

## THESAURUS - TREE STRUCTURES

ORGANIC CHEMICAL COMPOUNDS	C
ALDEHYDES	C.6
..ALICYCLIC ALDEHYDES	C.6.6
..ALIPHATIC ALDEHYDES	C.6.11
..ACETALDEHYDE	C.6.11.6
..ACETAMIDE	C.6.11.11
..ACROLEIN	C.6.11.16
..ALLANTOIN GALACTURONIC ACID	C.6.11.21
..BUTYRALDEHYDE	C.6.11.26
..CHLOROACETALDEHYDE	C.6.11.31
..CITRAL	C.6.11.36
..CITRONELLAL	C.6.11.41
..CROTONALDEHYDE	C.6.11.46
..DICHLOROACETALDEHYDE	C.6.11.51
..DODECANAL	C.6.11.56
..FORMALDEHYDE	C.6.11.61
..FORMALIN	C.6.11.61.6
..SUFONAMIDE-FORMALDEHYDE RESIN	C.6.11.61.11
..GLYOXAL	C.6.11.66
..HEPTANAL	C.6.11.71
..HYDROXYCITRONELLAL	C.6.11.76
..ISOBUTYL ALDEHYDE	C.6.11.81
..METHYLMERCAPTOPROPIONALDEHYDE	C.6.11.86
..NONYL ALDEHYDE, N-	C.6.11.91
..SUCCINALDEHYDE	C.6.11.96
..TRICHLOROACETALDEHYDE	C.6.11.101
..UNDECANAL	C.6.11.106
..AROMATIC ALDEHYDES	C.6.11

## 4. ábra Fa struktúrájú kiírás

A teljes hierarchia kidolgozása után elkészül a tezausz géppel olvasható formája. Szövegét két formában nyomtatják ki. A *betűrendes elrendezésben* (3. ábra) minden fő címszó megjelenik az összes keresztutalással és szinonimával együtt.

A *fa-szerkezetű elrendezésben* (4. ábra) minden kifejezés a legbővebb tartalmú összefoglaló kifejezés alatt jelenik meg. A szinonimákat ebben az elrendezésben nem tüntetik fel. A kifejezések mellett a struktúrában elfoglalt helyre utaló alfanumerikus kód is található.

Ez a tezausz az egyik legnagyobb létező – keresztutalásokkal bőven ellátott – hierarchikus kémiai tezausz.

/LANDAU, H. B. – BYER, W. L.: *Production of a hierarchical chemical thesaurus = Journal of Chemical Information and Computer Sciences*, 16. köt. 3. sz. 1976. p. 141–146./

(Valkó Péter)



## Az egyetemes vezérfonal és a tezauszok kompatibilitása

Mind sürgetőbbé válik a tezauszok olyan kompatibilitásának megteremtése, hogy az egyik tezausz meghatározott részeinek a másikba való beolvasztása ne követelje meg a befogadó tezausz paradigmatis struktúrájának alapvető módosítását.

Eredményeket eddig főként a tezauszok formai egységesítése területén értek el (a GOSZT 18383–73 szabványa, az ISO/IS 2788–74 szabványa, az NTMIK normatívái). A fogasabb kérdés megoldása azonban még hátra van, nevezetesen az, hogy *miként egységesíthető a tezauszok tartalmi struktúrája*.

Erre többféle lehetőség kínálkozik. A legígéretesebb az, hogy meg lehet találni és *össze lehet gyűjteni azokat az egyetemes kategóriákat, amelyek egymással kompatibilis tezauszok alapját képezhetik*. E kategóriák adják az egyetemes vezérfonalat, a tezauszok tartalmát feltáró fazetták és lexikai blokkok sémáját, azaz az egyetemes vezérfonal a meghatározott tezausz tartalma a kategóriák szintjén. Valamennyi tezausz számára a közös egyetemes vezérfonaltól fejleszthetők ki a nagy tudománycsoportok típus-vezérfonalai, majd belőlük az ágazati-tudományági rész-vezérfonalak.

A cikk szerzője által javasolt egyetemes vezérfonal a következő:

## 1. A meghatározott szakterület

1.1 Formai fazetták (a kidavány típusa, nyelve, kiadási éve stb.)

1.2 Általános fazetták

1.21 Idő

1.22 Tér

1.23 Alkalmazás

1.24 Általános tulajdonságok

1.3 Folyamatok

1.31 Technológiai folyamatok

1.311 Termelési folyamatok

1.312 Megismerési folyamatok

1.32 Jelenségek és események

1.321 Feltételek

1.322 Nem kívánatos jelenségek

1.323 Egyéb jelenségek és események (a típus-vezérfonalakban fogalmazódik meg értelmük)

1.4 Tárgyak

1.41 Anyagi tárgyak

1.411 Technikai eszközök

1.412 Anyagok

1.413 Egyéb anyagi tárgyak (a típus-vezérfonalakban fogalmazódik meg értelmük)

1.42 Szellemi eszközök

1.421 Konceptiók, elméletek stb.

1.422 Módszerek

A tudományterületi és ágazati vezérfonalak kimunkálása közben táruznak fel előttünk a *terminológia átadásá-*

*nak szabályai.* Más és más szabályok érvényesek az alaptudományokban, a komplex tudományokban, az alkalmazott tudományokban és az integrált tudományokban. Az alaptudományok területén például

- a matematikai teauruszba matematikai,
- a fizikaiba fizikai + matematikai,
- a vegyészetibe kémiai + fizikai + matematikai,
- a biológiaiba biológiai + kémiai + fizikai + matematikai,
- a társadalomtudományiba társadalomtudományi + biológiai + kémiai + fizikai + matematikai terminológia is bekerül.

A szerző a javasolt egyetemes vezérfonal alapján sikerrel fejlesztett ki több kompatibilis teauruszt (társadalmi kommunikáció, papíripar, hajózás, atomenergetika). Az egyetemes vezérfonal alkalmasságát egy, az USA-ban készült és fazettás építkezésű oktatásügyi teaurusz összehasonlító elemzésével is eredményesen bizonyította be.

/SZOKOLOV, A. V.: *Ob odnom vozmozsno podhode k obeszpečeniju szovmesztimoszti IPT. Univerzsaľ naja fabula informacionno-poiskovüh teauruszov.* = *Naucsno-Tehnicsezskaja Informacija*, 2. sor. 1. sz. 1977. p. 19–24./

(Futala Tibor)

## A TÁJÉKOZTATÁS GÉPESÍTÉSE

### Automatikus profilszerkesztést segítő módszer

A számítógépes információkeresés munkaigényes megelőző feladata a *keresőprofilok szerkesztése*. Cél szerű ezért olyan eljárást keresni, amellyel a ráfordított idő és a manuális munka csökkenthető, illetve automatizálható legyen.

A végső cél olyan módszer kifejlesztése, amelyben a felhasználó egyetlen feladata a kikeresett dokumentumok relevanciájának az elbírálása, míg a kiválasztott dokumentumokat a számítógép már automatikusan elemzi, értékeli.

#### Módszertan

A releváns dokumentumokat az irreleváns dokumentumoktól a *felhasznált terminológia alapján lehet megkülönböztetni*. Bizonyos szakkifejezések, szókombinációk vagy szótöredékek a releváns dokumentumokban gyakrabban fordulnak elő, mint az adatbázis egészében, és ezek képezik a potenciális keresőkifejezések halmazát. Tehát egy szó vagy kifejezés hasznosságát úgy definiálhatjuk, mint annak előfordulási gyakoriságát a releváns anyagban, viszonyítva a teljes adatbázisban való előfordulási gyakoriságához. Erre jellemző szám a kifejezés *specifikussága*, amelynek definíciója az előzőek szerint:

$$S = \frac{R}{N}, \quad 0 \leq S \leq 1,$$

ahol  $R$  a kifejezést tartalmazó *releváns* dokumentumok száma,  
 $N$  a kifejezést tartalmazó *összes* dokumentum száma.

*Hasznos kifejezések* azok, amelyek nagy gyakorisággal jelennek meg a releváns kis gyakorisággal az irreleváns dokumentumokban. Hasznos kifejezésekre  $S$  értéke 1-hez közeli szám. Viszont pontosan 1 lehet az értéke akkor is, ha az egész adatbázisban csak egyetlen egy

releváns dokumentum van és csak ebben, egyszer jelenik meg a kifejezés, hiszen ebben az esetben  $R = N = 1$ . A specifikusság rendjében készült szójegyzékben éppen ezek a csak egyszer előforduló kifejezések helyezkednek el legfelül, torzítva az eredményeket. *Az ellentmondás feloldására a kifejezés specifikusságát másképpen célszerűbb definiálni:*

$$S = \frac{R^2}{N}$$

Ha  $R > 1$ , és az irreleváns anyagban való megjelenés gyakorisága ( $X = N - R$ ) kicsi,  $S > 1$  adódik. Tehát minél hasznosabb egy kifejezés a keresés szempontjából, azaz minél több releváns dokumentumban fordul elő, annál nagyobb  $S$  érték tartozik hozzá, és a specifikussági rangsorban annál magasabban helyezkedik el.

A módszer kipróbálására mindegyik minta-kérdésre három CAC mágnesszalag összesített adatbázisából keresték ki a releváns dokumentumokat. Kiszámították az ezekben szereplő kifejezések  $S$  értékét, rangsorolták azokat csökkenő értékük szerint, majd az így adódó rangsor alapján a fenti adatbázisból ismét kikeresték a releváns dokumentumokat és ezeket hasonló módon értékelték ki. Ezt az egyre kevésbé specifikus kifejezésekkel való keresésen alapuló közelítést mindaddig megismételték, amíg több új releváns dokumentum nem jelentkezett. Majd minden kifejezésre *kiszámították a keresési pontosságot (precision) és a relatív lehívási értéket (recall)* (a kifejezés alapján kikeresett releváns tétel/összes kikeresett releváns tétel).

Kezdetben e módszerrel csak *egy kifejezéses* specifikussági listákat készítettek. Számos fogalom azonban nem adható meg egy kifejezéses listákkal. Ezek a fogalmak a kifejezések logikai ÉS kapcsolatával határozhatók meg, így keletkeznek a *kifejezés-párok*. Megtörténhet, hogy az alacsony specifikusságú kifejezések esetleg magasabb specifikusságú kifejezés-párokat adnak. Ennek megfelelően előállították a kifejezés-párok specifikussági listáját.