

STN[®]

Analyzing CAS Data

John Zabilski

Senior Application Specialist, CAS

CAS is a division of the American Chemical Society

Session Agenda

- Overview
- Analyzing search results
 - To improve a search strategy
 - To determine competitive trends
 - To determine research trends

The analysis of search results can facilitate many work processes:

- Improving a search strategy
- Determining technology-related business trends
- Identifying research opportunities

Competitive data can be obtained from the CAplusSM file

For this type of analysis	Consider these data
Top organizations	PA, CS.ORG
Top researchers	AU, IN
Assessing technology	CT, IC, NCL, MC
Discovering trends	PY, PY.B, PRY PRC, PC PNC

Competitive data can be obtained from the REGISTRYSM file

For this type of analysis	Consider these data
Analysis of chemical substances	IN, CN, MF RID, ES SEQ

STN tools are available for search results analysis

Availability	Tool
Online	<ul style="list-style-type: none">◆ ANALYZE command◆ TABULATE command
STN Express [®] with <i>Discover!</i> [™] wizards	<ul style="list-style-type: none">◆ Analyze wizard◆ R-group analysis wizard

Session Agenda

- Overview
- Analyzing search results
 - to improve a search strategy
 - ANALYZE command
 - ANALYZE displays
 - to determine business trends
 - to determine research trends

Analyzing answer sets to improve a search strategy

Search Question: Locate research on membrane separation, which is used in the purification of water or treatment of wastewater.

Run a preliminary search using known terminology:

=> FILE CAPLUS

=> S MEMBRANE SEPARATION AND (WATER PURIF? OR WASTEWATER TREAT?)

L1 4587 MEMBRANE SEPARATION AND (WATER PURIF? OR WASTEWATER TREAT?)

=> D SCAN

•••

TI Membrane separation apparatus and its operation method

ST membrane permeation solid liq sepn app; water membrane filtration solid liq sepn

IT Membrane filters

(app.; membrane sepn. app. for solid-liq. sep

IT Water purification

(filtration; membrane sepn. app. for solid-li

IT Separation

(liq.-solid; membrane sepn. app. for solid-li

IT River waters

(purifn. by membrane sepn. app.;

membrane sepn. app. for solid-liq. sepn.)

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):END

Database indexing can be used to enhance the search strategy. Relevant terms may be identified by review of no-cost displays.

Extract relevant indexing:

With ANALYZE, up to 5 fields may be extracted. LEN 7 extracts the main group of the IPC.

```
=> ANALYZE L1 1- IC LEN 7 CT
```

```
L2 ANALYZE L1 1- IC CT LEN 7 : 1754
```

```
=> D
```

```
L2 ANALYZE L1 1- IC CT LEN 7 : 1754
```

TERM #	# OCC	# DOC	% DOC	IC CT
1	5687	2421	52.78	WATER PURIFICATION
2	5577	2572	56.07	WASTEWATER TREATMENT
3	2133	1329	28.97	C02F001
4	773	718	15.65	MEMBRANES, NONBIOLOGICAL
5	768	500	10.90	C02F003
6	684	569	12.40	B01D065
7	608	407	8.87	B01D061
8	523	418	9.11	B01D063
9	432	402	8.76	MEMBRANE FILTERS
10	267	213	4.64	C02F009

The default display of extracted terms shows the top 10 terms based on occurrence.

Display specific subject-relevant indexing:

=> D CT WITH "WATER"

L2 ANALYZE L1 1- IC CT LEN 7 : 1754

TERM #	# OCC	# DOC	% DOC	IC CT
1	5687	2421	52.78	WATER PURIFICATION
2	5577	2572	56.07	WASTEWATER TREATMENT
12	218	208	4.53	WASTEWATER TREATMENT SLUDGE
14	163	130	2.83	WASTEWATER
21	121	119	2.59	DRINKING WATERS
39	79	78	1.70	INDUSTRIAL PROCESS WATERS
52	58	58	1.26	RIVER WATERS
69	41	41	0.89	GROUNDWATERS
73	38	38	0.83	SEAWATER
78	35	34	0.74	WATERS

Index terms containing keywords of interest can be displayed easily using the syntax

WITH "TERM"

Display the top IPCs:

```
=> D OGT50 IC
L2          ANALYZE L1 1- IC CT LEN 7 :    1754 TERMS
```

TERM #	# OCC	# DOC	% DOC	IC CT
3	2133	1329	28.97	C02F001
5	768	500	10.90	C02F003
6	684	569	12.40	B01D065
7	608	407	8.87	B01D061
8	523	418	9.11	B01D063
10	267	213	4.64	C02F009
13	184	130	2.83	B01D071
17	142	95	2.07	C02F011
22	116	97	2.11	B01D013
26	103	91	1.98	B01D069
46	71	48	1.05	B01D021

IPCs with >## occurrences
(where ## is a number) can
be displayed easily using
the syntax

OGT##

Isolate specific terms from the analysis display:

```
=> SELECT L2 3 5 6 7 8; D SELECT  
E1 THROUGH E5 ASSIGNED
```

E1	2133	C02F001/IC
E2	768	C02F003/IC
E3	684	B01D065/IC
E4	608	B01D061/IC
E5	523	B01D063/IC

```
=> SET EXPAND CONTINUOUS
```

The SELECT command is useful for extracting a small number of terms.

Determine IPC code definitions:

```
=> E E1 5
E#    FREQUENCY    AT    TERM
--    -
E6        1        C02E011-14/IC
E7      69085      2      C02F/IC
E8      46886      --> C02F001/IC
E9       3827      2      C02F001-00/IC
E10     684        2      C02F001-02/IC
```

```
=> E E9+INDEX
```

```
E11      3827      --> C02F001-00/IC
          Treatment of water, waste water, or sewage
          (C02F003-00 to C02F009-00 take precedence) (3)
```

```
***** END **
```

The INDEX relationship code provides a definition for the IPC.

Determine definitions (cont'd):

=> E B01D063-00+NT/IC

E12	683	-->	B01D063-00/IC Apparatus in general for separation processes using semi-permeable membranes (5) (IPC EDITION: 5-7)
E13	1322	NT1	B01D063-02/IC . Hollow fibre modules (5) (IPC EDITION: 5-7)
E14	166	NT2	B01D063-04/IC . . comprising multiple hollow fibre assemblies (5) (IPC EDITION: 5-7)
E15	321	NT1	B01D063-06/IC . Tubular membrane modules (5) (IPC EDITION: 5-7)
E16	551	NT1	B01D063-08/IC . Flat membrane modules (5) (IPC EDITION: 5-7)
• • •			

The NT relationship code shows the subgroup headings.

Enhance the search strategy with controlled terminology and IPC codes:

=> S E1-E2 OR ?WATER? (S) (TREATMENT OR PURIF?)

L3 378485 (C02F001/IC OR C02F003/IC) OR ?WATER?(S)(TREATMENT OR PURIF?)

=> S E3-E5 OR MEMBRANE (S) (SEPARAT? OR FILT?)

L4 74176 (B01D065/IC OR B01D061/IC OR B01D063/IC) OR MEMBRANE(S)(SEPN OR SEPARAT? OR FILT?)

=> S L3 AND L4

L5 13974 L3 AND L4

=> S L5 AND PY.B>1998

L6 6114 L5 AND PY.B>1998

With the enhanced strategy, recall increased by ca. 1500 answers. Recall: 4500 answers were obtained in the initial search.

Session Agenda

- Overview
- Analyzing search results
 - to improve a search strategy
 - to determine business trends
 - TABULATE command
 - to determine research trends

Analyzing answer sets to determine business trends

***Search Question:* Analyze the literature obtained on membrane separation to determine**

- **Major organizations filing patents**
- **Countries where innovation is occurring**
- **Major authors/inventors**

Important business information can be gleaned using CPlus data

For an analysis of	Consider extracting these data
Top organizations or patent assignees	CS.ORG PA
Development countries	PRC
Authors/inventors	AU
Publication years	PY.B

Extract the data of interest:

```
=> ANALYZE L6 1- CS.ORG PA AU PRC PY.B
```

```
L7 ANALYZE L6 1- CS.ORG PA AU PRC PY.B : 13405 TERMS
```

Display top organizations conducting research:

```
=> D PA TOP20 DOC
```

```
L7          ANALYZE L6 1- CS.ORG PA AU PRC PY.B
```

TERM #	# OCC	# DOC	% DOC	CS.ORG PA AU PRC PY.B
9	200	200	3.27	KURITA WATER INDUSTRIES, LTD., JAPAN
10	185	185	3.03	TORAY INDUSTRIES, INC., JAPAN
11	182	182	2.98	JAPAN ORGANO CO., LTD., JAPAN
13	105	105	1.72	MITSUBISHI RAYON CO., LTD., JAPAN
14	84	84	1.37	KUBOTA, LTD., JAPAN
15	84	84	1.37	USA
16	79	79	1.29	EBARA CORP., JAPAN
17	76	76	1.24	KURITA KOGYO CO., LTD., JAPAN

• • •

The top ## (where ## is a number) can be displayed easily using the syntax

TOP##

Display countries where innovation is occurring:

=> D PRC 1-

L7 ANALYZE L6 1- CS.ORG PA AU PRC PY.B : 13405 TERMS

TERM #	# OCC	# DOC	% DOC	CS.ORG PA AU PRC PY.B
1	3426	2727	44.60	JP
7	1172	354	5.79	US
8	662	278	4.55	WO
12	276	138	2.26	DE
20	73	71	1.16	CN
26	87	47	0.77	FR
27	61	47	0.77	KR
32	54	40	0.65	RU
34	65	38	0.62	EP
49	67	28	0.46	GB

...

Display major authors:

=> D AU

L7 ANALYZE L6 1- CS.ORG PA AU PRC PY.B : 13405 TERMS

TERM #	# OCC	# DOC	% DOC	CS.ORG PA AU PRC PY.B
29	43	43	0.70	KIHARA, MASAHIRO
30	43	43	0.70	MINEGISHI, TORATARO
31	41	41	0.67	IZUMI, KIYOSHI
33	39	39	0.64	YAMADA, YUTAKA
40	34	34	0.56	MIZUNO, KENICHIRO
41	34	34	0.56	SAWADA, SHIGEKI
42	32	32	0.52	NAKATSUKA, SHUJI
47	30	30	0.49	HIROTA, SHINYA
48	29	29	0.47	SATO, NOBORU
52	27	27	0.44	KOBAYASHI, MASUMI

Trend analysis establishes relationships between 2 data fields

- Accomplished with the TABULATE command
- Terms previously extracted using the ANALYZE command are used
- A 2-dimensional display is created

Important business trends can be gleaned using CPlus data

To assess	Consider correlating these data
Trends in an organization's research	IC, CT, NCL, MC vs. PY, PY.B, PRY
Comparative technologies being pursued by competitors	IC, CT, NCL, MC vs. PA
Top authors	AU vs. PY

Specify the data to be correlated:

=> TABULATE L7

DISPLAY AS GRID FORMAT (N), Y, OR ? : N

ENTER PRIMARY DISPLAY CODE OR (?) : AU

ENTER SECONDARY DISPLAY CODE OR (?) : PY.B

DISPLAY PRIMARY (TOP 10), ENTIRE OR ? : DGT15

DISPLAY SECONDARY (TOP 10), ENTIRE OR ? : ENTIRE

PRIMARY SORT ORDER (CURRENT), DOC, ALPHA, OR ? : DOC

PRIMARY SORT DIRECTION (DEFAULT), A, D, OR ? : D

SECONDARY SORT ORDER (CURRENT), DOC, ALPHA, OR ? : ALPHA

SECONDARY SORT DIRECTION (DEFAULT), A, D, OR ? : A

A FEE WILL BE CHARGED. PROCEED? (Y), N, OR ? : Y

L7 ANALYZE L6 1- CS.ORG PA AU PRC PY.B : 13405 TERMS

Create a 2-dimensional display:

TERM #	# DOC	% DOC	AU	PY.B
29	43	0.70	KIHARA, MASAHIRO	
	9	0.15	--	2003
	8	0.13	--	2002
	19	0.31	--	2001
	5	0.08	--	2000
	2	0.03	--	1999
30	43	0.70	MINEGISHI, TORATARO	
	4	0.07	--	2003
	9	0.15	--	2002
	6	0.10	--	2001
	17	0.28	--	2000
	7	0.11	--	1999
31	41	0.67	IZUMI, KIYOSHI	
	1	0.02	--	2003
	9	0.15	--	2002
	15	0.25	--	2001
	7	0.11	--	2000
	9	0.15	--	1999

• • •

Session Agenda

- Overview
- Analyzing search results
 - to improve a search strategy
 - to determine business trends
 - TABULATE command
 - to determine research trends

Analyzing answer sets to determine research trends

***Search Question:* What ring-containing small molecules used for the treatment of leukemia have been reported in recent pharmaceutical patents?**

Locate recent pharmaceutical patents covering leukemia:

=> FILE CAPLUS

=> S LEUKEMIA/CT AND P/DT AND PY.B>2002 AND
(1/SC OR 63/SC OR 64/SC)

L1 241 LEUKEMIA/CT AND P/DT AND PY.B>2002 AND
 (1/SC OR 63/SC OR 64/SC)

Isolate ring-containing small molecules:

```
=> FILE REGISTRY
```

```
=> TRANSFER L1 1- RN
```

```
L2          TRANSFER L1 1- RN :    7353 TERMS
```

```
• • •
```

```
L3          7353 L2
```

```
=> S L3 NOT SEQUENCE/FS AND C<50 AND RSD/FA
```

```
L4          3063 L3 NOT SEQUENCE/FS AND C<50 AND RSD/FA
```

Extract substance identification information:

```
=> ANALYZE L4 1- RID IN ES
```

```
L5 ANALYZE L4 1- RID IN : 3678
```

```
=> D RID
```

```
L5 ANALYZE L4 1- RID IN ES : 3678 TERMS
```

TERM #	# OCC	# DOC	% DOC	RID	IN	ES
2	1709	1709	55.79	46.150.18		
6	183	183	5.97	46.156.30		
8	162	162	5.29	46.195.28		
10	156	156	5.09	46.150.1		
11	153	153	5.00	46.169.3		
12	142	142	4.64	16.138.1		
17	108	108	3.53	46.383.1		
18	108	108	3.53	591.49.57		
20	103	103	3.36	46.157.1		
24	88	88	2.87	591.49.51		

Extract the ring identifiers, index names, and elemental sequences.

The most common RID (46.150.18) is benzene.

46 = 6 membered rings

16 = 5 membered rings

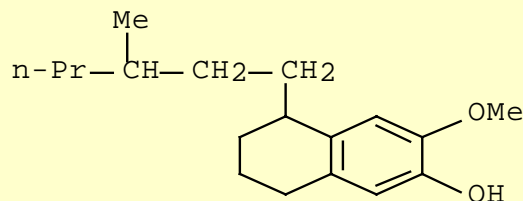
591 = ?

Identify substances using RIDs:

=> S 591/RID AND L4; D SCAN
L6 446 591/RID AND L4

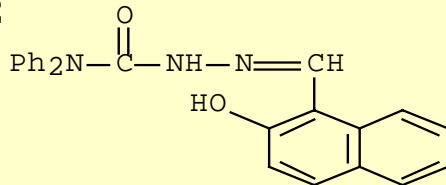
The first part of the RID describes the ring skeleton.

L6 446 ANSWERS REGISTRY COPYRIGHT 2003 ACS on STN
IN 2-Naphthalenol, 5,6,7,8-tetrahydro-3-methoxy-5-(3-methylhexyl)- (9CI)
MF C18 H28 O2



HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):1

L6 446 ANSWERS REGISTRY COPYRIGHT 2003 ACS on STN
IN Hydrazinecarboxamide, 2-[(2-hydroxy-1-naphthalenyl)methylene]-N,N-diphenyl- (9CI)
MF C24 H19 N3 O2



HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):0

Identify rings with elemental sequence:

=> D L5 ES TOP 10

L5 ANALYZE L4 1- RID IN ES : 3678 TERMS

TERM #	# OCC	# DOC	% DOC	RID IN ES
--------	-------	-------	-------	-----------

1	1885	1755	57.30	C6
3	271	268	8.75	NC5
4	237	237	7.74	NCNC3
5	214	214	6.99	C6-C6
7	167	163	5.32	OC4
9	157	157	5.13	N2C4
13	132	126	4.11	NC4
14	129	129	4.21	NCNC2
15	116	116	3.79	NC2NC2
16	113	113	3.69	NC4-C6

=> S NC4-C6/ES AND L4

678529 NC4-C6/ES

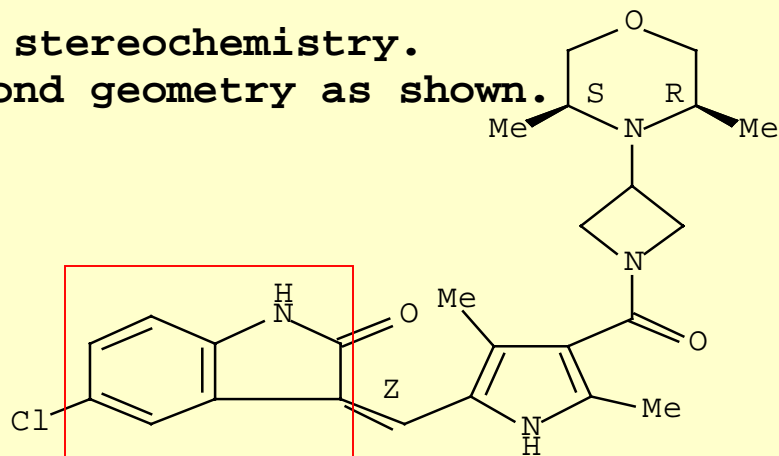
L7 113 NC4-C6/ES AND L4

Identify rings (cont'd):

=> D SCAN

L7 113 ANSWERS REGISTRY COPYRIGHT 2003 ACS on STN
IN Azetidine, 1-[[5-[(Z)-(5-chloro-1,2-dihydro-2-oxo-3H-indol-3-ylidene)methyl]-2,4-dimethyl-1H-pyrrol-3-yl]carbonyl]-3-[(3R,5S)-3,5-dimethyl-4-morpholinyl]-, rel- (9CI)
MF C25 H29 Cl N4 O3

Relative stereochemistry.
Double bond geometry as shown.



HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1): 0

Identify substances using index names:

=> D L5 WITH "PYRIMIDINE"

L5 ANALYZE L4 1- RID IN ES : 3678 TERMS

TERM #	# OCC	# DOC	% DOC	RID IN ES
1044	1	1	0.03	COPPER, (6-FLUORO-1-(TETRAHYDRO-2-FURANYL)-2,4(1H,3
1045	1	1	0.03	COPPER, (6-FLUORO-1-(TETRAHYDRO-2-FURANYL)-2,4(1H,3
1871	1	1	0.03	PYRIDO(3,2-D)PYRIMIDINE-2,4-DIAMINE, 6-(1-NAPHTHALE
1875	1	1	0.03	PYRIDO(3,4-D)PYRIMIDINE-2,4-DIAMINE, N2-1H-INDAZOL-
1876	1	1	0.03	PYRIDO(3,4-D)PYRIMIDINE-2,4-DIAMINE, N2-1H-INDAZOL-
1877	1	1	0.03	PYRIDO(3,4-D)PYRIMIDINE-4,6-DIAMINE, N4-(3-BROMOPHE
1880	1	1	0.03	PYRIMIDINE
1881	1	1	0.03	PYRIMIDINE, 2-((2-CHLOROETHYL)DITHIO)-4-METHYL-
1882	1	1	0.03	PYRIMIDINE, 2-((3-CHLOROPROPYL)DITHIO)-4-METHYL-
1883	1	1	0.03	PYRIMIDINE, 2-(BUTYLDITHIO)-4-METHYL-

140 MORE TERMS WITH AN OCCURRENCE COUNT OF 1

Identify substances using index names:

=> S L4 AND PYRIMIDIN? NOT M/ELS NOT NR>1

L8 32 L4 AND PYRIMIDIN? NOT M/ELS NOT NR>1

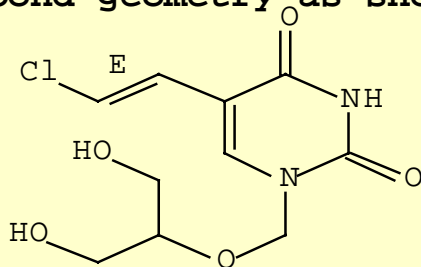
=> D SCAN

L8 32 ANSWERS REGISTRY COPYRIGHT 2003 ACS on STN

IN 2,4(1H,3H)-Pyrimidinedione, 5-[(1E)-2-chloroethenyl]-1-[[2-hydroxy-1-(hydroxymethyl)ethoxy]methyl]- (9CI)

MF C10 H13 Cl N2 O5

Double bond geometry as shown.



HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):1

Search for pyrimidine derivatives, remove metal-containing compounds and limit to one ring.

Find recent patents using pyrimidine compositions for treatment of leukemia:

=> **FILE CAPLUS**

=> **S L8/THU**

L9 **5717 L8/THU**

=> **S L1 AND L9**

L10 **24 L1 AND L9**

Search for therapeutic use and combine with original search.

Find recent patents (cont'd):

=> D L10 BIB ABS HITSTR 9

L10 ANSWER 9 OF 24 CAPLUS COPYRIGHT 2003 ACS on STN

AN 2003:609872 CAPLUS [Full-text](#)

DN 139:154909

TI Compositions for delivery of antitumor drug combinations

• • •

IT 51-21-8, 5-FU

RL: PAC (Pharmacological activity); THU (Therapeutic use); BIOL
(Biological study); USES (Uses)
(compns. for delivery of drug combinations)

RN 51-21-8 CAPLUS

CN 2,4(1H,3H)-Pyrimidinedione, 5-fluoro- (9CI) (CA INDEX NAME)



Summary

- Results analysis is an important final step in the search process. It delivers search results in a meaningful way.
- Several STN features facilitate results analysis:

Online	<ul style="list-style-type: none">◆ ANALYZE command◆ TABULATE command
STN Express with <i>Discover!</i> wizards	<ul style="list-style-type: none">◆ Analyze wizard◆ R-group analysis wizard

STN[®]

Analyzing CAS Data

Visit CAS at Stand 90

CAS is a division of the American Chemical Society