

Tips from the Helpdesk

STN User Days June 2008

Tips from the Helpdesk

How do I.....?

- Deal with L-number limits
- Search for chemical substances in non-CA files
- Search for unassigned patents

Queries may reach system limits

- **System limits for STN queries**
 - Answers per file in one session: 8,000,000
 - L-numbers per session: 999
 - L-numbers in a single query: 390
 - Number of characters in a single query: 20,000
 - Number of characters in a search statement: 256
 - E-numbers per session: 999

Substance search in WPINDEX

Search Question:

What patents deal with acetylsalicylic acid? We want to exclude publications which are already present in our internal database?

Search strategy?

How to do the search... in theory

- Step 1. Perform the substance search
- Step 2. Search the known publication numbers
- Step 3. Subtract the known numbers from the
main search

BUT: System limits will prohibit this strategy

How to upload 8,000 publication numbers

- Search for each number in a single line?
 - **NO**: 8,000 L-numbers would be needed
- Combine all PN by OR operator ?
 - **NO**: maximum number of characters in a search statement: 256
- Gather PNs in several search statements, combined by OR?
 - **NO**: Maximum number of L-numbers in a query: 390
- Upload several blocks of PNs
 - Search and SAVE the answer sets (xxx/A)
 - Search and SAVE the query (xxx/Q) - **recommended**

Search strategy

How to do the search... in practice

- Step 1. Search the known publication numbers in blocks and SAVE as QUERY.
(xxx1/Q; xxx2/Q; xxx3/Q;...)
- Step 2. Delete history
- Step 3. Perform the main search
- Step 4. SEARCH (not ACTIVATE!) the queries
- Step 5. Subtract the results of step 4 from step 3

The publication numbers are uploaded

=> File WPINDEX

```
=> que ( EP316434 or EP101294 or ... or AU9886052 )/PN
L1  QUE ( EP316434 OR EP101294 OR ... OR AU9886052 )/PN
=> que ( AU2002257147 or EP103836 or ... or AU9886672 )/PN
L2  QUE ( AU2002257147 OR EP103836 OR ... OR AU9886672 )/PN
...
=> que ( AU2002352073 or EP103945 or ... or AU9886800 )/PN
L199  QUE ( AU2002352073 OR EP103945 OR ... OR AU9886800 )/PN
=> que ( DE69635754 or EP112130 or ... or AU9886801 )/PN
L200  QUE ( DE69635754 OR EP112130 OR ... OR AU9886801 )/PN

=> que L1-L200
L201  QUE (L1 OR L2 OR ... or L199 OR L200)

=> Save L201 pn200a/q
QUERY L201 HAS BEEN SAVED AS 'PN200a/Q` Save PNs as QUERY.
..
=> Save Lxxx pn200x/q
=> Save ...
```

Only a few L-numbers are created

```
=> Del his y
=> s aspirin/cn
L1      1 ASPIRIN/CN
=> s l1/dcr
L2      2841 L1/DCR
=> s aspirin or acetylsalicylic acid
L3      4417 ASPIRIN OR ACETYLSALICYLIC ACID
=> s l2 or l3
L4      5082 L2 OR L3
=> s PN200a/Q
L5      617 PN200AQUERY/Q
...
=> s PN200e/Q
L9      705 PN200eQUERY/Q
=> s l4 not l5-l9
L9      3159 L4 NOT (L5 OR L6 OR L7 OR L8 OR L9)
```

Delete history to keep L-numbers low.

Substance search via DCR and keyword search

Search of /Q does NOT create excessive number of L-numbers

Helpful for large strategies

- **Searching with long L-number lists can be simplified**
 - Use saved queries (xxx/Q)
 - Single file queries only
 - Use DELETE HISTORY
 - Use SEARCH xxx/Q

Tips from the Helpdesk

- Dealing with L-number limits
- Searching for chemical substances in non-CAS bibliographic files
- Searching unassigned patents

Non-CAS bibliographic files

- Some non-CAS files include CAS RNs for indexed chemical substances
- May not be systematic or comprehensive
- WPI files do not include CAS RNs
- Need to search by all available synonyms to maximize retrieval
- Utilize CAS Registry nomenclature

Classic approach: SELECT CHEM

Two-step approach from Registry record:

- => *File Registry*
- => *E CLARITIN/CN*
- => *S CLARITIN/CN*
- => *SELECT CHEM Lx 1 (Extract names)*
- => *D SELECT (List extracted names & E-no)*
- Enter bibliographic/patent files of interest
- Search on relevant E-numbers
(**Search names**)
e.g. => *S E1-E5 OR E6-E10*

Alternative approach: TRANSFER CHEM

One-step approach from Registry record:

- => *File Registry*
- => *E CLARITIN/CN*
- => *S CLARITIN/CN*
- Enter bibliographic/patent files of interest
- => *TRANSFER CHEM Lx 1*
(**Extract & search names**)

What's the difference?

SELECT CHEM:

- 2-step from substance record
- Names extracted in Registry file
- Preview extracted names
 - can pick by E-number
- Scrolling E-numbers
- Hit set size: n

TRANSFER CHEM:

- 1-step from substance record
- Names extracted in destination file
- No preview/choice :
 - 'Just do it!'
- No E-numbers
- Hit set size: the same
(more, if names include common words)

Structure searching in WPI

- WPI structure-searchable 1999-date
- Draw structure and upload query, **or**,
- In Registry file, model on known RN:

=> *STRUCTURE 79794-75-5*
ENTER (DIS), GRA, NOD, BON OR ? : end
L5 STRUCTURE CREATED
- In WPI, => **Search L5 (Exact or Family)**
- Hit set Lx (**Substance records**)
- => **S Lx/DCR (Literature references)**

Finding unique records from WPI and CA

- => *DUPLICATE REMOVE* works on **basic** patents (first family member into database)
- => *TRANSFER PN, APPS L2* extracts **all** patent numbers terms from L-number answer set L2
- Searches extracted data in target file
- Can be used to *TRANSFER* patent numbers from *CAPLUS* (L2) to WPI (**Hit set L3**)
- Suppose direct WPI search produced **L4**
- Unique answers in WPI from =>S L4 **NOT L3**
- Finish with *FSORT* to identify extended families

Tips from the Helpdesk

- Dealing with L-number limits
- Searching for chemical substances in non-CAS bibliographic files
- Searching unassigned patents

Create an answer set with unassigned patents

- An answer set may contain several documents that are not assigned to a company or organization
- US Provisional applications are not required to have a company name listed
- Other patents may be assigned only to an individual
- These documents will not show up in an analysis by company
- How can we create an answer set with only these records in CPlus

The search problem

FILE 'CAPLUS' ENTERED AT 01:26:30 ON 17 FEB 2008

L1 1704 S NANOSPHERES+OLD/CT

L2 307 S L1 AND P/DT

L3 11680 S NANOWIRES+OLD/CT

L4 1340 S L3 AND P/DT

L5 1631 S L2 OR L4

=> ANALYZE L5 PA 1-

L6 ANALYZE L5 1- PA : 663 TERMS

=> D

L6 ANALYZE L5 1- PA : 663 TERMS

TERM #	# OCC	# DOC	% DOC	PA	
1	154	154	9.44	USA	
2	44	44	2.70	S KOREA	
3	44	44	2.70	THE REGENTS OF THE UNIVERSITY OF CALIFORNIA USA	
4	37	37	2.27	TSINGHUA UNIVERSITY PEOP REP CHINA	
5	33	33	2.02	SAMSUNG ELECTRONICS CO LTD S KOREA	
6	32	32	1.96	NATIONAL INSTITUTE OF MATERIALS SCIENCE JAPAN	
7	28	28	1.72	HEWLETT PACKARD DEVELOPMENT COMPANY L P USA	
8	28	28	1.72	NANOSYS INC USA	
9	28	28	1.72	SAMSUNG SDI CO LTD S KOREA	
10	25	25	1.53	KONINKLIJKE PHILIPS ELECTRONICS N V NETH	

Ingredients for a solution

- Searching USA/PA will also find all the assigned patents from US companies
- The CO field doesn't include the address information from the PA field
- CPlus doesn't copy the inventor names in the PA field when assigned to an individual
- For unassigned patents the CO field is empty
- The SORT command indicates how many records do not have an entry in the sorted field
- So....

The solution

```

=> SORT L5 CO,D 1-
      261 ANSWERS DID NOT HAVE 'CO' SORT FIELD
PROCESSING COMPLETED FOR L5
L7      1631 SORT L5 1- CO D

=> SORT L7 1-261 PY
PROCESSING COMPLETED FOR L8
L8      261 SORT L7 1-261 PY

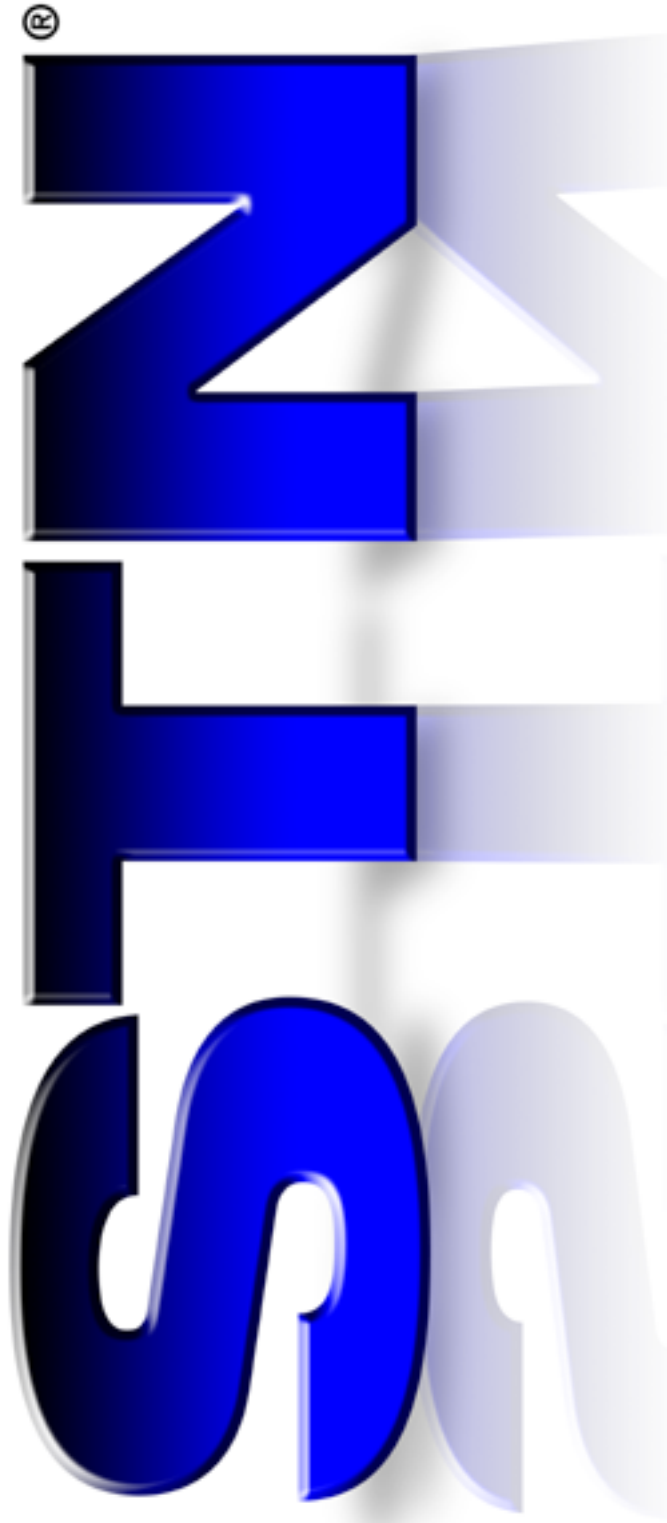
=> ANALYZE PA L8 1-
L10     ANALYZE L9 1- PA :      15 TERMS
  
```

TERM #	# OCC	# DOC	% DOC PA	
1	154	154	59.00	USA
2	44	44	16.86	S KOREA
3	11	11	4.21	PEOP REP CHINA
4	11	11	4.21	TAIWAN
5	9	9	3.45	GERMANY
6	8	8	3.07	JAPAN
7	4	4	1.53	UK
8	2	2	0.77	CAN
9	2	2	0.77	RUSSIA
10	2	2	0.77	SINGAPORE
11	2	2	0.77	SWED
12	1	1	0.38	AUSTRALIA

Tips from the Helpdesk

- Dealing with L-number limits
- Searching for chemical substances in non-CAS bibliographic files
- Searching unassigned patents

Questions?



Tips from the Helpdesk

STN User Days June 2008