

# AZ IDEÁLIS PARANCSNYELV: A DIALOG, A DATA-STAR, AZ ORBIT, A PERGAMON INFOLINE ÉS AZ STN PARANCSNYELVÉNEK ÖSSZEHOSONLÍTÁSA

Novák Teréz

Országos Műszaki Információs Központ és Könyvtár

## 1. Bevezetés

Az Országos Műszaki Információs Központ és Könyvtár (OMIKK) online számítógépes információkereséseket végző részlegében jelenleg hat külföldi online szolgáltatóközponthoz van hozzáférési lehetőségünk. Ez a viszonylag nagy szám azzal magyarázható, hogy az OMIKK-hoz mint országos információs központhoz szinte valamennyi műszaki-tudományos szakterületről fordulhatnak információkeresési kéréssel. A legtöbbet a *Dialog* és a *Data-Star* szolgáltatóközpontot használjuk [1]; a *Dialog* rendszert rendkívül gazdag adatbáziskínálata és leginkább "felhasználóbarát" jellege miatt, a *Data-Star* pedig elsősorban földrajzi közelsége (a kisebb távközlési költség és az offline printek gyorsabb megérkezése) miatt. Az *Orbit* és a *Pergamon InfoLine* rendszereken főleg az exkluzív, azaz a csak általuk szolgáltatott és az adott tárgykörben egyedülálló adatbázisokat szoktuk keresni. Az *STN International* rendszert eddig kizárólag a CAS Online által a Chemical Abstracts (CA) adatbázis keresésére nyújtott speciális lehetőségei miatt használtuk.

A felhasználókért folyó, erősödő piaci verseny a felhasználói igényeket minél inkább figyelembe vevő újdonságok bevezetésére készíti a szolgáltatóközpontokat. Szinte havonta meglepnek bennünket valami egészen új vagy tökéletesített keresési lehetőséggel, illetve adatbázissal. Az online piacon folyó verseny igazi nyertesei mi, felhasználók vagyunk. Az utóbbi években megfigyelhető, hogy az egyes rendszerek igyekeznek parancsnyelvükbe beépíteni a többi rendszer előnyös keresési sajátosságait. Erre talán a legjellemzőbb példa a *Dialog* parancsnyelv új változatának, a *DIALOG 2*-nek a bevezetése volt. A *DIALOG 2* bevezetése előtt a CA adatbázis kere-

sésére majdnem mindig a *Data-Star* választottam, mivel ennek volt néhány, különösen a CA kereséséhez előnyösen használható parancsnyelvi sajátossága (lásd később), ami a *Dialog* parancsnyelvének régi változatából (*DIALOG*) hiányzott. A *DIALOG 2*-nek azonban már megvan valamennyi, korábban csak a *Data-Star* által nyújtott keresési lehetősége, ugyanakkor megtartotta az egyébként kifinomultabb *DIALOG* parancsnyelv egyéb előnyös jellemzőit [2].

A címben szereplő öt online rendszer (szolgáltatóközpont) öt különböző parancsnyelvet használ. A használatuk során szerzett tapasztalatok alapján szeretném összehasonlítani az egyes rendszerek legfontosabb keresési sajátosságait, azok előnyeit és hiányosságait, valamint a szolgáltatóközpont kiválasztásának szempontjait, különös tekintettel a kémiai tárgyú keresésekre. A kritikai összehasonlítás alapszempontja, hogy mi az ideális megoldás, keresési lehetőség a felhasználók számára, és azt az egyes rendszerek mennyire elégítik ki.

## 2. A legfontosabb keresési sajátosságok

### 2.1 Helyzeti operátorokkal való keresés (proximity)

A helyzeti operátorok a keresőszavak egymáshoz viszonyított helyzetét szabják meg a rekordon belül. Ezek alkalmazásával növelhető a szabadszöveges keresés pontossága, és módot adnak a keresési stra-

## Helyzeti operátorok

<i>A keresőszavak egymáshoz viszonyított helyzete</i>	<i>Dialog</i>	<i>Data-Star</i>	<i>Orbit*</i>	<i>Pergamon InfoLine</i>	<i>STN</i>
Egymás mellett, adott sorrendben	+	+	+	-	+
Maximum <i>n</i> szó távolságra, adott sorrendben	+	-	+	-	+
Egymás mellett, tetszőleges sorrendben (NEAR)	+	-	-	+	+
Maximum <i>n</i> szó távolságra, tetszőleges sorrendben (NEAR)	+	-	-	+	+
Egy mondatban vagy almezőben	+**	+	+	+**	+**
Egy adatmezőben	+	+	+	-	+**
NOT típusú helyzeti operátorok	+	-	+**	-	-

\* Nem minden adatbázisban használható helyzeti operátor.

\*\* Korlátozott, adatbázisonként változó alkalmazás.

tégia rugalmas kidolgozására és módosítására. Az 1. táblázat mutatja az egyes rendszereken alkalmazható helyzeti operátorok típusait.

A helyzeti operátorok legnagyobb választékát a Dialog biztosítja, a legszegényesebb a Pergamon InfoLine kínálata. A Dialog rendszeren az egyetlen hiányosság, hogy az *egy mondaton belüli* előfordulást megszabó (S) operátor nem alkalmazható általánosan valamennyi adatbázisban. Ez a lehetőség különösen a teljes szövegű adatbázisok keresésére lenne jó, de a kivonatban való keresést is megkönnyítené. Az (S) operátor helyettesítésére jól alkalmazható azonban a keresőszavak tetszőleges sorrendű egymás melletti előfordulását megszabó NEAR operátor, ha a két szó között elég nagyszámú, egyéb szó közbejöttét is megengedjük (nN). Használható viszont az (S) operátor a kémiai kereséseknél legfontosabb két adatbázisban, a CA és a World Patents Index (WPI) adatbázisokban. A CA esetén ez azért különösen jelentős, mert a deskriptormezőt egy kötött címszóból (General Subject Index Heading vagy a vegyületeket azonosító Registry Number) és a hozzájuk kapcsolódó, nem kötött szavakból álló, magyarázó szövegrészt tartalmazó *mondatok* alkotják. A szabadanyag- vagy azonosítómező is több mondatból áll (Keyword Index), amelyek a publikáció lényegi mondanivalóját kifejező szavakat tartalmazzák felsorolásszerűen (1. ábra). Az egymással szoros tartalmi kapcsolatban lévő fogalmak tárgyszavai nagy valószínűséggel egy mondaton belül fordulnak elő, és az (S) operátor lehetőséget nyújt a címben, egy azonosítómondatban, egy deskriptor-

mondatban vagy egy szekciócímben előforduló tárgyszavak összekapcsolására. Az (L) operátor a deskriptormező egy mondatára való korlátozást teszi lehetővé.

A WPI adatbázis esetén az (S) operátorral kapcsolhatók össze a kémiai szabadalmak indexelésére használt, külön almezőkbe sorolt, összetartozó kémiai kódok. Alkalmazható azonban az (S) operátor a címet és a kivonatot tartalmazó adatmezőben, ill. az alapszótárban tárgyszavak összekapcsolására is. Ez utóbbi lehetőségnek a WPI adatbázisban azért van nagy jelentősége, mert a kémiai és a villamosági szabadalmak kivételével, a Nemzetközi Szabadalmi Osztályozás viszonylag durva rendszerén kívül nincs kötött keresőnyelvi eszköz. A kémiai és villamosági szabadalmak indexelésére használt speciális kódokat azonban csak az adatbázis-készítő Derwent cég előfizetői érhetik el. A WPI adatbázisban így a legtöbb esetben szabad tárgyszavas keresésre kényszerülünk, amelynek a pontossága — a teljesség jelentősebb csökkenése nélkül — lényegesen növelhető a keresőszavak egy mondaton belüli előfordulását megszabó (S) operátorral.

A deskriptormezőn belül, az egy almezőben előforduló szavak összekapcsolására a Dialog rendszeren előnyösen alkalmazható az (L) operátor a többszintű deskriptorokat (fődeskriptor—aldeskriptor párok) tartalmazó (pl. Medline, Compendex, FSTA) adatbázisokban, vagy a Predicasts adatbázisokban az összetartozó termékek, események és országok neveinek és/vagy kódjainak összekapcsolására is.

## Azonosítómunkatok:

resoln mandelic acid liq chromatog  
hydroxymandelic acid resoln liq chromatog

## Deszkriptormunkatok:

Chromatography, column and liquid ← 1  
high performance, for resoln of mandelic  
and hydroxymandelic acids } ← 2  
90-64-2 1198-84-1 17119-15-2 ← 3  
high performance liq chromatog resoln of,  
using L-phenylalanine-copper sulfate  
complex eluent } ← 2

- 1 Kötött címszó
- 2 Nem kötött, magyarázó szövegrész
- 3 Registry Numerek

1. ábra Egy CA rekordhoz tartozó azonosító-  
és deszkriptormező részlete

A keresőszavak egy mondaton belüli előfordulása a legszélesebb körben, valamennyi adatbázis szinte valamennyi adatmezőjében a Data-Star rendszeren kereshető.

A NEAR operátor, amely a keresőszavak tetszőleges sorrendű egymás melletti (vagy minimum  $n$  szó közbejöttével) előfordulását írja elő, csak a Dialog, a Pergamon InfoLine és az STN parancsnyelvének a része. Ennek az operátornak a használata a CA adatbázisban különösen fontos, mivel nagyon sok deszkriptorkifejezés a szaknyelvben használt kifejezés invertált alakja. Például a természetes gumira a *Natural rubber* helyett deszkriptorként a *Rubber, natural* kifejezést kell használni. Ha a teljes alapszótárra kiterjedő keresőprofilunk akarunk megszerkeszteni, a NEAR operátorral mindkét forma kereshető egyetlen keresőkifejezéssel, például a Dialog rendszeren: *S Natural (N) Rubber* formában.

A Pergamon InfoLine és az Orbit rendszeren több szóból álló kifejezések egy már kiválasztott rekordhalmazon belül, a soros állományban (a rekordok "végigolvasásával") kereshetők kifinomultan (SCAN, ill. STRINGSEARCH vagy SENSEARCH, lásd 2.7 pont). Ez azonban időigényes művelet.

A helyzeti operátorokkal való kereséstől elvárják a felhasználók, hogy a helyzeti operátorok alkalmazhatók legyenek a halmazszámokkal is, ami nagyon megkönnyíti és rugalmassá teszi a keresési stratégia utólagos módosítását. Ez a lehetőség az STN kivételével valamennyi rendszerben megvan, a Dialog esetén a DIALOG 2 parancsnyelv egyik új vonásaként. Az STN rendszeren a találati halmazok nem kapcsolhatók össze helyzeti operátorokkal (ez évről tervezzük a lehetővé tételét), de az S STEPS parancs használatakor az egyes, közbenső keresőszavakhoz rendelt L-számok (nem valódi halmazok!) összekapcsolhatók.

A több keresőszó kombinációjából álló keresőkérdés bevitelkor az S STEPS keresőparancs alkalmazásakor a Dialog rendszer minden egyes keresőszóra külön halmazt képez S-számok hozzárendelésével (2. ábra), az STN pedig L-számokat rendel hozzájuk; mindkettő később újra felhasználható. A Pergamon InfoLine SELECT parancsának alapértelmezés szerinti sajátja a közbenső halmazok képzése. Ezzel szemben a Data-Staron és az Orbiton nem képződnek közbenső találati halmazok, amelyek pedig nagyon sokszor szükségesek lehetnek a keresési stratégia módosításakor. Ha egy újabb keresőkérdésben akarunk felhasználni egy adott keresőszót, ismételt beírása helyett elegendő a halmazszámra hivatkozni.

## Keresőkérdés bevitele SELECT (S) parancssal

(nem képződnek halmazok az egyes keresőszavakra)

```
? S Petroleum AND Prices AND OPEC AND PY = 1979
    2711 Petroleum
    2683 Prices
    151 OPEC
    101655 PY = 1979
S1 83 Petroleum AND Prices AND OPEC AND
    PY = 1979
```

Keresőkérdés bevitele SELECT STEPS (S STEPS,SS) parancssal  
(az egyes keresőszavakra is képződnek halmazok)

```
? SS Petroleum AND Prices AND OPEC AND PY = 1979
S2 2711 Petroleum
S3 2683 Prices
S4 151 OPEC
S5 101655 PY = 1979
S6 83 Petroleum AND Prices AND OPEC AND
    PY = 1979
```

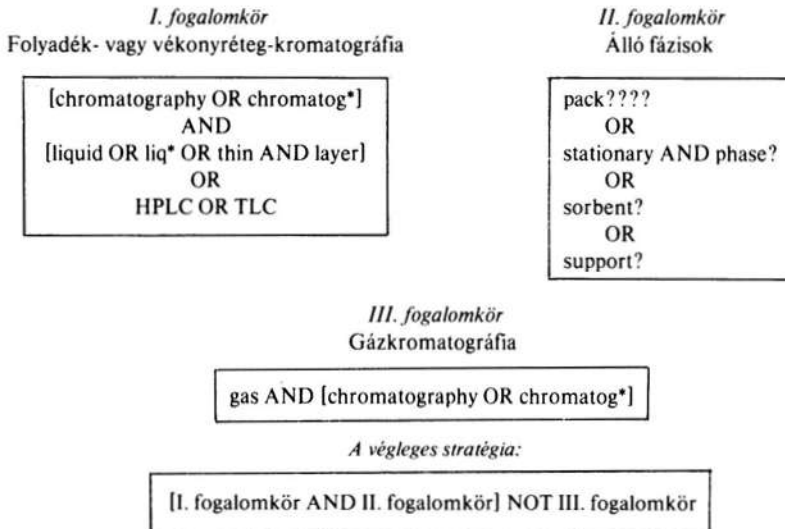
2. ábra Több keresőszóból álló keresőkérdés bevitele  
a Dialog rendszeren SELECT és SELECT STEPS  
parancssal

Szabadszöveges keresésnél, amikor egy fogalomra minél több lehetséges szinonimát kell figyelembe venni, gyakran felvetődik több, OR logikai operátorral kombinált keresőszóból álló keresőkifejezések helyzeti operátorokkal való összekapcsolásának a szükségessége (distributed proximity). Ezt mind az öt, itt tárgyalt rendszer lehetővé teszi, a Dialog rendszeren ez is a DIALOG 2 parancsnyelv új vonása. A helyzeti operátoroknak ez az alkalmazása igen gyakori a CA adatbázisban való kereséseknél, hiszen itt még a deszkriptormező nagyobbik része (a deszkriptorokhoz kapcsolódó magyarázó mondatok) is szabad szavakból áll.

A helyzeti operátorok alkalmazását a Dialog, a Data-Star és az STN rendszeren a CA adatbázisban végzett keresési példán mutatom be. A feladat a

folyadék- és a vékonyréteg-kromatográfiában alkalmazható álló fázisokra vonatkozó publikációk megkeresése. A keresési stratégia elvi vázlatát a 3. ábra szemlélteti. Maga a keresés a 4. ábrán látható, a 4.a ábra a Dialog, a 4.b ábra az STN, a 4.c pedig a Data-Star rendszeren végzett keresést mutatja.

A keresőprofilok bevitelkor a 3. ábrán szereplő AND helyett a legtöbb esetben a megfelelő helyzeti operátort alkalmaztam. A "chromatography OR chromatog" és a "liquid OR liq OR thin(W)layer", ill. a Data-Staron "thin ADJ layer" keresőkifejezések vagy halmazszámaik összekapcsolására az egy



\* A CA szabványos rövidítései.

? = a csonkolás jele (egy ? maximum egy karaktert jelent).

3. ábra A folyadék- és vékonyréteg-kromatográfias állófázisok témára a CA adatbázisban kidolgozott keresési stratégia vázlat

mondaton belüli előfordulást megszabó (S), (L), ill. WITH helyzeti operátort használtam a minél nagyobb teljesség kedvéért, hogy pl. a *Chromatography, column and liquid* deskriptorkifejezést is megtaláljuk. Ezt a keresőkérdés-sort a Dialog esetén (4.a ábra) nem az S STEPS paranccsal, egy állításban vittem be, mivel a "liquid OR liq OR thin(W)layer" keresőszavak halmazára külön később nem volt szükségem. A "chromatography OR chromatog" halmazra viszont szükségem volt később, és mivel nagy halmazok képzése a Dialog rendszeren – különösen olyan nagy adatbázisokban, mint a teljes CA adatbázis (1967-től napjainkig több mint hétmillió rekord) – nagyon időigényes, külön halmazt képeztem rá (S1), amelyet azután (S) helyzeti operátorral külön lépésben kapcsoltam a "liquid... stb." szavakkal (S2). A gázkromatográfia szövege jöhet keresőkifejezések – gas chromatog(raphy); chromatography, gas; gas liquid chromatog(raphy) stb. – előhívására a "gas" szót kapcsoltam (1N) helyzeti operátorral az S1 halmazhoz: "chromatog(raphy)", és ezt zártam ki (S6) a folyadék- vagy vékonyréteg-kromatográfia (S3) ÉS az állófázisokra (S4) vonatkozó rekordhalmazból (S5). Az STN-en (4.b ábra) hasonló a keresőprofil,

azzal a különbséggel, hogy az első keresőkérdést egy lépésben, S STEPS paranccsal vittem be, mivel itt nem képződnek közbenső halmazok, viszont az egyes keresőszavakhoz rendelt L-számok felhasználhatók további keresésre (nem kell újra beírni, de az L-számoknak megfelelő keresőszóra ismételt keresést fog végezni a rendszer). A Data-Staron (4.c ábra) azokra a keresőszavakra, amelyekre külön is szükségem volt (pl. 1-es és 9-es halmaz), illetve amelyek önmagukban nagy találatsszámot eredményeznek (pl. 2-es és 3-as halmaz), külön halmazt képeztem, és kombináltam őket a már ismertetett logika szerint. A Data-Star rendszeren nem a nagy találatsszámú halmazok képzése, hanem ezek OR-ral való kombinálása ("összeadása") az időigényes lépés. Ezért a "chromatog(raphy)" halmaznak (1-es) a "liquid" (2-es), a "liq" (3-as) és a "thin ADJ layer" (4-es) halmazokkal WITH operátorral külön-külön képeztem kombinációit kapcsoltam össze a logikai OR-ral (5-ös halmaz). Mivel a Data-Starban nincs NEAR operátor, a gázkromatográfia fogalmára két formában kerestem: "chromatog(raphy)" (1-es) ADJ "gas" (9-es) OR "gas ADJ chromatog(raphy)", másképpen 9 ADJ 1 (10-es halmaz).

```

File 399:CA SEARCH 1967-1986 UD=10420
<Coop. 1986 by the Amer. Chem. Soc.>

      Set  Items  Description
      ---  -
?S CHROMATOGRAPHY OR CHROMATOG
Processing
      86140  CHROMATOGRAPHY <SEE ?IGNOTE>
      120194 CHROMATOG
      S1 129024 CHROMATOGRAPHY OR CHROMATOG
?S S1<S><LIQUID OR LIQ OR THIN<W>LAYER>
Processing
Processing
      129024 S1
      209473 LIQUID
      130391 LIQ <LIQUID>
      51379  THIN
      67903  LAYER
      19263  THIN<W>LAYER
      S2 55087 S1<S><LIQUID OR LIQ OR THIN<W>LAYER>
?S S2 OR HPLC OR TLC
Processing
Processing
      55087  S2
      7578   HPLC
      3738   TLC
      S3 56581 S2 OR HPLC OR TLC
?S PACK???? OR SA TATIONARY<W>PHASE? ? OR SORBENT? ? OR SUPPORT???
Processing
      15342  PACK????
      7679   STATIONARY
      255377 PHASE? ?
      4715   STATIONARY<W>PHASE? ?
      5268   SORBENT? ?
      27605  SUPPORT???
      S4 51034 PACK???? OR STATIONARY<W>PHASE? ? OR SORBENT? ? OR
      SUPPORT???

?C3+4
Processing
      56581  3
      51834  4
      S5 3956 3+4
?S S5 NOT GAS<1N>S1
Processing
Processing
      3956   S5
      237375 GAS
      129024 S1
      48299  GAS<1N>S1
      S6 2896 S5 NOT GAS<1N>S1
?LOGOFF
      06Jun86 03:56:41 User033285
      19.81   0.109 Hrs File399

```

4. ábra Keresési példa a CA adatbázisban (Dialog 399-es fájl)

## 2.2 Csonkolás

A Pergamon InfoLine kivételével valamennyi rendszer lehetővé teszi a jobb oldali (szóvégi) csonkolást meghatározott vagy tetszőleges számú karakterre. A Pergamon InfoLine csak a tetszőleges számú karakterre való csonkolást engedi meg, ill. speciális végződésekkkel rendelkező szótöve kereshetünk, pl. col\_ (our OR or). (A soros állományban való kereséskor jóval több csonkolási lehetőség van, még bal oldali csonkolás is.) A Data-Staron a jobb oldali csonkolás kissé nehézkes lehet, ha túl erősen csonkoltunk, mivel a rendszer csak száz, az

adott szótóval kezdődő keresőszót tud megkeresni egyszerre, és a keresés folytatását további százszávként kell mindaddig kérni, amíg a szótárállományban ilyen szótóvű szavak találhatók.

A Data-Star kivételével valamennyi rendszer lehetővé teszi a szón belüli karaktermaszkolást.

Különösen a kémiai kereséseknél lenne fontos a bal oldali (szóeleji) csonkolás, amely lehetővé teszi bizonyos vegyületfragmentumokat, csoportokat tartalmazó vegyületek megkeresését. A bal oldali csonkolás általános érvényességgel egyik rendszeren sem megengedett. A vegyületnevek esetén (ill. a Dialog rendszeren a CA adatbázisban általános

```

FILE 'CA' ENTERED AT 02:45:29 ON 06 JUN 86
COPYRIGHT 1986 BY THE AMERICAN CHEMICAL SOCIETY

=> S STEPS <CHROMATOGRAPHY OR CHROMATOG><L><LIQUID OR LIQ OR THIN<W>LAYER>
L1 < 88449>CHROMATOGRAPHY
L2 < 120447>CHROMATOG
L3 < 140788>LIQUID
L4 < 130910>LIQ
L5 < 51320>THIN
L6 < 68229>LAYER
L7 55302 <CHROMATOGRAPHY OR CHROMATOG><L><LIQUID OR LIQ OR THIN<W>LA

=> S L7 OR TLC OR HPLC
      3765 TLC
      7692 HPLC
L8 56816 L7 OR TLC OR HPLC

=> S PACK*** OR STATIONARY<W>PHASE# OR SORBENT# OR SUPPORT***
      15409 PACK***
      7730 STATIONARY
      162097 PHASE#
      4760 STATIONARY<W>PHASE#
      5227 SORBENT#
      27741 SUPPORT***
L9 52030 PACK*** OR STATIONARY<W>PHASE# OR SORBENT# OR SUPPORT***

=> S L8 AND L9
L10 3997 L8 AND L9

=> S <L1 OR L2><1A>GAS
      88449 CHROMATOGRAPHY
      120447 CHROMATOG
      238045 GAS
L11 48430 <L1 OR L2><1A>GAS

=> S L10 NOT L11
L12 2933 L10 NOT L11

=> LOGOFF Y
COST IN POUNDS STERLING      SINCE FILE      TOTAL
                               ENTRY      SESSION
FULL ESTIMATED COST          4.76      4.95

DISCOUNT AMOUNTS <FOR QUALIFYING ACCOUNTS>
CA SUBSCRIBER                -0.56      -0.56

STN INTERNATIONAL LOGOFF AT 02:49:44 ON 06 JUN 86

```

4. b ábra Keresési példa a CA adatbázisban (STN CA-fájl)

tárgyszavaknál is) a bal oldali csonkolás pótlására szolgál a vegyületnevek fragmentálása, ami azt jelenti, hogy az adatbázis szótárába belekerül a teljes név is, valamint a központozási jeleknél (“-”, “,”, “.”, “( )”, “[ ]”), ill. a kémiai jelentésű szótöredékeknel való hasítással kapott névfragmentumok is (5. ábra). Ezek helyzeti vagy logikai operátorokkal való összekapcsolásával bármely vegyület vagy rész-szerkezet egyszerűen és flexibilisen leírható.

A vegyületnevek fragmentálására alkalmazott algoritmus a Dialog rendszeren a Chemical Abstracts bibliográfiai (CA Search) és a vegyületszótár-(Chemname, Chemsis, Chemsis, Chemzero) adatbázisok mellett a Claims és a WPI szabadalmi adatbázisokra is kiterjed, jelentősen megnövelve a vegyületekre vonatkozó keresések teljességét, ami a

szabadalomkutatásoknál különösen fontos. A Data-Star is alkalmazza ezt a fragmentálási rutint a CA adatbázisokban, az STN viszont csak a vegyületszótárnak megfelelő Registry File-ban, a CA bibliográfiai adatbázisban viszont nem.

Az Orbit és az STN rendszerek néhány adatbázisában a bal oldali csonkolás is megengedett, így az Orbiton a CA, a Compendex és az Ei Engineering Meetings adatbázisokban. A Dialog és a Pergamon InfoLine védjegy-adatbázisaiban (Trademarkscan, ill. UK Trade Marks) is van lehetőség a szavakon belüli karakterlánc keresésére. Az Orbit és a Pergamon InfoLine nyújtotta közvetlen, soros keresési lehetőség valamennyi adatbázisban biztosítja a bal oldali csonkolást, azonban sok gépidő-felhasználással.

```

D-S/CHZZ/CHEM ABS 1967-V104.I20/1986      SESSION 1414
COPYRIGHT BY AMERICAN CHEMICAL SOCIETY, COLUMBUS/OHIO, U.S.A.

D-S - SEARCH MODE - ENTER QUERY
1_: (CHROMATOGRAPHY OR CHROMATOG).DE.I SI.TI.
   RESULT  128463

2_: LIQUID
   RESULT  140459

3_: LIO
   RESULT  134195

4_: THIN ADJ LAYER
   RESULT  19257

5_: 1 WITH 2 OR 1 WITH 3 OR 1 WITH 4
   RESULT  52012

6_: 5 OR HFLC OR TLC
   RESULT  54310

7_: (PACK#4 OR STATIONARY ADJ PHASE#1 OR SORBENT#1 OR SUPPORT#3).DE.SI.TI.
   RESULT  51666

8_: 6 AND 7
   RESULT  3706

9_: GAS
   RESULT  245622

10_: 1 ADJ 9 OR 9 ADJ 1
   RESULT  45074

11_: 8 NOT 10
   RESULT  2729

12_: ..0
*CONNECT TIME CHZZ: 0:05:06 HH:MM:SS 0.085 DEC HRS. SESSION 1414*

```

4.c ábra Keresési példa a CA adatbázisban (Data-Star CHZZ adatbázis)

1,4-butanediol ← eredeti név  
 1  
 4 } ← a központozási jeleknél való  
 butanediol } hasítással kapott fragmentumok

butane, diol } ← kémiai jelentésű fragmentumok  
 di, ol }

5. ábra Az 1,4-butanediol fragmentálása a Dialog rendszeren

### 2.3 Halmazszámok utólagos minősítése adatmezőre

Ezt a lehetőséget csak a Data-Star és a Dialog új változata, a DIALOG 2 nyújtja. A halmazszámok utólagos minősítésével az alapszótárban végzett keresés eredményeként kapott rekordhalmazt a keresési stratégia módosításakor igen egyszerűen korlátozhatjuk utólagosan pl. a deskriptor- vagy címmezőre a pontosság növelése érdekében.

### 2.4 A szótárállomány felépítése

A Data-Star kivételével valamennyi vizsgált rendszerben több szótárállomány van: egy alapszótár, amely a téma szerinti kereséseknél fontos, ún. tematikus mezők (cím, deskriptorok, szabad tárgyszavak, kivonat, osztályozási szekció címe) egyes szavait tartalmazza, valamint az egyéb, inkább a bibliográfiai adatelemeket tartalmazó ún. kiegészítő szótárak. A Data-Star adatbázisoknak egyetlen szótárállományuk van, ami sokszor még a valamennyi tematikus mezőre kiterjedő keresésnél is szükségessé teszi a mezőkre való minősítést (4.c ábra), mert bizonyos keresőszavak előfordulhatnak pl. a folyóirat címében vagy a szerző nevében.

A szótárállomány kapcsán szeretnék kitérni a deskriptorok keresésére. A Data-Star kivételével valamennyi rendszer valamennyi adatbázisában az összetett deskriptorkifejezések eredeti formájukban (helyközzel, vesszővel stb.) is és szavanként is bekerülnek az alapszótárba (double posting = kettős szóelőfordulás), és mindkét módon kereshetőek. Ez lehetővé teszi, hogy egyetlen deskriptorszó keresésével előhívjuk valamennyi, az adott szótár

talmazó deskriptorkifejezéssel indexelt rekordot, pl. az *Alloys* szó keresésével az *Aluminium Alloys*, *Gallium Alloys* stb. deskriptorokat. Ezenkívül a több szóból álló deskriptorkifejezések az egyes szavak helyzeti operátorokkal való összekapcsolásával is kereshetők.

A Dialog és az Orbit rendszerben az is megkülönböztethető, hogy egy deskriptorként keresett szó egyszavas deskriptor vagy egy összetett kifejezés része-e (DF, DE, ill. IT, IW).

A Pergamon InfoLine rendszeren a deskriptorkifejezések egyes szavai szerepelnek ugyan az alapszótárban, de a deskriptormezőre korlátozva csak kifejezésként kereshetők (CT=).

A Data-Star esetében egyes adatbázisokban, pl. INSPEC, Medline az összetett deskriptorok csak teljes kifejezésként kerülnek be a szótárba, és egyes szavaik alapján nem kereshetők, ami nagyon megnehezíti a keresést. Az ötvözetek példáját véve: ha valamennyi fém ötvözetét meg akarjuk keresni, akkor az összes olyan kifejezésre keresnünk kell, amelynek második szava az "alloys" szó, és ilyen szép számmal akad.

A legrugalmasabb keresést a Dialog adatbázisok szótárállománya teszi lehetővé. Sok, tezaurusszal rendelkező adatbázisba még a tezausz is be van építve, és ennek tanulmányozásával (EXPAND) online is megkereshetők az adott témát leíró deskriptorok. Végezetül az egyes adatbázisok szótárállományának felépítése és az adatmezők jelölése is itt a legegységesebb, ami – több adatbázisra kiterjedő kereséskor – megkönnyíti a keresési stratégia átvitelét egyik adatbázisból a másikba. Ez legkevésbé a Data-Starra mondható el, ahol pl. a CA adatbázis deskriptormezőjének a mezőazonosítója az általánosan elterjedt *DE* helyett *SI*.

## 2.5 Megőrzési parancsok

A keresési stratégia mind az öt rendszeren ideiglenesen (általában az adott hét végéig) és tartósan megőrizhető, későbbi újrafuttatásra. A Pergamon InfoLine rendszeren azonban egyidejűleg több keresőprofil nem őriztethető meg ideiglenesen.

Az STN parancsnyelv ezenkívül egyedülálló lehetőséget nyújt a *találati halmazok* megőrzésére is, ami lehetővé teszi a rekordok későbbi kinyomtatását. Erre pl. olyankor lehet szükség, ha a keresés során nem tudjuk azonnal eldönteni, hogy relevánsak-e a találataink. Ilyenkor a keresést megszakítva és konzultálva a végfelhasználóval, majd ismét visszkapcsolódva a rendszerbe, a hét folyamán bármikor kiírathatjuk a megőriztetett találati halmazt. Vagy pl. a rendszer felső határainak (max. megenge-

dett rekord- vagy halmazszám) túllépése esetén megőrizhetjük a feltétlenül szükséges rekordhalmazokat, majd a nemkívánatos halmazok törlése és a megőrzött halmazok előhívása után tovább folytathatjuk a keresést. A találati halmazok megőrzésének a lehetősége arra is módot ad, hogy kérjük a folyamatos témafigyelés (SDI) eredményének megőrzését helyi nyomtatás céljára.

Valamennyi rendszerben van lehetőség a megőrzött keresőprofil "megtisztítására", vagyis azoknak a halmazoknak a törlésére, amelyekre a későbbi futtatásnál nincs szükségünk. A Dialog és a Pergamon InfoLine rendszereken erre külön szövegszerkesztő program (EDIT) szolgál, míg a másik három rendszeren ez egyszerű parancsokkal végezhető el.

## 2.6 Találati rekordok adatalemeinek automatikus felhasználása keresésre

Az Orbit, az STN és a Pergamon InfoLine parancsnyelvvel valamennyi adatbázisban lehetővé teszi, hogy adott rekordhalmaz tetszőleges adatmezőjében található egyes adatalembeket (pl. deskriptorokat, a szerzők nevét) újra-beírásuk nélkül, automatikusan felhasználjuk további keresésre, ugyanabban vagy egy másik adatbázisban. Erre az Orbit rendszeren a PRINT SELECT, a Pergamon InfoLine-on pedig a GET parancs szolgál. A kigyűjtött adatalembeket az eredetitől eltérő adatmezőre is átminősíthetjük (QUALSELECT, ill. a GET parancs speciális argumentuma). Az STN-en csak az elmúlt évben vezették be ezt a lehetőséget (SELECT, ill. EDIT az adatmező átminősítésére). A Dialog rendszer MAP parancsa csak meghatározott adatbázisok bizonyos adatmezőiben működik, tehát nem annyira általános érvényű, mint az előbb említett rendszereken. A Data-Staron viszont egyáltalán nincs ilyen lehetőség.

Az adatalemek kigyűjtésének és keresésre való újrafelhasználásának néhány tipikus esete:

- ◆ a CA vegyületszótár-adatbázisban megtalált Registry Numerek automatikus átvitele a CA bibliográfiai adatbázisba és a rájuk vonatkozó hivatkozások megkeresése (erre a Dialog rendszeren a MAPRN parancs szolgál, lásd 6. ábra);
- ◆ a CA vegyületszótár-adatbázisban megkeresett vegyületek szinonim elnevezéseinek kigyűjtése, majd felhasználásuk más adatbázisokban való keresésre. A Dialog MAPSY parancsával kigyűjtött, több szóból álló vegyületnevek más adatbázisba való átvitelekor a Dialog például automatikusan a (W) helyzeti operátorral (egymás mellett, adott sorrendben) kapcsolja össze az egyes szavakat. Ez rendkívül megkönnyíti a vegyületek keresését



```

File 301:CHEMNAME(tm) 1967-SEP85 1,547,123 subs
(Copr. Dialog Inf.Ser.Inc.1986)

Set Items Description
-----
?SELECT SY=ALUMINA
S1 1 SY=ALUMINA
?TYPE 1/2

1/2/1
CAS REGISTRY NUMBER: 1344-28-1
FORMULA: Al2O3
REPLACED CAS REGISTRY NUMBER(S): 12522-88-2
12737-16-5 39354-49-9 53809-96-4 67853-35-4 67894-14-8
67894-42-2 68189-68-4 68389-42-4 68389-43-5 90669-62-8
CA NAME(S):
HP=Aluminum oxide (Al2O3) (9CI)
HP=Aluminum oxide (8CI)
SYNONYMS: Faserton; Fasertonerde; Alumina; Exolon
XW 60; .alpha.-Alumina; .gamma.-Alumina; .gamma.-Alum-
inum oxide; .alpha.-Aluminum oxide; Lucalox;
.eta.-Alumina; Ludox CL; Alon; Aluminum trioxide;
Martoxin; Conopal; Alumite; Alumite(oxide); Aluminum
oxide (Brockmann); .beta.-Aluminum oxide; LA 6;
Dialuminum trioxide; Alundum 600; Dispall M; PS
1(alumina); Cab-O-Grip; Almite; KHP 2; Aloxit; Hypalox
II; KA 101; Aluminum sesquioxide; Catapal S; Poraminar
Dispall; Ketjen B; Aluminite 37; PS 1; Alcoa F 1; Alon C
; Compalox; Dotment 324; Dotment 358; Q-loid A 30;
Neobead C; Microgrit WCA; .delta.-Aluminum oxide;
Alumogel A 1; A 1; A 1(sorbent); F 360; F 360(alumina)
Jubenon R; G 0(oxide); G 0; G 2(oxide); GK(oxide); G 2
GK; RC 172DBM; Alumina C.gamma.; Aluminasol 100; Aero
100; Kimal; Linde A; CKA; AQ 50; AQ 25; AQ 10; MM 21;
MA 11(metal oxide); MA 11; Kyoward 200; JRC-ALO 4; AL
13(oxide); AL 13; KHD 46

*7MAPRN S1
1 select statement(s)
serial SDD37

?BEGIN 309
File 309:CA Search - 1972-1976
(Copr. 1984 by the Amer. Chem. Soc.)

Set Items Description
-----
?EXECUTE STEPS SDD37
10956 RN=1344-28-1 (SEE ?IGNOTE)
28 RN=12522-88-2
0 RN=12737-16-5
1 RN=39354-49-9
49 RN=53809-96-4
0 RN=67853-35-4
0 RN=67894-14-8
0 RN=67894-42-2
0 RN=68189-68-4
0 RN=68389-42-4
0 RN=68389-43-5
0 RN=90669-62-8
S1 11007 RN=1344-28-1 + RN=12522-88-2 +
RN=12737-16-5 + RN=39354-49-9 +
RN=53809-96-4 + RN=67853-35-4 +
RN=67894-14-8 + RN=67894-42-2 +
RN=68189-68-4 + RN=68389-42-4 +
RN=68389-43-5 + RN=90669-62-8

```

6. ábra A MAPRN parancs használata a Dialog rendszeren

azokban az adatbázisokban, amelyekben nincsenek deskriptorok vagy egyéb kötött keresőnyelvi eszközök (pl. speciális kódok) a vegyületkeresésre (7. ábra);

- ♦ a CA adatbázisban talált szabadalmak szabadalomszámának vagy prioritási számának kigyűjtése, majd a megfelelő rekordok megkeresése

ezek alapján a Derwent WPI adatbázisában, amely a CA-val ellentétben kivonatot (mégpedig nagyon jót) és a szabadalomcsaládra vonatkozó információt is tartalmaz (erre a Dialogban a MAPPN, ill. a MAPAN parancs szolgál);

- ♦ egy ismert, releváns szabadalom rekordjáról a nemzetközi és/vagy az amerikai szabadalmi osztályozás jelzeteinek kigyűjtése, majd felhasználásuk az adott témakörben végzett téma szerinti keresésnél (a Dialogban a MAPIC vagy a MAPCL parancs segítségével);
- ♦ a WPI adatbázisban, a nemzetközi szabadalmi szervezetek (EPO és WIPO) szabadalmaiban hivatkozásként szereplő szabadalmak számának kigyűjtése, majd ezekre keresve a hivatkozott vagy a rájuk hivatkozó, későbbi szabadalmak rekordjainak megkeresése (a Dialog MAPCT parancsa);

```

? BEGIN 301
File 301:CHEMNAME(TM) 1967-Jun86 1,689,212 SUBS
(Copr. Dialog Inf.Ser.Inc.1986)

Set Items Description
-----
? SELECT SY=LIDOCAINE
S1 1 SY=LIDOCAINE
? TYPE 1/5

1/5/1
CAS REGISTRY NUMBER: 137-58-6
FORMULA: C14H22N2O
RING SYSTEM DATA:
(01) (nr=01; sr=6; ar= fr=C6.01; ir=46-150-18)
REPLACED CAS REGISTRY NUMBER(S): 8059-42-5 8059-66-3 91484-71-8
CA NAME(S):
HP=Acetamide (9CI). SB=2-(diethylamino)-N-(2,6-dimethylphenyl)-HP=2',6'-Acetoxyilidide
(8CI). SB=2-(diethylamino)-
SYNONYMS: alpha.-Diethylamino-2,6-acetoxyilidide; 2-(Diethylamino)-2',6'-acetoxyilidide;
Isicaine; Leostes in; Lidocaine; Lignocaine; Maricaine; Rucaina; Xycaine; Xylocaine; Xylocitin;
Duncaine; Xylestesin; Gravocain; Anestacon; Solcain; Xylocain; Isicaina; Xilina; Xylina; Xylina;
Remicaine; Medicaine

? MAPSY TEMP
1 select statement(s)
serial TD342
? BEGIN 16
File 16 PROMT - 72-86/Aug. Week 4
(Copr. Predicasts Inc. 1986)
** Monthly SDI service for this file is now available.
** Weekly SDI service is only available on DIALOG 1.

Set Items Description
-----
? EXS TD342
0 ANESTACON
0 DUNCAINE
0 GRAVOCAIN
0 ISICAINE
0 ISICAINE
0 LEOSTESIN
31 LIDOCAINE
9 LIGNOCAINE
0 MARICAINE
1 MEDICAINE
0 REMICAINE
0 RUCAINA
0 SOLCAIN
0 XILINA
0 XYCAINE
0 XYLESTESIN
1 XYLINE
1 XYLOCAIN
7 XYLOCAINE
0 XYLOCITIN
S1 45 ANESTACON + DUNCAINE + GRAVOCAIN + ISICAINE +
ISICAINE + LEOSTESIN + LIDOCAINE + LIGNOCAINE +
MARICAINE + MEDICAINE + REMICAINE + RUCAINA +
SOLCAIN + XILINA + XYCAINE + XYLESTESIN + XYLINE +
XYLOCAIN + XYLOCAINE + XYLOCITIN

? TYPE 1/5/3
1/5/3
1350863
Pharmaceuticals:Anaesthetic replacement from Astra.
Manufacturing Chemist & Aerosol News April 0, 1986 p. 20
Astra Pharmaceuticals (US) has developed a local anaesthetic cream, Emla, which
could replace local anaesthetic injection. An eutectic mixture of prilocaine and lignocaine
has been developed to overcome the problem of a formulation containing anaesthetic in
sufficient concentration for it to be absorbed. The cream is applied for a minimum of 1
hour under an occlusive dressing. When the cream is wiped off, anaesthesia lasts for 30
min.
*1USA *United States *2834280 *Anesthetic Preparations *33 *products; *Astra
Pharmaceutical Prod

```

7. ábra A MAPSY parancs használata a Dialog rendszeren

- ◆ a WPI adatbázisban a szabadalmak prioritási számának ismételt felhasználása keresésre az ún. eltérő, nem uniós elsőbbségű, de az adott szabadalommal tartalmi kapcsolatban lévő (related) szabadalmak feltárására (MAPAN vagy MAPPR a Dialogban);
- ◆ a téma szerinti kereséssel kapott találati rekordokon szereplő cégekre (pl. szabadalomtulajdonos, szerző munkahelye) vonatkozó információk megkeresése a céginformációs adatbázisokban.

## 2.7 Közvetlen keresés a soros állományban

Már a helyzeti operátorokkal való keresés (2.1 pont) és a csonkolás (2.2 pont) sajátosságainak tárgyalásakor utaltam erre a keresési lehetőségre, amelyet csak az Orbit (STRINGSEARCH, SENSERSEARCH) és a Pergamon InfoLine (SCAN) nyújt. (Korlátozott mértékben a teljes szövegű adatbázisokban az STN-en is lehetséges, a DISPLAY BROWSE parancson belüli SEARCH parancssal). A soros keresés kifinomult módot nyújt arra, hogy tetszőleges karaktersort, kifejezést keressünk egy már kiválasztott, lehetőleg nem túl nagy rekordhalmaz rekordjain belül. Így alkalmazhatjuk a bal oldali csonkolást, tiltott szavakat (stopwords, pl. of, for, by) tartalmazó kifejezésekre is kérdezzük stb.

## 2.8 Korlátozási (LIMIT) parancs

Ezzel a parancssal egy már kiválasztott rekordhalmaz tovább szűkíthető év, nyelv, dokumentumtípus stb. szerint. Erre a legegyszerűbb és leggyorsabb lehetőséget a Dialog biztosítja. A Dialog rendszeren például valamennyi adatbázisban lehetőség van az adatbázisba való bekerülés éve szerinti korlátozásra a rekordok azonosítószáma (Accession Number=AN) alapján, amit pillanatok alatt elvégez a rendszer:

LIMIT n/650001-999999

↑ halmazszám     ↑ alsó határ     ↑ elméleti felső határ  
(megfelel valamely év első AN-jének)

A többi korlátozási lehetőség adatbázisonként változik. A DIALOG 2 lehetőséget ad arra is, hogy a LIMIT parancs paramétereit a SELECT parancssal használjuk a keresőszó vagy halmazszám után bevitt utótag (suffix) formájában: pl. *S Laser/ENG* a lézerrel kapcsolatos angol nyelvű publikációk keresésére. A keresés elején bevitt LIMITALL pa-

rancssal pedig az egész további keresés korlátozható bizonyos szempont szerint.

A Data-Star LIMIT parancsa nem terjed ki minden adatbázisra, így pl. a CA adatbázisban nem alkalmazható, és adatbázisonként változnak a korlátozható adatmezők. Itt a LIMIT parancsban mindig meg kell adni a mezők azonosítóját (amely szerint a korlátozás történik), egy relációs operátort (=, >, <) és az adatelem értékét:

..LIMIT/N YR >82 az *n*-edik sorszámú halmaz korlátozása az 1982 után publikált dokumentumokra.

Az STN RANGE parancsa csak időszak szerinti szűkítést tesz lehetővé.

Az Orbit és a Pergamon InfoLine parancsnyelve nem rendelkezik korlátozó parancssal. Az Orbiton azonban lehetőség van a keresőmódban évszám szerinti korlátozásra relációs operátorok alkalmazásával: pl. *n AND PY >82* (*n* = halmazszám). Az Orbiton és a Pergamon InfoLine-on — nem túl nagy halmazok esetén — leghatékonyabban soros kereséssel szűkíthetünk a korlátozni kívánt adatelem aktuális értékére. Például az angol nyelvű publikációkra való korlátozás:

STRINGSEARCH/LA English (Orbit)

SCAN LA=English (Pergamon InfoLine)

## 2.9 Nyomtatási formátumok

Valamennyi rendszer lehetőséget nyújt bizonyos adatmezők kombinációjából álló, "előre gyártott" formátumok vagy tetszőleges adatmezők online vagy offline kinyomtatására. A Data-Staron és a Dialogon a felhasználó által gyakran használt, adatbázis-specifikus formátumok is definiálhatók. Tetszőleges adatmezők kijelölésének a lehetőségét, mint nyomtatási formátumot, a Dialog és az STN rendszeren csak nemrég vezették be. Ennek segítségével a CA adatbázisban pl. mód van arra, hogy olyan formátumokat jelöljünk ki, amelyek az indexmezők közül a relevancia megítéléséhez általában elegendő információt tartalmazó és nem túl hosszú azonosítómezőt tartalmaznak egyéb mezőkkel (pl. csak a cím- vagy a bibliográfiai adatokat tartalmazó mezők) kombinálva, de a gyakran egy-két oldal hosszúságú deskriptormező nélkül. Ennek különösen az online nyomtatásnál van jelentősége.

A Dialog, a Data-Star és az STN teljes szövegű adatbázisaiban, valamint az Orbit néhány adatbázisában kiíratható a rekordnak az a része, amelyben a keresőszavak előfordulnak. Erre a Dialog rendsze-

ren a *KWIC* (KeyWord-In-Context), a Data-Staron a *HITS* és az *OC* (Occurence), az Orbiton a *HIT* paraméter nyomtatási parancson belüli alkalmazásával, az STN-en pedig a *DISPLAY BROWSE* parancson belül van lehetőség. Az STN-en ezenkívül automatikusan, a Dialog rendszeren pedig a *HIGHLIGHT* lehetőség (SET HIGHLIGHT ON) alkalmazásával a szövegen belül a találatot eredményező keresőszavak, megkülönböztetve, kiemelten jelennek meg.

A Dialog egyes, elsősorban céginformációkat tartalmazó adatbázisokban speciális kinyomtatási lehetőséget kínál a *REPORT* paranccsal. A találati halmaz rekordjainak néhány adatalemét (pl. a cég neve, létszáma és éves forgalma) ezzel táblázatba rendezve írathatjuk ki. Ez a nyomtatási forma egyrészt áttekinthetőbb, másrészt olcsóbb, mint a teljes rekord kinyomtatása.

A Pergamon InfoLine egyedülálló, statisztikai analízis végzésére alkalmas *GET* parancsának segítségével egy adott rekordhalmaz valamely adatmezőjében szereplő adatalemeket előfordulási gyakoriságuk szerint rangsorolva gyűjthetjük ki (pl. azoknak a cégeknek a nevét az INPADOC adatbázisban, amelyek egy adott témakörben a legtöbb szabadalmat nyújtják be). A kigyűjtött adatalemek ismételt keresésre való felhasználásáról már a 2.6 pontban szó volt.

## 2.10 Grafikus keresési lehetőség

Erre csak az STN-en elérhető CAS Online rendszer Registry File-jában van lehetőség. Grafikus terminálon közvetlenül felrajzolható, alfanumerikus karakterkészletű terminálon pedig egyszerű utasításokkal felépíthető a keresni kívánt vegyület szerkezeti képlete. Lehetőség van adott szerkezeti fragmentumokat (egyes kötések, vegyértékek nyitva hagyva) tartalmazó vegyületek családjának megkeresésére (szubstruktúra-keresés), több szubsztituens-variáns egyidejű megadására (Markush-képlet) is. A vegyületkeresés legegyszerűbb és leghatékonyabb módja a szerkezeti képlet alapján való grafikus keresés, ami a kémiai elnevezések alapján gyakran még a vegyész szakembereknek sem könnyű feladat. Az STN ezzel a keresési lehetőséggel egyedülálló helyet foglal el az öt rendszer között.

Érdekes megfigyelni, hogyan reagáltak a CA adatbázist szintén szolgáltató központok az STN-nek arra a kihívására, amit a grafikus keresés bevezetése és a CA-kivonatok online elérése (kizárólag az STN-en) jelentett. A Pergamon InfoLine "kiszállt" a versenyből, és visszavonta a CA adatbázist a rendszeréről. A Data-Star megtartotta ugyan a vegyületszótár-adatbázisát (CNAM), de csak korlá-

tozott keresési lehetőséggel: nemcsak hogy felhagyott az adatbázis továbbfejlesztésével, hanem törölte például a kifinomultabb keresést lehetővé tevő, a gyűrűrendszerekre vonatkozó adatalemeket is. A Dialog ezzel szemben olyan kifinomult vegyületkeresési lehetőségeket dolgozott ki, amelyek szinte felveszik a versenyt a grafikus kereséssel. Eredményes alkalmazásuk azonban — az egyszerűbb kérdéstípusokat leszámítva — alapos kémiai ismereteket kíván meg.

## 3. Adatbázis-kínálat

A Dialog nyújt az adatbázisok legnagyobb választékához hozzáférést, szinte valamennyi tudományos, műszaki, gazdasági, társadalomtudományi és humán szakterületen. 250-nél több adatbázisával a világ legnagyobb szolgáltatóközpontja. A műszaki adatbázisok kínálatából talán egyedül a gépészeti, bányászati és építészeti adatbázisok választéka nem kielégítő.

A második helyen a Data-Star áll több mint száz adatbázisával. Elsősorban az orvosbiológiai és az üzleti adatbázisok területén foglal el kiemelkedő helyet. Előnye az is, hogy sok nagy adatbázishoz, amelyek például a Dialog rendszeren csak több részállományban kereshetők, egyetlen állományban, az ún. szuperállományban (ZZ végződésű adatbázis-címkevel) is hozzáférést biztosít. Ilyen adatbázisok pl. a CA (CHZZ), a Medline (MEZZ), az Embase (EMZZ); a Biosis (BIZZ), a Predicasts cég PTS PROMT, PTS Forecasts és PTS Time Series adatbázisait egyesítő (PTZZ) adatbázis. 1986-tól a CA és a Biosis adatbázist a Dialog is elérhetővé tette egyetlen állományban. Az összevont adatbázisokban való kereséssel a keresési idő, ezzel a költségek is csökkenthetők (lásd 4.1 pont).

Az STN-en elérhető adatbázisok száma 30 körül van, de ez a szám tovább bővül majd, ahogy az INKA rendszer adatbázisai fokozatosan átkerülnek az STN-re, illetve ha bekapcsolódik a hálózatba a Japán Tudományos és Műszaki Információs Központ (JICST) is, mint a hálózat harmadik csomópontja. Az STN azon kívül, hogy itt kereshető leghatékonyabban a CA, néhány egyedülálló adatbázishoz is hozzáférést nyújt; pl.: DECHEMA (vegyipari technológia), ICONDA (építészet), BIOMASS (a biomassa energetikai felhasználása), CJACS (az Amerikai Kémiai Társaság 17 folyóiratának teljes szövege), CAOLD (az 1967 előtti CA-hivatkozások, amit 1920-ig visszamenően terveznek kiterjeszteni).

Az Orbit rendszer mintegy 60 adatbázisa közül néhány kizárólag rajta keresztül érhető el: API-

LIT/APIPAT (petrolkémia), TULSA (kőolajbányászat), PESTDOC (növényvédő szerek), VETDOC (állatgyógyászati szerek), CRDS (szerves kémiai reakciók), Biotechnology (biotechnológia), CEH Online (vegyianyag-árak). Az Orbit különösen a szabadalmi keresésekhez rendelkezik sok adatbázissal és előnyös keresési lehetőségekkel (pl. PRINT SELECT).

A Pergamon InfoLine csaknem 60 adatbázisának nagyobbik része exkluzív, és főleg az ilyen adatbázisok számát növelik. Közöttük szerepelnek olyan fontos adatbázisok, mint a RAPRA (műanyag- és gumiipar), a PIRA (papírgyártás, csomagolás, nyomdaipar), az Aqualine (vízügy) és az INPADOC (a világ legteljesebb körű szabadalmi adatbázisa). Az Orbit rendszer megvásárlása az SDC-től és a BRS, valamint az ESA-IRS szolgáltatóközpontokkal tervezett kapuszolgálat (gateway) minden valószínűséggel jelentősen ki fogja szélesíteni a Pergamon InfoLine felhasználóinak táborát. Az erősödő piaci versennyel való lépéstartás jellemző módja az utóbbi években az erőforrások egyesítése, cégek megvásárlása, kapuszolgálatok megnyitása, vagy — mint az STN esetén — hálózat létrehozása.

Az olyan témák keresésénél, amelyek több adatbázis igénybevételét is megkívánják, természetesen azok a rendszerek élveznek előnyt, amelyekben lehetőleg valamennyi, a témához szükséges adatbázis megtalálható. Van egy speciális lehetőség, amely segít a megfelelő adatbázisok kiválasztásában. A Dialog, a Data-Star és az Orbit rendszereken létezik egy-egy olyan adatbázis, amely a rendelkezésre álló valamennyi adatbázis szótárállományait tartalmazza. Ebben az adatbázisban a témát leíró egyetlen keresőkérdés lefutatható egyszerre több adatbázis szótárában is, és a kapott találatok alapján kiválaszthatók a legmegfelelőbb adatbázisok. A Dialog rendszeren az erre szolgáló *DIALINDEX*-ben és a Data-Staron a *CROS* adatbázisban a teljes szótárállomány kereshető, és a keresőkérdésben logikai és helyzeti operátorokat is használhatunk. Az Orbit *DBI* (Data Base Index) összevont szótár-adatbázisába csak az alapszótár adatai kerülnek, és egyedi szavakra vagy kifejezésekre keresve, annak az öt adatbázisnak a nevét kapjuk meg, amelyekben a kérdéses keresőszó a legtöbbször előfordul.

## 4. Időtényező

### 4.1 A keresési idő

A keresési idő függ az adatbázis nagyságától, a keresőprofil bonyolultságától, a csonkolás mértékétől, gyakran még attól is, hogy a nap melyik szakában

végeztük a keresést. A 2.1 pontban már kitértem arra, hogy a Dialogon például az S STEPS parancs alkalmazása egy összetett keresőkérdés bevitelére — amikor minden egyes keresőszóra külön halmaz képződik — időigényesebb, mint a sima SELECT esetén. A Data-Star, a Pergamon InfoLine és az Orbit rendszeren a nagy halmazokkal végzett logikai műveletek végrehajtása tart nagyon sokáig.

A 2.1 pontban bemutatott, a Dialog, a Data-Star és az STN összevont CA állományában végrehajtott mintakeresés (4. ábra) keresési időit összehasonlítva, az STN bizonyult a leggyorsabbnak, ezt követi a Data-Star, majd a Dialog. A keresési idő az STN-en mintegy 35%-kal volt rövidebb, mint a Dialogon, és 16,5%-kal rövidebb, mint a Data-Staron; a Data-Star pedig mintegy 22–23%-kal bizonyult gyorsabbnak a Dialognál [2]. Az öt rendszer közül leglassúbbnak az Orbitot és a Pergamon InfoLine-t találom, de ezt számadatokkal nem tudom alátámasztani.

Azoknál az adatbázisoknál, amelyek több részállományban és egy állományban is kereshetők, a válaszidő megnő ugyan az összevont állományban, de a keresési idő mégis rövidebb, mintha ugyanazt a keresőprofil minden egyes részállományban végig kellene futtatni. A CA adatbázis esetén a 4. ábra szerinti keresés mind a Dialogon, mind a Data-Staron mintegy 20%-kal rövidebb ideig tartott az összevont állományban, mint az egyes kollektív indexperiódusoknak megfelelő részadatbázisokban kapott idők összege [2].

### 4.2 Az offline printek megérkezésének ideje

Az STN (a karlsruhei csomópontozat kapcsoldva), a Data-Star és a Pergamon InfoLine előnye számunkra, hogy az offline printek általában egy hét alatt megérkeznek, az STN-től néha még hamarabb. Az USA-beli Dialog és Orbit szolgáltatóközpontok esetén ez az idő átlagosan két hét. A hosszú postai átfutási idő kiküszöbölésére mind a Dialog, mind az Orbit bevezette az offline printek elektronikus továbbítását (electronic mailbox). A Dialog *DIALMAIL* rendszerén keresztül a keresés utáni napon viszonylag alacsony óradíjjal (12 \$/óra) online kiirathatók az offline megrendelt találatok, a nyomtatási parancs argumentumába mindössze annyit kell beírni, hogy *VIA DIALMAIL*. Az Orbiton pedig a *PRINT ELEC(TRONIC)* paranccsal megrendelt találatok legkésőbb két óra múlva kiirathatók online a *PRINTS* nevű adatbázisból, hasonlóan alacsony óradíj mellett. Ezzel a lehetőséggel általában csak akkor élünk, ha a találatok száma nem

túl nagy, mert a magyar NEDIX-hálózat által megengedett 300 baud adatátviteli sebesség mellett az online nyomtatás sok időt vesz igénybe, és a ki-nyomtatott anyag nem olyan jó minőségű, mint a hagyományos offline nyomtatásnál.

## 5. Költségtényező

Az adatátviteli költségeket tekintve egyértelmű, hogy azok az európai szolgáltatóközpontok esetén kisebbek. A NEDIX-hálózat használatának díja Európába 15 Ft/perc, az USA-ba pedig 25 Ft/perc.

A szolgáltatóközpontoknak fizetendő költségeket (adatbázis-használati díj, nyomtatási díj) illetően már nehezebb az összehasonlítás. A költségeket még a valutaárfolyamok alakulása is befolyásolja, így ha a dollár erős, a nem dollárelszámolású központok az olcsóbbak és fordítva. Ha egy adatbázis több szolgáltatóközponton is elérhető, általában nincsenek nagy különbségek az óradíjakban (kivéve, ha valamelyik központ az előfizetők árkedvezményt kapnak, mint a CA esetén a STN-en).

A tényleges költségeket inkább a rendszer válaszideje és a kereső személy gyakorlottsága szabja meg. A kettő közül általában a gyakorlottság a meghatározóbb. Így például én a CA adatbázist legtöbbször a Dialog rendszeren keresem, jóllehet az STN-en sokszor gyorsabb a válaszidő. A Dialog parancsnyelvet azonban jobban ismerem, és eredményesebben tudom alkalmazni [2]. A 4. ábrán bemutatott mintakeresés teljes költsége (beleértve az adatátviteli költséget is) az STN-en keresztül volt a legalacsonyabb, a Dialoghoz képest mintegy 40%-kal, a Data-Starhoz képest pedig csaknem 14%-kal volt olcsóbb [2], hasonlóan a 4.1 pontban leírt időarányokhoz. Az említett példa azonban nem egy tipikus keresés, amikor menet közben változtatni, módosítani kell a keresési stratégiát. Az utóbbi gyors és hatékony elvégzésében az adott parancsnyelvben való jártasság, "otthonosság" igen sokat számít. Egy kevésbé ismert parancsnyelv esetén az erre fordított hosszabb idő miatt nem biztos, hogy ki tudjuk használni a gyorsabb válaszidőt.

Végezetül, véleményem szerint a költségeknél is fontosabb, hogy az adott rendszerben megvannak-e azok a keresési sajátosságok, amelyek az adott téma leghatékonyabb lekérdezését elősegítik, és várhatóan a legjobb eredményt adják.

## 6. A felhasználókat támogató szolgáltatások

A Data-Star kivételével valamennyi rendszeren kérhetünk online információkat a rendszer paran-

csairól, adatbázisairól stb. (HELP, EXPLAIN). A Dialog, az Orbit és az STN azt is lehetővé teszik, hogy online üzenetet küldjünk a szolgáltatóközpontnak, vagy pl. a rendszer használatára vonatkozó keresési segédleteket rendeljünk meg (DIALMAIL, DIALORDER, COMMENT, SEND). Valamennyi szolgáltatóközpontban vannak a rendszer használatát ismertető keresési segédletek (a parancsnyelvet leíró kézikönyv, adatbázis-leírások, newsletter). Összehasonlítva az öt rendszer keresési segédleteit és a felhasználók tájékoztatásának gyorsaságát, alaposágát a különféle változásokról, a Dialog messze kiemelkedik az öt rendszer közül, leghátul pedig a Data-Star áll. Ugyanez vonatkozik a felhasználók képzésére is. A sokféle és rendszeresen megtartott tanfolyamok mellett a felhasználók képzését nagymértékben elősegíti a Dialog olcsó oktató adatbázisainak (ONTAP = Online Training And Practice) gazdag választéka és az ún. Classroom Instruction Program (CIP), amelynek keretében oktatási célokra a legtöbb adatbázis teljes állománya kereshető minimális – 15 \$/óra – óradíj ellenében.

## 7. Összefoglalás

Az ideális parancsnyelv az lenne, amely a szövegesen megfogalmazott kérdés alapján elvégezni mindazt, ami ma még nagyrészt a közvetítő szakemberre hárul (az adatbázis kiválasztása, a keresőprofil megszerkesztése és a keresés végrehajtása stb.). Ennek egy közelítő megoldását nyújtja a gazdasági adatbázisok keresésére a Dialog menü által vezérelt Business Connection szoftverje és még néhány, Magyarországon nem használt rendszer. Maradjunk azonban a valóság talaján, és nézzük meg, hogy az öt jelentős szolgáltatóközpont parancsnyelvére mennyire felel meg a hagyományos értelemben vett parancsnyelvvvel szemben támasztott igényeknek.

A 2. táblázatban összefoglaltam, hogy a cikkben tárgyalt, általam fontosnak tartott keresési jellemzők mely rendszereken található meg, és melyeken nem.

A DIALOG 2 bevezetése óta a Dialog parancsnyelvre tekinthető a legfejlettebbnek, különösen akkor, ha figyelembe vesszük a rendszer megbízhatóságát és a felhasználókat támogató szolgáltatásokat is. Amit még szívesen látnék: az egy mondaton belüli keresést biztosító (S) operátor és a MAP parancs szélesebb körű, valamennyi adatbázisra kiterjedő alkalmazhatósága; a bal oldali csonkolás; a GET-hez hasonló statisztikai analízis program; grafikus keresési lehetőség.

A Dialog, a Data-Star, az Orbit, a Pergamon InfoLine és az STN rendszerek fő keresési sajátosságainak összehasonlítása

Keresési sajátosság	Dialog	Data-Star	Orbit	Pergamon InfoLine	STN
<b>Helyzeti operátorok</b>					
egymás mellett adott sorrendben	+	+	+*	-	+
max. <i>n</i> szó távolságra adott sorrendben	+	-	+*	-	+
NEAR	+	-	-	+	+
LINK	+*	+	+*	+*	+*
alkalmazás halmazszámmal	+	+	+	+	+**
alkalmazás OR-os keresőkifejezésekkel	+	+	+	+	+
Közbenső halmazok képzése	+	-	-	+	+**
<b>Csonkolás</b>					
szóvégi, tetszőleges karakterre	+	+	+	+	+
szóvégi, meghatározott karakterre	+	+	+	-	+
maszkolás	+	-	+	+	+
bal oldali csonkolás	-	-	+*	-	+*
Vegyületnevek fragmentálása	+*	+*	-	-	+*
Halmazszámok utólagos minősítése	+	+	-	-	-
Double Posting	+	+*	+	+*	+
Online tezaurusz	+	-	-	-	-
Keresési stratégia megőrzése	+	+	+	+*	+
Találati halmazok megőrzése	-	-	-	-	+
Találati rekordok adatalemeinek felhasználása keresésre	+*	-	+	+	+
Keresés a soros állományban	-	-	+	+	+*
Korlátozó parancs	+*	+*	-	-	+*
<b>Nyomatás</b>					
tetszőleges adatmezők kiválasztása	+	+	+	+	+
keresőszavakat tartalmazó szövegrész	+***	+***	+*	-	+***
keresőszavak kiemelése a szövegben	+***	-	-	-	+***
táblázatszerkesztés	+	-	-	-	-
<b>Statisztikai analízis</b>					
Grafikus keresés	-	-	-	+	-
Összevont szótár-adatbázis	+	+	+*	-	-
Adatbázisok körülbélüli száma	250	100	60	60	30
Offline printek megérkezésének ideje (nap)	14	7	14	7	7
Offline printek elektronikus továbbítása	+	-	+	-	-
Rendszerinformációk online kérése	+	-	+	+	+
Online üzenetküldés	+	-	+	-	+

\* Korlátozott alkalmazás.

\*\* Nem találati halmazok, csak hozzárendelt L-számok.

\*\*\* Elsősorban a teljes szövegi adatbázisokban.

Egy adott téma keresésére a szolgáltatóközpont kiválasztásának szempontjai között meghatározó szerepük van a parancsnyelv keresési lehetőségeinek, amelyekkel eredményes keresést lehet végezni.

### Irodalom

[1] VÁLÁS György: Online információkeresés a nagyvilágban és az OMIKK-ban. = Tudományos és Műszaki Tájékoztatás, 33. köt. 3. sz. 1986. p. 103–114.

[2] NOVÁK Teréz: Searching the Chemical Abstracts database: a comparison of Dialog, Data-Star and CAS Online. = 10th International Online Information Meeting, London, 2–4 December 1986., Learned Information, Oxford and New Jersey, 1986. p. 353–364.

[3] Az online szakirodalmi információkeresés kézikönyve. 1–2. rész. Szerk.: Novák Teréz, Roboz Péter. Tudományos-Műszaki Információ Szakkönyvtára 16., 17. sz. Bp. OMIKK, 1985. 550 p.

*NOVÁK Teréz: Az ideális parancsnyelv: a Dialog, a Data-Star, az Orbit, a Pergamon InfoLine és az STN parancsnyelvének összehasonlítása*

Az online keresés hatékonyságát biztosító feltételek között alapvető fontosságuk van a parancsnyelv által nyújtott keresési lehetőségeknek. A szerző összehasonlítja a Dialog, a Data-Star, az Orbit, a Pergamon InfoLine és az STN rendszerek parancsnyelvének legfontosabb sajátosságait, azok előnyeit és hátrányait, különös tekintettel a kémiai tárgyú keresésekre. Az összehasonlítás szempontjai: a helyzeti operátorokkal való keresés, a csonkolás, a halmazszámok utólagos minősítése, a szótárállomány felépítése, a megőrzési parancsok, találati rekordok adatelemeinek automatikus felhasználása keresésre, keresés a soros állományban, a korlátozási parancs, nyomtatási formátumok, grafikus keresési lehetőség, adatbázis-kínálat, idő- és költségtényezők, a felhasználókat támogató szolgáltatások. A Dialog parancsnyelv új változatának, a DIALOG 2-nek a bevezetése óta a Dialog parancsnyelve tekinthető a legfejlettebbnek a tárgyalt öt parancsnyelv közül.

\* \* \*

**НОВАК, Т.: Идеальный командный язык: сравнение командных языков Dialog, Data-Star, Orbit, Pergamon InfoLine и STN**

Среди условий, обеспечивающих эффективность дистанционного поиска, важными являются возможности, предоставляемые командным языком. Автор дает сравнение главных особенностей командных языков систем Dialog, Data-Star, Orbit, Pergamon InfoLine и STN, их преимуществ и неблагоприятностей, особенно в отношении поисков химической тематики. Аспекты сравнения: поиск с оператором контекстной близости, усечение, последующее ограничение множеств, построение словаря, команды сохранения, автоматическое использование элементов данных найденных записей для дальнейшего поиска, поиск в последовательном файле, команда ограничения, формы распечатки, возможности графического поиска, перечень баз данных, показатели времени и расходов, услуги для облегчения работы потребителя. Со времени внедрения новой версии командного языка DIALOG 2, Dialog считается самым развитым среди пяти рассмотренных командных языков.

*NOVÁK, T.: The ideal command language: a comparison of the command languages of Dialog, Data-Star, Orbit, Pergamon InfoLine and STN online services*

The most important criteria of the effectiveness of online searching are associated with the search possibilities offered by the command language of the service. A comparison of the main features of the command languages of Dialog, Data-Star, Orbit, Pergamon InfoLine and STN systems is presented, their advantages and drawbacks are analyzed, with special attention to subject searches in chemistry. The aspects of the comparison are proximity searching, truncation, postqualification of set numbers, inverted file structure, save commands, automatic extraction of data elements of retrieved records for subsequent use, searching the direct file, limiting possibilities, print formats, graphic search, database availability, time and cost factors, user services. Since the introduction of the DIALOG 2 command language, the new version of Dialog service, it can be regarded as the most advanced one of the five compared.

\* \* \*

*NOVÁK, T.: Die ideale Befehlsprache: Vergleich der Befehlsprachen Dialog, Data-Star, Orbit, Pergamon InfoLine und STN*

Unter den Bedingungen, die die Wirksamkeit der On-Line-Suchung sichern, haben die Suchungsmöglichkeiten der Befehlsprachen eine grundlegende Bedeutung. Der Verfasser vergleicht die wichtigsten Charakteristika der Befehlsprachen der Systemen Dialog, Data-Star, Orbit, Pergamon InfoLine und STN, deren Vor- und Nachteile, mit besonderer Rücksicht auf die Suchungen im Bereich der Chemie. Die Hinsichten des Vergleichs: die Suchung mit Situationsoperatoren, die Absetzung, die nachträgliche Qualifikation der Mengenzahlen, der Aufbau des Wörterbuchbestandes, die Befehle der Aufbewahrung, die automatische Verwendung der Datenelemente der Trefferrekorde für die Suchung, Suchung im Reihenbestand, Limitationsbefehle, Druckformate, Möglichkeiten der graphischen Suchung, Datenbasisangebot, Zeit- und Kostenfaktoren, Dienstleistungen für die Verbraucher. Seit der Einführung der neuen Variante der Dialog-Befehlsprache, der Dialog 2, kann die Dialog-Befehlsprache als die entwickeltste der erwähnten fünf Befehlsprachen betrachtet werden.