



# Beszámolók ◦ Szemlék ◦ Közlemények

## TÁJÉKOZTATÁSI MUNKA ÁLTALÁBAN

### A tudományos szakirodalom szerkezeti dinamikája

A tudomány művelésének – új ismeretek létrehozásának vagy közvetítésének – folyamata struktúrákat hoz létre, amelyekkel kapcsolatban a legnagyobb probléma az, hogy ezek láthatatlanok.

A kérdés ezek után az, hogyan fedezhetjük fel ezeket a struktúrákat, hogyan kísérhetjük figyelemmel változásaikat. A leginkább hozzáférhető és legkevésbé torz forrást alighanem a tudományos folyóiratirodalom és a hivatkozások nyújtják, bár kétségtelen, hogy a hivatkozások a tudományos tevékenységnek csak melléktermékei. Weinstock kimerítően feltárta: miért hivatkoznak a szerzők másokra. Az okok és motivációk éppen olyan összetettek és változók, mint a tudományos gondolkodás. A legtöbb hivatkozás mégis a tudományos dokumentumok közötti valós fogalmi kapcsolatokat tükrözi.

A struktúra felismerésének talán legjobb megközelítése egy tudományterképező rendszernek nevezhető eljárás volna, amely a Science Citation Index (SCI) kumulációjából indulhatna ki. A kapcsolódások alapvető mértéke az együttes hivatkozás lehetne, azaz két dokumentum együttes hivatkozása egy későbbi dokumentumban. Hipotézisünk, hogy az együttes hivatkozás az esetek többségében a cikkek által reprezentált eszmék vagy személyek közti elemi asszociációt jelzi.

Az első kísérletek alapvető tanulsága, hogy a tudomány elsődleges strukturális eleme a tudományos szakterület. A szakterületek olyan laza hálózattá kapcsolódnak össze, amely a tudomány teljes építményét átfogja. A jelen tanulmány célja, hogy a statikus, keresztmetszeti modell helyett dinamikus, időben változó tudomány szerkezeti modell kialakításához járuljon hozzá.

### A módszer

Mint hogy a tudományterképező rendszer célja az adott időben észlelhető struktúrák azonosítása mellett, ezek időbeni változásának mérése, arra van szükség, hogy a hivatkozási adatok elvont időegységeivel dolgozzunk. Az SCI éves kumulációi megfelelő időegységeket jelentenek. Az SCI multidiszciplináris hivatkozási index, a természettudományok minden területéről kb. 2400 folyóiratot dolgoz fel. A jelen cikkben ismertetett vizsgálathoz az 1973-as és 1974-es évfolyamokat használták fel:

kiválasztottak minden olyan tételt, amelyre az adott évben egy meghatározott küszöbértéknél (a mi esetünkben 15-nél) többször hivatkoztak;

megállapították az ezekből azonosított párok együttes hivatkozottsági gyakoriságát;

ezek alapján tételcsoportokat (clustereket) képeztek megfelelő algoritmussal;

meghatározták a clusterek egymásközti kapcsolatait az eltérő clusterekbe tartozó, de együttesen hivatkozott tételek száma alapján;

összevetették az egymást követő években, egymástól függetlenül nyert clustereket, meghatározva az új, a kiesett, ill. a folytatódó clustereket.

A kiválasztási küszöbérték megállapítása igen lényeges, ettől függ ugyanis a clusterek tartalmi részletessége. A kísérletek szerint nincs olyan küszöb, amely valamennyi tudományágra nézve optimális lenne. Az alkalmazott érték mellett azonban a nagyobb tudományterületek interdiszciplináris szerkezetére vonatkozólag lehet bizonyos megállapításokat tenni.

Az együttes hivatkozási párokat a forrás (hivatkozó) cikkek alapján határozták meg, majd rendezték, s ezután meghatározták gyakoriságukat. Ezt a gyakoriságot a két összetevő tétel saját gyakorisága alapján (százalékos értékre hozva) normalizálták, s a clusterképző algoritmus során itt is alkalmazták küszöbértéket (16%). A cluste-

rek közti kapcsolat vizsgálatára szolgáló asszociációs mértékként azon tételek számát használták, amelyek, bár különböző clusterben helyezkednek el, előfordulnak együttesen hivatkozott tételekként.

**Az 1973-as és 1974-es cluster-statisztika összevetése**

1973-ról 1974-re a forrásdokumentumok (hivatkozási dokumentumok) száma kis mértékben (1,5%) csökkent, ugyanakkor a hivatkozások (4,3%) és a hivatkozott dokumentumok száma (2,3%) növekedett. *Az egy hivatkozott tételre jutó hivatkozások átlagos száma 1,81-ről 1,84-re, és az egy hivatkozó forrásban szereplő hivatkozások száma 12,1-ről 12,8-ra emelkedett.* Az eltérések azonban inkább az SCI-re, mint a szakirodalomra jellemzőek.

*A gyakran hivatkozott, vagyis 15-nél többször előforduló tételek száma 8%-kal emelkedett, s ennek megfelelően emelkedett az ilyen tételekre is hivatkozó források mennyisége is.*

A gyakori hivatkozások közül *az együttes hivatkozási párok* mennyisége értelemszerűen még nagyobb ütemben nőtt: 25%-kal. Növekedett *a hivatkozott dokumentumok kapcsolatainak* szorossága is: egyrészt az összes lehetséges páros kapcsolatok közül a ténylegesen előfordulók aránya, másrészt pedig az adott hivatkozott tételhez tartozó párok átlagos száma. Mindebből azonban még nem vonhatjuk le azt a következtetést, hogy a természettudományok egyre jobban összefüggnek egymással.

A cluster-képzéshez használt 16%-os küszöb mellett 1973-ban 1610 két vagy több tételt tartalmazó csoportot, 1974-ben 1702 csoportot találtak. A clusterek átlagos tételszáma 5-ről 5,2-re emelkedett.

A kiválogatott tételek mintegy 51%-a került be – ilyen küszöb mellett – valamelyik clusterbe. A forrásdokumentumok esetében hasonlóan alakult ez az arány. Minthogy a kiválasztott forrástételek az összesnek kb. 36%-át teszik ki, azt mondhatjuk, hogy az ilyen küszöbszintű csoportosítás csak a tételek 18%-át sorolja be. A küszöb csökkentésével természetesen ez az arány növelhető.

Az egyes forrástételek 1973-ban átlag 1,5, 1974-ben 1,6 clusterben voltak megtalálhatók, másszóval: ha egy hivatkozott tétel pontosan egy clusterbe kerül, akkor a forrásdokumentumok átlagosan egynél több helyen lelhetők fel. *A különböző clusterekbe tartozó tételek együttes hivatkozása a két évben csaknem azonosan gyakori, s az átlagos kapcsolatszám valamennyi cluster esetében minimális mértékben emelkedett.*

**A változás mértéke**

Az 1973-as és az 1974-es szerkezet meglehetősen hasonló, ez a relatív állandóság azonban számos belső



1. ábra

változást takar, ha a clustereket alkotó tételeket tekintjük. A belső változás mérésére *azt vizsgáltuk, hogy a clustert alkotó tételekből mennyi nem változott egyik évről a másikra.*

A clusterbe sorolt tételek száma 1973-ban 8042, 1974-ben 8917 volt, a mindkét évben előfordulóké 4046; az 1973-ban szereplő tételek közül 3996 kiesett, viszont 1974-ben 4871 új tétel került be (1. ábra). Eszerint tehát *a clusterek számának csaknem változatlan volta mellett az azokat alkotó tételek jelentősen változtak.*

Figyelemre méltó, hogy *Price és Gursey* hasonló eredményre jutott, amikor azt vizsgálta, hogy egymást követő években *mennyire állandó a publikáló szerzők névsora* (2. ábra).



2. ábra

A clusterekben megmaradó dokumentumok vizsgálata elvezet a folytatódó clusterek vizsgálatához, utóbbin olyan clustert értve, amelynek egy vagy több alkotó-eleme megmarad. Ilyen értelemben az 1973-as 1610 cluster közül 1004 (62%) folytatódott 1974-ben, és megfordítva, az 1974-es 1702 cluster közül 1051 (62%) 1973-ról folytatódott. A változás üteme lassúbb, mint a hivatkozott dokumentumoké, de még így is jelentős (3. ábra).



3. ábra

Amint az várható is, a folytatódó clusterok átlagos mérete nagyobb, mint az újaké, vagy a kiesőké. Ha azonban az új clusterokat úgy tekintjük, mint amelyek az ismeretek új területeit tükrözik, azt kell mondanunk, hogy a tudományok elég gyors ütemben változnak.

Említésre méltó, hogy amíg az új tételek aránya valamennyi clusterben 55%, addig a folytatódó clusterokban csak 40%. Az a folytatódó cluster, amelyben az átlagosnál jóval több új dokumentum bukkan fel, a téma jellegének viszonylag *lényeges változására utalhat*.

A cluster folytatódásának többféle formája lehet. A legegyszerűbb és leggyakoribb az adott dokumentum egyszerű továbbélése ugyanazon clusterben. Az 1973 évi clusterok közül 1004 volt folytatódó, s ebből 837 (83%) nem vált szét két vagy több clusterre. Hasonlóan, az 1974-ben folytatódott clusterok 86%-a *nem két vagy több korábbi clusterből keletkezett*. A nem túl nagy mennyiségű szétválás, ill. összeolvadás mégis vizsgálatra volna érdemes, mert ezek lényeges változásokra mutathatnak.

#### Változások a cluster-konfigurációban

Mint már jeleztük, minden cluster-párhoz meghatároztunk egy *asszociációs mérőszámot*, amellyel a cluster-rendszer időbeli változása is követhető.

Átlagosan minden cluster 32 további clusterhez kapcsolódott kettős hivatkozás révén. Ha maga a cluster folytatódó volt, más, szintén folytatódó clusterokhoz átlagosan 19 ízben kötődött, de maga a kapcsolat folytatódása nem szükségszerű. A helyzetet tovább bonyolítja, hogy ha két cluster asszociációja is folytatódott, ennek erőssége még mindig lehet változó.

A szerkezeti változások elemzéséhez először is egy *kiinduló clustert (centroidot) választottunk ki, és meghatároztuk a hozzá kapcsolódó clusterokat*. A centroid körül így clusterokból álló clusterok (másodrendű clusterok) keletkeztek. E folyamat során olyan küszöbértéket alkalmaztunk, hogy e másodrendű clusterok mérete kezelhető maradjon. Ezek a clusterok (és a bennük tükröződő struktúrák) azután időben előre és visszafelé követhetők. Azokat a clusterokat, amelyeknek sem elődjük, sem követőjük nincs, kizártuk. Az elemzés ennek megfelelően a folytatódó clusterokra szorítkozott.

/SMALL, H. G.: *Structural dynamics of scientific literature = International Classification*, 3. köt. 2. sz. 1976. p. 67–74./

(Sárdy Péter)



## Az informatikai alap kutatások és az információs technológia fejlődése

Az utóbbi két év során a tudományos információk terjesztésének szervezésében az alábbi *tendenciák* váltak jellemzővé:

szervezetek sora egyesült közös információs célkitűzések megvalósítására;

nemzeti és nemzetközi információs hálózatok alakultak ki;

ugyanakkor azonban a kiadók, az adatbázisok létrehozói, a könyvtárak és az információs szolgáltatások közötti kapcsolatok bizonytalanná, egyenlőtlené váltak.

Ez a három tényező együttesen azt valószínűsíti, hogy *a nemzeti vagy nemzetközi együttműködés feltétlenül új technológiák és módszerek alkalmazását követeli meg*.

### 1. A módszerek és eljárások fejlődése

#### 1.1 A technika változása

Az utolsó öt év folyamán évente megkétszereződtek a számítógépek kapacitását és az on-line tárolási kapacitást jelző paraméterek. Éppen ezért az információkezelés technológiája is nagy változásokon ment keresztül. Ma már elterjedten használják erre a célra az olyan korszerű számítástechnikai eszközöket is, mint például a *display* és az *optikai karakterolvasó*; emellett pedig rohamosan fejlődik a *reprográfia* és a *mikrofilmtechnika* használata (becslések szerint a világon ma már mintegy 1,5 milliárd fényképmásolatot készítenek naponta). Megváltozott a mikrofilm szerepe is: a korábbi archiválási eszközből a kiadás és az információs szolgáltatás eszövévé vált.

Megfigyelhető azonban még a hagyományos eszközöket alkalmazó *kommunikációs tevékenység* jelentős mennyiségi növekedése is, így például

a tudományos könyvkiadás az utóbbi hat évben 40%-kal nőtt az Egyesült Államokban;

a tengerentúli telefonbeszélgetések száma 1965 óta megtízszereződött.

Jelentősen kiszélesedett az információs rendszerekben az információkezelésre alkalmazható *eszközök választéka* és a működő berendezések száma. A lehetőségeket tovább bővíti, hogy ezek az eszközök (mágnesszalagos berendezések, mikrofilmtechnikai eszközök, ultrarövidhullámú vagy koaxiális kábeleket alkalmazó távközlési technikák) nemcsak önállóan, hanem célszerűen kialakított kombinációkban is alkalmazhatók.

Megjegyzendő másrésről az is, hogy az információs technika azon ritka területek közé tartozik, ahol *a fajlagos költség csökken*: így például az az információkezelés (adatfeldolgozás), amely 1960-ban 2000 dollárba került, 1975-ben 1 dollár költséggel volt végrehajtható.