

KULCSSZAVAK

A tudományos folyóiratcikk-címek információtartalmának alakulása az idő és a szakterület függvényében

Napjainkban már számos, címszavakból összeállított KWIC-index áll rendelkezésre (Chemical Titles, Biological Abstracts tárgyszóindexe stb.) és több olyan adatbázis is működik, amelyekben a keresés címszavak alapján végezhető. A tudományos cikkek címszavainak információtartalmát már sokan vizsgálták, azzal a kérdéssel azonban, hogy a címek alapján milyen hatékonysággal végezhető a keresés a nem-természettudományi (orvostudomány, pszichológia, történelem stb.) területeken, eddig még nem foglalkoztak.

A tanulmány egy felmérés előkészületeit, eredményeit és tanulságait ismerteti, amelynek során 15 folyóirattól kiválasztott címek információtartalmának változását vizsgálták. A cél az volt, hogy megállapítsák, mennyire alkalmasak a címek KWIC-index összeállítására és számítógépes keresésre.

A vizsgálathoz 11 angol, 2 német és 2 francia folyóiratot használtak fel különböző szakterületekről. Az időbeli változás megállapítására az 1947, 1962 és 1973 év adatait vizsgálták. Az adott folyóirattól a három évben megjelent címekből kiválasztottak 100–100 címet és ezek információtartalmát vizsgálták a címben előforduló kulcsszavak alapján.

A vizsgálat előtt meg kellett határozni, hogy melyek a kulcsszónak tekinthető szavak – sőt továbbmenve – mit tekinthetünk egyáltalán szónak.

A szerzők minden olyan szót kulcsszónak fogtak fel, amely nem szerepelt az általuk összeállított 230 szavas ún. tiltó-jegyzéken. A tiltott szavak főleg névelők, névmások, elöljárók stb. voltak, de szerepeltek köztük általánosan használt szavak is, mint pl. *probléma*, *módszer*, *különböző* stb.

Nagyon nehéz szabályt felállítani a címszavak elkülönítésére. A szerzők a kémiai vegyületeket, elemeket (*NaCl*, *Fe* stb.), továbbá a számmal írt évszámokat, a rövidítéseket, az értelemszerűen összetartozó szóösszetételeket (pl. *non-conventional*) egy szónak tekintették. Ezzel szemben az összetett kémiai elnevezések (pl. *hydrogen cyanide*), valamint az olyan szóösszetételek, amelyeknek tagjai különállóan is lényegesek lehetnek, külön-külön kerültek felvételre.

A vizsgálat eredményét az 1. táblázat mutatja. Ebből kitűnik, hogy a három kémiai és a két biológiai folyóirat esetében a címek információtartalma (a kulcsszavak száma) 1962 és 1973 között lényegesen megnövekedett. Ezt a növekedést az említett tudományterületeken bevezetett KWIC-indexek kiadásának lehetne tulajdonítani, de kiderült, hogy a cím információtartalmának lényeges

növekedése volt észlelhető az *Economica*, az *English Historical Review* és a *Journal of the Optical Society of America* folyóiratokban is, pedig e területeken még nem léteznek KWIC-indexek.

A kulcsszavak mennyiségi növekedésének okai tehát nem feltétlenül a KWIC-indexek és a címszavakon alapuló gépi keresés megjelenésében keresendők, hanem a jelenségnek más, a folyóiratok számának növekedésével kapcsolatos okai is vannak. Ha a felhasználó pl. filozófiai kutatást végez, ahol egy folyóiratban mindössze kb. 20 cikk jelenik meg évente, kényelmesen áttekintheti az összes címet (a táblázatból kitűnik, hogy a *Philosophy* c. folyóiratban a kulcsszavak mennyisége nem is változott lényegesen). Gyakorlatibb területeken azonban a folyóiratcikkek állandó növekedése miatt az lehetetlen, a felhasználónak az a legfontosabb, hogy minél hamarabb megtalálja az őt érdeklő cikket, amihez pedig az kell, hogy a címek minél több kulcsszót tartsanak.

A vizsgálat arra is kitért, hogy milyen jellegű szavakból kerül ki az észlelt növekedés. Az *Analytical Chemistry* esetében a kulcsszavak szaporodása 1947 és 1973 között 2,39 volt, aminek legnagyobb részét (1,62) az új berendezésekkel, technológiákkal kapcsolatos szavak alkották. Természetesen társadalomtudományi és művészeti területeken sokkal ritkábbak az új szempontok vagy módszerek, ezért a címek információtartalmára gyakorolt hatásuk is sokkal kisebb.

Érdemes megvizsgálni a kulcsszavak szerepét a gépi keresés vonatkozásában is. A szerzők minden kulcsszót egyenlő értékűnek fogtak fel, pedig a kulcsszavak hasznossága között a különböző szakterületeken óriási különbségek lehetnek. Például, ha egy kémiai cikk címe és egy filozófiai cikk címe egyaránt hat-hat kulcsszót tartalmaz, nem kétséges, hogy a kémiai kulcsszavak a cikkről sokkal több információt közölnek, hiszen azok többnyire elemeket, vegyületcsoportokat, eljárásokat jelentenek.

A gépi keresés szempontjából nem tekinthetjük egyenértékűeknek a kulcsszavakat. Tegyük fel, hogy a felhasználót az angol mezőgazdaság története érdekli. A témából szerkesztett keresőkép valószínűleg nem fogja kiválasztani a *Gabonatermesztés Észak-Angliában a 14. században* c. cikk címét, pedig ez több kulcsszót tartalmaz, mint a keresett cím.

A *Nuclear Science Abstracts*-re épülő szelektív információterjesztés során G. Olive és munkatársai megállapították, hogy a rövidebb címek nagyobb pontosságot (*precision*), míg a hosszabbak nagyobb teljességet (*recall*) eredményeznek. Sokszor előfordul, hogy a rövid alcímek, sorozati címek sokkal hasznosabbak lennének a keresésre, mint a cikkek fő címei.

A gépi kereséssel kapcsolatban a továbbiakban a

1. táblázat

A vizsgált folyóiratok adatai

Folyóirat	Év	1 címre eső címszavak száma		1 címre eső kulcsszavak száma		A kulcsszavak aránya		1 kulcsszóra eső írásjelek száma	
		Átlag	Szórás	Átlag	Szórás	Átlag	Szórás	Átlag	Szórás
Transactions of the Faraday Society	1947	11,30	5,07	6,44	2,78	0,59	0,12		
	1962	9,91	4,02	6,51	2,19	0,68*	0,10		
Journal of the Faraday Society.									
Faraday Transactions	1973	13,27*†	5,64	8,43*†	3,06	0,66*	0,09	7,45	2,96
Analytical Chemistry	1947	8,64	3,29	5,59	2,02	0,66	0,11		
	1962	9,75	3,54	6,68*	2,35	0,70	0,11		
	1973	11,58*†	4,29	7,98*†	2,38	0,72*	0,12		
Journal of Organic Chemistry	1947	7,18	3,68	4,73	2,03	0,65	0,17		
	1962	7,40	3,32	4,92	1,99	0,69	0,14		
	1973	10,18*†	4,23	6,55*†	2,72	0,66	0,12		
Annals of Botany	1946–9	15,47	6,17	7,66	3,11	0,51	0,12		
	1962–3	11,76*	4,66	6,82	2,81	0,59*	0,12		
	1973	13,58*†	4,81	8,40†	2,81	0,63*	0,08		
Lancet	1947	6,24	3,06	4,15	1,84	0,71	0,16		
	1962	8,20*	3,17	5,46*	1,83	0,69	0,12		
	1973	9,04*	3,20	6,29*†	2,19	0,71	0,10		
Journal of Social Psychology	1947–9	9,50	3,74	5,44	2,12	0,59	0,13		
	1962	9,64	3,98	5,89	2,32	0,63	0,12		
	1973	9,82	3,62	6,05	2,21	0,63	0,11		
Economica	1946–50	6,72	3,04	3,82	1,58	0,60	0,16		
	1960–3	7,39	3,43	4,50	1,97	0,64	0,16		
	1971–4	8,90*†	3,23	5,17*	1,85	0,60	0,14		
British Journal of Sociology (1950-ben indult)	1950–4	7,42	3,55	4,46	1,97	0,63	0,14		
	1960–4	7,36	3,48	4,37	1,86	0,63	0,15		
	1971–3	8,42	3,70	4,89	2,15	0,60	0,12		
English Historical Review	1946–50	7,51	3,05	4,64	1,71	0,65	0,16		
	1960–3	8,81*	3,61	5,56*	2,06	0,65	0,13		
	1971–4	9,26*	3,37	5,79*	1,86	0,64	0,12	7,14	2,61
Philosophy	1945–50	4,82	2,36	2,82	1,09	0,64	0,17		
	1960–3	5,03	2,65	2,99	1,45	0,63	0,17		
	1971–4	4,47	1,91	2,84	1,13	0,66	0,19		
Journal of the Optical Society of America	1947	9,50	4,03	5,67	2,06	0,63	0,15		
	1962	8,34	3,26	5,53	1,75	0,69*	0,14		
	1973	9,33	3,71	6,55*†	2,19	0,73*	0,12		
Zeitschrift für Physikalische Chemie (Leipzig)	1973	11,78	4,76	6,62	2,40	0,58	0,10	11,02	4,80
Angol fordítás		14,25	5,25	8,18	2,68	0,59	0,10		
Journal de Chimie Physique	1974	16,11	6,38	8,39	3,29	0,53	0,09	9,04	3,87
Angol fordítás		14,36	5,76	8,29	3,21	0,59	0,11		
Historische Zeitschrift	1970–4	8,84	4,21	5,14	2,30	0,60	0,13	9,63	4,39
Angol fordítás		11,39	5,50	6,05	2,74	0,55	0,12		
Annales	1972–3	10,76	4,00	5,68	2,13	0,54	0,11	7,59	2,61

* 1947-hez képest lényeges változás

† 1962-höz képest lényeges változás

szerzők azt vizsgálták, hogy *miért könnyebb a címek szerinti keresés a természettudományok, mint a társadalomtudományok, művészetek stb. irodalmában.* Ennek kapcsán a következő okokat állapították meg:

a kémiában – kisebb részt a biológiában és az orvostudományban – sok esetben a *szótöredékek is alkalmasak* lehetnek keresésre. Például az -áz végződés a kémiában több enzim kiválasztását eredményezheti (*oxidáz, ureáz, hidrogenáz* stb.). Ugyanígy a bonyolult vegyületek elnevezéseit megbontva, keresésre alkalmas szótöredékeket kapunk. Ilyen jellegű szóbontás a történelem, a pszichológia stb. területén lehetetlen;

a történelmi vonatkozású irodalomban igen gyakran kell *dátum szerint keresnünk.* Mivel a természetes nyelv sokféleképpen meghatározhat egy periódust, és egy bizonyos időpont, időszakasz a keresetnél tágabb időhatárokat jelentő sokféle fogalomba beleesik, a keresés igen nehézkes;

különösen a filozófiában, de a művészetekben, szociológiában stb. is sok *szimbolikus jelentésű* cím található. Olyan címek például, mint *Soha ne mosolyogj a krokodilra* (Journal for the Theory of Social Behavior, 1973), vagy *Tehén a háztetőn* (Journal of Philosophy, 1973) bár igen megkapóak, keresésre teljesen alkalmatlanok.

Az előbbieken ismertetett megállapítások összefoglalásaként kijelenthetjük, hogy a címszavak információtartalmát tekintve a társadalomtudományi folyóiratcikkek nem maradnak el lényegesen a természettudományok mögött, mégis, a szisztematikus szószervezetek hiánya miatt címszavaik gépi keresésre kevésbé alkalmasak.

/BUXTON, A. B. – MEADOWS, A. J.: *The variation in the information content of titles of research papers with time and discipline = Journal of Documentation*, 33. köt. 1. sz. 1977. p. 46–52./

(Novák István)



S Z A B V Á N Y

Új bibliográfiai szabvány jelent meg a Szovjetunióban

A szokásos szabványosítási eljárást követően 1978. január 1-én megjelent az új szovjet bibliográfiai szabvány. Ez a 7.0–7.7 jelű szabvány a 16448–70 jelűt váltotta fel. A szabványosítási eljárás a megszokottnál nehezebb volt, mivel több terület (információ, bibliográfia, könyvtárügy, könyv- és folyóiratkiadás) volt érdekelt benne, s ennek következtében igen nagyfokú terminológiai bizonytalanságot kellett leküzdeni elkészítése közben. Az említett nehézségek az elkészült szabványban is tükröződnek. Részint azért, hogy az új szabvány viszonylag kevés fogalmat és fogalommeghatározást tartalmaz: mindössze 82-t. Részint pedig azért, hogy némely fogalommeghatározás kompromisszumra utal. Ennek ellenére az új szabvány fontos lépés a bibliográfiai terminológia egységesítésében.

Az új szabvány egyes fejezeteiben a *következő fogalmak szerepelnek:*

1. Általános fogalmak: Bibliográfia (Bibliográfia), Bibliograficeszkaja informacija (Bibliográfiai információ), Bibliograficeszkaja rabota (Bibliográfiai munka), Bibliograficeszkaja zapisz' (Bibliográfiai címleírás), Bibliograficeszkoe poszobie (Bibliográfiai segédeszköz).
2. Vidü bibliografii (Bibliográfiai műfajok): Goszudarsztvennaja bibliografija (Nemzeti bibliográfia), Naucsno-vszipomogatel'naja bibliografija (Tudományos segédeszköz-bibliográfia), Rekomendatel'naja bibliografija (Ajánló bibliográfia), Otraszlevaja bibliografija (Tudományági vagy ágazati bibliográfia), Kraeved-

cseszkaja bibliografija (Helyismereti bibliográfia), Izdatel'szko-knigotorgovaja bibliografija (Kiadói-könyvkereskedelmi bibliográfia), Tekucsaja bibliografija (Kurrens bibliográfia), Retroszpektivnaja bibliografija (Retrospektív bibliográfia), Bibliografija bibliografii (Bibliográfiák bibliográfiája).

3. Bibliograficeszkaja rabota (Bibliográfiai munka): Bibliografirovanie (Bibliografizálás), Bibliograficeszkij poizsk (Bibliográfiai keresés), Bibliograficeszkij otbor (Bibliográfiai szelekció), Szosztavlenie bibliograficeszkij opiszaniy (Bibliográfiai leírások készítése), Annotirovanie (Annotálás), Bibliograficeszkaja szisztematizacija (Bibliográfiai osztályozás), Bibliograficeszkaja predmetizacija (Bibliográfiai tárgyszavazás), Bibliograficeszkaja gruppirovka (Bibliográfiai csoportosítás), Bibliograficeszkoe redaktirovanie (Bibliográfia-szerkesztés), Bibliograficeszkoe obszluzsivanie (Bibliográfiai szolgálat), Szpravocsno-bibliograficeszkoe obszluzsivanie (Tájékoztató-bibliográfiai szolgálat).
4. Bibliograficeszkije poszobija (Bibliográfiai segédeszközök): Bibliograficeszkij ukazatel' (Bibliográfiai mutató), Bibliograficeszkij szpizsok (Bibliográfiai jegyzék), Bibliograficeszkij obzor (Bibliográfiai szemle). A bibliográfiai segédeszközök ún. kis formái (olvasóterv, emlékeztető stb.) a teljes terminológiai bizonytalanság következtében nem kerültek szabványosításra.
5. Vidü bibliograficeszkij poszobij (A bibliográfiai segédeszközök műfajai): Itt a szabványban összesen 32 fogalom szerepel 13 csoportba osztva. A csoportosítás