Cambridge Scientific Abstracts a többi online szolgáltatóhoz képest igen kifinomultan kezeli a hivatkozásokat. 2005 áprilisában hivatkozásokkal bővítette szociológiai kivonatait, és néhány más adatbázisát. A kiadók digitális archívumaikban gyorslinkkel ellátott „élő” hivatkozásokat kínálnak.

A *HighWirePress*, amely élen jár kiadók archívumainak digitalizálásában, automatikusan megjeleníti azon cikkek rövid listáját, amelyek idézik az aktuálisan a képernyőre hívott cikket. A hivatkozó ugrópontok kétirányúvá válnak, amint egy következő folyóiratcikk idézi a korábbi, s ekként bővül egyre a hálózat, pontosan úgy, ahogy fél évszázaddal ezelőtt ezt *Eugene Garfield* megálmodta.

Rövid találati lista

Jacsó régóta „gyötri” az online információszolgáltatókat, hogy jól láthatóan jelenítsék meg, hány cikk hivatkozik a rövid találati listán megjelenő közleményekre. Ez a legbiztosabb módja annak, hogy a felhasználó ráébredjen az idéző hivatkozások jelentőségére, s így azok közül kiválaszthatja a legígéretesebbeket, vagyis a legtöbbször idézett cikkeket. A CSA kezdetektől fogva felkínálja ezt a lehetőséget, és megtartotta a 2005-ben bevezetett *Illumina* szoftverben is. Az online szolgáltatók zöme azonban még mindig hátraveti ezt a fontos információt a teljes rekordmegjelenítésben.

A felhasználóknak többnyire nincs más választásuk, végig kell futniuk az egész találati listán. Az *Annual Reviews (AR)* dicséretes módon közlésezi, hogy az *ISI Citation Index*-ekben hány hivatkozás található egy AR-fejezetre, ám ez az információ csak akkor jelenik meg, amikor a felhasználó kiválaszt egy tételt a találati listából, hogy részletesen megnézze az adatokat.

A Google Scholarban egyelőre túl kevés információ jelenik meg a felhasznált forrásokról, s az idézett és idéző hivatkozások kezelése még elnagyolt: nem különítik el világosan az idézett és idéző folyóiratokat, az idézett és idéző éveket, valamint nem aknázták még ki eléggé a tudományos cikkek, tanulmányok strukturált, metaadatokban gazdag megakollekcióit, amelyeket a kiadók sokasága ezüsttálcán kínált fel a Google-nak. Pozitívuma, hogy nemcsak a dokumentum idézettségi indexét jeleníti meg, hanem ezt a számot arra is felhasználja, hogy a rövid találati listán eszerint határozza meg a cikkek megjelenítésének sorrend-

jét. (Más rangsorolási opciót gyakorlatilag nem is kínál fel.)

A Scopusnál kitűnő a rövid találati halmaz tömörsége és elrendezése, s az idézettségi tényező is szembeötlő az eredményrácsban.

A Web of Science korábban csak a részletes rekordban jelenítette meg az idézettségi tényezőt, ma már mindez a rövid listában is szerepel. Próbakeresés során látható a Google Scholar és a Web of Science élen álló tételeinek magas idézettségi mutatója, ami ékesszólóan bizonyítja a multidiszciplináris adatbázisok előnyeit az egy tudományágra specializált adatbázisokkal szemben.

Töredékek a találati listában

Az általános keresőmotorok találati listáin szereplő töredékek kevésbé informatívak és szervezettek, mint a professzionális adatbázisokban olvasható kiragadott idézetek, „snippetek”. Mivel ezek többnyire automatikusan választódnak ki a szövegből, gyakran irrelevánsak vagy nehezen érthetők, még a legjobb tudományos adatbázisokban is.

Az ingyenes és szuperintelligens *CiteSeer* rendszer (korábban: *Research Index*) azt bizonyítja, hogy tervezői jól oldották meg feladatukat, mindent tudnak a tudományos közlemények szerkezetéről, elrendezéséről, metaadatairól, valamint jól ismerik az idézés stílusát és a formátum trükkjeit. Ennek eredménye, hogy a *CiteSeer* idézettség-indexelése és metaadat-kibontása relevánsabb töredékeket hoz létre, mint a Google Scholar példái. A *CiteSeer* feltárja a tartalmat az idéző tanulmányban, és egyértelműen azonosítja az idéző forrásokat.

Rendezés az idézett hivatkozások szerint

Mivel az idézettség – különösen mások által – fontos tényező, nem elhanyagolható a tudományos közlemény idézettségi gyakorisága szerinti találati halmaz rendezése sem. Ennek ellenére a rendezési alternatívák vagy teljesen hiányoznak, vagy nagyon korlátozottak. Csak a három legnagyobb online szolgáltató csoportosítja idézettség szerint a talált cikkeket.

A Google Scholar „eszi, nem eszi, nem kap mást” alapon egyetlen rendezési formát ajánl. A Web of Science-ben már valódi választás lehetséges négy opció közül: ezek egyike az idézettség száma szerinti rendezést kínálja föl. A WoS-nál szokásos módon a találati lista limitje 300 rekord. Igaz, hogy ez nem komoly korlátozás a végeredményhez képest (ami általában nem több 20-40 találatnál), ám az idézettség szerinti csoportosítás remek eszköz lehet az első, nagyszámú találati halmaz szűkítésére.

A képzetlen felhasználót elárasztják a találatok, ha például a „depression” szóra keres; még rosszabb a helyzet, ha a keresés multidiszciplináris, teljes szövegű adatbázisban történik, mivel a „depression” gazdasági, fémipari és geodéta értelmű rekordjait is a halmazba gyűjti. A Scopusban erre a szóra több mint 190 000 találatot kapunk. Az első 2495 rekord 2005-ben publikált cikk, többségük még idézetlen. A halmaz csoportosítható szerző, folyóiratcím, relevancia szerint. A szerző vagy folyóiratcím szerinti rendezés nem sokat segít az átlagos egyetemi vagy diplomás hallgatónak, de a relevancia szerinti csoportosítás sem, ahol az első két találat 1998-as hivatkozásokat takar.

A tökéletes kulcsot a kiváló tervezésű találati rács rejti, amelyben az idézettségi fejlécnél indított gyors, 20-25 másodperc alatt lefutó rendezés idézettség szerint csökkenő sorrendbe rendezi a találatokat. A keresésben járatos kutató egy kevésbé szembeötlő, ám annál szofisztikáltabb eszközhöz nyúl: az együtthivatkozási analízis (co-citation analysis) az idézett hivatkozáslistán együtt szereplő cikkeket gyűjti össze és rangsorolja. A Web of Science által bevezetett, és a Scopus által is átvett eszköz jóval megbízhatóbb forráslistát ad a kutató kezébe, mint a széles körben elérhető relevancia-rangsor.

JACSÓ Péter: Options for presenting search results. Part 1. Common options. = *Online Information Review*, 29. köt. 3. sz. 2005. p. 311–319.

JACSÓ Péter: Options for presenting search results. Part 2. Options for citation searching. = *Online Information Review*, 29. köt. 4. sz. 2005. p. 412–418./

(Vida Andrea)