Explore “Big Data” with new STN®

Agenda

• Big Data
• Search example
“**Big Data** *is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.***”


“**New STN truly brings the power and functionality of Big Data to the study of chemical, and patent information. Searches that were nearly unthinkable in the past can now be done in record time.**”

“Even Broader Structure Query on Sitagliptin Core”

Agenda

- Big Data
- Search example

New STN workflow is oriented around projects

Projects allow you to:
- Easily return to previous work
- Reuse common queries
- Update searches with the most current information

To create a project, click the + icon.
New STN puts query, history and results at your fingertips

- Query Builder panel
- Results panel
- History panel

Select Databases
Create a Structure
Explore Search Indices or Thesaurus for Terms
Explore Help
Manage Term Lists

Use Filters for integrated, real time analysis

Click blue arrows to see additional options

Setup Update

Create Term List, Save Sets, or Export Results
Search Question:
Find polymer patents involving controlled-release anti-inflammatory compounds of structure (I)
One or more databases can be selected
- The selection can be changed any time
- You can also use the classic FILE command in the query builder

Throughout this search example CAplus, REGISTRY, DWPI and DCR are used.

Search steps:
- All specific and general polymers.
- Controlled release.
- Anti-inflammatory.

Questions:
Structure searching in new STN

- Create structure queries using the Structure Editor
- Use a right mouse-click to change bond/node attributes, and verify using the attributes panel
- Structure queries are added to the Structures tab of the History panel and assigned STR numbers
- Exact (EXA), Family (FAM), Closed Substructure (CSS), and Substructure (SSS) options are available
- For example:
  - STR1/SSS,FULL (default)
  - STR1/SSS,SAM
- Structure displays include substructure highlighting
- Tags can be added to structures in the history panel

Query Builder Panel – Structure Editor

Mouse over the node/bond to verify which attribute(s) it has in the attribute panel (in yellow).
Query Builder Panel – Structure Editor

Click OK to add the structure query to the “Structures” tab of the History panel.

Mouse over the attribute panel to verify which nodes/bonds have a given attribute (in yellow).

Structure search workflow in new STN

- Structure queries are assigned STR numbers for searching
- Hit portion of the structure is highlighted for easier evaluation
- Use REFX to retrieve references
Hit structures in detailed displays

DCR hit structures are automatically included as a part of DWPI detailed displays.

Search steps

+ All specific and general polymers. + Controlled release. + Anti-inflammatory.

8,036 substance records in REGISTRY
967 substance records in DCR

8,888 patent records in CAplus
6,371 patent records in DWPI
Retrieve all specific polymers in REGISTRY

Use POLYMER/CI (class identifier) to retrieve all polymers in REGISTRY.

>1.4 million records are retrieved.

Retrieve all general polymers in CAplus

Browse the CA Lexicon with Broader, Narrower and Other tabs.

Apply to add the selected terms to the Query Builder panel. You can continue to browse additional terms.
Retrieve all general polymers in CAplus (cont.)

>1.9 million records are retrieved.

Combine the REGISTRY and CAplus answers

REFX L3 retrieves all CAplus records for the POLYMER/CI search in REGISTRY.

>4.9 million total CAplus records are retrieved (L5).

>2.1 million are CAplus patent records (L6).
Corresponding steps to retrieve polymers in DWPI

**POLYMERS/CC** retrieves the small set of polymers present in DCR (L7).

**REFX L7** retrieves the corresponding DWPI records.

This query retrieves records with any Class, Manual Code, or Subscriber Indexing for polymers in DWPI (L9).

> 3.6 million total DWPI records are retrieved (L10).

Search steps

- All specific and general polymers.
- Controlled release.
- Anti-inflammatory.

- 8,031 substance records in REGISTRY
- 967 substance records in DCR
- >1.4 million substance records in REGISTRY
- 1,873 substance records in DCR
- 8,825 patent records in CAplus
- 6,299 patent records in DWPI
- >2.1 million patent records in CAplus
- >3.6 million patent records in DWPI
Explore controlled-release and anti-inflammatory terminology using the Term Explorer

As before, the Term Explorer can be used to explore the CA Lexicon (/CT thesaurus)

The narrow and other terms help find the best terminology for a basic index search in both DWPI and CAplus

Additional classification codes should be included to extend the search, especially in DWPI

Retrieve controlled-release concepts

/FCL, /M2,M6 and /MC are only searched in DWPI.
Retrieve anti-inflammatory concepts

Search steps

+ All specific and general polymers. + Controlled release. + Anti-inflammatory.

8,031 substance records in REGISTRY
967 substance records in DCR

>1.4 million substance records in REGISTRY
1,868 substance records in DCR*

>181k patent records in CAplus
>183k patent records in DWPI

>183k patent records in CAplus
>132k patent records in DWPI

8,825 patent records in CAplus
6,299 patent records in DWPI

>2.1 million patent records in CAplus
>3.6 million patent records in DWPI

/ MC is only searched in DWPI.
Bringing it all together….

1,037 total CAplus patent records are retrieved (L13).

1,118 total DWPI patent records are retrieved (L14).

New STN tools for working with search results

- Use Filters to explore the results
- Access the STN Extended Patent Family
- Explore specific polymers in REGISTRY – SUBX
- Create Term Lists to help with duplicates
Use Filters to explore the results

Choose the level of detail for review: Brief, Basic Invention, Full.

Access the the STN Extended Patent Family (next slide...).
Access the STN Extended Patent Family

- Extended Patent Family enables users to
  - Access patent family information for extended families
  - Export extended family tables
  - View detailed records for selected family members
  - Select the preferred source for family information

Extended Patent Family

[Image of Extended Patent Family interface]

View Summary
View Table

Export in .xlsx format.

The Extended Patent Family integrates all patent files into a single table.
Explore specific polymers in REGISTRY – SUBX

Cross File Search with SUBX to find all substance records indexed to references.

1,195 specific polymers are retrieved in REGISTRY (L15) from the 1,037 patent records in CAplus (L13).

Use Filters to explore the results

Click on the name to open the detailed view.
Create Term Lists to help with duplicates

- Use Create Term List to extract data and transfer terms to other files for searching
- Main focus is on patent information
  - PN, PNK, PRN, AP available in all patent files
  - Basic versions (.B) available in patent family files
  - RN, CN and DOI are also available
- Term Lists are identified by Q#
  - Permanent asset, project independent
  - Can be searched in combination with other terms
  - Can be re-qualified with one or more field codes
Search Term Lists via their assigned Q-numbers

Q10 = patent number/kind taken from CAplus (L13).

L14 = DWPI combined search results.

576 patent records in DWPI are identified (L16) in addition to the CAplus results (L13).

Learn more via the STN help menu

What's New – lists enhancements from recent releases.

Quick Tour – an overview video of the new platform that enables users to get started easily.

Help – a detailed reference tool to all features and content, including videos, examples, and tutorials.
Summary

• Big Data
• Search example

We always welcome your feedback!

For more information …

CAS
help@cas.org
Support and Training:
www.cas.org

FIZ Karlsruhe
helpdesk@fiz-karlsruhe.de
Support and Training:
www.stn-international.de