

ANALYZE and TABULATE for statistical analysis

With ANALYZE, you can easily extract up to 50,000 terms from an answer set for statistical analysis.

Use the DISPLAY command to view your analyzed terms as a one-dimensional list.

Use the TABULATE command to correlate analyzed terms from two fields.

TABULATE is used mainly in patent and engineering databases. All extraction fields, except patent number, application number, priority number, CAS Registry Number[®], accession number, Basic Index, and similar fields, may be used.

Use ANALYZE to find top authors publishing on artificial collagen.

1 Enter and search in the appropriate databases.

2 (Optional) Eliminate duplicates.

3 Enter ANALYZE and the L-number of the answer set (L2), answers to analyze (1-), and field code to analyze (AU).

4 Display the 10 most frequently occurring terms.

```

=> FILE CAPLUS BIOSIS MEDLINE

=> S (ARTIFICIAL OR SYNTHETIC) (W)COLLAGEN
L1          462 (ARTIFICIAL OR SYNTHETIC) (W) COLLAGEN

=> DUP REM L1
PROCESSING COMPLETED FOR L1
L2          348 DUP REM L1 (114 DUPLICATES REMOVED)
           ANSWERS '1-309' FROM FILE CAPLUS
           ANSWERS '310-337' FROM FILE BIOSIS
           ANSWERS '338-348' FROM FILE MEDLINE

=> ANALYZE L2 1- AU
L3          ANALYZE L2 1- AU :      1044 TERMS

=> D
L3          ANALYZE L2 1- AU :      1044 TERMS

TERM #      # OCC  # DOC  % DOC  AU
-----
   1         35    22    6.32  REPRINT AUTHOR
   2         20     4    1.15  INVENTOR
   3         13    13    3.74  RAINES, RONALD T.
   4          7     7    2.01  IKADA, YOSHITO
   5          6     6    1.72  GOODMAN, MURRAY
   6          6     6    1.72  MIYATA, TERUO
   7          6     6    1.72  YANNAS, I. V.
   8          5     5    1.44  FENG, YANGBO
   9          5     5    1.44  KOIDE, MIKIO
  10         5     5    1.44  MORODER, LUIS
    
```

Display options after ANALYZE

Option	Explanation	Example
Default	Display the 10 most frequently occurring terms in the current sort order	=> D => DISPLAY TOP 10
Sorting	Change sort to alphabetical order	=> DISPLAY ALPHA
	Change sort to occurrence order (default)	=> DISPLAY OCC
Terms to display	Display all of the terms, in the current sort order	=> DISPLAY ENTIRE => DISPLAY 1-
	Display a range or number of terms, in the current sort order	=> DISPLAY 1-20
	Display the top "n" terms, in the current sort order	=> DISPLAY TOP 20
	Display terms with occurrence counts greater than "n", in the current sort order	=> DISPLAY OGT 5
	Display terms with document counts greater than "n", in the current sort order	=> DISPLAY DGT 4
	Display terms with percentage counts greater than "n", in the current sort order	=> DISPLAY PGT 10 => DISPLAY %GT 10
	Display only terms containing a specified character string (in quotes)	=> DISPLAY WITH "US"
	Display only terms that do not contain a specified character string	=> DISPLAY NOT "US"
Additional requirements	Display the answer numbers with the terms	=> DISPLAY ANS
	Display the terms in the delimited format for post-processing	=> DISPLAY DELIMITED
	Display the terms with the field code appended	=> DISPLAY DETAIL

Use TABULATE to find information on patenting trends for a company. Correlate main IPCs for its patents with publication years.

1 Enter the database and search the name of a company of interest.

2 With the answer set (L1), ANALYZE two fields (ICM and PY).

3 Enter TABULATE and the ANALYZE L-number. Choose the GRID format. Enter the primary and secondary display codes and sort requirements.

Enter Y to proceed.

The top 10 main IPCs are listed on the vertical axis. The publication years are listed on the horizontal axis. The document count is shown at each intersection of ICM and PY.

4 To reformat, choose the non-grid format (N).

Enter the display requirements.

```

=> FILE WPINDEX
      .
      .
=> ANALYZE L1 1- ICM PY
L2          ANALYZE L1 1- ICM PY :          76 TERMS
=> TABULATE L2
DISPLAY AS GRID FORMAT (N), Y, OR ? :Y
ENTER PRIMARY DISPLAY CODE OR (?) :ICM
ENTER SECONDARY DISPLAY CODE OR (?) :PY
DISPLAY PRIMARY (TOP 10), ENTIRE OR ? :.
PRIMARY SORT ORDER (CURRENT), DOC, ALPHA, OR ? :.
PRIMARY SORT DIRECTION (DEFAULT), A, D, OR ? :.
SECONDARY SORT ORDER (CURRENT), DOC, ALPHA, OR ? :.
SECONDARY SORT DIRECTION (DEFAULT), A, D, OR ? :.
A FEE WILL BE CHARGED. PROCEED? (Y), N, OR ? :Y

L2          ANALYZE L1 1- ICM PY :          76 TERMS
                                                PY
ICM          2003 2002 1999 1995 1994 1992 1990 1991
C12N015-32      13  15  10  19  12  11   7   4
A01N063-00      10  14  11  14  13   8   6   2
C12N015-09       7  12   5  14  10   4   5   5
C12N001-20       5   7   7   8   8   4   5   3
A01N063-02       5   8   5   4   6   4   2   3
C12N015-82       6   6   3   3   4   1   4   1
A01N037-02       7   3   4   3   3   2   0   0
C07K014-325     5   3   2   4   3   1   1   1
C12N001-21       3   3   2   4   2   4   4   3
C12N015-00       2   4   3   2   3   0   3   1
      .
      .
REFORMAT USING SAME DISPLAY FIELDS? (N), Y, OR ? :Y
DISPLAY AS GRID FORMAT (N), Y, OR ? :N
EXCHANGE PRIMARY AND SECONDARY DISPLAY FIELDS (N), Y,
OR ? :N
DISPLAY PRIMARY (TOP 10), ENTIRE OR ? :.
DISPLAY SECONDARY (TOP 10), ENTIRE OR ? :.
PRIMARY SORT ORDER (CURRENT), DOC, ALPHA, OR ? :.
PRIMARY SORT DIRECTION (DEFAULT), A, D, OR ? :.
SECONDARY SORT ORDER (CURRENT), DOC, ALPHA, OR ? :.
SECONDARY SORT DIRECTION (DEFAULT), A, D, OR ? :.

```

The top 10 ICMs are listed in the ICM column. For each ICM, the publication years are listed in the PY column. For each term, the number and percent of documents are displayed.

5 To reformat the same fields in the delimited format, enter **TABULATE DELIMITED**. Enter the display requirements.

Terms are displayed in the delimited format for downloading.

Enter **N** to exit the **TABULATE** command without reformatting.

```

L2              ANALYZE L1 1- ICM PY :          76 TERMS
TERM # # DOC    % DOC    ICM              PY
-----
      7      30    28.04    C12N015-32
          19    17.76    -
          15    14.02    -
          13    12.15    -
          12    11.21    -
          11    10.28    -
          10     9.35    -
           9     8.41    -
           7     6.54    -
           4     3.74    -
           4     3.74    -
      11     23    21.50    A01N063-00
          14    13.08    -
          14    13.08    -
          13    12.15    -
          11    10.28    -
          10     9.35    -
          .
          .
          .
REFORMAT USING SAME DISPLAY FIELDS? (N), Y, OR
?:TABULATE DELIMITED
DISPLAY AS GRID FORMAT (N), Y, OR ?:N
EXCHANGE PRIMARY AND SECONDARY DISPLAY FIELDS (N), Y,
OR ?:.
          .
          .
          .
L2              ANALYZE L1 1- ICM PY :          76 TERMS
19;17.76;C12N015-32;1995
15;14.02;C12N015-32;1996
          .
          .
          .
REFORMAT USING SAME DISPLAY FIELDS? (N), Y, OR ?:N

```

For more information

Refer to STNote No. 17, *ANALYZE and TABULATE Commands*, available at www.cas.org.

Enter **HELP ANALYZE** or **HELP TABULATE** at an arrow prompt.