

sek, állami beszerző hivatalok, kormányzati szervek, valamint külföldi és nemzetközi szabványosítási intézmények adtak ki; ilyenek az iparági és országos szabványok, állami árubeszerzési specifikációk, katonai szabványok stb. A külföldi szabványok gyűjteménye 61 országból származó 156 ezer szabványból áll.

A Gaithersburg-ban (Maryland állam) működő Szolgálat bárki számára nyitva áll. A könyvtárban 600 kézikönyv, enciklopédia és egyéb referenzs könyv, mintegy 100 szabványosítási folyóirat és más szakirodalom áll az olvasók rendelkezésére.

A Szolgálat nemcsak eredeti formában, papíron tárolja a szabványokat, hanem a szövetségi specifikációkat és szabványokat, a katonai szabványokat és kézikönyveket, mintegy 6000 amerikai vállalat katalógusait, az építőipari szabványokat, továbbá az ISO, az IEC (*International Electrotechnical Commission = Nemzetközi Elektrotechnikai Bizottság*) és 9 amerikai szabványosítási szervezet szabványait mikrofilmen is őrzi.

A Szolgálat nem terjeszti a birtokában lévő szabványokat. Az intézmény célja információszolgáltatás a szabványokról, pl. a szabványokat kibocsátó források közlése, a szabványok beszerzésének lehetőségeiről érdeklődők átirányítása a megfelelő kiadó szervekhez, általános szabványügyi információ nyújtása. Az alábbi típusú kérdésekre kaphatnak választ a szolgáltatás igény-

bevevői: milyen szabványok léteznek vízierőművekre? Léteznek-e szabványok meghatározott háztartási gépekre? stb.

A Szolgálat munkatársai számítógéppel előállított KWIC index segítségével évenként több mint 6000 kérdésre adnak választ. A legtöbb kérdés az USA szövetségi kormányzati szerveitől érkezik, azonban a szolgáltatások minden egyéb állami és helyi kormányzati szerv, ipari, gazdasági, kereskedelmi szervezet, oktatási intézmény stb., valamint a lakosság rendelkezésére állnak.

Mivel az információkeresés KWIC indexszel történik, a kéréseket a lehető legrészletesebben körül kell írni, valamennyi lehetséges szakkifejezéssel. A kérésekre a helyszínen, telefonon vagy postán lehet választ kapni.

A Szolgálat publikációit nyomtatott formában és mikrofilmen, mérsékelt áron terjeszti; többi szolgáltatása ingyenes. Az intézményben felállított terminál on-line információkeresést is lehetővé tesz a központ szabványinformációs gyűjteményében. A terminállal más nagy adatbázisok (MEDLINE, NTIS, Chemical Abstracts stb.) is hozzáférhetőek.

*SLATTERY, W. J.: Information center profile: Standards Information Service = Information Hotline, 8. köt. 9. sz. 1976. p. 30-31./*

(Roboz Péter)

○○○○○○○○○○○○○○○○

## KÓDOLÁSI RENDSZEREK, TEZAUROSZOK

### Kódolási rendszerek tulajdonságainak vizsgálata

Az 1972-ben a varsói Gyógyszeripari Intézetben kifejlesztett *Signa excerpta* elnevezésű és a drezdai VEB *Arzneimittelwerk* vállalatban kidolgozott gyógyszerkészítményeket ismertető figyelőkarton-szolgálatot az említett intézmények a szófiai *Gyógyszeripari Kutatóintézet* közös gondozásában terjesztik. A két szolgáltatás kódolási rendszere azonos.

Ez egy átfogó jellegű dokumentációs szolgáltatás a gyógyszeripar számára, amely szolgáltatás a kémia (szintézis, analízis), a gyógyszerészet, a gyógyszerteran, az orvostudomány, a biológia, a mikrobiológia, a biokémiai tárgykörökbe tartozó közlemények adatait tartalmazza, áttekinthető, rendszerezett, kellő mélységű kódolással, a feldolgozók és a felhasználók által egyaránt könnyen kezelhető.

A két szolgáltatás összesen évente kb. 17 500 közleményt dolgoz fel. A folyóiratokat vegyészeti, gyógyszerészeti, mikrobiológusok referálják, és a referátumokat ők kódolják is. A kémiai vegyületek kódolására a vegyületeket alkotó atomcsoportokra jellemző *számokódokat* alkalmaznak; a természetes anyagok molekuláit alapszerkezetük (pl. indol, chinolin) alapján kódolják.

Más szakterületekre egymással kombinálható tárgyszavak rendszerét alkalmazzák; az egyes gyógyszerek általános vagy specifikus hatásait, aktivitását is jelzik a kódban.

A két szolgáltatás használhatóságának ellenőrzésére 7000 dokumentumot tartalmazó halmazban kerestek választ a felhasználók által feltett 33 kérdésre. A válaszok és a 7000 dokumentum elemzése nyomán a következő mutatószámokat határozták meg: *relevanciátényező*, *megtalálási tényező* (recall factor), *zavartényező* és *hibatényező*.

Relevanciátényező: 
$$\frac{a}{R} = \frac{a}{a + b};$$

Megtalálási tényező: 
$$\frac{a}{P} = \frac{a}{a + c};$$

Zavartényező: 
$$\frac{b}{R} = \frac{b}{a + b};$$

Hibatényező: 
$$\frac{c}{P} = \frac{c}{a + c};$$

ahol

a = a formailag és tartalmilag releváns eredmények (találatok) száma;

- b = a csak formailag releváns (tartalmilag nem releváns) eredmények (ballaszt) száma;  
 c = a tartalmilag releváns, de formailag nem releváns és így a keresés számára elvesztett adatok száma;  
 R = a + b = valamennyi formailag releváns eredmény száma;  
 P = a + c = a gyűjteményben lévő valamennyi tartalmilag releváns dokumentum száma.

Az elemzett példák alapján a mutatószámok a következő értékeket mutatták:

	átlag	alsó szélső érték	felső
relevanciatényező	0,93 ± 0,06	0,26	1,0
megtalálási tényező	0,67 ± 0,14	0,17	1,0

A két rendszer relevanciájának értékei jók, a rendszerek használhatók. Az egyes közlemények kódolására fordított 10–20 perc idő elfogadható, és sokkal rövidebb, mint a kémiai szerkezetek topológiai módszerű kódolásához szükséges idő. A keresésnél ebben az esetben fellépő zaj okozta hátrányokat kiegyenlíti a dokumentumok feldolgozásának ezzel a módszerrel elérhető nagyobb sebessége.

Az egyes vegyületek keresésénél kapott kiugróan rossz eredmények, valamint a nagy szóródás azonban arra mutat, hogy a keresőrendszereket még javítani kell.

Így például: a kódolás és a kérdésfeltevés során egyaránt célszerű több adatot közölni: a kémiai szerkezet és a gyógyszeres aktivitás együttes megadása stb. a tapasztalat szerint lényegesen javítja a rendszer hatását. Különösen hasonló vagy közel azonos kémiai szerkezetű vegyületek megnevezéseinek kódolásakor a fogalmakat mélyebben kell bontani, és célszerű valamilyen járulékos információt is adni (pl. természetes anyagok molekuláinál a szénatomok számát közölni). A relevanciatényező növelhető negatív kódolások alkalmazásával is, ez azonban járulékos kódolási munkát kíván, és megnöveli a tárolókapacitás-igényt. Sok tévedés és hiba okozója a kódoló személy; pl. szükséges deszkriptorok elhagyása, a kódolásra vonatkozó szabályok helytelen alkalmazása, lehetséges modifikációk figyelmen kívül hagyása, a kódjegyzék időszakos változásainak, ill. módosításainak figyelmen kívül hagyása esetében.

A tapasztalt hiányosságok alapján az alábbi következtetéseket vonták le.

Egy keresett kémiai szerkezet megtalálásának relevanciája a vegyület funkcionális csoportjainak számával nő; kevésbé különböző szerkezetű, azonos hatású vegyületcsoportok esetén csökken, az alkalmazott kódolás mélységével és a kémiai szerkezeten kívül megadott további jellemzőkkel nő. Ezért a kódolási rendszert a tapasztalatok, észrevételek alapján, új típusú vegyületek bevezetésekor stb. rendszeresen módosítani kell.

A kódolás és a keresés módszertanát és szervezési formáját gondosan ki kell dolgozni; ez a kódolókat

megvédi az újonnan bevezetett változások részleteivel és időpontjával kapcsolatos tévedésektől. A kódolókat folyamatosan oktatni kell. A felmerült problémákat a feldolgozó-kódoló személyek együttesen beszéljék meg, és a megoldást adó kódolási formát dolgozzák ki. A kódolásokat célszerű időnként kölcsönösen ellenőrizni.

A *Signa excerpta* elnevezésű és a gyógyszerkészítményeket ismertető *figyelőkarton-szolgálat* fontos információs segédesszékői Lengyelország, Bulgária és az NDK gyógyszeriparának szakirodalmi információkkal való ellátásának.

/ROSZKOWSKA, E. – MARTIN, U. – LESIAK, M.:  
*Analyse der Leistungsfähigkeit einer Codierungssysteme = Informatik, 23. köt. 5. sz. 1976.*  
 p. 25–31./

(Pálincás János)

000000000

### Információkereső teauruszok kompatibilitásának elmélete

Az „egyszeri feldolgozás – többszöri felhasználás” elvének gyakorlati megvalósítása jelenleg az egyik legidősebb kérdés a szakirodalmi tájékoztatás terén. Ez egymással kompatibilis információkereső rendszerek működését feltételezi. Azok viszont csak akkor lesznek kompatibilisek, ha

meg lehet valósítani az információkereső nyelvek kompatibilitását;

szabványosítva lesz a bibliográfiai leírás és dokumentum formátum;

sikerül egységes gépparkot kialakítani, illetve az anyagi információhordozók egységesítését megvalósítani;

egységes struktúrájú és technológiájú információkereső rendszerek jönnek létre.

A feltételek közül az információkereső nyelvek kompatibilitásának megvalósítása a legfontosabb, minthogy kihat az összes többi feltétel pozitív alakulására is.

A számítógépes információkereső rendszerek számára a deszkriptoros keresőnyelvek a legalkalmasabbak. Ezek meghatározó eleme a teaurusz. Napjainkban egyre több teaurusz készül, de e teauruszok egyre kevésbé kompatibilisek egymással. E problémán többféleképpen próbálnak úrrá lenni, úm.

a meglévő teauruszok kompatibilitásának utólagos megteremtésével;

a teauruszokról való lemondással, azaz a szabad indexelés gyakorlatára való áttéréssel;

eleve kompatibilis teauruszok, sőt egész teauruszrendszerek egyidejű kidolgozásával.

A tapasztalatok azt mutatják, hogy csak az utóbbi módszerrel lehet kiutat találni az inkompatibilitás útvesztőjéből.