

A Science Citation Index CD-ROM-on, avagy a világ legnagyobb szakértő rendszere?*

A szerző ismerteti a CD-ROM-on megjelent Science Citation Indexet (SCI), melyet kereskedelmi adatbázisban megvalósított bibliográfiai kapcsolatnak tekint. Tárgyalja az együttlidézési kapcsolat és a bibliográfiai kapcsolat közötti különbséget. Mivel az SCI a publikációk irodalomjegyzékének felhasználásával készül, és ezek összeállítása jellegzetes szakértői feladatnak tekinthető, a CD-ROM-on megjelent SCI egy hatalmas szakértő rendszernek fogható fel.

Bevezetés

Aki ilyen címmel ír tanulmányt, annak fel kell készülnie az azonnali tiltakozásra. Mit jelentsen ez: a világ legnagyobb szakértő rendszere? A válasz: ez egy olyan szakértő rendszer, amelyben több ezer, sőt több százezer szakértő vesz részt. A tiltakozás: ez el-entmondást jelent a szakértő rendszerek értelmezésével szemben, mivel "éppen a szakértők ritka volta miatt kell szakértő rendszereket létrehozni" [1].

Nem áll szándékomban a szakértő rendszerek definícióját forradalmasítani, lehet, hogy a cikk címe csak egy metafora, amely bizonyos hasonlóságot emel ki, de az is lehetséges, hogy e hasonlóság mögött több rejlik, ezért a szakértő rendszerek új szemlélete nem zárható ki.

A szakirodalmi informatika egyik vívmányára szeretnék utalni, mely azáltal jött létre, hogy a *Science Citation Index (SCI)* CD-ROM-on is hozzáférhető. Megvalósították az ún. *bibliográfiai* kapcsolatot (bibliographic coupling) egy kereskedelembe kapható adatbázisban. Ezzel összefüggésben bemutatjuk az *együttlidézési* (co-citation) kapcsolat és a bibliográfiai kapcsolat közötti különbséget, majd egy példa segítségével tárgyaljuk a bibliográfiai kapcsolat hasznosítását az SCI adatbázisban. Végül megvizsgáljuk, hogy az egésznek van-e a szakértő rendszerekre jellemző vonása.

A bibliográfiai kapcsolat és az együttlidézési kapcsolat

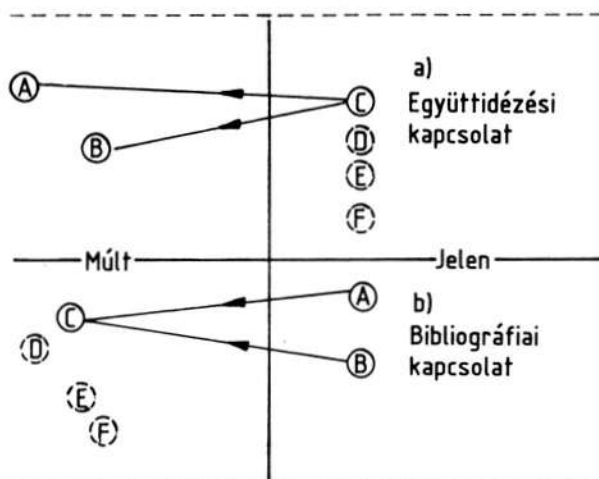
Két tudományos közlemény többféle módon hozható egymással kapcsolatba, pl. úgy, hogy összehasonlítjuk a szerzőket, az intézményeket, a tartalmat vagy az alkalmazott módszereket. Vizsgálatunk szempontjából két kapcsolatfajta jelentős, és mindkettőnek köze van a szakirodalmi idézéshez.

*Az *SCI* auf CD-ROM oder das größte Expertensystem der Welt? c. cikk (Informatik, 37. köt. 1. sz. 1990. p. 37–40.) rövidített fordítása. (Az SCI CD-ROM szakirodalmi szolgáltatásként hozzáférhető a MTA Könyvtárában.)

Két azonos korú közlemény hivatkozásainak összehasonlítása kimutathatja, hogy található-e bennük egy vagy több közös hivatkozás. A közlemények jelenkoriak, azaz most jelentek meg, ezzel szemben a hivatkozások a múlt munkáit képviselik. A hivatkozások egyezése bizonyos rokonságot mutathat ki a két vizsgált közlemény között. Ennek alapján két publikációt, ha ilyen bibliográfiai kapcsolatot mutat, "related papers"-nek, formailag rokon közleményeknek nevezünk. Könnyen belátható, hogy milyen hatalmas munkára van szükség ahhoz, hogy megállapíthassuk, mely közlemények állnak egymással bibliográfiai kapcsolatban. Ennek érdekében minden közleményt páronként kell hivatkozásai alapján összehasonlítani. Ez csak számítógéppel lehetséges, előzőleg azonban fáradságos munkával minden hivatkozást be kell táplálni a számítógépbe. A "bibliográfiai kapcsolat" fogalma mintegy negyed évszázados [2].

A közlemények kapcsolatba állításának másik módja "idézettségük" (citation) vizsgálatán alapul, vagyis azon közlemények alapján, amelyek a szóban forgó cikkeket idézték. Két közlemény kapcsolata annál szorosabb, minél többször idézik őket együtt más közlemények (együttlidézés). Az együttlidézés a jelenben történik, az idézett közlemények viszont a múltban jelentek meg. Az együttlidézési módszer alkalmazásával [3, 4] – ha ezt olyan (lehetőleg gyakran idézett) közleményekből álló halmazra használjuk fel, amelyben a közleményekhez tartozó minden idézet (citation) megvan (Citation Index) – együttlidézési klaszterek állíthatók elő, amelyek de facto a tudomány "feltérképezését" jelentik. Ezek a klaszterek a kutatási fronterületeket jelenítik meg az illető téma-terület sokat idézett, összetartozó publikációiból és azon közleményekből, amelyek a sokat idézett közleményeket idézik. A számítógép segítségével ez a kartográfiai folyamat automatizálható. A bibliográfiai kapcsolattal szemben a számítógépes ráfordítás "valamivel" kisebb, mert mint ismeretes, korántsem idéznek minden közleményt, és a gyakran idézett publikációk még ritkábbak. Ezért az együttlidézési

klaszterek ötletét már a 80-as évek elején megvalósították, míg a bibliográfiai kapcsolat módszerének kiaknázásához szükséges műszaki előfeltételek csak sokkal később álltak rendelkezésre. A "related papers" keresése ezért csak napjainkban valósulhatott meg. Az 1. ábra a két kapcsolattípus közötti különbséget mutatja be.



a) A és B co-cited paperek, melyeket C (+ D,E,F...) együttesen idéznek.

b) A és B related paperek, melyek együttesen idézik C (+ D,E,F...)-t.

1. ábra A két kapcsolattípus közötti különbség

Egy információ-tömeg együttidézési elemzése folyamán kristályosodnak ki a kutatási fronterületek. Ehhez bizonyos időnek kell eltelnie, amely alatt a kutatási fronterületeket képviselő közleményeket idézhetik, vagyis az idézeteknek össze kell gyűlniük. Ez az elemzési módszer, amely tehát csak a sokat idézett publikációkat használja fel, témafigyelésre (current awareness) alkalmatlan, erőssége sokkal inkább a tudomány "forró" tématerületeinek automatikus felfedezésében és követésében van.

Más a helyzet a bibliográfiai kapcsolatelemzéssel. Ez a témafigyelés hagyományos módszereit azokkal a lehetőségekkel gazdagítja, melyeket az "idézés nyelve" kínál, és egy nagyon is jelenkori információs tömegből választja ki mindazokat a rokon közleményeket, amelyek "párosával" a közös hivatkozás ismérvével rendelkeznek. Az 1. ábra azt mutatja, hogy két közlemény együttidézési kapcsolatánál a kapcsoló tag, az idéző harmadik közlemény, a jelenben, azonos időben a kapcsolt munkák után jelent meg, míg két publikáció bibliográfiai kapcsolata esetén a kapcsoló tag, az általuk idézett közlemény, a múltban, vagyis időben a kapcsolt publikációk előtt látott napvilágot.

Rokon közlemények az SCI CD-ROM-ban

Az SCI CD-ROM-ban lévő rokon közlemények (related papers) keresését a következőkben ismertetjük. A példa, amely a "földrengés-előrejelzés"

témában megjelent közlemények keresését mutatja be, erre a célra nagyon le van egyszerűsítve (az ISI által kiadott demonstrációs lemezről* [5] vettük át). Az SCI CD-ROM-ban számítógéppel leolvasható *Permuterm Subject Index*, vagyis a közlemények címszavaiból összeállított tárgymutató is található. Ez a mutató (az egyes szavak gyakoriságának előzetes lekérdezése után) pl. lehetőséget ad arra, hogy a rendszernek a következő keresőkérdést tegyük fel: (1) (EARTHQUAKE OR EARTHQUAKES) AND PREDICT.

E keresőkérdésre a válasz nyolc közlemény bibliográfiai adatait tartalmazza. Ezek közül kettőt megemlítenek (2. ábra). Itt az adalék különösen érdekes, hogy hány related records tartozik az egyes közleményekhez. A számítógépigényes bibliográfiai kapcsolatelemzés ezt az adatot járulékos információs elemként tartalmazza. (Ha az információ reprezentált ismeretet jelent [6], akkor a related papers számának megadása olyan új ismeretet képviselő információ, amelyet a bibliográfiai kapcsolat elemzése útján az idézetek szakértői ismeretéből nyertünk.)

| | |
|---|--------------------|
| Record: 1 of 8 | (2) |
| SPEZLER-H MIZUTANI-H | |
| PREDICTING FAILURE IN ROCKS AND ITS IMPLICATIONS FOR PREDICTING EARTHQUAKES | |
| TECTONOPHYSICS Vol 138 No1 pp 109-113 1987 | |
| References: 27 | Related Records: 6 |

| | |
|---|--------------------|
| Record: 2 of 8 | (3) |
| WYSS-H BURFORD-RO | |
| OCCURRENCE OF A PREDICTED EARTHQUAKE ON THE SAN-ANDREAS FAULT | |
| NATURE Vol 329 No 6137 pp 323-325 1987 | |
| References: 17 | Related Records: 7 |

2. ábra Példák a feleletekre és a keresőkérdésre

A rendszer tehát lehetővé teszi, hogy minden egyes rokon közleménynek, minden related record-nak pl. a (3) közleményhez való kapcsolódását külön-külön megvizsgálhassuk.

A (4), (5) és (6) rokon közlemények adatbázisbeli megjelenésében az a figyelemre méltó (3. ábra), hogy related recordjuk száma mellett még "rokonsági fokuk", nevezetesen a közös hivatkozások, vagyis a kapcsolóelemek száma is szerepel. Feltűnő továbbá, hogy ezek közül a közlemények közül egyiket sem találtuk volna meg címszó szerinti keresés során (1). Rokonságban álló közleményeknek nincs minden esetben közös címszavakból álló címük, ezek "rokonsága" kizárólag a közös hivatkozásokon alapul. A related paper keresés tehát olyan közleményeket hoz napvilágra, melyek különben rejtve maradtak volna. Éppen ebben a pozitívumban van az új szakirodalmi keresési lehetőség jelentősége.

*Az MTA Könyvtára kérésre – beküldött hajlékonylemeze – átmásolja a demonstrációs lemezt.

| |
|---|
| <p>Related Record: 1 of 7 (4)</p> <p>HABERMANN-RE</p> <p>HABERMANN METHOD FOR DETECTING SEISMICITY RATE CHANGES</p> <p>JOURNAL OF GEOPHYSICAL RESEARCH-SOLID EARTH AND PLANETS Vol 92 No 89 pp 9446-9450 1987</p> <p>References: 20 Related Records: 7 Shared References: 5</p> |
|---|

| |
|--|
| <p>Related Record: 2 of 7 (5)</p> <p>KHATIRI-KN</p> <p>GREAT EARTHQUAKES, SEISMICITY GAPS AND POTENTIAL FOR EARTHQUAKE DISASTER ALONG THE HIMALAYA PLATE BOUNDARY</p> <p>TECTONOPHYSICS Vol 138 No1 pp 79-92 1987</p> <p>References: 76 Related Records: 16 Shared References: 3</p> |
|--|

| |
|---|
| <p>Related Record: 3 of 7 (6)</p> <p>YOSHIDA-A</p> <p>PRECURSORY ACTIVITY IN SEISMIC BELTS CONTAINING THE FOCAL REGION OF FORTHCOMING LARGE INTRAPLATE EARTHQUAKES</p> <p>TECTONOPHYSICS Vol 140 No 2-4 pp 131-143 1987</p> <p>References: 35 Related Records: 6 Shared References: 2</p> |
|---|

3. ábra A rokon közlemények megadása

Ezt a lehetőséget pedig minden nehézség nélkül kihasználhatjuk, ha egy tetszőleges közleményből vagy related recordból kiindulva ennek related recordjait jutunk el, és így mondhatni az egész "rokon szövedéket" felderítjük. Végezetül ezt egy példával illusztráljuk.

SPETZLER, H. és MIZUTANI, H. publikációjához (2) megtalálhatjuk a "leányközleményeket" (4. ábra) és az "unokaközleményeket" (5. ábra). Hangsúlyozzuk, hogy a "leánypublikációk" és az "unokapublikációk" alatt nem időben egymás után következőt értünk. A közlemények között a bibliográfiai kapcsolódási útvonalakat a keresés közben mindkét irányban leírhatjuk.

A világ legnagyobb szakértő rendszere?

Térjünk most vissza vizsgálódásunk elejére, közleményünk címére, és tegyük fel újra a kérdést, vajon az SCI CD-ROM-változatának vannak-e a szakértő rendszerre jellemző vonásai.

Amikor a hallgatók az első durva áttekintést kapják azokról a különböző CASCO (COMPUTER AIDED SCIENTIFIC COMMUNICATION) rendszerekről, melyek szakirodalmi információt nyújtanak, akkor többnyire egy háromcsoportos beosztást ismernek meg (6. ábra). A besorolási jellemzők egyrészt az intellektuális ráfordítás, amelyet az információk előkészítéséhez mindig be kell fektetni, másrészt a rendszer által kiszolgált felhasználók száma.

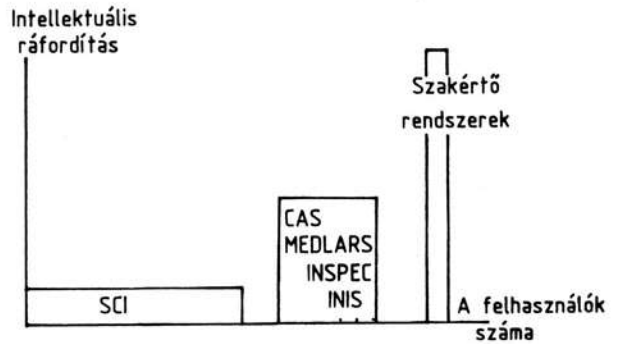
A középső csoportba tartoznak a nagy "klasszikus" szakirodalmi információs rendszerek: a CAS, INSPEC, INIS stb., amelyeket az információ közepes

| |
|--|
| <p>Related Record: 1 of 6 (7)</p> <p>MAHPEI-KD MAUK-FJ</p> <p>SEISMIC WAVE MOTION FOR A NEW MODEL OF HYDRAULIC FRACTURE WITH AN INDUCED LOW-VELOCITY ZONE</p> <p>JOURNAL OF GEOPHYSICAL RESEARCH-SOLID EARTH AND PLANETS Vol 92 No 89 pp 9293-9309 1987</p> <p>References: 71 Related records: 17 Shared References: 1</p> |
|--|

4. ábra "Leányközlemény" a (2) közleményhez

| |
|---|
| <p>Related Record: 1 of 17 (8)</p> <p>FEHLER-M KAIEDA-H HOUSE-L</p> <p>DETERMINING PLANES ALONG WHICH EARTHQUAKES OCCUR - METHOD AND APPLICATION TO EARTHQUAKES ACCOMPANYING HYDRAULIC FRACTURING</p> <p>JOURNAL OF GEOPHYSICAL RESEARCH-SOLID EARTH AND PLANETS Vol 92 No 89 pp 9407-9414 1987</p> <p>References: 12 Related Records: 3 Shared References: 1</p> |
|---|

5. ábra "Unokaközlemény" a (2) közleményhez



6. ábra A szakirodalmi információ felosztása

mérvő intellektuális előkészítése után (pl. referátumok készítése) széles felhasználói kör rendelkezésére bocsátanak.

A legdrágább, legtöbb ráfordítást igénylő rendszerek a szakértő rendszerek, amelyek csak szűken határolt területeket és csak kevés felhasználót szolgálnak ki, és amelyeknél magas képesítésű, jártas "knowledge engineer"-ek, kognitológusok a szakértők, akiknek a tudását elő kell csalogatni [7].

A legolcsóbb rendszer az SCI, itt az információt bizonyos fokig laikusok készítik elő, és felhasználói köréhez az egész világ tudományos közössége tartozik. Természetesen mellékesen megjegyezhetjük, hogy az ilyen beosztással semmiféle értékelést nem kapcsolunk össze. Mindezeknek a rendszereknek megvan a jól megérdemelt helyük a tudományos kommunikáció összrendszerében, valamint társadalmi megbízhatóságuk és hálás használói. Ha azonban az SCI egyrészről, és a szakértő rendszerek másrészről ezen a ráfordítás/felhasználói skálán egymástól

ennyire távol esnek, akkor hogyan viselhetik egy szakértő rendszer vonásait?

Senki sem tagadhatja egy olyan tapasztalt kutató szakértői rangját, aki szakmai eredményeit és ismereteit publikációkban közli. A publikációiban szereplő hivatkozások szakértői ítéleteknek foghatók fel. A fenti értelmezés azonban nem szokványos, ugyanis manapság a hivatkozásokat legjobb esetben egy információs nyelv elemeinek szokták tekinteni [8].

Ha olyan problémák léteznek, amelyeknek megoldási feltétele minden korabeli idézet ismerete, és létezik egy olyan rendszer, amely ezeket a problémákat – az egyes szakértőktől függetlenül és elkülönítve – meg tudja oldani, akkor ez a rendszer talán megérdemli a szakértő rendszer elnevezést. Itt most három ilyen kérdést említünk meg:

1. Kik a legidézettebb személyek a tudományban, vagy melyek a legidézettebb tudományos közlemények?
2. Melyek a kutatás frontierületei, és ezeken kik helyezkednek el?
3. Mely korabeli közlemények alkotják egy adott tudományterület "szövedékét"?

Anélkül, hogy részletekbe mennénk, világos, hogy a három kérdésfelvetés közül egyik sem oldható meg egyetlen olyan kutató közreműködésével, aki nem rendelkezik az idézetek adatbázisával és egy nagyméretű számítógéppel, enélkül még azt sem tudná megfelelő gyorsasággal és pontossággal tisztázni, hogy ki idézi saját közleményeit.

Ha azonban ezek az előfeltételek – mint az ISI-nél – fennállnak, akkor a követendő szabályok egyszerűek. Az első kérdésnél: a hivatkozások hozzárendelése az idézett szerzőkhöz, megszámlálás, összehasonlítás. A második kérdésnél: minden sokat idézett szerző páronkénti összehasonlítása az együttidézés ismerve szerint, az együtt idézett szerzők klaszterezése. A harmadik kérdésnél: minden korabeli publikáció páronkénti összehasonlítása a bibliográfiai kapcsolat ismerve szerint, a related papers nyilvántartása.

A szabályok egyszerűségét a szakértők rendkívül nagy számával kell ellensúlyozni. De létezik-e egyáltalán arany szabály a szakértő rendszerek számára? Az SCI esetében, a bemutatott elemzési és információs lehetőségek fényében indokoltnak látszik az a megállapítás, hogy egy kivételes szakértő rendszerrel van dolgunk.

Marad azonban még egy fenntartás, amely az egyes szakértők jelentőségét és ezzel a rendszernek mint szakértő rendszernek a jelentőségét látszólag

kisebbiteni kívánja. Eszerint elég csekély az a hozzájárulás, amelyet az egyes "szakértő", vagyis a kutató az SCI óriási rendszerébe bevihet, azaz közleményeinek irodalomjegyzéke. De: ahogyan egy tudományos publikáció elkészítése, a hozzá tartozó irodalomjegyzéké is jellegzetesen szakértői teljesítmény, amelyet a kutató helyett sem egy automatizált rendszer, de még egy knowledge engineer sem végezhet el. A kutató maga bocsátja további felhasználásra, azaz a szakértő rendszer rendelkezésére a szükséges ismereteket az oly sok tisztázatlan, megmérhetetlent hordozó idézéseken keresztül.

Alig hiszem, hogy egy ilyen szemlélet a "klaszikus" szakértő rendszerek dicsőségét vagy exkluzivitásának mítoszát csökkentené, sokkal inkább kiterjeszti azon problémák szemléletét, amelyek csak több tízezer kutató – akik szakértői tudásukat viszik ebbe a tevékenységbe – közös és felelős működésének eredményeképpen oldhatók meg. Így értendő tehát ennek a közleménynek a szokatlan címe.

Irodalom

- [1] ROTH, M.: Die intelligente Maschine. Leipzig, Jena, Berlin, Urania Verlag, 1988.
- [2] KESSLER, M. M.: An experimental study of bibliographic papers. = IEEE Transactions on Information Theory, 9. köt. 1. sz. 1963. p. 50–51.
- [3] MARSÁKOVÁ, I. V.: Sistema svazej među dokumentami postroennaá na osnove ssylok: (po dannym SCI). = Naučno-tehničeskaá informáciá, Ser. 2. 6. sz. 1973. p. 3–8.
- [4] SMALL, H.: Co-citation in the scientific literature: a new measure of the relationship between two documents. = Journal of the American Society of Information Science, 24. köt. 1973. p. 256–269.
- [5] Science Citation Index. Compact Disc Edition, Institute for Scientific Information, Philadelphia, 1988. (Demonstration disc)
- [6] BONITZ, M.: Information – Wissen – Informatik. = 16. Internationales Kolloquium über Information und Dokumentation, 18–21. Dezember, 1989. Vortrag.
- [7] MIRIMANOVA, M. S.: Ekspertnye sistemy: Izlečenie ekspertnyh znaniy. = Teoria i praktika obščestvenno-naučnoj informacii, 1. köt. 1989. p. 138–151.
- [8] GARFIELD, E.: Citation indexing, its theory and application in science, technology and humanities. ISI Press, Philadelphia, 1979.

Fordította: Zsindely Sándor

Beérkezett: 1990. V. 23-án.