



STN[®] AnaVist[™]

Frequently Asked Questions

2014



Table of Contents

| | |
|---|---|
| Software | 3 |
| Answer Set Creation | 4 |
| Visualization | 5 |
| Document Display | 6 |
| The Research Landscape and Charts | 7 |
| Results Sharing | 9 |



Software

Q. *Are there any preference settings that should not be customized?*

A. Yes. To ensure good performance, keep the setting for new Matrix Charts at its default setting of 50 rows and columns. To see more data a matrix chart, right-click the chart and select Properties. Under the View tab, modify the number of rows and columns to include in the matrix chart.

Q. *What are the advantages of STN AnaVist 2.1 over 2.0/2.01?*

A. STN AnaVist 2.1 provides compatibility with newer Windows operating systems, including Windows 7, 8 and 8.1. It increases the memory allocated to STN AnaVist projects which improves performance. It also provides the current copyright year for charts, matrices and the research landscape created under STN AnaVist 2.1.

Answer Set Creation

Q. *What software can I use to create answer sets for STN AnaVist?*

A. All document sets must be created with the Save for STN AnaVist wizard available in STN Express® or with the STN AnaVist Assistant in STN® on the Web™.

Q. *Which databases and clusters are available for document set creation in STN AnaVist?*

A. The following databases/clusters are available for answer set creation:

- CAplusSM family
- EPFULL
- PCTFULL
- USPATFULL/USPAT2
- WPINDEX/WPIDS/WPIX
- 2ANAVIST/2HANAVIST clusters

Q. *Can I create a document set for STN AnaVist in CASREACT®, MARPAT®, the CASM family of databases, and/or the CASLINK cluster?*

A. Yes. Conduct a search in one or more of these databases or the CASLINK cluster and then search the L-number in CAplus. Use the Save for STN AnaVist wizard with the resulting L-number created in CAplus. For example:

FILE CASREACT

```
=> S vinyl chloride
    22325 VINYL
    115824 CHLORIDE
L1  482 VINYL CHLORIDE
    (VINYL(W)CHLORIDE)
```

FILE CAPLUS

```
=> S L1
    L2  482 L1
```

Q. *Why do I get an error message in STN Express when I use the Save for STN AnaVist wizard with more than 20,000 records?*

A. The record limit for projects in STN AnaVist is 20,000.

Q. *Is highlighting retained when I import answers into STN AnaVist that have been saved in STN Express?*

A. No. Highlighting is not retained. Hit term highlighting is not available in STN AnaVist.

Visualization

Q. *What data fields can I use for custom visualization?*

A. The following data fields may be used for clustering: Titles/Abstracts, Claims (Exemplary/First Claim and All Claims), Technology Indicators, and International Patent Classification (IPC) Codes. Visualizations may also be created using combinations of text fields. Research Landscapes created with IPC codes include IPC codes only.

Q. *Can I adjust the Concept Frequency when creating Research Landscape based on IPC codes?*

A. No. Concept Frequency is always set to 100% and cannot be altered.

Q. *Will I be charged another visualization fee if I visualize again based on a different set of data fields?*

A. Yes.

Q. *When selecting data fields for clustering, what is a Backup field?*

A. If a document does not include data in the selected clustering field; the Backup field is used for clustering instead. For example, including a backup field is important when clustering on Claims alone in a document set that includes patent and non-patent documents.

Q. *Why do clustering options sometimes change?*

A. Available clustering options are based on the source databases of the imported documents. For example, the option to cluster using Technology Indicators is available only when documents from CAplus, USPATFULL, and USPAT2 are source databases. Technology Indicators are CAS indexing terms that are only available in documents from CAplus, and in chemistry-relevant USPATFULL and USPAT2 documents that have been enhanced with this indexing.

Q. *Does a custom visualization with Technology Indicators include indexing terms from WPINDEX?*

A. No. Technology Indicators are CAS indexing terms that are only available in documents from CAplus, and in chemistry-relevant USPATFULL and USPAT2 documents that have been enhanced with indexing.

Q. *Will I be charged more for visualization if my document set includes records from multiple databases?*

A. No. The price to create a visualization project is determined by the total number of records included in the project, regardless of which database or databases are included.

Q. *How long does it take to visualize a results set?*

A. The time it takes for a visualization to complete depends on several factors, including how many documents there are to be visualized and what else the system is doing when the visualization initiates. The larger the document set to be visualized, the longer the visualization will take.

Q. *When working with the visualization results, are there any activities that may take longer than expected to complete?*

A. In general, visualization activities happen fairly quickly. Some requests, however, may take longer to complete:

- Displaying an answer with a very large number of CAplus index terms.
- Displaying a very long patent document.
- Opening the list of cluster concepts for term grouping if the result set contains a large number of documents, i.e., several thousand answers.
- Displaying all entries for a chart or matrix when the total number of entries is in the thousands.

Document Display

Q. *What information is included in documents displayed in STN AnaVist?*

A. Documents are displayed in a condensed format that provides bibliographic data and the abstract, if available. The format also includes the data available to create bar and matrix charts. A minimum of three additional display options per database are available in STN Express via the Display from STN AnaVist wizard.

Q. *Why are documents displayed from STN AnaVist in STN Express not reflected in the STN AnaVist session summary?*

A. Documents displayed using the Display from STN AnaVist wizard are reflected in the STN Express expenditures and do not appear in the STN AnaVist session summary.

Q. *If I display a document in STN AnaVist, will I be charged to display it again if I use the Display from STN AnaVist wizard?*

A. Yes. STN Express does not account for previous displays in STN AnaVist.

Q. *As a subscriber to WPIDS/WPIX, can I display WPIDS/WPIX documents in STN AnaVist?*

A. No. The source database for displayed DWPISM documents is WPINDEX. To display documents in WPIDS or WPIX formats, use the Display from STN AnaVist or Create L-Number from STN AnaVist wizards in STN Express.

Q. *Can I display both invention- and member-level documents from WPINDEX in STN AnaVist?*

A. No. To display member-level documents, use the Display from STN AnaVist wizard in STN Express.

Q. *Why is USPAT2 indicated as a source database for U.S. patent documents in STN AnaVist, Version 2.1, when in previous versions, only USPATFULL was indicated as the source?*

A. Previous versions of STN AnaVist do not indicate USPAT2 as a source database because USPATFULL and USPAT2 were combined into a single database.

| | |
|-------------------------------|------------------------------|
| | Version 2.1 |
| Combined as a single database | No |
| Source identification | USPATFULL or USPAT2 |
| Contents | First and latest stages only |

When using the Display from STN AnaVist wizard in STN Express, select USPATFULL and USPAT2 for more comprehensive results.

Q. *What file formats are available when I save documents?*

A. Documents can be saved in .rtf or .pdf formats.

Q. *Can I save full-text patent documents from STN AnaVist to evaluate in STN Viewer?*

A. Yes. When saving a document set, select the .xta file format. A fully-functional L-number can be created with the create L-number from STN AnaVist wizard in STN Express. The L-number can be used to:

- Display answers from STN AnaVist in any format.
- Move full-text patent documents from STN AnaVist to STN[®] Viewer[™] via the Evaluate with STN Viewer wizard.

The Research Landscape and Charts

Q. *Why is there a distant peak in the corner of my Research Landscape?*

A. STN AnaVist creates Research Landscapes based on data available in the user-defined or default clustering fields. Documents that do not have data in any of the clustering fields (including the Backup field) are identified as different from the other documents and are placed in a peak away from the others. This peak may also include documents with clustering field data that is truly unique compared to the other documents.

Q. *Can I identify cells within a matrix chart that are applicable to more than one highlighting group?*

A. Yes. With the multiple-color highlighting feature in Version 2.1, certain cells within a matrix chart may be associated with more than one highlighting group. Such cells appear with black triangles in each corner of the cell. Hovering over the cell provides additional information.

Q. *Can I identify records within a bar that are applicable to more than one highlighting group?*

A. Yes. With the multiple-color highlighting feature in Version 2.1, any documents belonging to more than one highlighting group are grouped together in a single gray segment of the bar. The documents represented by this segment are members of multiple highlighting groups, but not necessarily the same highlighting groups.

Q. *Why do I get a message that there was not enough system memory to create a matrix chart?*

A. Certain matrix charts may require a significant amount of system memory. Try one or more of these solutions to avoid the system memory error:


- Contact your STN Service Center for assistance with editing the anavist.ini file and its JRE memory settings. If you have a system with sufficient memory (RAM) you may be able to expand the memory available to STN AnaVist
- Reduce the number of highlighting colors used at one time.
- Reduce the number of rows and columns in the matrix by modifying Chart Properties. To do this, right-click the chart and select **Properties**. Under the View tab, reduce the number of rows and columns to include in the matrix chart. Default matrix chart settings may be modified in the Charts menu available in **Tools > Preferences**. (A default setting of 50 cells on each axis is optimal for performance.)
- Create two bar charts for a similar result. A matrix chart displays the number of documents where terms co-occur. Co-occurring terms can also be displayed using bar charts with the highlighting feature. When a term is highlighted in one bar chart, co-occurring terms are highlighted in the other bar chart.
- By selecting **Sort Descending by Highlighted Count**, the co-occurring terms will automatically sort to the top each time you highlight another term. You may set a higher number of bars on a bar chart than rows or columns on a matrix chart without system memory issues.

Q. *Did CAS develop the clustering algorithm that creates the Research Landscape in STN AnaVist?*

A. CAS licensed visualization software developed by Sandia National Laboratories as the base software for generating the Research Landscape. The software was significantly enhanced to improve visualization results, utilizing the expertise of our database building staff and scientists:

- CAS vocabulary to standardize the clustering concepts
- A stopword list to improve cluster results for sci-tech searches

These enhancements allow for the software to produce more scientifically relevant clusters that are focused on scientific and intellectual property information.



STN AnaVist uses force-directed placement for the text clustering. Force-directed placement is a fast clustering method, most notably with larger answer sets. To learn more about the clustering method used in STN AnaVist, refer to:

- Visualizing an Auto-Generated Topic Map by Nadine Amende and Stefan Groschupf (2003, technical report, Halle/Saale)
- Domain Visualization Using VxInsight for Science and Technology Management by Kevin W. Boyack, Brian N. Wylie, and George S. Davidson (2001, preprint from *Journal of the American Society for Information Science and Technology*)

Q. *Why do certain clustering concepts in the Research Landscape seem to be mistakes or misnomers?*

A. Use the Term Editor to change the name of any clustering concept. You can also report suspected anomalies by using the Contact Us button within STN AnaVist. The list of clustering concept terms is continually updated.

Q. *Why don't researcher and company names in the bibliographic data always correspond to the charts?*

A. Researcher names are grouped together under a standard representation whenever possible prior to the creation of charts with Key Researcher data.

Prior to creating charts with Key Organization/Assignee data, company names are standardized and grouped using an improved version of the CAS Company Name Thesaurus.

The names appearing in the chart are the standardized names to which a number of variations may be mapped.

Q. *Can I use my own customized dictionary to standardize concepts?*

A. Not at this time.

Q. *What are the Technology Indicator terms in CAplus documents?*

A. The terms used to create the Technology Indicators chart are based on controlled indexing in CAplus documents. When a CAplus document provides both the common name and the Latin name for an organism, the common name (e.g., Pea) is used in the Technology Indicators chart in preference to the Latin name (e.g., *Pisum sativum*).

Results Sharing

Q. *How can I print all documents in a visualization project?*

A. To print all documents in a visualization project:

- Click and drag on the Research Landscape to highlight all documents.
- Click the Print icon and select the desired print options.

Only 1,000 documents at a time can be printed.

Q. *Can I print the entire set of documents after importing but before visualization?*

A. No. An entire results set can only be printed after it has been visualized. Prior to visualization, individual documents can be printed from the detailed view of that result.

Q. *Can I create multiple copies of a visualization project?*

A. Select **File > Save Copy of <Project>**. The project will be saved as a .shx file. When the saved project is opened, highlighting, additional charts, and edited terms, are available just as in the original project.

Q. *Can I share projects across service centers with STN AnaVist?*

A. It is not possible to share .shx files across service centers at this time. However, .xta files may be shared between users with full-access STN login IDs.

Q. *Are visualization projects saved to large files?*

A. No. The .shx file created when a copy is saved is only a link to the visualization project maintained on the CAS server.

Q. *Can I edit the Detailed Report and the Summary Report?*

A. Both the Detailed Report and the Summary Report are available as either .rtf or .pdf documents. Both can be modified with an appropriate editor.

Q. *Can I create my own charts with the visualization data?*

A. Yes. Chart data can be saved in comma delimited (.csv) format for use in Microsoft Excel.

Q. *What is the difference between the PNG and JPEG image formats available for saving an image of the landscape?*

A. The PNG format is much more compact and provides a better image than the JPEG format.