

tájékozódik folyamatosan a gépileg kezelt információ-tömegben on-line kapcsolat révén. A használó és az információ közötti kapcsolatban csak a visszamenőleges keresésnél lesz szerepe az információs szakembernek, de most már mint az információ kiértékelőjének és értelmezőjének. Ebből következően nem is annyira az informatikában vagy a könyvtártudományban, hanem az illető szakterületen kell szakképzettséggel rendelkeznie. Ahogy a szerző mondja: *kis jövő vár arra a kollégára, aki csak könyvtáros vagy informatikus lesz.* Más szóval, az önálló információs funkció magasabb szinten ismét visszatér oda, ahonnan kivált, megszűnik különállósága és ismét integrálódik a használó tevékenységébe.

Térjünk azonban vissza a jövőből, s *Wilson-Davis*-szel együtt vessünk egy pillantást a használók tömegére. Angliában a lakosság 70%-a soha nem használ könyvtárat, 50–60%-a soha nem vásárol könyvet. Ha az előző bekezdés jövőképe a használóknak csak egy szűkebb rétegére is érvényes, mégis részét képezi az egész társadalom információs struktúrájának, amelyben nyilvánvaló integrációs tendenciák nyilvánulnak meg. Az információ társadalmi mozgását, a tájékozódási szokásokat vizsgáló kutatások egyazon tárggyal foglalkoznak, legyen szó futballeredményekről vagy a társadalomtudósok könyvtárhasználatáról. *Az információs igényeket és szükségleteket széles társadalmi beágyazottságukban kell kutatni*, kiterjedve az egyén és társadalmi csoportok életmódjára is. Csakis ezen a módon lehet megállapítani az információs rendszerek hatékonyságát, s azt, miképpen lehet és kell fejleszteni őket.

A kutatás tárgyának és témáinak kijelölése mellett hasonlóképpen fontos a módszerek kidolgozása, amelyre az olvasói szokásokkal foglalkozó kutatások központja nagy súlyt helyez. Éppen módszertani tekintetben hagytak sok kívánnivalót maguk után a korábbi vizsgálatok, s ezért nem vethetők egymással össze, erősen vitatható mintavételi eljárásokat alkalmaztak, szűkös és egyoldalú módszertani apparátusuk. A központ tervei között ezért egy *módszertani segédkönyv kiadása* is szerepel, amelyben többek között helyet kapnának a különféle vizsgálati eljárások leírásai, minta-kérdőívek, a szabványosított szakkifejezések szótára stb.

\* \* \*

Elhamarkodott dolog lenne egy konferencia három előadása nyomán bármiféle általánosító következtetést levonni egy szerény ismertetésben. A legtöbb, amit mondhatunk az, hogy az előadók gondolatai és megállapításai *az informatika használó-orientáltságának az erősödését jelzik*, amely részben bizonyos technikai problémák megoldásának, még inkább a társadalmi-gazdasági tényezők befolyásának a következménye. S úgy látszik, ez kedvező alapot nyújthat más-más társadalmi rendszer-

ben működő szakemberek között termékeny párbeszéd folytatásához

*WILSON-DAVIS, K.: The centre for research on user studies: aims and functions = Aslib Proceedings, 29. köt. 2. sz. 1977. p. 67–76.*

*SAUNDERSON, M.: The information function in relation to user departments: an excursion into the future = Aslib Proceedings, 29. köt. 2. sz. 1977. p. 77–90.*

*SWEENEY, G. P.: The use of national resources to encourage the more effective use of information by industry = Aslib Proceedings, 29. köt. 2. sz. 1977. p. 91–103.*

(Papp István)



### Gondolatok a dokumentumok avulási idejének meghatározásával kapcsolatban

A kutatók és a gyakorlati szakemberek is foglalkoznak a tudományos információ egyik fontos tulajdonságával, az avulással. Avuláson azt értjük, hogy az információ az idő múlásával elveszti értékét. Az avulás kutatása az információs tevékenység gyakorlata szempontjából fontos, mert megismerése alapul szolgálhat az információérés és állománygyarapítás visszamenőleges értékének meghatározására, információtárak állományfejlesztési elveinek, a retrospektív bibliográfiakészítés szabályainak felállítására stb.

Az információ avulása a dokumentumok avulásában mutatkozik meg, a dokumentumok avulása viszont abban, hogy csökken az irántuk való érdeklődés, csökken használatuk intenzitása és gyakorisága. Tehát *a tudományos-műszaki irodalom avulása nem más, mint a tudósok és szakemberek által történő felhasználás csökkenése a megjelenés időpontja óta eltelt idő függvényében.*

Már többen megkísérelték különböző típusú és tartalmú dokumentumok avulását meghatározni, amihez természetesen a felhasználás intenzitásának meghatározására volt szükség. Vagy azt vizsgálták, hogy *hányszor hivatkoztak a konkrét dokumentumokra*, vagy azt, hogy *hányszor rendelték meg az eredetiket vagy mikromásolatukat a tájékoztató intézményeknél.*

Az avulási folyamat jellemzésére több kritériumot javasoltak és alkalmaztak. A leggyakoribb ilyen mérőszám a *felezési idő*, ami alatt felére csökken az adott kiadvány igénybevételi eseteinek száma.

Az avulás ütemének értékelésére gyakran az *avulás évi tényezőjét* számítják ki. Ezt úgy kapják, hogy az egyik év folyamán elért felhasználások számát ( $C_n$ ) elosztják az előző év hasonló ( $C_{n-1}$ ) adataival:

$$a = \frac{C_n}{C_{n-1}}$$

Brookes a könyvtári gyakorlat számára a folyóiratok célszerű megőrzési idejének kiszámításához bevezette az általa *általános hatásfoknak* nevezett paramétert:

$$u = \frac{1}{1-a}$$

Ritkábban fordulnak elő az olyan paraméterek, mint a *hivatkozások átlagos kora* vagy az *újdomsági tényező*.

Az egyes ilyen mutatószámok alkalmazásának a jogsultságát csak akkor állapíthatjuk meg, ha megvizsgáljuk a tapasztalati adatok előállításának modelljét.

Egy bizonyos  $t$  évben megjelent valamilyen dokumentumhalmaz avulásának megfigyelése ideálisan úgy történhetne, ha felmérnénk ennek a halmaznak a felhasználási gyakoriságát az első, a második ( $t+1$ ), a harmadik ( $t+2$ ) és következő években egészen addig, míg igénybevételük teljesen meg nem szűnik. A mérés egy gyakorisági sorozatot eredményezhetne:

$$C_t, C_{t+1}, C_{t+2}, \dots, C_{t+i}, \dots, C_{t+n} \quad (1)$$

A sorozat elemzésekor esetleg észlelünk statisztikai törvényszerűséget (ha van ilyen) a dokumentumhalmaz felhasználási gyakorisága és az idő között, és ha ez a függvény monoton csökkenő, akkor meg lehet határozni a felhasználási gyakoriság közepes csökkenési együtthatóját, az *évi átlagos avulási tényezőt*:

$$a = \frac{\sum_{i=1}^n C_{t+i}}{n C_{t+1}}$$

Az  $a$  érték csak az adott dokumentumhalmazra és a  $t$ -től  $t+n$ -ig terjedő időközre érvényes, mert a felhasználás megszűnése a  $t+n$  évben még nem bizonyítja, hogy nagyobb időköz elmúltával nem kerül-e a dokumentumhalmaz még felhasználásra. A technikatörténet tudományos és műszaki eszmék (és természetesen a dokumentumok) egész tömegének újjászületéséről tud.

A megfigyelt folyamatoknak a *felezési idővel* történő jellemzése tehát jogosulatlan, mert a felhasználási gyakoriság mérése egy bizonyos időre szorítkozik. Hiszen ha lehetséges újra felvillanni a függvény monoton csökkenése után, akkor egyedül a függvény ideiglenesen csillapított szakasza alapján lehetetlen az összes felhasználási esetnek a felét kiszámítani. Ugyanez áll az *általános hatásfokra* és a többi kritériumra is.

A felhasználási gyakoriság változását jellemző függvény ingadozásainál is súlyosabbak azok a hibák, amelye-

ket a legtöbb ismert kutató a tapasztalati adatok feldolgozásakor elkövet. Minthogy a fenti modell szerinti mérések végrehajtása igen bonyolult, ezért minden kutató úgy szerez adatokat, hogy a kutatás időpontjában rögzíti valamely konkrét tematikájú dokumentumok felhasználásának összes esetét; amennyiben a vizsgált dokumentumok más és más évben keletkeztek, akkor a mérések eredményeként más és más korú (1 éves, 2 éves, 3 éves, . . .  $i$  éves, . . .  $n$  éves) dokumentumok felhasználási gyakoriságát kapják:

$$C_1, C_2, C_3, \dots, C_i \dots C_n \quad (2)$$

Azután ezt a gyakoriság-sorozatot vizsgálják az (1) sorozat helyett, amikor is már az első tekintetre megmutatkozik az elvi eltérés. Míg ugyanis az ideális modell szerint az időben változatlan dokumentumhalmaz felhasználásának időbeli gyakoriságait kapjuk, addig a reális vizsgálatokban különböző halmazokra határozzák meg a gyakoriságot, s (1) felcserélése (2)-vel csak akkor engedhető meg, ha a különböző években létrehozott *dokumentumok halmazai mennyiségileg és minőségileg is azonosak*. Az első feltétel a legtöbb ismeretágban nem teljesül, mert különböző években különböző számú dokumentum születik, számuk exponenciálisan növekszik. Az ebből adódó helyzetet egy példával lehet érzékeltetni:

Tegyük fel, hogy valamely állomány 55 dokumentumból áll, közülük 1 dokumentum tízéves, 2 kilencéves, 3 nyolcéves, 4 hatéves stb. A folyó évi dokumentumok száma 10. Ha a vizsgálat ideje alatt minden dokumentum egyszer került felhasználásra (pl. kikölcsönözték), akkor az állomány felhasználási gyakorisága 55.

Az avulási idő általánosan elfogadott kiszámítási módszere szerint kapjuk, hogy a tízéves dokumentumok kiadásának (kölcsonzésének) gyakorisága 1 (sűrűség 1/55), az ideiek gyakorisága 10 (a sűrűség 10/55). Nyilvánvaló ugyan, hogy mindegyik dokumentumot ugyanolyan gyakran adták ki az olvasóknak, mégis maga a számítási módszer azt sugalmazza, mintha a tízéves dokumentumok felhasználása tízszerre kisebb volna, mint az idei dokumentumoké.

A valóságban természetesen mások a korcsoportok számarányai, ezért a számítási hibák nem olyan nagyok, mint az előbbi példánkban. Sokszor azonban mégis számottevők. Ha például a kémiai könyvek avulására vonatkozó adatokat nézzük, amelyek a folyóiratokban előfordult hivatkozások százalékarányából és abból a törvényszerűségből adódtak, hogy a kémiai irodalom volumene 8,2 év alatt megduplázódik, az *1. táblázat* alapján kiderül, hogy a 40–50 éve megjelent könyvekre való hivatkozások aránya egynegyede a legfrissebb könyvekre való hivatkozások százalékának, és nem 1/64 része, ahogy azt korábban számították.

A példa Mihajlov – Csernűj – Giljarevszkij: Az informatika alapjai c. könyvében közölt számításokat tartalmazza. Ennek alapelvei a következők:

az egyszerűség kedvéért csak a monográfiákra való hivatkozásokat veszik számításba;

megvizsgálták, hogy a folyóiratokban található összes olyan hivatkozás közül, amely monográfiákra vonatkozik, hány százalék utal 0–10 éves, 11–20 éves stb. monográfiára, így kapták a  $C_i$  százalékos értékeket;

az egyszerűség kedvéért feltételezték, hogy a kémiai irodalom volumene 10 évenként duplázódik meg. Így, ha a 0–10 éves művek száma  $A$ , akkor a 11–20 éves műveké  $0,5 A$ , a 21–30 éveseké  $0,25 A$  stb.;

az összes hivatkozás  $B$  számából az egyes korcsoportokra esik:

$$B_i = \frac{1}{100} C_i B;$$

végül azt számították ki, hogy egy-egy korcsoportban relatíve, vagyis a megjelenések darabszámára vonatkoztatva mennyi a hivatkozás (azaz milyen mérvű a felhasználás):

$$f_i = \frac{1}{100} \frac{C_i \cdot B}{A_i}$$

A kémiai monográfiák körében tehát a félévszázados könyvekre aránylag csak négyszer ritkábban hivatkoznak, mint a legújabbban megjelent művekre.

1. táblázat

A kémiai irodalom tényleges hivatkozási gyakoriságának számítása

Korcsoportok sorszáma (i)	Művek életkora év	A folyóiratcikkekben megjelent hivatkozásoknak hány %-a jut ezekre a könyvekre ( $C_i$ ) %	Hány műből áll a korcsoport ( $A_i$ )	Az i-edik korcsoport könyveire való hivatkozások száma a felmérés pillanatában ( $B_i$ )	Hogy aránylik a könyvekre való hivatkozások száma az i-edik korcsoport könyveinek számához ( $f_i$ )
1	0–10	64	$A_1$	$0,64 B_1$	$\frac{B_1}{A_1}$
2	11–20	24	$0,5 A_1$	$0,24 B_1$	$\frac{B_1}{A_1}$
3	21–30	6	$0,25 A_1$	$0,06 B_1$	$\frac{B_1}{A_1}$
4	31–40	2	$0,125 A_1$	$0,02 B_1$	$\frac{B_1}{A_1}$
5	41–50	1	$0,062 A_1$	$0,01 B_1$	$\frac{B_1}{A_1}$

Hasonló képet mutatott a szilárdtest-fizikai cikkekre való bibliográfiai hivatkozások elemzése (2. táblázat). E cikkek száma évenként  $g = 1,134$ -szeres mértékben növekszik. Ha az évi növekedési tényezőt változatlanul tekintjük, akkor a

$$g_i = g^{t_i - t_0}$$

összefüggés adja a valamely  $t_i$  évben megjelent cikkek  $g_i$  növekedését a kiindulási  $t_0$  évben megjelent cikkek számához viszonyítva.

2. táblázat

A szilárdtest-fizikai folyóiratcikkek tényleges hivatkozási gyakoriságának kiszámítása

$t_i$	$B_i$	$g_i$	$A \cdot f_i$	$\frac{f_i}{f_{\max}}$
1969	39	13,984	—	—
1968	204	12,332	16,542	—
1967	248	10,875	22,804	1
1966	209	9,590	21,793	0,955
1965	153	8,457	18,091	0,793
1964	150	7,458	20,112	0,881
1963	147	6,577	22,350	0,980
1962	102	5,800	17,585	0,771
1961	85	5,115	16,617	0,728
1960	50	4,511	11,084	0,485
1959	67	3,978	16,842	0,738
1958	48	3,508	13,683	0,600
1957	36	3,094	11,635	0,510
1956	38	2,729	13,924	0,610
1955	29	2,407	12,048	0,528
1954	18	1,122	8,478	0,371
1953	21	1,873	11,211	0,491
1952	25	1,652	15,133	0,663
1951	26	1,457	17,844	0,782
1950	21	1,285	16,342	0,716
1949	16	1,134	14,109	0,618
1948	12	1	12,000	0,526
1947	3			
1946	2			
1945	2			
1944	1			
1943	—			

Mint ahogy a második világháború alatt  $g$  értéke valószínűleg más (kisebb) volt, ezért a kutatók az 1948. évet vették  $t_0$  kiindulási évné. A *Philosophical Magazine* 1969-ben megjelent 19. és 20. kötetében található  $B_i$  hivatkozásokat használva, meghatározták a relatív hivatkozási gyakoriságot tetszőleges  $t_i$  évre vonatkozóan, ugyanez év publikációinak számára vetítve:

$$f_i = \frac{B_i}{A \cdot g_i}$$

ahol  $A$  az 1948-ban megjelent publikációk száma.

Ha elosztjuk az évenkénti gyakoriságot ( $f_i$ ) a maximális gyakorisággal, amelyet történetesen 1967-ben tapasztaltak ( $f_{\max}$ ), akkor megkapjuk azt a szintet, ahova a hivatkozások a dokumentumok szaporodása közben leestek. Így látható, hogy az arány az 1948. évi és az 1967. évi publikációkra való hivatkozások között csak 2 (pontosabban 1 : 0,52), ez a szám jellemző a húsz éves irodalom hasznosságára. A hagyományos eljárással 21-szer kisebb értéket kapnánk (248 : 12), mert az nem vette figyelembe, hogy 1967-ben csaknem 11-szer annyi (1,134<sup>19</sup> = 10,875) volt a publikációk száma, mint 1948-ban.

Az információáramlás dinamikáját figyelembe lehet venni a felezési idő módszerének alkalmazása során is, ha bevezetjük a *helyesbített felezési idő* (corrected half life) fogalmát, vagyis évenként helyesbítünk az évenkénti irodalom volumenét tekintve. E módszer nem ad elég szemléletes képet az irodalom avulásáról, és a könyvtárak vezetői számára érdektelen.

Elvileg le kell szögezni, hogy a legutóbbi időkhöz uralkodó volt az a nézet, hogy a szakirodalom gyorsan avul, s ez a nézet igen veszélyes lehet.

A valóságban valamely szakterület egy év alatt megjelent műveire való *hivatkozások száma* (tehát a használat mérőszáma) az évek múlásával sokkal lassabban csökken, mint ahogy eddig hittük, a fentiek szerint 10–15-ször lassabban. Az állománygyarapítás mélységének, a visszakeresés időhorizontjának, az állományapasztás időszakainak a megállapítása szempontjából éppen ennek a dinamikának van nagy jelentősége, nem pedig az olyan elvont fogalmaknak, mint a frissesség, hasznosság, felezési idő.

A vizsgálatokat végző kutatók gyakran mellőznek fontos tényeket. Például felmérték a tudományos könyvtárak állományrészeinek olvasottságát, meghatározták, hogy az 1917–1940, 1941–1945, 1946–1950 és a további öt éves időszakok alatt beszerezett műveket hány-szor kérték ki, s ezt a számot az időszak állományegységeihez viszonyították. Az első két időszakra vonatkozóan 0,4 adódott, majd 1946–1950-re és 1951–1955-re 0,8. További számításokból 1956–1960-ra 1,0 és 1966–1970-re 0,7 adódott. Maximumot, 1,5 arányt kaptak az 1961–1965 időszakra, amiből az látható, hogy a harmincévesnél régebbi irodalom használata csak 3,75-ször kisebb, mint a leginkább használt időszaké (pontosan 1,5 : 0,4), mégis a kutatók a felmérésből azt a következtetést vonták le, hogy számszerűleg az összes kérés 90%-át ki lehet elégíteni az utóbbi 15–20 év anyagából, tehát a műszaki irodalom gyorsan avul.

A használat vizsgálatok a folyó év kiadványaival kapcsolatban figyelembe kell venni, hogy ezek nem

egyszerre válnak ismertekké. Tegyük fel, hogy a vizsgált évben  $D$  számú dokumentum jelent meg, az ezekre való hivatkozásokat vizsgáljuk. Amennyiben feltételezhető, hogy ez a  $D$  mennyiség egyenletesen növekedett 0-tól (dokumentumok száma az év elején) kezdve, akkor az

információfelhasználók az év folyamán csak  $\frac{D+0}{2} = 0,5$

$D$  dokumentummal rendelkeztek. Ez azonban csak akkor igaz, ha a vizsgálatot az év végén végezzük. Az év közepén ez a szám csak  $0,25 D$ . Ha még azt is figyelembe vesszük, hogy az anyag elkészültétől a kiadásáig legalább fél év eltelik, akkor a folyó év anyagaira hivatkozó közlemények szerzői legfeljebb e dokumentumok egyne-gyedével ismerkedhettek meg, azaz legfeljebb ennyire hivatkozhattak. Ezért a hivatkozásokat ábrázoló görbén a folyó évre vonatkozóan kapott adatokat legalább négydel kell szorozni.

Még az ilyen korrekció után is meg kell vizsgálni, hogy *érvényes-e a következő két axióma*:

a dokumentumok használatának gyakorisága csak koruktól és tartalmuktól függ;

a dokumentumokra való hivatkozásokban, valamint az információtárakhoz beérkező kérésekben valóban megmutatkozik a tényleges használat.

Az első feltevés ellen az szól, hogy a használatot vizsgáló felmérések más és más eredményt adnak, ha a vizsgálat az év közepén, illetve év végén történt; másfel-szeres, sőt több mint kétszeres eltérések is tapasztalhatók. *Tehát az első axióma nem érvényes.*

Nyilván a felhasználókat tájékoztató rendszer állapota is lényeges hatással van a használat gyakoriságára, mert megjelenhetnek új információk kiadványok vagy szolgáltatások stb.

A hivatkozások és kérések száma valóban a használatot, a hasznosítást mutatja-e? Sokszor fordulnak elő hivatkozások elavult művekre, éppen azért, hogy elavulásukat tudtul adják. Sokszor a ténylegesen használt anyagokra nem a szokásos technikával hivatkoznak, hanem csak az ismert szerző nevét említik a szövegben.

Az ismertett eljárásokról kimondható, hogy *az elavulás megállapítása céljára nem elég objektívek*, legfeljebb arra jók, hogy az állományt aktív, passzív és letéti raktári részekre lehessen bontani.

Az avulás megállapítása céljára objektívebb kritériumokat kell keresni. A technikában ilyen kritérium lehet például az, hogy egy-egy találmányt ténylegesen alkalmaznak-e, az elméleti tudományok területén pedig a különböző korú dokumentumokra való hivatkozások összeszámolása mellett a hivatkozások okát is elemezni kell, s a rejtett hivatkozásokat is fel kell kutatni.

## Irodalom

- [1] CSERNŰJ, A. I.: O nekotorüh central'nuh problemah informatiki. Problemü informatiki. Moszkva, VINITI, 1973. p. 3–21.

- [2] BURTON, R. E. – KEBLER, R. E.: The „half life” of some scientific and technical literatures = American Documentation, 71. köt. 1. sz. 1960. p. 18–22.
- [3] LINE, M.B.: The „half life” of periodical literature: apparent and real obsolescence = Journal of Documentation, 26. köt. 1. sz. 1970. p. 46–52.
- [4] BROOKES, B. C.: Obsolescence of special library periodicals: sampling errors and utility contours = Journal of the American Society for Information Science, 21. köt. 5. sz. 1970. p. 320–329.
- [5] POPILOV, L. JA.: K voproszu o kriterijah kacsesztvennoj i kolicesztvennoj ocenki informacionnüh potokov i trudoemkoszti obrabotki isztochnikov informacii = Naucsno-Tehnicoszkaja Informacija, 2. sor. 3. sz. 1968. p. 3–6.
- [6] DOBROV, G. M.: Nauka o nauke. Kiev, Naukova Dumka, 1966. 270 p.
- [7] MIHAJLOV, A. I. – CSERNÜJ, A. I. – GILJAREVSZKIJ, R. SZ.: Osznovü informatiki. Moszkva, Nauka, 1968. 756 p.
- [8] BRUKSZ, B. SZ.: Sztarenie naucsnoj literaturü. Problemü informatiki. Moszkva, VINITI, 1973. p. 74–102.
- [9] VICKERY, B. C.: Editorial note = Journal of Documentation, 26. köt. 1. sz. 1970. p. 53–54.
- [10] BROOKES, B. C.: The growth, utility, and obsolescence of scientific periodical literature = Journal of Documentation, 26. köt. 4. sz. 1970. p. 283–294.
- [11] VJAL'JAOTSZ: Vozrasztnoj aszpekt ocenki fondov naucsnoj literaturü = Knygotyra, 11. köt. 4. sz. 1974.
- [12] KOZACSKOV, L. SZ.: Szisztémü potokov naucsnoj informacii. Kiev, Naukova Dumka, 1973. 197 p.
- [13] JUNISZISZT. Doklad ob iszszledovanii vozmoznosztej szozdanija vszemirnoj szisztémü naucsnoj informacii. Moszkva, VINITI, 1972. 192 p.
- [14] Perszpektivü razvitija fondov naucsnuh bibliotek. Szbornik naucsnuh trudov, 15. sor. Moszkva, Goszudarsztvennaja Biblioteka SZSZSZR im. V. I. Lenina, 1974. 82 p.
- [15] IVANOV, R. N.: Metodü analiza v upravlenii proceszszami naucsno-tehnicoszkaj informacii. Moszkva, Sztatistika, 1973. 112 p.
- [16] KASAFUTDINOVA, E. SZ.: O racional'noj organizacii dokumental'nüh fondov = Naucsno-Tehnicoszkaja Informacija, 2. sor. 3. sz. 1975. p. 7–11.

*MOTÜLEV, V. M.: Ob opredelenii vremeni sztarenija dokumentov = Naucsno-Tehnicoszkaja Informacija, 2. sor. 12. sz. 1976. p. 3–7./*

(Dezsö László)



A dokumentumok korának figyelembevétele a keresési folyamat során

Az információkeresés fő célja a felhasználó témájának megfelelő dokumentumok kikeresése. Ennek két fő oldala: a dokumentumokhoz valamely módszerrel tárgyszavak hozzárendelése (indexelés), és a felhasználó keresőképének, profiljának megszerkesztése. További, látszólag ezektől független szempontként felmerülhet az irodalom avulása, ami azt jelenti, hogy a felhasználók megkövetéseket tesznek a dokumentumok korára vonatkozólag.

Az itt ismertetett módszer lehetőséget nyújt a dokumentumok tartalmi információinak és korának együttes kezelésére, aminek alapján hatékonyabb visszakeresési algoritmusok alkothatók meg. Ennek alapja olyan matematikai modell, amelyben két változó szerepel: egyrészt a dokumentum és a keresőkép index-kifejezéseinek nyelvészeti azonossága, másrészt a dokumentum kora a kérdés feltevésének időpontjában.

*A nyelvészeti azonosság és a dokumentum korának közös kezelése*

Az információkereső gyakorlatban általában nem szorítkoznak kizárólag az index-kifejezések egyeztetésével kapott találatokra. Más kritériumokat is szoktak megszabni a relevanciára, így élnek a dokumentum típusára, nyelvére, publikálásának időpontjára stb. vonatkozó megszorításokkal is, amelyek mind az output volumenét csökkentik.

Másfajta keresőstratégia is alkalmazható bizonyos keresőrendszereknél. Ennek lényege az adatbázis összes dokumentumának rangsorolása egy, valamennyi dokumentumhoz hozzárendelt szám szerint, amelyet a kérdező által tett lehetséges megkötések szerint két vagy több változó alapján határoznak meg. Ez lehetővé teszi a kikeresendő dokumentumok halmazának definiálását egyetlen küszöbérték szerint, szemben az előbbi módszerrel, ahol ehhez többkomponensű vektor szükséges.

A kétféle általános módszerrel kapcsolatban felvetendő kérdések:

*a két módszer egyenértékű-e, vagy pedig – adott lehívási vagy pontossági értékek mellett – az adatbázis különböző értékeléséhez vezet-e?*

*létezik-e egyértelmű kritérium annak eldöntésére, hogy meghatározott igények kielégítésére melyik módszert válasszuk?*

*hogyan lehetne olyan elméletet felállítani, amely gyakorlatban segíthet a két módszer közötti választásban?*

A fenti kérdésekre adandó válaszokat a bevezetőben említett két változó összevetésével lehet megfogalmazni: az index- és a keresőkifejezések nyelvészeti azonossága és a dokumentum kora jelenti a két változót. E két változó két súlyozó tényezőt rendel minden dokumentumhoz, a tényezők kombinációjával egyetlen súlyérték képezhető.

*Az irodalom avulása*

*Az irodalom avultságának kettős értelme lehet:*

a) a dokumentumban foglalt információk értéke csökken a dokumentum korának növekedésével;

b) a dokumentum felhasználója úgy viselkedik, mint ha a) igaz volna.