



A web mint a tudományos információ egyik forrása: összehasonlítás fizetős adatbázisokkal

A tudományos haladás nagyrészt a kutatók közti kommunikációtól függ. Ennek két útja alakult ki:

- *közvetlen*: a szakemberek közötti személyes kommunikáció (konferenciák, találkozók, személyes kapcsolatok);
- *közvetett*: a szakirodalomhoz való hozzáférés.

A szakirodalomhoz a hozzáférést hagyományosan az intézményeken keresztül elérhető bibliográfiai adatbázisok és folyóiratok tették lehetővé. A web – a földrajzi korlátok lebontásával, az intézményektől való függetlenségével, a késések eltüntetésével, az ingyenességgel – új dimenziókat nyitott az információ terjesztésében.

A szolgáltatók a következőképpen használják ki a web adta új lehetőségeket:

- kutatóintézetek, egyetemek a weben teszik közzé kutatási jelentéseiket;
- a weben keresztül érhető el a könyvtári katalógusok;
- különböző szervezetek itt helyezik el a tevékenységükről szóló jelentéseket, konferenciák dokumentumait;
- a kiadók hozzáférhetővé teszik folyóirataik archívumát;
- ingyenes adatbázisok kereshetők a weben.

A web hátrányai: megoldatlan a hosszú távú hozzáférhetőség és az információ megbízhatósága.

A tanulmány célja, hogy minőségi és mennyiségi szempontból értékelje a webet tudományos információforrásként, a fizetős adatbázisokból visszakeresett információval való összehasonlításban. A kutatás háttérében áll az a feltételezés, hogy a fizetős adatbázisok a teljesség és a megbízhatóság szempontjából elfogadható minőséget képviselnek, s ezek összehasonlítási alapul szolgálhatnak a weben elérhető információk értékeléséhez.

A tanulmányban tudományos információknak tekintettek minden publikációt, szakirodalmi írást, vagy bibliográfiai információt. Az értékelést egy szűk tudományterületen belül végezték el. Konkrét kutató

tási példa alapján folytattak keresést adatbázisokban és a weben. A „web” terminus ebben a tanulmányban az ingyenesen elérhető oldalakra és szolgáltatásokra vonatkozik (megkülönböztetve ezeket a fizetős szolgáltatásoktól).

Módszer

A kiválasztott tudományterületről a következő keresőkérdést fogalmazták meg: „Mely dokumentumok láttak napvilágot 2000 és 2003 decembere között a molekuláris jelek a szójabab mennyiségi jellemzőinek vizsgálatára való alkalmazásáról?” A kutatásnak három fázisa volt mind az adatbázisok, mind a web értékelése kapcsán:

- a kutatások előkészítése,
- információkeresés,
- az eredmények értékelése.

Előkészületi szakasz

Ebben a szakaszban három fázist különböztettek meg:

1. Az információs szükségletek elemzése, a keresési probléma tartalmi, szemantikai és szintaktikai szempontból való megragadása. A tartalmi elemzés során az információs szükségletek behatároltságát, a főbb témák félreérthetetlen leírását, a további szempontokat és a keresés dátum, nyelv és dokumentumtípus szerinti behatároltságát határozták meg.
2. A megfelelő adatbázisok és a webes keresőprogramok kiválasztása. A fizetős adatbázisok közül a téma alapján a *CAB Abstract* és a *Biosis Preview* szolgáltatásokat választották ki. Mindkettő lektorált cikkeket szolgáltat, a felületük nagyon hasonló a webhez, online hozzáférést kínálnak a teljes szöveghez, földrajzi és nyelvi tekintetben nem korlátozzák a dokumentumok körét, ahogyan a dokumentum típusa szerint sem, így jól összehasonlíthatók a webes keresőkkel.

A keresőmotorok közül a reprezentativitás és felhasználóbarát felület miatt a Google-t (www.google.com) és a Scirust (www.scirus.com) választották ki. (A Google Scholar a kutatás időszakában nem volt elérhető.) A Scirus a Google-lal szemben olyan adatbázisok eredményeit is „látja”, mint a *BioMed Central*, *LexisNexis*, illetve a *ScienceDirect*.

3. A keresési stratégia meghatározása. A stratégiát alapvetően meghatározza, hogy van-e tezaurusz, vagy csak szabad szöveges kereső áll rendelkezésre.

A kutatás parttalaná válásának érdekében, két feltételt szabtak meg:

- egyetlen doménneven belül maximum öt „kattintásnyi” távolságra lehet továbbhaladni,
- az eredmények elemzésénél csak az első 40 találatot veszik figyelembe.

A keresés

Először mind az adatbázisok, mind a webes keresőrendszerek esetében alaposan meg kellett ismerni a rendszer felépítését, a lekérdező nyelvet és a kínált keresési lehetőségeket. Ezután került sor a keresésre külön a weben és a kiválasztott adatbázisokban. A Biosis Preview nem használ tezauruszt, esetében szabad szöveges, míg a CAB Abstract adatbázisban kombinált (tezaurusz-alapú és szabad szöveges) keresés történt. Mind a web, mind az adatbázisok esetében a keresést a dokumentumok relevanciáját és megbízhatóságát elemző értékelés követte.

Eredmények

A visszakeresett dokumentumok mennyisége és fajtái

A két adatbázis összesen 65, a két webes keresőmotor együtt 85 releváns találatot adott. Az átfedések kiszűrése után 103 dokumentumot találtak a kétféle módszerrel együttesen. Ezek a következő csoportokba sorolhatók:

- Utalás folyóiratcikkekre. Az adatbázisok 45, a keresőmotorok 36 cikket találtak referátummal. Ez a kétféle eszköz találatai közötti 50%-os átfedés miatt összesen 54 cikket jelent.
- Folyóiratcikkek teljes szöveggel. A kiválasztott bibliográfiai adatbázisok nem szolgáltatnak teljes szövegű cikkeket, viszont 14 esetben külső csatolásokkal ingyenesen elérhetővé tették őket. A web 25 szabadon hozzáférhető teljes szöveges

cikket talált, főleg kutatók, intézmények honlapján keresztül. 34%-os átfedéssel összesen 29 cikket találtak az adatbázisok és a keresőmotorok.

- Konferenciaanyagok. Az adatbázisok 4, a web 16 konferenciakötetben megjelent cikkekre való utalást talált.
- Szürke irodalom. Projektek, mérések zárójelentései, értékelése, kutatási jelentések, hivatalos kiadványok. A keresőmotorok hatot, az adatbázisok egyet sem találtak.
- Szabadalmak. Az adatbázisok és a web két szabadalmat találtak, mindkettőt az Egyesült Államok Szabadalmi és Védjegy Hivatalának szabad hozzáférésű adatbázisában.

Az eredmények azt mutatják, hogy míg az adatbázisok jórészt folyóiratcikkekre való hivatkozásokat találtak, a weben való keresés a kutatóintézeti és egyéni kutatói oldalokról bőven eredményezett irodalmat. A webes keresés főleg egyesült államokbeli találatokat hozott, az adatbázisok földrajzi értelemben szélesebb kört fogtak át. Az adatbázisokban a találatok és a referátumok nyelve az angol, a más nyelven írt anyagok nem érhetőek el a weben az adott nyelv ismerete nélkül.

Az információ szervezettsége, megbízhatósága

Az adatbázisok jól strukturált formában adják meg a találatokat, és az adott dokumentum relevanciájának megítéléséhez referátumok is rendelkezésre állnak, szemben a webbel. A CAB Abstracts szolgáltatásban visszakeresett rekordok 96%-a, a Biosis Previewsban találtaknak 92%-a bizonyult relevánsnak. Az első 40 találatban a Google esetében 43%, a Scirusnál 35% volt a releváns tételek aránya. A Google találatai közt 29%, a Scirus esetében 22%-ban fordult elő ismétlődés. Az adatbázisok garantálják a találatok megbízhatóságát, míg a keresőmotorok esetében a keresőnek kell utánajárnia, hogy az adott tétel kellően megbízható-e vagy sem.

A befektetett idő és az esetleges költségek

A teljes keresés a két adatbázisban 12 és fél órát, a weben a két keresőmotorral 17 órát vett igénybe. A legtöbb időt ebből a keresési lehetőségeknek, az adatbázis struktúrájának és funkcióinak a megismerésére fordították. A web ilyen értelemben időtakarékosabb, tekintve a keresőmotorok egyszerű struktúráját, a tezauruszok és az indexelési eljárások nélkülözését.

A konkrét keresés 2,5-2,5 órán át tartott az adatbázisokban, illetve 3,5 órát a Scirusnál, és 4 órát a Google-lal. Az időben a webes találatok ellenőrzése is benne van. (Csak az ismétlődések kiválogatása, és a találatok megfelelő formába rendezése ebből 3-3 órát igényelt a két keresőmotornál.)

A weben való keresés költsége az internet-hozzáférés díjára korlátozódik, az adatbázisoknál az előfizetési díjak, licenccdíjak, a letöltött rekordokért vagy teljes dokumentumokért fizetett díjak is növelhetik a költségeket.

A keresési stratégia

Az információs szükségletek keresést megelőző mélyreható elemzése segít a keresőkérdés pontos megfogalmazásában, a megfelelő eszközök kiválasztásában és használatában. Az adatbázisokban való keresés stratégiáját befolyásolja, hogy rendelkezésre áll-e teaurusz. Ha igen, akkor a szabad szöveges keresés, a teauruszalapú keresés és a kombinált keresés között választhatunk. Az eredmények azt mutatták, hogy a kombinált módszer vezetett a találatok legátfogóbb halmazához. A weben csak a szabad szöveges keresés jöhet számításba.

A webes keresőrendszerek esetében a második lépés beiktatása a keresésbe a Scirusnál 38, a Google-nál 60%-kal növelte a releváns találatok számát. Az, hogy párhuzamosan több keresőrendszerrel is rákerestek ugyanarra, az adatbázisok esetében 36, a keresőmotoroknál 17%-kal növelte a megtalált dokumentumok számát.

A kereső számára szükséges készségek

Az adatbázisoknál:

- az elérhető adatbázisok ismerete, hogy ki lehet válaszítani az információs igény szempontjából a legmegfelelőbbet;
- a kiválasztott adatbázisok felépítésének ismerete, hogy a legjobb keresési stratégiát tudjuk alkalmazni;
- a téma ismerete, hogy felismerjük az információs igény szempontjából releváns találatokat;

- az angol nyelv ismerete, mivel a legtöbb adatbázis ezen a nyelven érhető el.

A webes keresésnél:

- az elérhető webes keresőeszközök ismerete, hogy ki tudjuk választani a feladat szempontjából a legmegfelelőbb eszközt;
- a kiválasztott eszközök hatékony használatának képessége;
- a téma ismerete, hogy a kereső ki tudja választani a releváns találatokat;
- jártasság az információtechnológiában, hogy az adatokat és dokumentumokat a kereső konvertálni tudja a megfelelő formába;
- információs és dokumentációs jártasság a megtalált dokumentumok elrendezéséhez.

A web mint a tudományos információ forrása

A kiválasztott területen a web tudományos szempontból is hasznos forrásnak bizonyult. Az adatbázisokkal való összehasonlítás azt mutatta, hogy ha nem áll rendelkezésre adatbázis, akkor a weben is elfogadható színvonalú információkat találhatunk. Ha adatbázis is rendelkezésre áll, a megtalált dokumentumok körét bővíthetjük a weben való kereséssel.

Az olyan keresőmotorok használata, mint a Scirus, amely adatbázisok találatait is „látja”, különösen fontos szerepet játszhat abban, hogy színvonalas találatokat gyűjtsünk a weben. Összességében az a feltevés, hogy a web tudományos információ szempontjából értéktelen, megalapozatlannak bizonyult. Az információs szükségletek elemzésének alapvető fontossága van mindenféle keresést megelőzően, mivel a keresési probléma szűkítése, a megfelelő információkereső eszközök kiválasztása és a keresőkérdés megfogalmazása ezáltal válik lehetővé. A keresési stratégia nagymértékben függ a keresőeszköztől. A web több szabadságot enged a keresőnek a probléma megoldására, egyben több kreativitást, állhatatosságot igényel, mint az adatbázisok.

/DOLDI, Luisa M.–BRATENGEYER, Erwin: *The web as a free source for scientific information: a comparison with fee-based databases.* = *Online Information Review*, 29. köt. 4. sz. 2005. p. 400–411./

(Tóth Máté)