

Turn Information into Knowledge

www.TheVantagePoint.com



vantage point

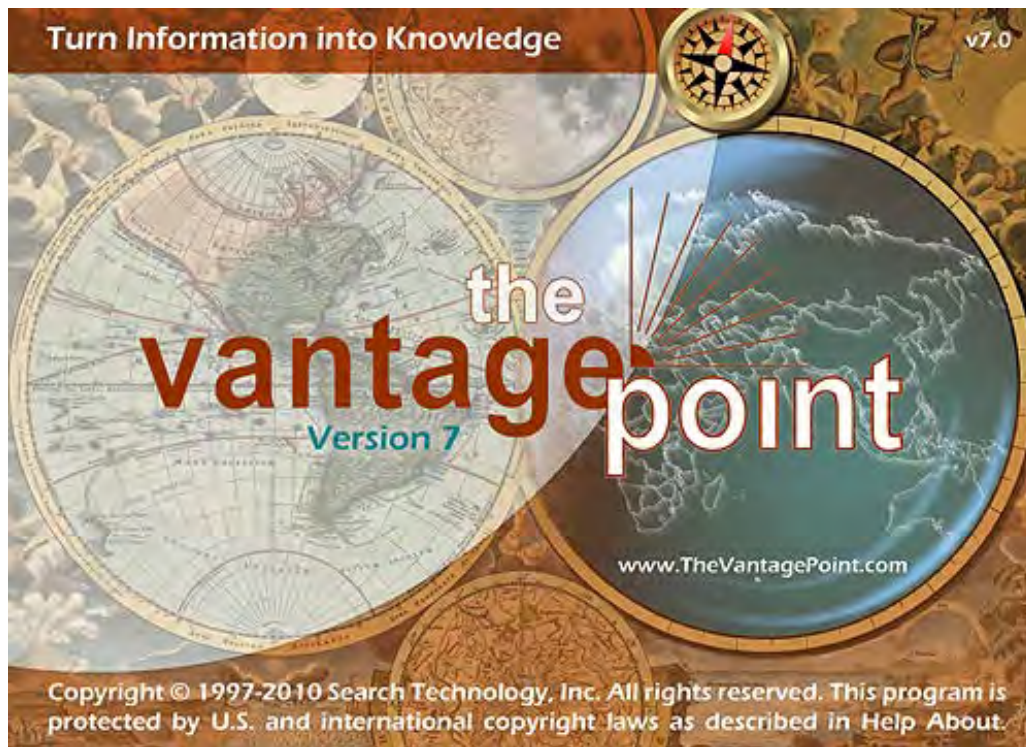


Search Technology
770.441.1457

VantagePoint@searchtech.com

Session 1

Introduction to VantagePoint



What is VantagePoint?

VantagePoint is desktop
Text Mining Software for
Discovering Knowledge in
virtually any Structured
Text Database

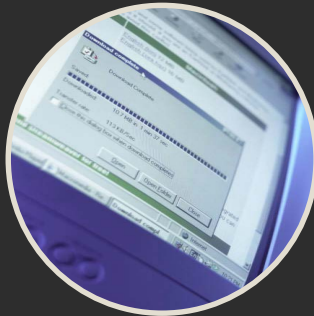


How does VantagePoint work?



Search

Run a search in
your Structured
Text Databases



Download

Retrieve the
search results
from the
database



Import

Use an import
filter to bring the
data into
VantagePoint



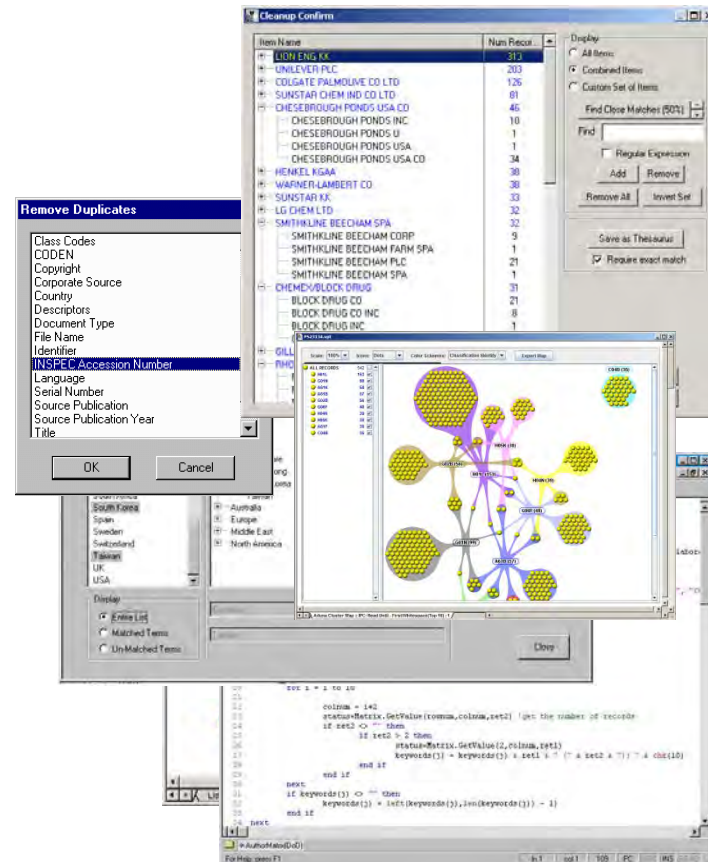
Analyze

Use the tools
within
VantagePoint to
explore the data



What can you do with VantagePoint?

- Analyze Information
- Map Relationships
- Identify Trends
- Develop Indicators
- Automate Your Analysis
- Discover Knowledge



The VantagePoint Tool Set

Import

RegEx
Pattern
Matching
Entity
Extraction
NLP

Clean

Fuzzy
Matching
Thesaurus
Grouping
Manual
Editing

Analyze

The List
Co-
occurrence
Mapping
Drill Down
List
Comparison

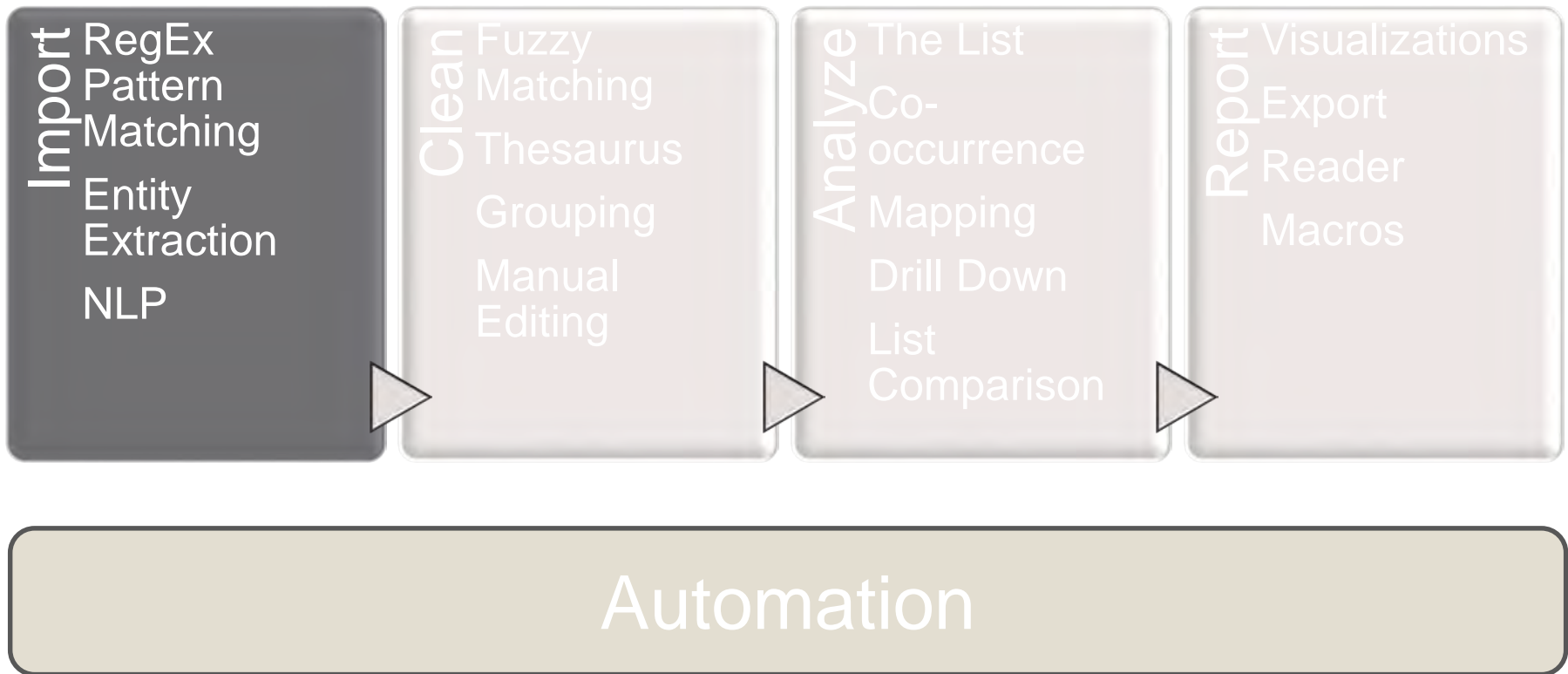
Report

Visualizations
Export
Reader
Macros

Automation

Session 2

Importing Data



The Search

- Where to Search
 - Any Structured Text Databases
 - TI, STN, CSA, EBSCO, Orbit, Pubmed, Excel, etc...
- How to Search
 - Standard Search Interfaces
- Search Strategy
 - Inclusive/Cast a broad net
- Downloading Search Results
 - The more fields the better
 - Dialog - format 9, Tags On, Highlighting Off
 - STN – IALL or ALL



What does the Raw Data Look Like?

The data is:

- Fielded
- Delimited
- Structured

The format can be:

- Text
- XML
- Excel

```
PMID- 16492546
OWN - NLM
STAT- MEDLINE
DA - 20060222
DCOM- 20060328
PUBM- Print
IS - 1054-3589 (Print)
VI - 52
DP - 2005
TI - Inhibition of insulin secretion.
PG - 173-92
AB - Reports of the effects of amylin and amylin agonists on insulin secretion
have varied widely. Some confusion can be attributed to the use of human
amylin, which has been shown to readily fall out of solution resulting in
low estimates of bioactivity. Some confusion can be resolved by assessing
the probability that this had happened. The view taken here, supported by
authors using reliable and well-characterized ligands (representing the
preponderance of recent studies), is that exogenously administered amylin
agonists inhibit insulin secretion, at least partly via activation of an
amylin-like receptor linked to Gi-mediated inhibition of cAMP in islets.
There may additionally be autonomic extrapancreatic effects of amylin on
insulin secretion that derive from its action at the area postrema.
Studies with amylin receptor antagonists, including human studies,
indicate that endogenously secreted amylin may physiologically inhibit
beta-cell secretion (insulin and amylin) via feedback inhibition that is
characteristic of many other hormones. Part of this inhibition may be
local (paracrine), as indicated by the amylin sensitivity of isolated
preparations and the fact that the concentration of secreted products in
the islet interstitium can be over 100-fold higher than in the circulation
(Bendayan, 1993).
AD - Amylin Pharmaceuticals, Inc., San Diego, California, USA.
FAU - Young, Andrew
AU - Young A
LA - eng
```

How the Import Process works

The import process maps the contents of a record to a field within VantagePoint.

So an author on a record would be mapped to list of authors in VantagePoint.


The mapping key is the “Import Filter” which tells VantagePoint how to treat the content of a record based on its source.

TI - Amylin and the integrated control of nutrient influx.
PG - 67-77
AB - The most potent actions of amylin that occur at physiological plasma concentrations include inhibition of food intake, gastric emptying, acid and digestive enzyme secretion, and glucagon secretion. These actions share a common outcome; they each help regulate the rate at which nutrients (including glucose) appear in the blood (Ra). Amylin physiologically orchestrates, via several parallel processes, the rate of entry of nutrient into the circulation, as shown schematically in Fig. 1. In this way, amylin's function may be viewed as complementary to that of insulin (secreted from the same ; the exit of nutrient from blood ; following discussion addresses th is co-secreted with an endocrine pancreatic islets), the target f relevant effects appears to be th be primarily regarded as a neuro AD - Amylin Pharmaceuticals, Inc FAU - Young, Andrew

VantagePoint1

	#Records	#Instances	Author
1	8	8	Mutoh, Seitaro
2	8	8	Young, Andrew
3	5	5	Baron, Alain D
4	5	5	Draghia-Akli, Ruxandra
5	5	5	Heinemann, Lutz
6	5	5	Kawamura, Ikuro
7	4	4	Fineman, Mark S
8	4	4	Hirosuji, Jiro

Summary





How it really works....

The primary tool within the import process is Regular Expression Pattern Matching (RegEx).

This pattern matching protocol allows an import filter to find and extract relatively complex data elements.

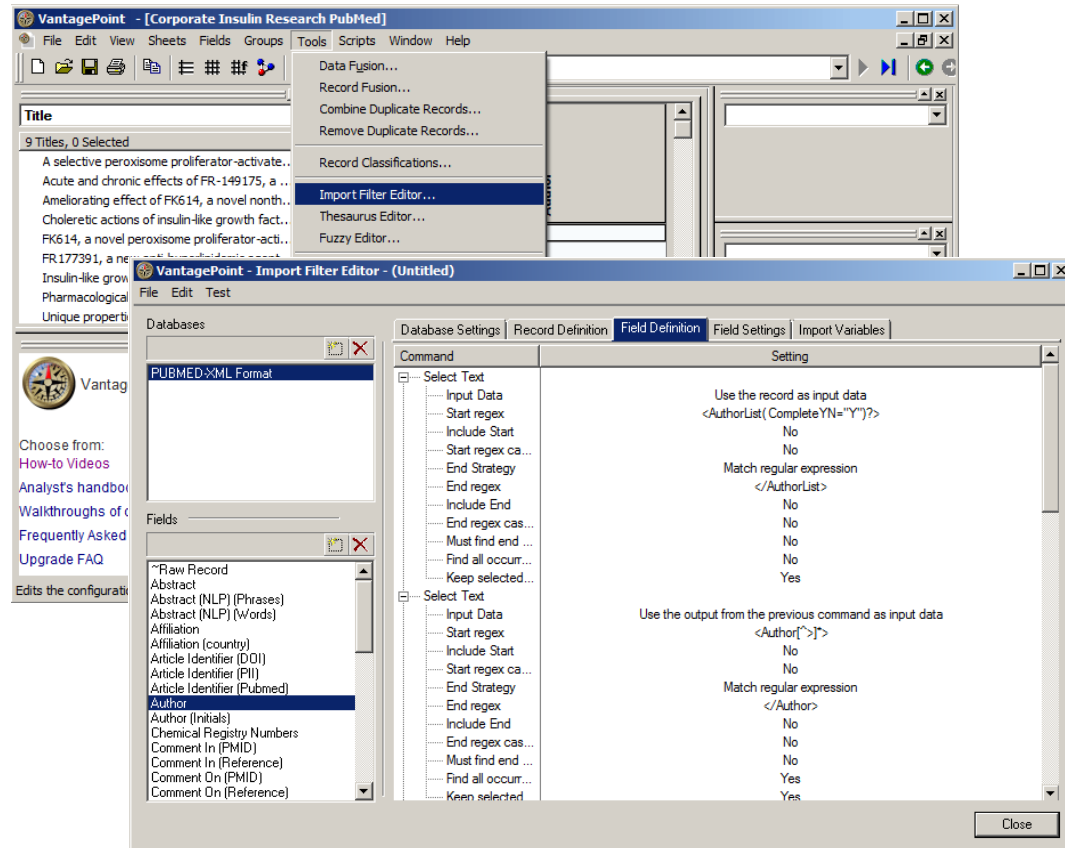
In addition to RegEx, the Import process also supports Natural Language Processing (NLP), Entity Extraction, Dictionary Lookup, Find and Replace (via Thesaurus) and some Mathematical Functions.

The import process can also incorporate script commands to allow macros to run during import.



Import Filter Editor

Import filters can be changed and new ones can be created with the Import Filter Editor (*Tools/Import Filter Editor*)



Downloading Import Filters

To download an import filter

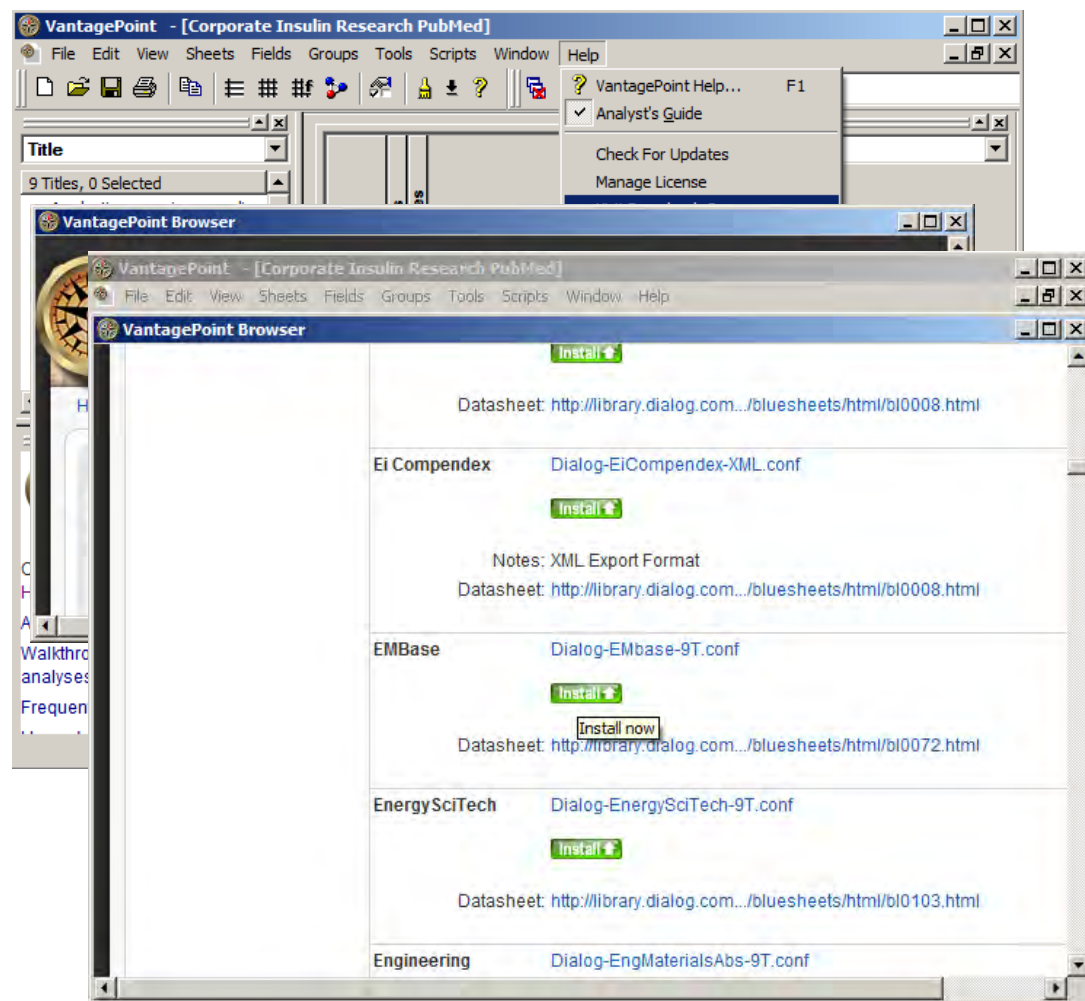
Go to *Help/Visit Downloads Page*

(This will open up a browser and automatically log you into the download site)

Select *Import Filters* from the download menu

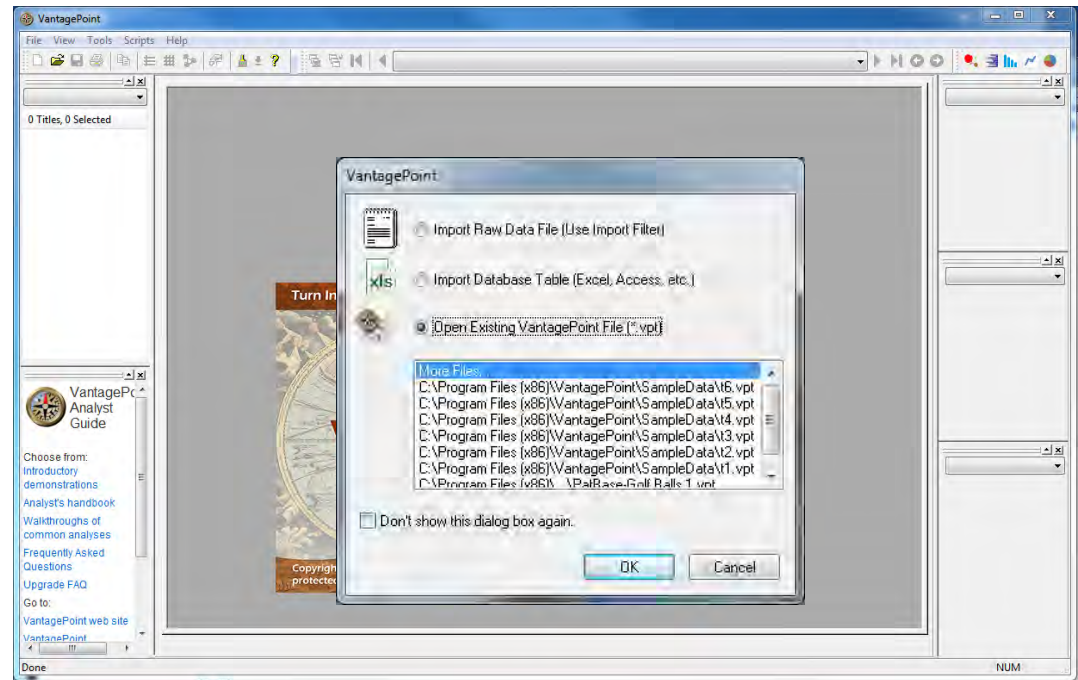
Pick the filter you want to download and click *Install*

This will install the filter in your Import Filter directory



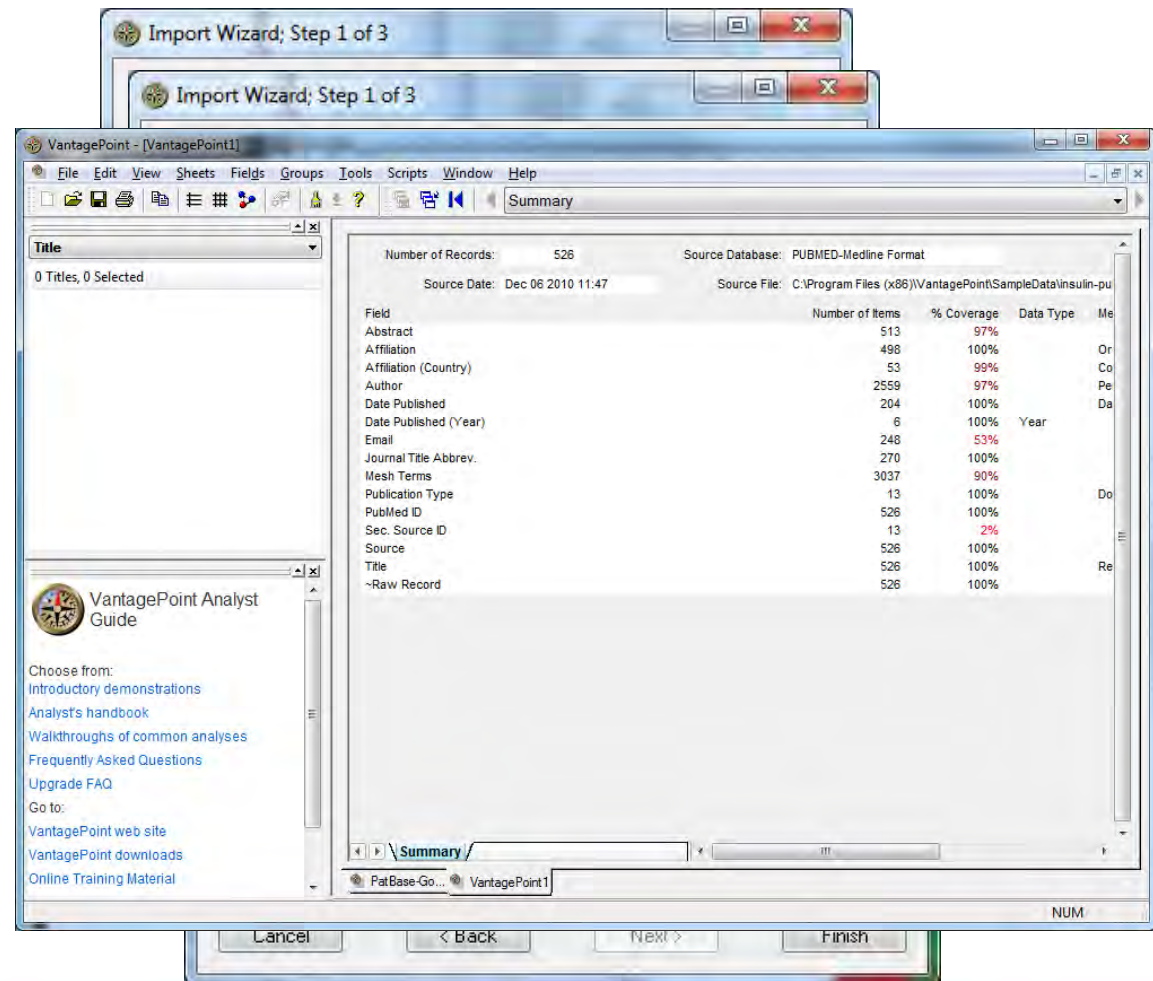
Starting VantagePoint

- From an existing VantagePoint file
 - Double click on a VantagePoint file (*.VPT)
- From the Start menu
 - Programs
 - VantagePoint
 - Dialog to import new or open existing file



Importing Search Results

- Initial Import
 - *Import Data*
- Selecting data
 - *.TXT, *.DAT, etc
 - Excel
 - XML
- Selecting Import Filter
 - *Select New Filter Directory*
- Picking fields
 - *Display Secondary Fields*



Summary Sheet

- Contents
 - Summary Info
 - Fields
 - Items
 - Coverage
 - Data Type
 - Meta Tags
- Right Click Menu
 - *View Statistics*

The screenshot displays the VantagePoint Analyst Guide application. The main window shows a summary report for a dataset named 'PatBase-Golf Balls 1'. The report includes the following information:

- Title:** PatBase-Golf Balls 1
- Number of Records:** 1705
- Source Database:** Minesoft Patbase Documents (XML)
- Source Date:** Sep 15 2010 14:52
- Source File:** C:\Users\Demo\Documents\PatBase-Golf Ball.xml

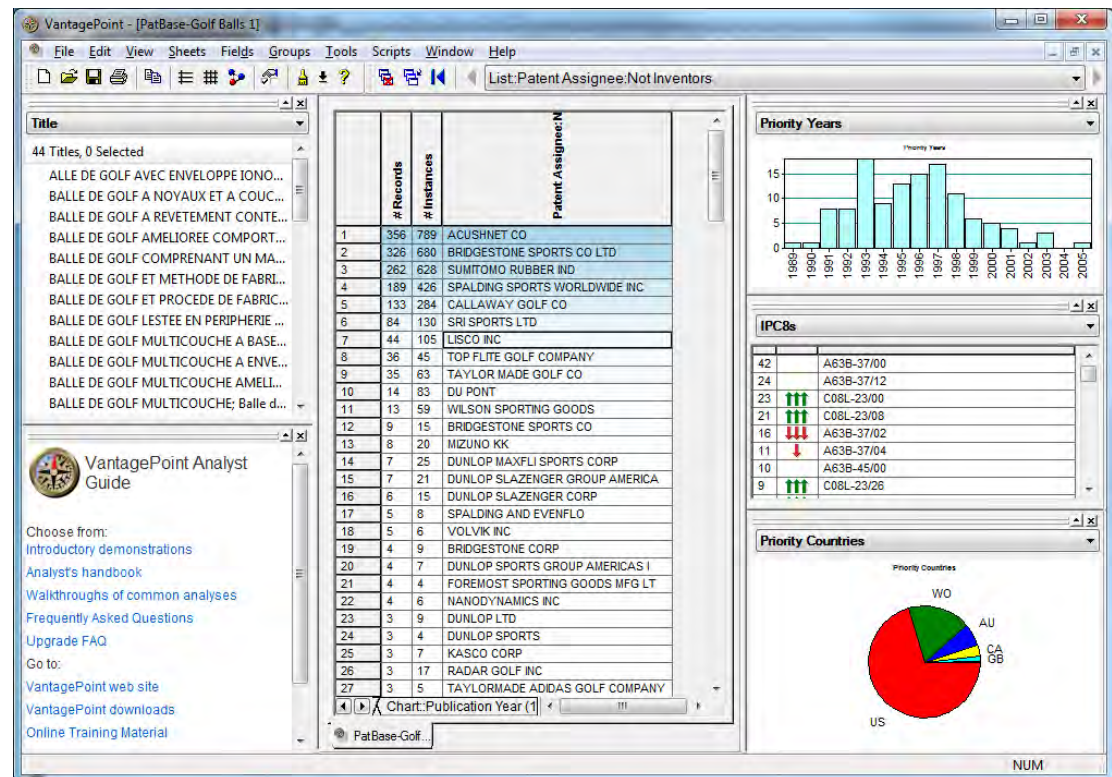
The report lists various fields and their corresponding statistics:

Field	Number of Items	% Coverage	Data Type	Meta Tags
::Advantage	1705	100%	Record Cla	
Abstract	2992	99%		
Abstract (English) (NLP) (Phrases)	13104	97%		
Application Countries	39	100%		Country
Application Dates	2171	100%		Date
Application Numbers	4764	100%		
Application Numbers (long)	4764	100%		
Application Years	21	100%	Year	
Citations	6358	80%		
DesignatedStates	141	18%		
ECLA Code	910	99%		
Family Accession Number	1705	100%	Number	Unique ID
Family Member Countries	29	100%		Country
Family Member Numbers	6932	100%		
Family Member Numbers (long)	6932	100%		
Image Link	1243	72%	File	
Inventors	821	99%		Person
Inventors (Cleaned)	634	99%		Person
PCBs	646	99%		
PCBs (definitions)	646	99%		
PCBs (subclass)	80	99%		
PCBs (subclass) (definitions)	80	99%		
PCs	9	0%		
PCs (definitions)	9	0%		
PCs (subclass)	4	0%		
PCs (subclass) (definitions)	4	0%		
Key	1705	100%	Key	
Patent Assignee	397	97%		Organization
Patent Assignee Not Inventors	153	89%		Organization

The interface also includes a sidebar with navigation links such as 'Introductory demonstrations', 'Analyst's handbook', 'Walkthroughs of common analyses', 'Frequently Asked Questions', and 'Upgrade FAQ'. The bottom of the window shows a search bar and a list of documents, including 'PatBase-Golf Balls 1'.

VantagePoint Navigation

- Work space
 - Sheets
 - Workbooks
- Title View
- Select Controls
 - Shift Select
 - Control Select
- Sizing/Turning Windows On and Off



The slide features a decorative border at the top and bottom, consisting of a historical map with various geographical labels and a grid pattern. The main content area is white.

Importing Data

- Start VantagePoint
- Go to the download site
- Find the WOS import filters
- Install the filter to your copy of VP
- Import the raw data file
- Save your .vpt file using: "Save as"



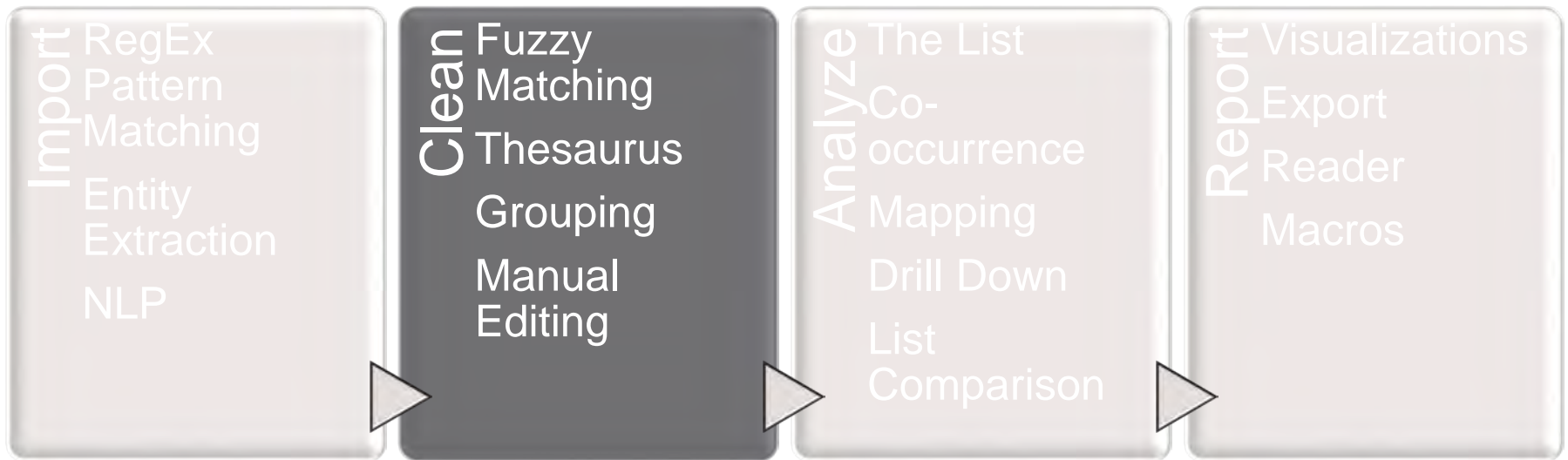
Quiz!

- What organization appears the most?
- What are the top 4 Web of Science Categories?



Session 3

Cleaning




Automation



Clean Your Data

- VantagePoint is primarily a statistical analysis tool, so...

Garbage In = Garbage Out

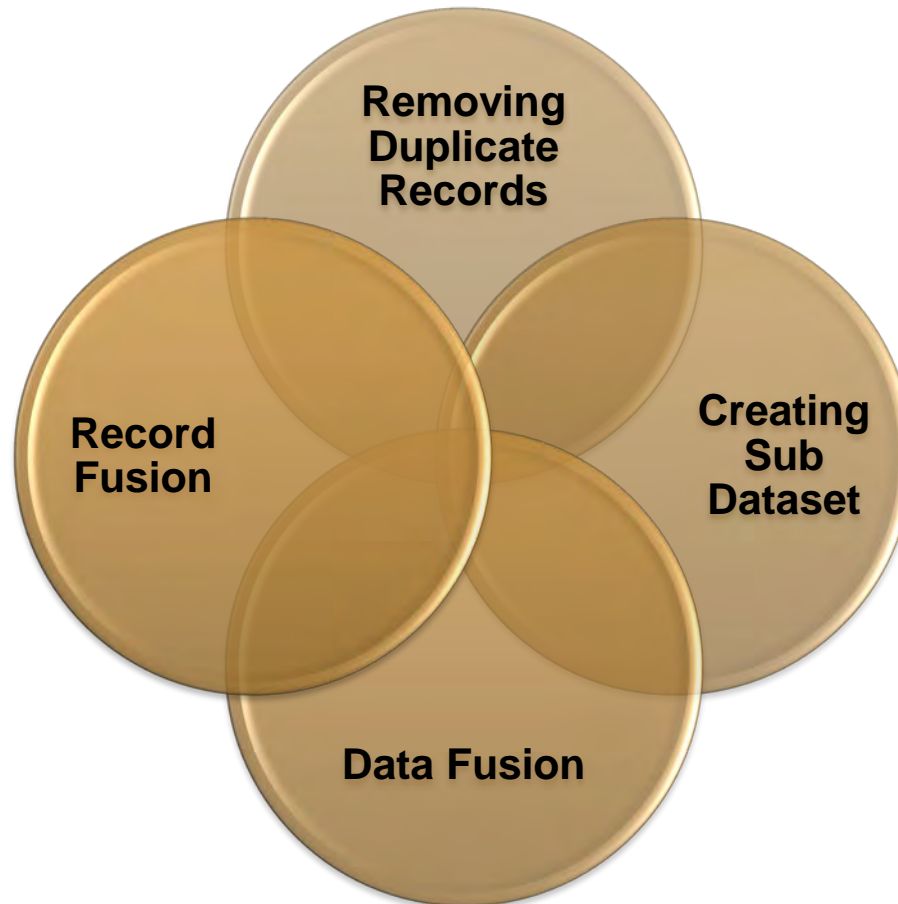
- Data needs to be clean before you run your analysis.
 - You might also find that additional manipulation is necessary as you conduct your analysis.
 - VantagePoint has several different data cleaning/data manipulation tools.
- 

Two Levels of Cleaning

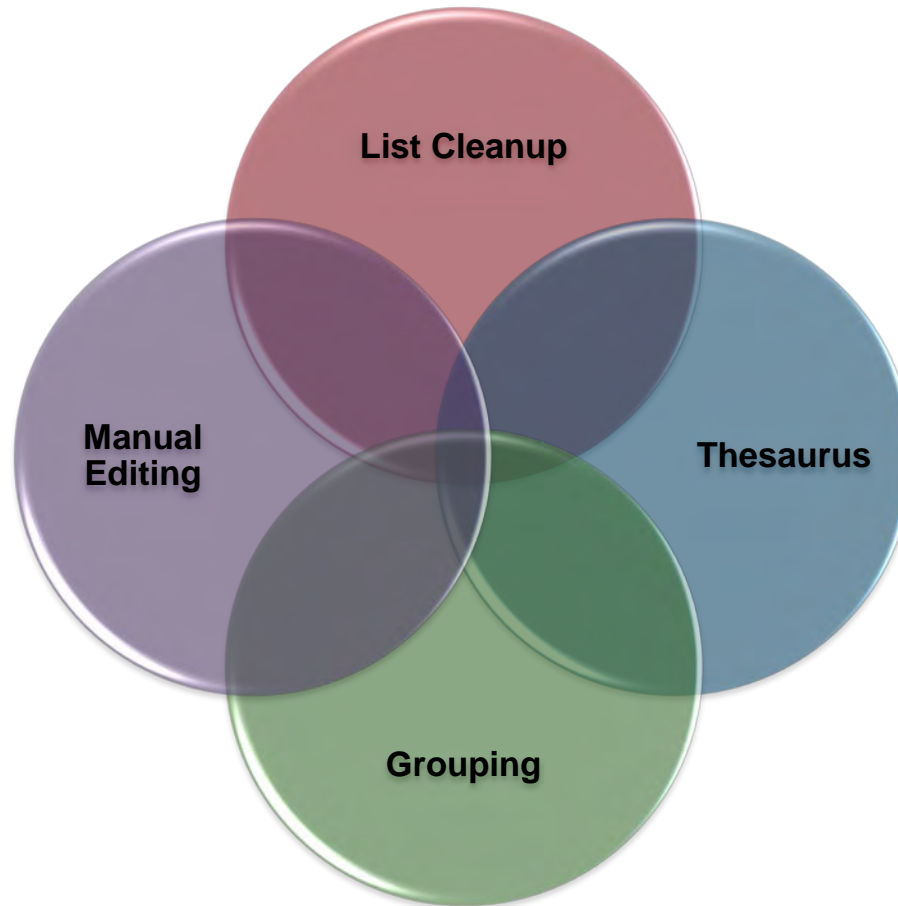
- VantagePoint files can be cleaned at two levels:
 - The record level
 - The field level



Four Ways to Manipulate Records



Four Ways to Clean Fields



List Cleanup

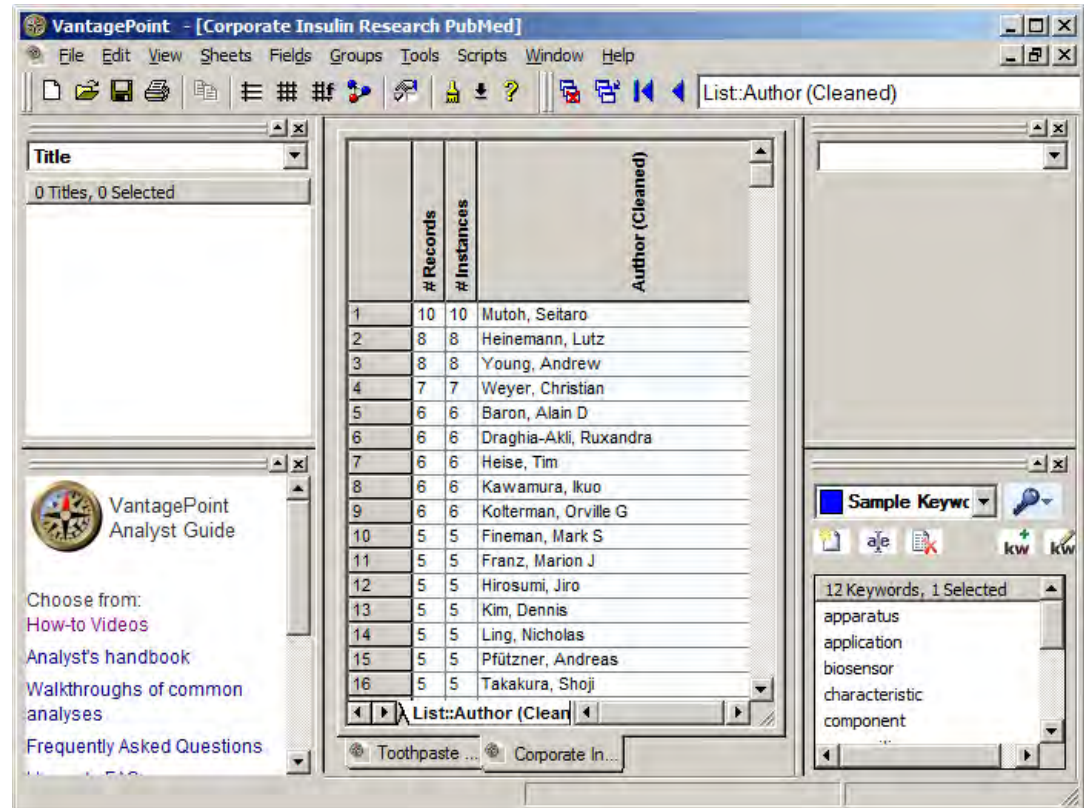
Small differences in the data can distort results

List Cleanup (*Fields/List Cleanup*) helps remove:

- Misspellings
- Hyphenation
- Capitalization
- Different Styles

Cleaned data can be saved as a thesaurus

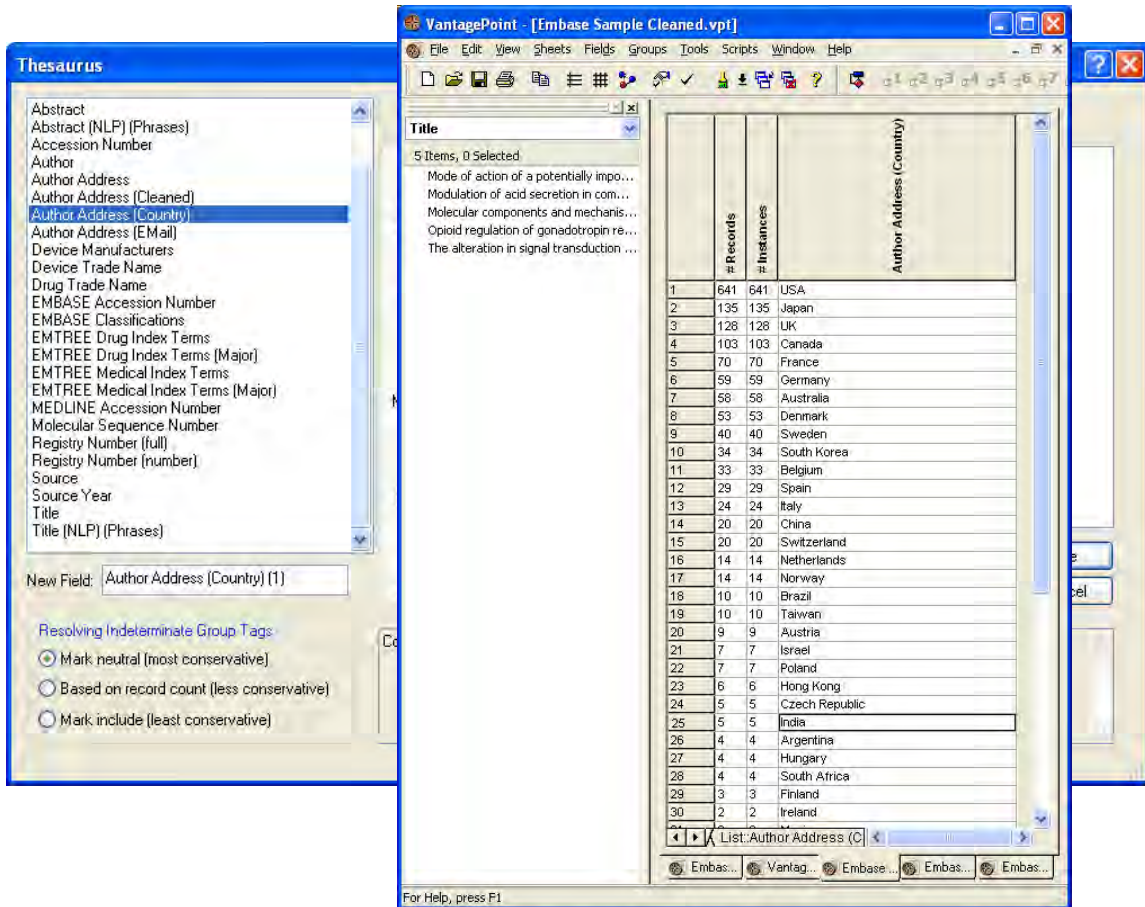
List Cleanup produces a new list. It does not change the original list.



Thesaurus

The Thesaurus
(*Fields/Thesaurus*)
helps you reduce
and categorize
data.

Like List Cleanup,
running a Thesauri
creates a new list.



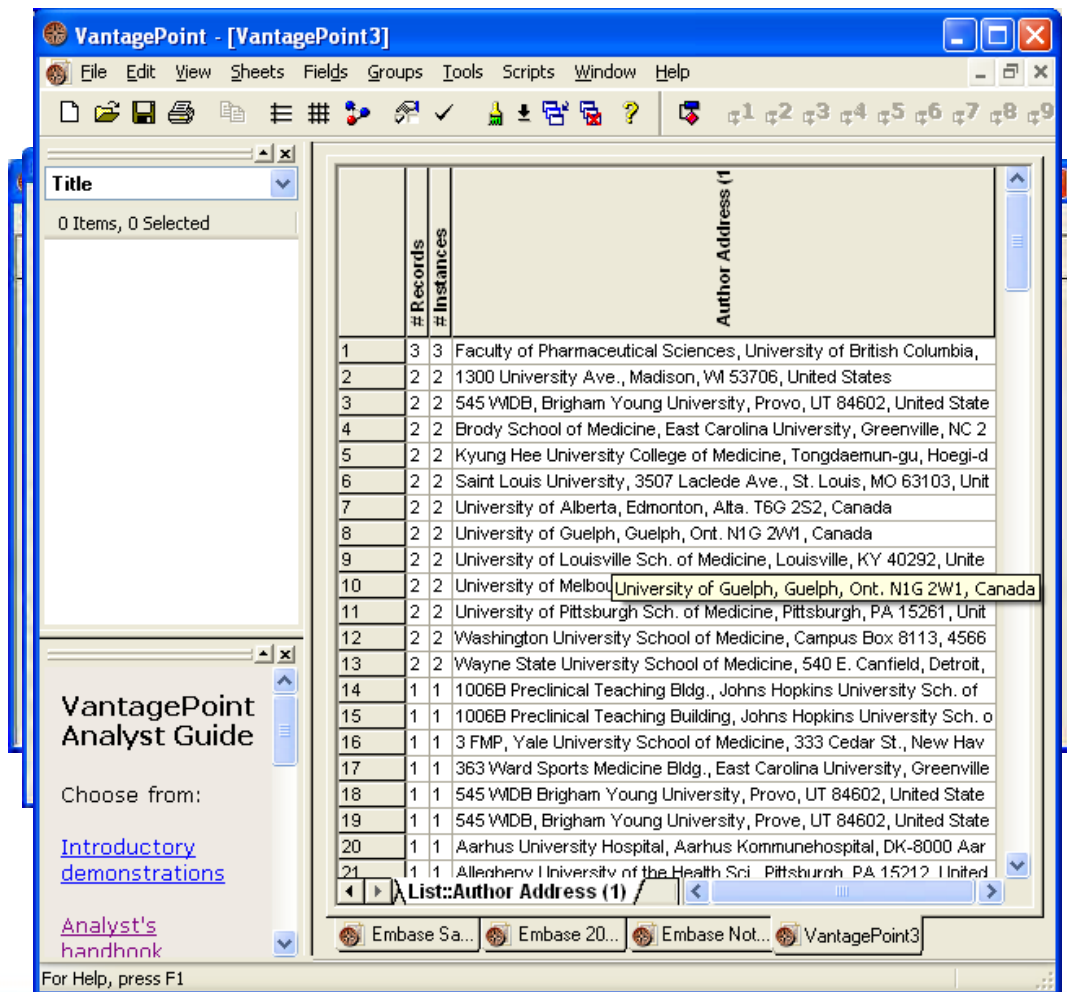
Find and Replace

A Thesaurus that acts as a find/replace within items

- Fill out abbreviations
- Match synonyms within phrases
- Translate American <-> British

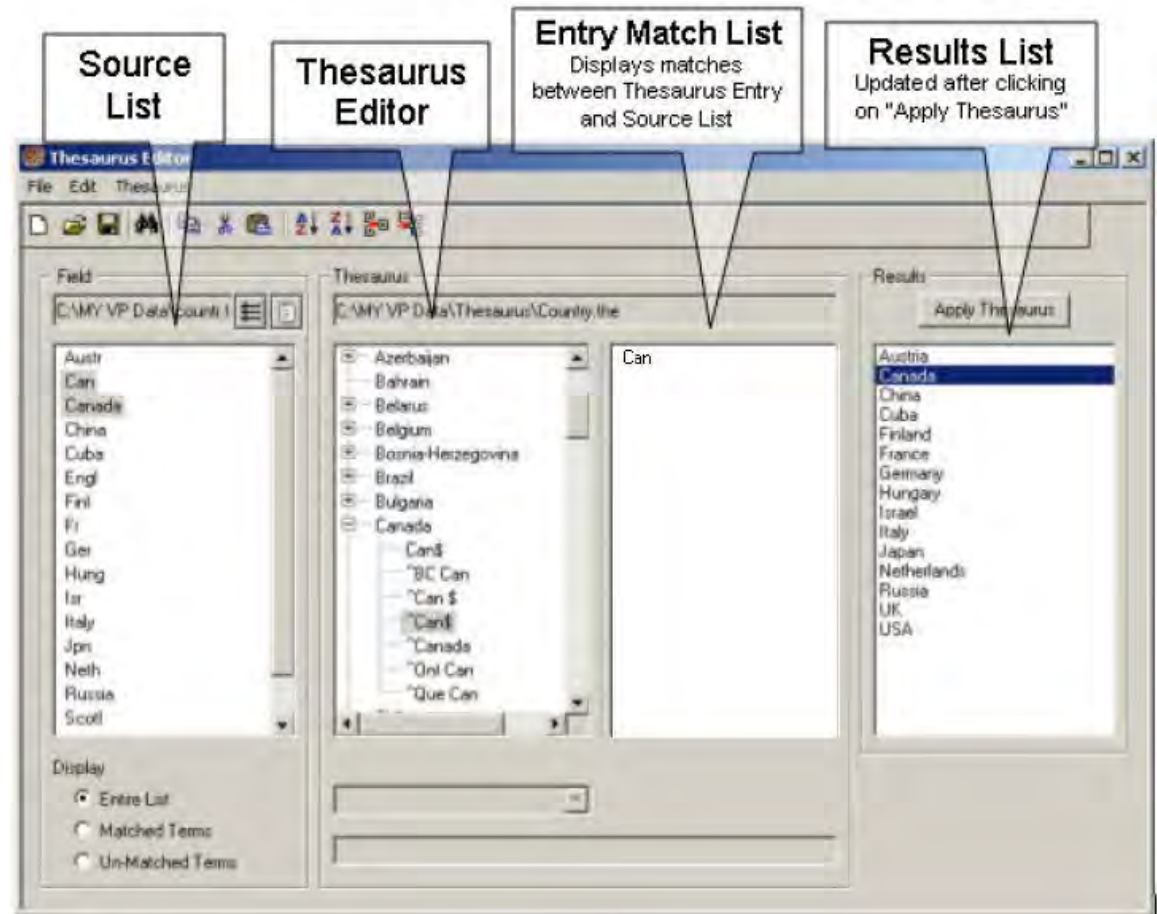
Uses standard thesaurus files

Can be used at import to improve matching



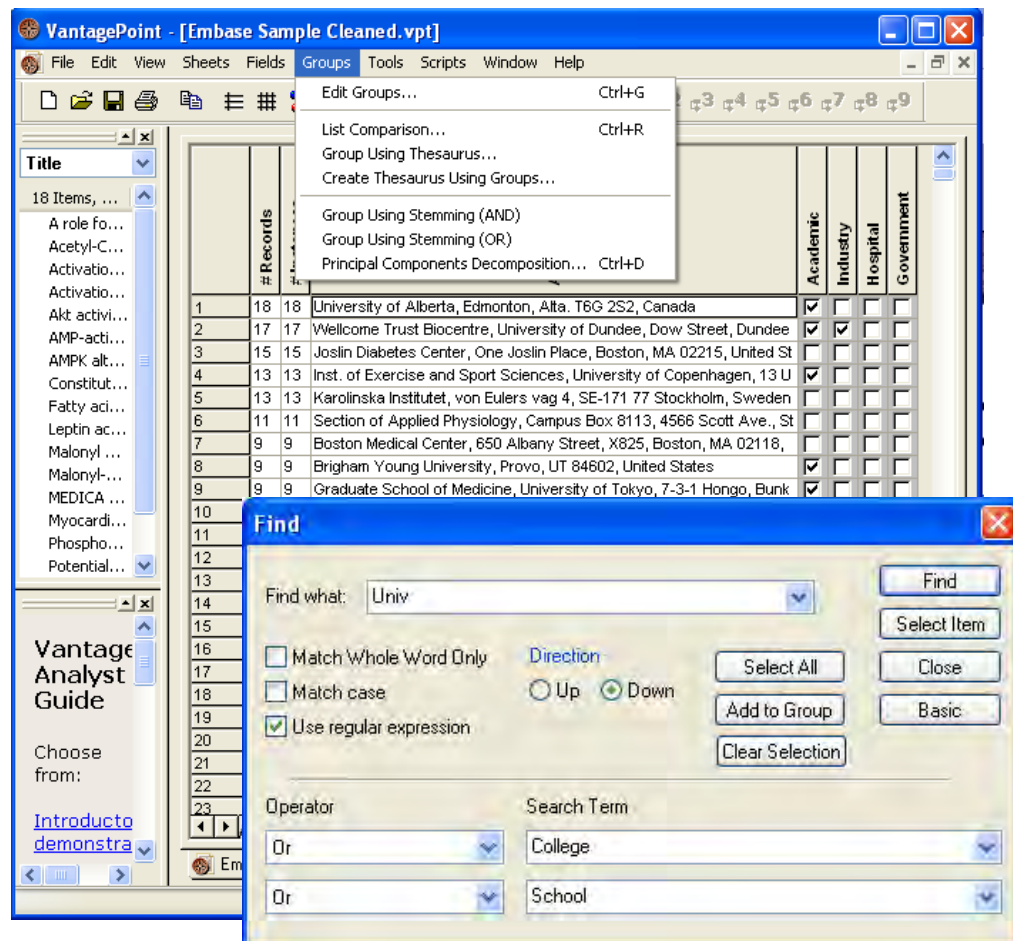
Thesaurus Editor

Thesaurus Editor
(*Tools/Thesaurus Editor*) allows you to customize thesauri.



Groups

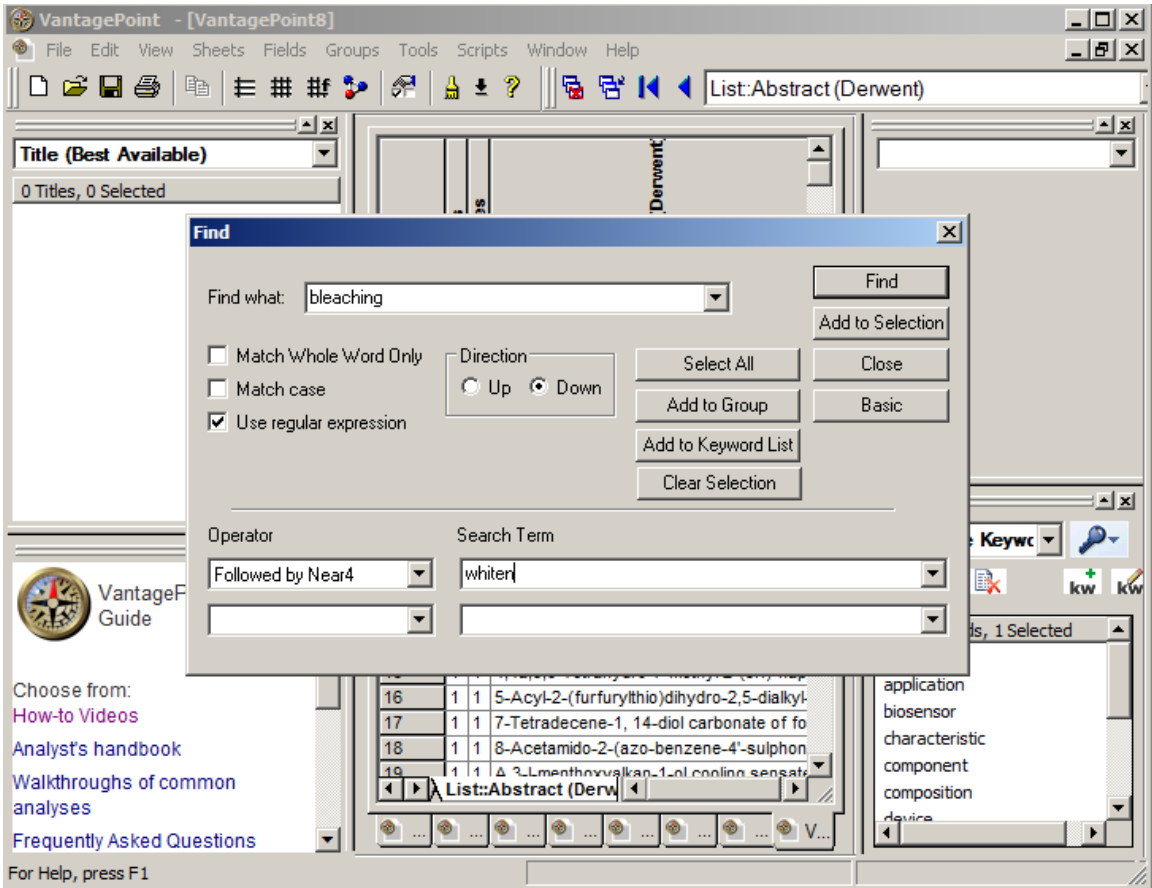
- Groups allow you to categorize field content
- Groups can be created manually, semi manually, or automatically
 - Check/X/Blank toggle
 - Find/Select All
 - Groups using Stemming And/Or



[illegible]

Find Command

The Find Command (*Edit/Find*) or Ctrl-F opens the find window which can be used to find terms or assign found items to groups



The screenshot displays the VantagePoint software interface. A 'Find' dialog box is open, showing search criteria. The 'Find what' field contains 'bleaching'. The 'Match whole word only' checkbox is unchecked, 'Match case' is unchecked, and 'Use regular expression' is checked. The 'Direction' is set to 'Down'. The 'Operator' is 'Followed by Near4' and the 'Search Term' is 'whiter'. The background shows a list of chemical compounds and a sidebar with navigation links.

Choose from:
[How-to Videos](#)
[Analyst's handbook](#)
[Walkthroughs of common analyses](#)
[Frequently Asked Questions](#)

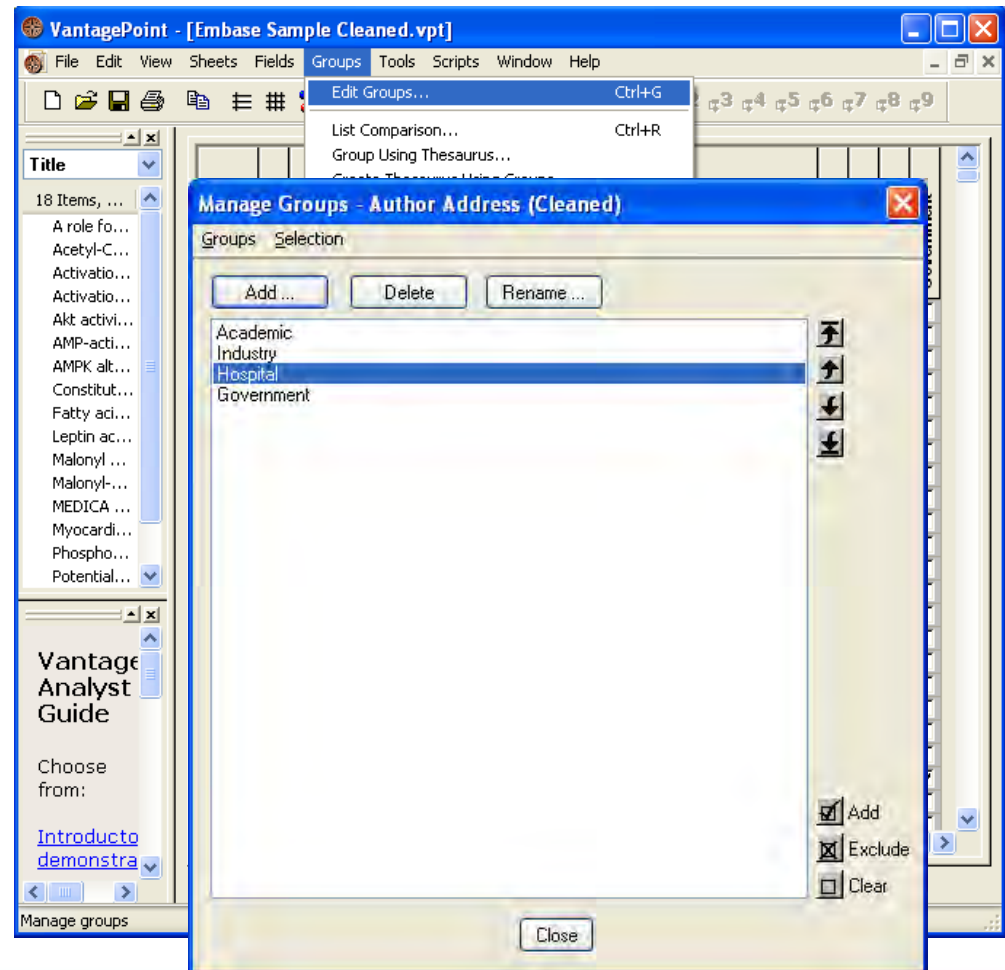
For Help, press F1





Group Management

- The Edit Groups Command (*Groups/Edit Groups*) enables you to manage your groups
 - Create new groups
 - Delete groups
 - Change state of entire group
 - Reorder Groups
 - Rename Groups





List Cleanup/Thesauri/Group are Related

Do note that during List Cleanup, you can save your cleanup as a Thesaurus. You can also turn your groups in to a thesaurus. And you can turn your thesaurus into groups.

These three related functions are designed to allow you to replace content within the database with something else more appropriate for your analysis.

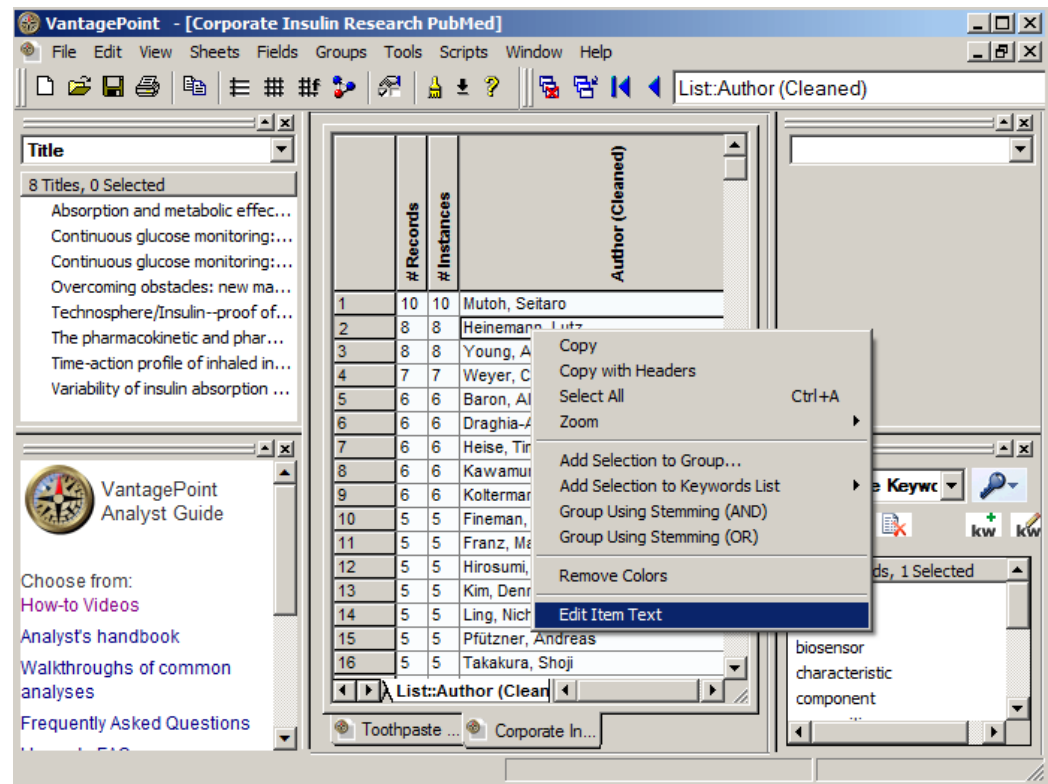
- List Cleanup relies on the software (Fuzzy Matching)
- Thesaurus relies on pre-recorded codified knowledge (Thesauri)
- Groups rely on you the user (Manual Changes)



The Fourth Way – Manual Editing

Right clicking on any item in a list will open up a menu that contains *Edit Item Text*

Selecting this command allows you to edit individual entries in a field



The slide features a decorative border at the top and bottom, consisting of a historical map with various geographical labels and a grid pattern. The main content area is white.

Data Cleaning

- Open the file
- Clean the Authors
- Apply a thesaurus to the author affiliations (organization only) to get groups (academic, industrial, etc.)



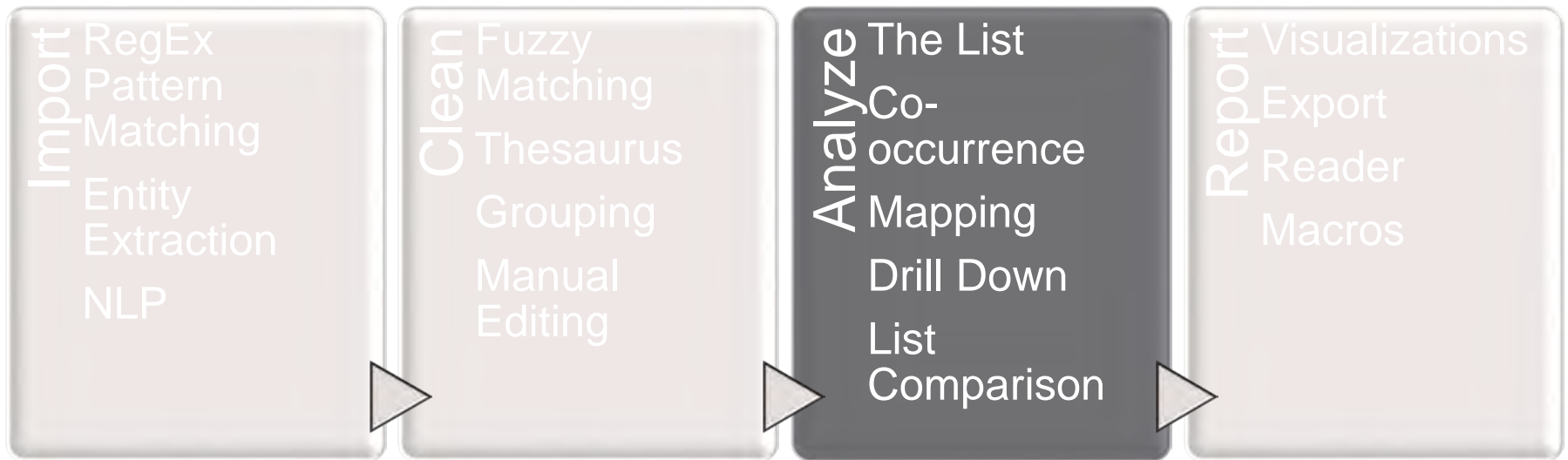
Quiz!

- How many records does Hussain have after you clean the data?
- Which is the leading corporate publisher in the dataset?



Session 4

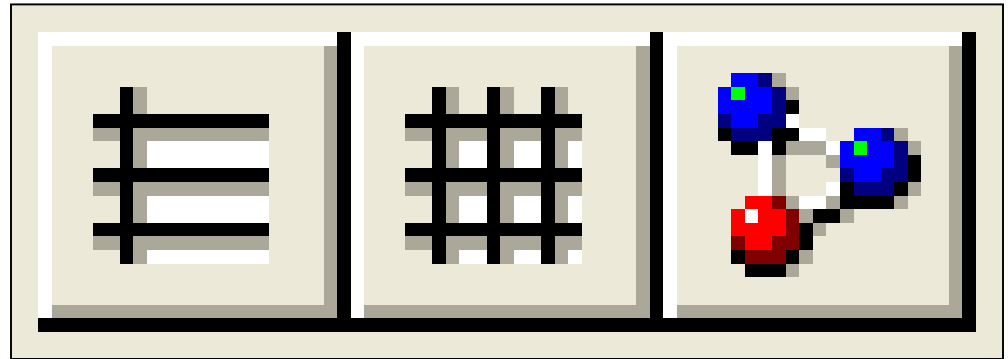
Analyzing Data



Automation

The Three Core VantagePoint Analytical Tools

- **The LIST**
- The CO-
OCCURRENCE
MATRIX
- The MAPS

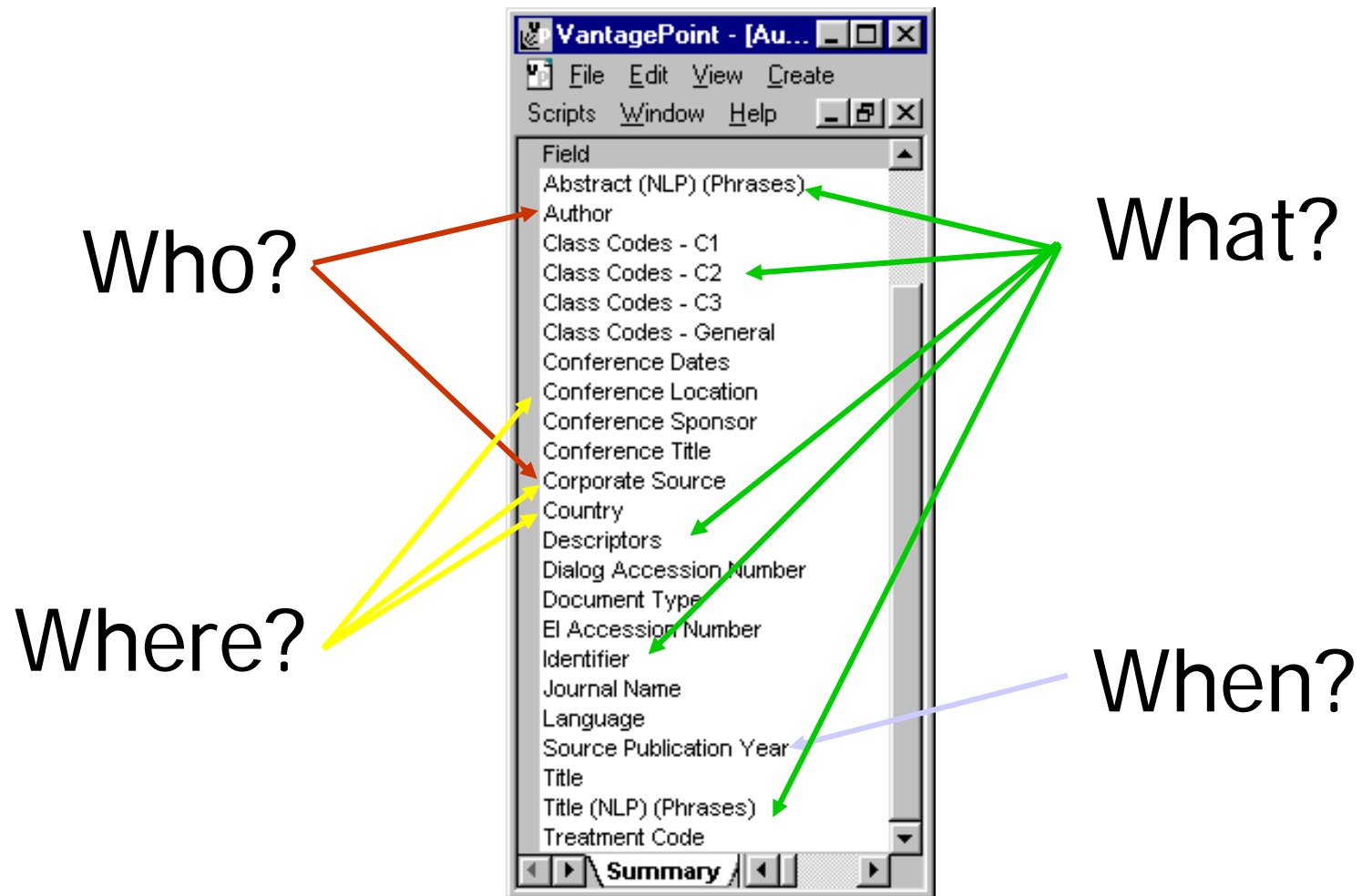


The List

- Lists show you all the items in a particular field across all of the records in the imported set
 - # Records
 - # Instances

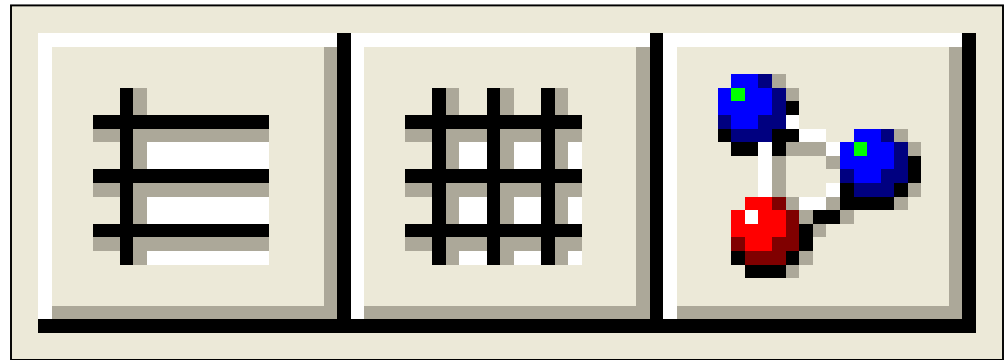
	# Records	# Instances	Inventors (C)
1	376	719	SULLIVAN MICHAEL J
2	161	233	LADD DEREK A
3	155	292	MURALI RAJAGOPALAN
4	140	365	NESBITT ROBERT DENNIS
5	136	281	KENNEDY THOMAS J III
6	116	301	BINETTE MARK L
7	90	183	YAMAGISHI HISASHI
8	84	180	TZIVANIS MICHAEL JOHN
9	79	190	HIGUCHI HIROSHI
10	75	146	ICHIKAWA YASUSHI
11	67	97	HERBERT EDMUND A
12	65	120	CAVALLARO CHRISTOPHER
13	64	106	BULPETT DAVID A

Analytical Uses of Lists



The Three Core VantagePoint Analytical Tools

- The LIST
- The CO-
OCCURRENCE
MATRIX
- The MAPS



The Co-occurrence Matrix

- The co-occurrence matrix shows the numerical relationship between two fields.
- It is the simplest form of co-word bibliometric data.
- With the co-occurrence matrix, you can see the relationship between one field and any other field.

		Field 2		
		Item A	Item B	Item C
Field 1	Item X	3	0	2
	Item Y	0	7	0
	Item Z	0	4	2

Creating a Co-occurrence Matrix

- From the Menu
 - *Sheets/Add Matrix*
- From the  Button

Create Matrix

Rows: Columns:

VantagePoint - [Embase Sample.vpt]

File Edit View Sheets Fields Groups Tools Scripts Window Help

3013 Items, 0 ...

Title

- Role of PDE3...
- Mining freque...
- Role of PKA a...
- Resistance ex...
- Development...
- Topiramate st...
- Phosphodiester...
- Protective eff...
- AMP-activate...
- AMPK regulati...
- Acetyl-L-carnit...
- Inhibition of g...
- Activation of ...
- Effects of ad...

VantagePoint Analyst Guide

Choose from:

[Introductory demonstration](#)

Reset

Author

Records

1 1950 **2** 1045 **3** 770 **4** 604 **5** 544 **6** 265 **7** 244 **8** 238 **9** 216 **10** 194

EMBASE Classifications

Records

Show Values >= 1

Cooccurrence # of Records

1 138 **2** 84 **3** 80 **4** 57 **5** 41 **6** 40 **7** 36 **8** 36 **9** 31 **10** 31 **11** 30 **12** 29

Hardie, D. G. 87 22 20 8 23 3 1 2 5

Kemp, B. E. 55 12 14 7 12 3 1 3 5

Carling, D. 60 10 13 6 2 8 3 5 3

Witters, L. A. 45 3 3 1 4 3 1 1 3 6

Ruderman, N. B. 20 12 17 10 7 3 1

Richter, E. A. 19 3 6 3 20

Winder, W. W. 24 2 6 2 26 1

Goodyear, L. J. 20 6 12 3 11 2

Saha, A. K. 13 10 13 8 6 2 1

Stapleton, D. 23 1 1 2 2 1 2 3

Lopaschuk, G. D. 18 5 5 4 7 5 16

Viollet, B. 14 4 12 3 1 4 1 3 2 1

List: Author Matrix: Author x EMBASE Clas

VantagePo... Embase Sam...

Navigating the Co-occurrence Matrix

- The functions within Co-occurrence include

- Sorting  

- Re-Sort 

- Flooding 

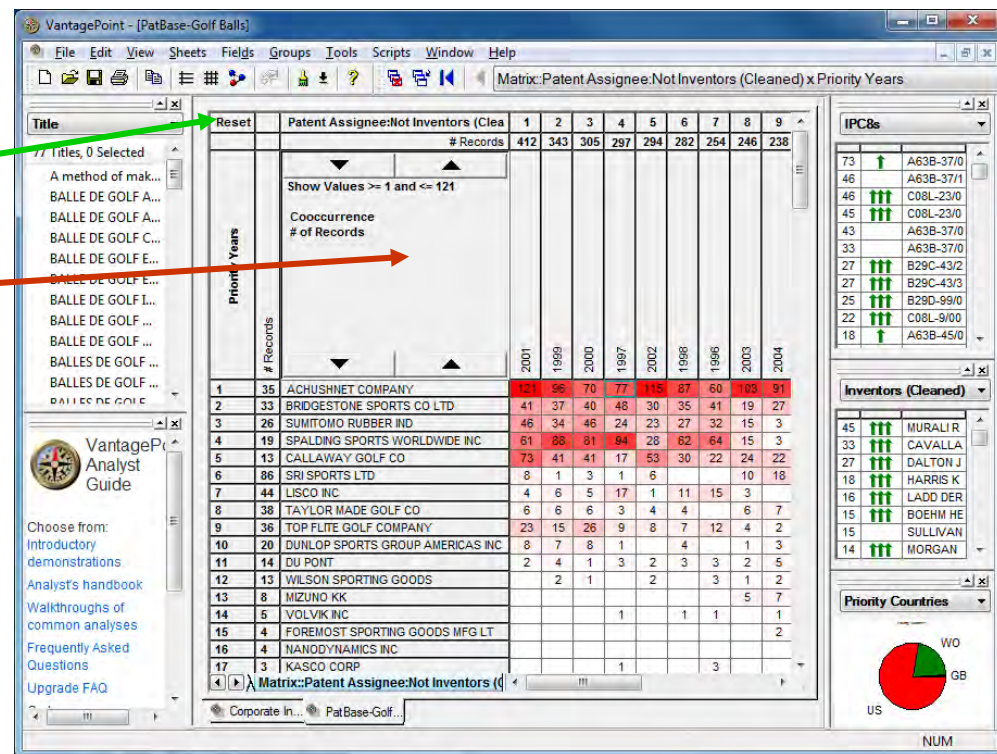
- Getting to the Record

- Painting 

- Heat Maps

- Copying/Pasting

- *Edit/Copy*



Analytical Uses of Co-occurrence Matrices

Who/Where?

Who/When?

The screenshot shows the VantagePoint software interface with a co-occurrence matrix. The matrix displays the relationship between 16 authors (rows) and 10 EMBASE classifications (columns). The top row shows the total number of records for each classification, and the first column shows the total for each author. The cells in the matrix represent the number of co-occurrences between a specific author and a specific classification.

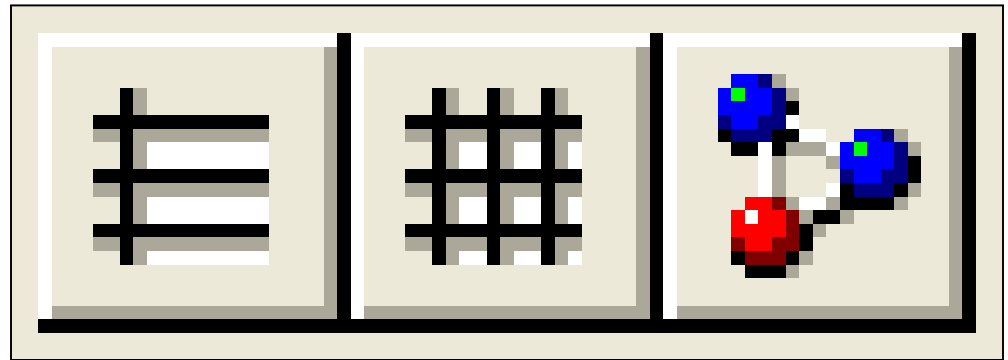
Reset	Author	1	2	3	4	5	6	7	8	9	10
	# Records	1950	1045	770	604	544	265	244	238	216	194
EMBASE Classifications	Show Values >= 1										
# Records	Cooccurrence # of Records										
1	Hardie, D. G.	87	22	20	8	23	3	1	2	5	
2	Kemp, B. E.	55	12	14	7	12	3	1	3	3	5
3	Carling, D.	80	10	13	6	2	8	3	5	3	
4	Witters, L. A.	45	3	3	1	4	3	1	3	6	
5	Ruderman, N. B.	20	12	17	10	7	3	1			
6	Richter, E. A.	19	3	6	3	20		2		1	
7	Winder, W. W.	24	2	6	2	26	1				
8	Goodyear, L. J.	20	6	12	3	11	2		2	2	
9	Saha, A. K.	13	10	13	8	6	2	1			
10	Stapleton, D.	23	1	1	2	2		1	2	3	
11	Lopaschuk, G. D.	18	5	5	4	7	5			16	
12	Violet, B.	14	4	12	3	1	4	1	3	2	1
13	Greengard, P.	13	6	1	4	4			8		1
14	Hawley, S. A.	15	7	3	3	1	1		1		1
15	Wojtaszewski, J. F. P.	18	4	5	3	16			1		1
16	Woods, A.	18	1		1			1			
17	Foutelle, F.	16	4	9	3	1			1		

Where/What?

etc.....

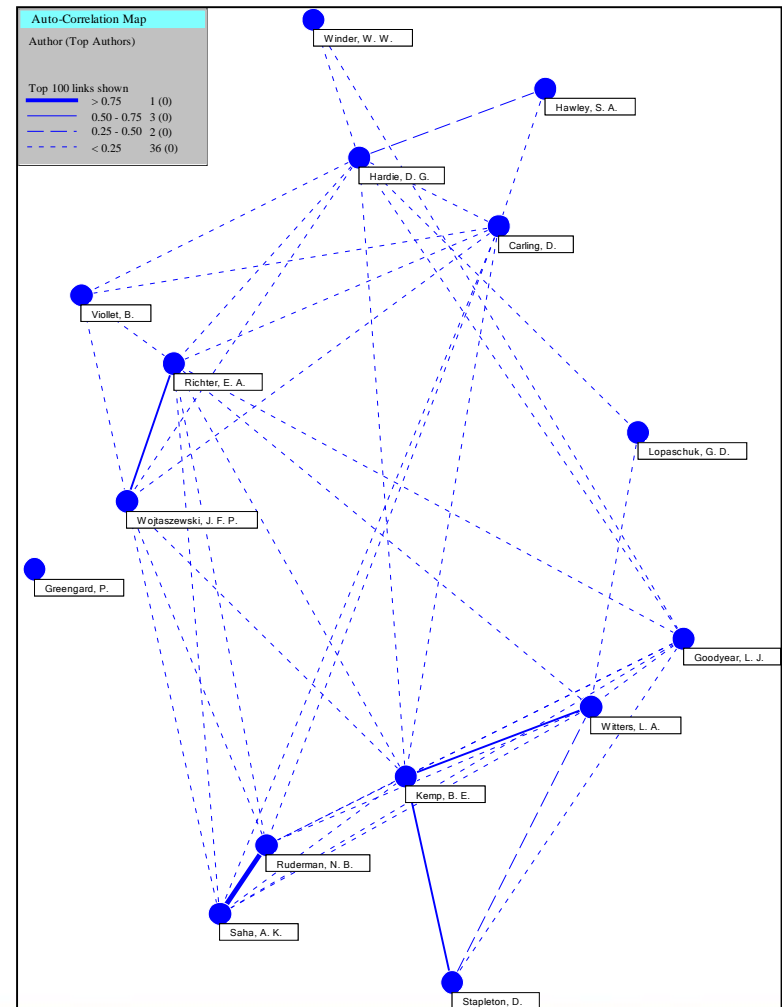
The Three Core VantagePoint Analytical Tools

- The LIST
- The CO-OCCURRENCE MATRIX
- **The MAPS**



What is a Map?

- A Map is a visualization of a multidimensional co-word analysis
- Maps show the relationships among selected items






What can you do with Maps

- Maps are used to visualize complex relationships
 - Auto-Correlation Maps are typically used to visualize knowledge networks
 - Mapping how authors co-publish
 - Cross-Correlation Maps are typically used to visualize differences between organizations and researchers
 - Mapping organizations or authors against descriptors
 - Factor Maps are typically used to reduce lengthy lists
 - Clustering and Mapping descriptors



Quiz!

- Make a auto-correlation map of country by country
 - Click on the  icon (3 connected balls)
 - Choose what to map: "Countries"
 - Click "Finish"

Navigating in a Map

- All VantagePoint Maps have similar layout and navigation
 - Navigating to raw record
 - *Click on a Ball*
 - *Click on a title*
 - The meaning of the ball
 - Represents item mapped
 - The meaning of stick
 - Represents linkage
 - Similarity Strength Legend
 - Drop down
 - *Roll over Ball*
 - *Double Click to Stick*

