

Az indexelő technika története

1958-ban Miles Conradnak, a *BioSciences Information Service* vezetőjének javaslatára az USA-ban országos szövetséget alapítottak *National of Abstracting and Information Services* (NFAIS) néven, azzal a céllal, hogy összefogja az indexelő és referáló szolgáltatásokat és a kooperáció révén fokozza hatékonyságukat. A szövetségnek jelenleg 44 tagja van, olyan szervezetek, amelyek indexelő és referáló kiadványokat, illetve ezekkel kapcsolatos szolgáltatásokat tartanak fenn, mint például a *Chemical Abstracts Service* és a *BioSciences Information Service*. Az évi közgyűlésen és az NFAIS különféle bizottságaiban a tagszervezetek képviselőinek alkalmanként a legkülönbözőbb szakmai kérdések megvitatására, véleménycserére, megismerkedhetnek a legújabb technikai vívmányokkal, szavaznak az indexelő szolgáltatások továbbfejlesztésére, szabványosítására vonatkozó javaslatokról. Az NFAIS számos amerikai és nemzetközi szervezetben képviseli tagjait és speciális kutatásokat is vállal. Tevékenységének javarésze az oktatással, az indexelő szolgáltatások népszerűsítésével kapcsolatos. Szemináriumsorozatot rendez, melyek történeti áttekintést nyújtanak az indexelés fejlődéséről, ismertetik a különféle indexelési formákat, bevezetnek az indexek használatába.

Az USA keleti államaiban működő néhány ismert indexelő szervezet:

BioSciences Information Service (BIOSIS), Philadelphia. Kiadványa a *Biological Abstracts*, amely kivonatokat és indexeket tartalmaz, géppel olvasható formában is készül.

Franklin Research Center, Philadelphia. A Nemzetközi Rákkutatási Adatbank fenntartója. Szerződésben áll az Országos Rákkutatási Intézettel, számítógépes fájlt állít elő a rákkutatásra vonatkozó feldolgozott szakirodalomból.

Institute for Scientific Information, Philadelphia. Három különböző citation indexet és több speciális kémiai indexet állít elő. Valamennyi nyomtatott kiadvány, de néhány adatbázisban számítógépes keresés is végezhető.

International Repertory of Music Literature (Nemzetközi Zenei Repertorium), New York. Kiadja a *RILM Abstracts* c. nyomtatott folyóiratot indexekkel, géppel olvasható hordozón is.

National Association of Social Workers (Közjóléti Dolgozók Országos Szövetsége), New York. Itt készül a *Social Work Research and Abstracts* c. kétrészes referáló folyóirat és indexei.

H. W. Wilson, New York. Igen sok indexet ad ki. Legismertebbek: *Library literature*, *Applied science and technology index*, *Business periodicals index*. Valamennyi nyomtatott, néhányat számítógép segítségével állítanak elő.

Ezek az intézmények indexeik előállításához különféle eljárásokat és technológiát használnak, a katalóguskártyákra való gépeléstől kezdve a közvetlenül számítógépbe történő indexelésig. A számítógép az indexelésben egyelőre segédeszköz. Jóllehet bizonyos típusú indexelés kizárólag számítógéppel, szellemi erőfeszítés nélkül is elvégezhető, ez a legtöbb esetben nem alkalmazható, s ahol alkalmazható, eredményei nem feltétlenül jobbak a manuális módszereknél. A legtöbb indexet ma még emberek állítják elő. Az indexkészítésben az összes körülménynek legjobban megfelelő technikát kell kiválasztani. A számítógép ugyan gyorsabb, magától nem téved, de drága, és minden megoldott probléma helyett újabbat termel.

Az indexelés története

A könyvindex története a középkort megelőző időszakra nyúlik vissza, amikor a magasabb művek elején felsorolták a fejezetcímeket vagy tartalmi összefoglalót adtak, hogy segítsék az olvasót az eligazodásban. Már a VIII. századból ismerünk olyan indexet, amely többszavas kifejezéseket a belőlük kiemelt kulcsszavak betűrendjében sorol fel, alighanem ez a KWIC-index legősibb példája.

A XIV. századra az európai egyetemek kialakulása, általában a magasabb műveltség terjedése már csírájában hordozta azt a folyamatot, amit ma információrobbanásnak nevezünk. Az „egyetemi kultúra” szerves részévé vált szenvedélyes „tudós viták” fokozták a szövegekre való hivatkozás igényét. A legelső indexeket – akárcsak sok későbbit – „már tegnapra kellett volna” feltételek születték, kézzel íródtak, szinte olvashatatlanul. A terminusokat, kifejezéseket többnyire a szövegből vették, ritkábban választottak szövegen kívüli fogalmat tárgyszónak. Mindkét eljárás ma is ismeretes. Néhány korai index használta már a *lásd* és a *lásd még* utalásokat is.

A XVII. sz. végére, az első tudományos folyóiratok megjelenésének idejére a mai értelemben vett név- és tárgymutató a könyvek vonatkozásában polgárjogot nyert. A kumulatív és kollektív folyóiratindexek később, a XVIII. sz. végén jelentek meg, a retrospektív indexek iránti igény pedig csak a XIX. században jelentkezett. *William Poole*, a Yale Egyetem könyvtárosa a múlt század közepén azt tapasztalta, hogy a folyóiratok régi évfolyamait szinte sohasem veszik igénybe, s arra a következtetésre jutott, hogy indexelésükkel növelni lehetne használatukat. 1848-ban kiadott egy 154 oldalas indexet, amely nagy sikert aratott. Az Amerikai Könyvtáros Szövetség 1876. évi első kongresszusán Poole javasolta, hogy a folyóiratok retrospektív indexelését önként vállalkozó könyvtárosokból álló munkacsoport folytassa. Így jelenhetett meg 5 évenként az *Index to periodical literature*. Ezt havonta megjelenő kooperáció-

ban készülő index követte a vezető folyóiratok kurrens számairól, amelyet később felváltott az *Annual literary index*, az *American library index*, majd az *American library annual*.

1901-től, a Wilson-féle *Readers' guide* megjelenésével az általános peridokumok indexelése szinte külön szakmává vált. Több kiváló indexelő szolgáltatás története szintén erre az időre nyúlik vissza, pl. az *Index Medicusé*, melyet az első világháború idején egy sebészorvos indított el.

Indexelő technikák

Az indexeket mindig is manuálisan állították össze, kartonok vagy papírcédulák betűrendbe sorolásával, majd az ezek alapján készült kéziratot kinyomtatták. Az évekre visszamenő kumulatív indexek összeállítása azonban igen időigényes és költséges volt. H. W. Wilson nevéhez fűződik az első, ma is használatos indexelési „újítás”. Ő a nagyobb kumulációk összeállításához felhasználta a korábbiakhoz készült szedéssorozatokat.

Az 1940-es években terjedt el a lyukkártya. Nem volt azonos a mai IBM lyukkártyával, pereme körül perforálták, szabad helyvel az idézet vagy összefoglaló számára. A peremen körbefutó lyukak pozíciója indexfogalmaknak felelt meg. Indexeléskor a megfelelő lyukat homyolni kellett, kereséskor pedig a megfelelő fogalom pozíciójánál tűt szúrtak át a cédulatömbön, melyről leestek a keresett tételek. A lyukak száma azonban korlátozta az indexfogalmak mennyiségét, sok kártya esetén csak részletekben lehetett keresni, és a kiesett kártyákat újra be kellett sorolni.

1951-ben hirdette meg *Mortimer Taube* új indexelési koncepcióját *Uniterm* néven, melyet általában (poszt) koordinált indexelésnek neveznek. E szerint a szerzőnek a címben és a dokumentumban használt eredeti szavakat kell használni az indexfogalmak kialakításához, és a fogalmakat lehetőleg egyetlen szóval kell kifejezni, „prekoordinálás” nélkül. Taube előtt a dokumentumok osztályozására, indexelésére szabványosított, nagy tárgyszókészleteket alkalmaztak, mint pl. a Library of Congress tárgyszójegyzéke vagy a Sears tárgyszójegyzék. Csaknem valamennyi amerikai könyvtár ma is használja egyiket-másikat. Ez a fajta indexelés angolul *indexing by assignment* (kb. hozzárendelő indexelés). Az indexelő meghatározott szókészletből hozzárendel a dokumentumhoz bizonyos fogalmakat a tartalom reprezentálására. Taube újításának lényege abban állt, hogy a dokumentum eredeti nyelvét kell használni, a fogalmakat összevonás nélkül, önálló tételekként kell kezelni, és a kereső személy feladata, hogy ezeket koordinálja, összekapcsolja.

Indexelő szókészletek

Ha az index címszavait mindennemű megkötöttség, szabályszerűség, irányelvek nélkül választják ki a címből vagy a dokumentumból, a feldolgozás következtelenséget eredményezhet, a keresésnél a rokon témák és fogalmak szóródását lehetetlen nyomon követni. Ezért az 50-es évek végén, a 60-as évek elején kialakultak az indexelő teauruszok. A teauruszokat általában szakterületek szerint szerkesztik (pl. napenergia, várostervezés), és a nyomtatott formátumban részletesen feltüntetik a mellé- és alárendeltségi kapcsolatokat a fogalmak között. A teaurusz szerkesztésekor az összes fogalmat fogalombokronként hierarchikusan elrendezik, és meghatározzák a köztük levő összefüggéseket. Végső elrendezésében a teaurusz egy sor hierarchikusan felépített fogalomcsaládból és betűrendes szójegyzékből áll, melynek minden tételénél utalnak a szűkebb, tágabb és rokon értelmű terminusokra. Az indexelésben mindhárom módszer (hagyományos, tárgyszójegyzék, kulcsszavak szabad kiválasztása, teaurusz) használatos, s vannak egyéb, „hibrid” megoldások is.

A számítógép

Ha az indexelést a dokumentumok címe alapján végzik, számítógéppel könnyen és gyorsan előállítható a KWIC-index. A számítógép a feldolgozandó tételcíméből elhagyja a lényegtelen, irreleváns szavakat (pl. a, az, és, mert stb.), a megmaradtakat indexfogalomként kezeli, és betűrendbe sorolva kinyomtatja. A KWIC-index korlátja, hogy a dokumentum címe gyakran nem fejezi ki annak lényegét, sőt néha igen félrevezető lehet. Az indexelők ezért gyakran „feljavítják” a címeket, kiegészítik a dokumentum tartalmát pontosabban kifejező szóval vagy kifejezésekkel.

A számítógéppel előállított indexek egyik típusa a permutált index, mely egyesíti a KWIC index és a hagyományos tárgyszójegyzék alapján készült index jellegzetességeit. A permutált index a dokumentumhoz tartozó összes indexfogalmat összetartozó egységnek, csoportnak tekinti, a betűrendben minden egyes tételnél felsorolja az adott dokumentumhoz tartozó összes többi fogalmat.

Ismét másik típus a hivatkozási index (*citation index*). Ez azon az egyszerű tényen alapszik, hogy a tudományos cikkek a tárgyukkal kapcsolatos más munkákra is hivatkoznak bibliográfiában vagy lábjegyzetben. A hivatkozási index a nevezett témával összefüggő egész dokumentumcsoporthoz biztosít hozzájutást, s mint-hogy csaknem teljesen számítógép állítja elő, gyorsan készül, szinte naprakészen tartható. Az említett philadelphiai Tudományos Információs Intézetben (ISC) a beérkező folyóiratok minden cikkét átnézik, magán a

folyóiraton kijelölik a címszavakat, megszerkesztik és kódolják a hivatkozásokat és utalásokat. A betáplált adatok alapján a számítógép rendezi az összes információt a hivatkozási indexhez.

Online információs szolgáltatások

Már az 1850-es évek végén felhasználták a számítógépet személyi bibliográfiák összeállításához, a *Biological Abstracts* első számítógépes KWIC-tárgyszóindexe 1961-ben készült. Az 1960-as években számos indexelő és referáló szolgáltatás tért át számítógépre a kiadványok fényszedéséhez és betűszedéséhez. Az indexek fényszedőgépből kikerült hasábkorrektúráit szintén számítógép ellenőrzi. Az első online rendszert a Lockheed cég tervezte meg a NASA-val kötött 1966-os szerződése keretében. Az online közvetlen távösszeköttetést jelent a számítógéppel. A mágnesszalagon tárolt hivatkozások, tömörítvények, indexfogalmak visszakereshetők a számítógépes írógépterminálon keresztül. Az összeköttetést telefon biztosítja, melyet kapcsolószervezet csatol az

írógépterminálhoz. 1972-ben a Lockheed kialakította a DIALOG rendszert, amelyben kezdetben három adatbázis – vagyis mágnesszalag-készlet – volt elérhető. A mágnesszalagokat azok az intézmények adták el, illetve bocsátották rendelkezésre, melyek számítógépet használtak indexelő kiadványaik előállításához. Ma már mintegy 120 adatbázis érhető el online összeköttetésben a DIALOG rendszeren keresztül.

A számítógép térhóhítása ellenére az indexelésben továbbra is kulcsszerepet játszik az ember. Ő tervezi meg a rendszert, ő írja meg az index összeállítását szolgáló programokat, ő készíti a teauruszokat, választja ki az indexfogalmakat, többnyire ő indexeli magát a dokumentumot. A számítógép ezt a munkát könnyebbé és gyorsabbá teszi, de a munka egészét az ember irányítja. Ebben nem sok változás történt a VII. század óta.

/CORNOG, M.: A history of indexing technology = The Indexer, 13. köt. 3. sz. 1983. p. 152–157/

(Fazekas Zsuzsa)

A cím mint a dokumentumtartalom leírásának eleme Vizsgálat az értékelési módszerekről

Minden hatékony dokumentációs rendszer megszervezésének alapja az, hogy jól válasszák meg a rendszerbe bekerülő információt. Meg kell keresni azt az elemet, amelynek segítségével a dokumentumot le lehet írni, nyilvántartásba lehet venni. A későbbiekben nagyrészt ez határozza meg az információterjesztés és az információkeresés lehetőségeit.

E kérdés fontossága tulajdonképpen csak a számítógép bevezetésekor (1958) merült fel. Kiderült u. i. hogy még a szakemberek által végzett dokumentumelemzés sem elég jó. Túl szubjektív, ahány ember, annyi eredmény; a nem megfelelő, eltérő szójegyzékek is zavart okoztak stb. Fölmerült mind az input, mind az output automatizálásának szükségessége. Ehhez az eredeti dokumentumból kellett kiemelni valami olyasmit, ami lehetőleg híven tükrözi a dokumentum tartalmát. A cím látszott erre a legalkalmasabbnak.

A cím azért elsődlegesen fontos dokumentációs elem, mert egyszerre tükrözi a dokumentum tartalmát és a szerzőnek munkája lényegéről alkotott véleményét. Feltevése, hogy a cím jó. Divatba jöttek ugyanis a fantázia-címek, amelyek a dokumentum tartalmáról semmit nem mondanak. Az Unesco ezért tartotta fontosnak külön

figyelmeztetni a szerzőket arra, hogy mindig informatív értékű, tartalomra utaló, pontos címet adjanak műveiknek [1].

Az információs rendszerekért felelős szakemberek közül elméletileg is sokan foglalkoztak a dokumentumtartalom leírásával, illetve ennek alapelemével, a címmel. Vizsgálataik két kérdésre összpontosultak:

- a cím adta információk elgendők-e ahhoz, hogy a dokumentum valóságos, hasznos tartalmát megértsük?
- megvan-e a címnek az automatizált dokumentációs rendszerekben nélkülözhetetlen visszakereső-értéke, tehát az information content értéken kívül az information retrieval capability értéke is?

A szerző nem e kérdésekre akar elsősorban válaszolni, hanem arra, hogyan merültek föl a címmel kapcsolatos problémák, és hogyan próbálták megoldani őket a szakemberek. A szakfolyóiratokban fellelhető 72, kifejezetten a címről szóló tanulmányból levonható következtetések, tanulságok kerülnek tehát ismertetésre az alábbiakban.