Software for
Business Intelligence

BizInt Smart Charts

Cookbook of Reports & Visualizations

www.bizint.com/Cookbook

Created with the BizInt Smart Charts Product Family

Patent/IP Sequence

OCTOBER 2020

USING THIS COOKBOOK

APPLYING THESE RECIPES TO YOUR DATA

We believe the methods in the recipes included here are applicable to other data sets and sources. However, because each recipe was built using a particular example dataset, you should take some caveats into consideration when applying recipe steps to your own data:

Caveat One: When you combine charts from different sources, the source selected first (the "Key Chart") will determine the names for common columns. For example, depending on the source, the column could be called Condition, Disease or Indications. While you may see different column names than those indicated in the recipe, the same steps should still apply.

Caveat Two: Some recipes contain steps and guidance based on the content of the example dataset. For example, we provide suggestions for changing Common Trial ID assignment in cases where the automatically generated Common ID combines trials we want to be treated separately. How much this step is necessary depends on the dataset.

Caveat Three: We revise these recipes based on user feedback, so check the latest edition of the Cookbook on our website to see if we have revised the recipe you're interested in.

STYLE CONVENTIONS USED

Commands are shown in **colored sans-serif type** with the pipe(|) symbol used to divide sequential menu selections. For example, **File | Combine** indicates that you should first click on the File menu and then select Combine.

Italics indicate a note, generally indicating steps that may be different depending on the dataset.

"Text in quotation marks" indicates something you should type in to a text box or text that appears in the software outside of the menus.

Note:

*This edition of the Cookbook is a selection of some of our favorite recipes, updated for **VantagePoint - Smart Charts Edition Version 11, and focused on patent and IP sequence reports and visualizations.***

If you are using an earlier version of VP-SCE, please refer to the February 2018 or earlier versions of the Cookbook, as many VP-SCE functions have been redesigned in Version 11.

THE JOURNEY CONTINUES...

TABLE OF CONTENTS

Analyzing Companies by Therapeutic Activity with a Bubble Chart [Patents PT-12]

This is one example of a bubble chart created with VantagePoint – Smart Charts Edition (VP-SCE) using data exported from BizInt Smart Charts for Patents (see other recipes in this Cookbook for bubble charts with pipeline and clinical trial data.) Here the bubble chart shows companies by therapeutic activity based on records from Derwent WPI, and PatBase. A single record for each patent family was created with BizInt Smart Charts Reference Rows and then exported to VP-SCE. We linked Probable Assignee from PatBase with Therapeutic Activity from DWPI to build the bubble chart in VP-SCE.

CAS9 - Top Therapeutic Activities for Top Assignees

	Title	Patent Family			Probable Assignee	Derwent Class	Therapeutic Activity	Application Category (VP-SCE)
		Patent	Kind	Date				
1.	CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING HERPES SIMPLEX VIRUS TYPE 1 (HSV-1)	WO 15153789	A1	2015-10-08	EDITAS MEDICINE INC	B04 D16	Virucide.	Virucide
2.	DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR HBV AND VIRAL DISEASES AND DISORDERS	WO 15089465	A1	2015-06-18	UNIV ROCKEFELLER	A96 B04 D16	Virucide. No biological data given.	Virucide
3.	CONFERRING RESISTANCE TO GEMINIVIRUSES IN PLANTS USING CRISPR/CAS SYSTEMS	WO 15048707 / WO 15048707	A2 / A3	2015-04-02 / 2015-05-21	UNIV MINNESOTA	A97 C06 D16; P13	Plant Protectant. No biological data given.	Plant Protectant
4.	A SOYBEAN U6 POLYMERASE III PROMOTER AND METHODS OF USE	US 2015059010 / US 2015082478 / WO 15026887 / WO 15026886 / WO 15026885 / WO 15026883 / WO 16007948 / AU 2014308898	A / A / A1 / A1 / A1 / A1 / A1 / AA	2015-02-26 / 2015-03-19 / 2015-02-26 / 2015-02-26 / 2015-02-26 / 2015-02-26 / 2016-01-14 / 2015-02-26	PIONEER HI BRED INT INC	C06 D16	Plant Protectant.	Plant Protectant
5.	CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING PRIMARY OPEN ANGLE GLAUCOMA	WO 15153780	A1	2015-10-08	EDITAS MEDICINE INC	B04 D16	Ophthalmological. No biological data given.	Ophthalmological
6.	CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING LEBER'S CONGENITAL AMAUROSIS 10 (LCA10)	US 2015252358 / WO 15138510	A / A1	2015-09-10 / 2015-09-17	EDITAS MEDICINE INC	B04 D16	Ophthalmological. Test details are described but no results given.	Ophthalmological

Software for
Business Intelligence

BizInt Smart Charts

Analyzing Companies by Therapeutic Activity with a Bubble Chart [Patents PT-12]

How did we create this?

SEARCHES:

- GenomeQuest: Identify sequences in GenomeQuest for CRISPR/CAS9.
- PatBase & Derwent World Patents Index: Search using the patents numbers from the GenomeQuest sequence results.

BIZINT SMART CHARTS FOR PATENTS

1. Import each set of search results into BizInt Smart Charts and use the **File | Combine** feature to create a combined chart.
2. Run the **Identify Common Patent Families** tool and sort by the Common Family field.
3. Display the Probable Assignee and Therapeutic Activity columns.
4. Save the chart. Select **File | Send to Reference Rows** to open the chart in BizInt Smart Charts Reference Rows.

BIZINT SMART CHARTS REFERENCE ROWS

5. Use **File | Export** to export the visible columns to VantagePoint - Smart Charts Edition.

VANTAGEPOINT - SMART CHARTS EDITION

6. If the file does not launch automatically in VP-SCE, then import the .vpxi file into VP-SCE using **Import from BizInt Smart Charts** under the **Home** tab. Save your .vpt file.
7. **Cleanup Assignees**: Right-click on Probable Assignee and select List Cleanup... On the List Cleanup panel, under Matching ruleset, select Organization Names (Pharma) and click OK.
8. Use the Cleanup Confirmation panel to further cleanup company names and then click Accept. *See "Technique: Using List Cleanup" recipe for more details.*
9. Select the top assignees, then right click to select Add Selection to Group.... In the Add Items panel, type the word "top" and click OK. *To focus on industry, you can uncheck individuals and academic institutions.*
10. **Cleanup Therapeutic Activity**: Return to the Summary Sheet. Double-click on Therapeutic Activity to see values for this field. Under the **Refine** tab, select **Further Processing | Divide at | Semicolon**. Then apply **Further Processing | Read Until - First | Period**.
11. Select the top therapeutic activities, then right click to select Add Selection to Group.... In the Add Items panel, type the word "top" and click OK.
12. **Create the Bubble Chart**: Save the file. Under the **Report** tab, select **Bubble Chart** and a window will open with options for the Bubble Chart.
13. For "Step 1: Select Field for Rows", select the cleaned assignees field - Probable Assignee (Cleaned). Choose the "Select a group" radio button and then select the "top" group.
14. For "Step 2: Select Field for Columns", select your cleaned therapeutic activity field - Therapeutic Activity: Divide at/ semicolon: Read Until - First/Period from the drop-down. Choose the "Select a group" radio button and then select the "top" group.
15. For Step 3, you can enter a template name if you want to save these settings.
16. Click "Run" and the a bubble chart showing assignees by therapeutic activity will be created.
17. **Edit the Bubble Chart**: You can click the "#" at the top left of the bubble chart to display the number of records in each bubble. Click on the chart title to edit it.
18. In VP-SCE, you can click on a bubble and see information about the records in each bubble in detail windows.
19. **Export the Bubble Chart**: Under the Report tab, click on Export Image to export the bubble chart in a variety of formats.
20. **Export new fields to BizInt Smart Charts:** Under the **Report** tab click on **Export to BizInt Smart Charts.** In the Export Wizard, click Deselect All and then click the checkboxes to select your new filtered fields (such as the cleaned up therapeutic activities.) Click OK to export.

BIZINT SMART CHARTS FOR PATENTS

21. When BizInt Smart Charts for Patents comes to the foreground, select the original chart. It will redisplay with your new cleaned-up fields as new columns.

Summarizing Sequence Locations in a Patent Family [Patents PT-8]

In this example, we've used two new features in BizInt Smart Charts Reference Rows -- the "Create Subtable from Columns" tool and "Summarize All Values" rule -- to create a subtable summarizing key columns for sequences grouped by the Common Patent Family tool.

SUBTABLE SUMMARIZING RELATED SEQUENCES

	Title	Database	Patent Assignee	Sequence Summary				
				Sequence ID	% Identity	Length	Location	
1.	New diagnostic and therapeutic polynucleotides and polypeptides, useful in diagnosing a condition, disease or disorder associated with human molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or in gene mapping.	1.1 GENESEQ | *link* 1.2 GPATPRT | *link*	INCYTE CORP.	WO20040023973-5059	91.59	642	Claim 27; Page; 190pp; English.	1.1
				WO2004023973-5059	91.59	642	claim: 27; 28	1.2
	1.1 GENESE(		1.1 GENESE(					
2.	New isolated PRO nucleic acid, useful for diagnosing, treating, or preventing, e.g. systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis, systemic sclerosis, psoriasis, allergic disease, or asthma.	2.1 GENESEQ | *link* 2.2 GENESEQ | *link* 2.3 GPATPRT | *link* 2.4 GPATPRT | *link* 2.5 GPATPRT | *link* 2.6 GPATPRT | *link* 2.7 GPATPRT | *link* 2.8 GPATPRT | *link*	GENENTECH INC.	WO20050019258-1490	100.00	642	Claim 8; SEQ ID NO 1490; 158pp; English.	2.1
				WO20050016962-1490	100.00	642	Claim 8; SEQ ID NO 1490; 158pp; English.	2.2
				EP2014675-1490	100.00	642	probable disclosure (not found by automated parsing)	2.3
				US20070184444-1490	100.00	642	claim: 1; 2; 8; 19; 20; 21; 22; 23; 25; 26; 27	2.4
				US20110245090-1490	100.00	642	claim: 20; 23; 28; 31; 36; 38; 40	2.5
				WO2005016962-1490	100.00	642	claim: 2; 8; 19; 20; 21; 22; 23; 25; 26	2.6
				WO2005019258-1490	100.00	642	claim: 1; 2; 7; 19; 20; 21; 22; 23; 25; 26; 27	2.7
				EP2182006-1490	100.00	642	claim: 1; 3; 4; 5; 6; 7; 9; 10; 11	2.8
	2.1 GENESE(		2.1 GENESE(					
3.	Improved yeast cellular assay system for detecting e.g. polycyclic aromatic hydrocarbons, dioxins, PCBs comprising additional cellular proteins from the AHR signaling pathway which improve the sensitivity and maximal	3.1 GENESEQ | *link* 3.2 GENESEQ | *link*	WISCONSIN ALUMNI RES FOUND.		97.04	642	Example; Fig 6C; 26pp; English.	3.1
					99.84	642	Example; Fig 6C; 26pp; English.	3.2

Summarizing Sequence Locations in a Patent Family [Patents PT-8]

How did we create this?

SEARCHES:

- GenomeQuest – sequence search from GQPAT and/or GENESEQ.

BIZINT SMART CHARTS FOR PATENTS

1. Import each set of results into BizInt Smart Charts for Patents to create .chp files.
2. Use the **File | Combine** feature to create a single combined chart. Run the Identify Common Patent Families tool.
3. Display the Seq. ID Number, Percentage Identity, Subject Length, and Patent Sequence Location columns (**View | Columns**).
4. Change the title on the Patent Sequence Location column to "Location" (by double-clicking on the column header). Repeat to rename the Seq. ID Number column to "Sequence ID", Percentage Identity to "% Identity", and Subject Length to "Length". *You need to give columns short names now, because you can't change them after creating the subtable in step 8, below.*
5. Save the .chp file.
6. Send the chart to Reference Rows using the toolbar icon (or **File | Send to Reference Rows**).

BIZINT SMART CHARTS REFERENCE ROWS

7. During the Create Reference Rows wizard, make GENESEQ the top-ranked database (if present). *This will cause the Derwent enhanced titles and normalized assignee to be preferred in the final report. This is faster than changing the database ranking in the columns one at a time*
8. When the wizard finishes, run **Tools | Create Subtable from Columns**.
9. Select the Sequence ID, % Identity, Length, and Location columns in that order and press Apply. A new column will be created.
10. Double click on the column header to rename the new column from Untitled Subtable to "Sequence Summary".
11. Use **View | Columns** to hide the original Sequence ID, % Identity, Length, and Location columns.
12. Use **Reference Rows | Column Rules...** to set the rule for the new column to Summarize All Values.
13. Use **File | Export** to export the report to HTML and see the subtable column summarizing the sequence data.

Software for
Business Intelligence
BizInt Smart Charts

Filtering the Patent Family by Authorities
[Patents PT-3]

In this example, we've used VantagePoint - Smart Charts Edition to filter the Patent Family subtable to include only US, EP, and WO patents. We've also filtered the patent numbers in the Patent Family to create a new column with only Chinese patents.

PATENT FAMILY FILTERED BY AUTHORITIES (US/EP/WO)

PATENT NUMBERS FILTERED BY AUTHORITIES (CN)

	Database	Patent Family			Patent Family:US/EP/WO			Chinese Patents
		Patent	Kind	Date	Patent	Kind	Date	
20	FAMPAT	WO 2006127537	A2	20061130	EP 1888785	A2	20080220	CN 101501214
		CA 2609214	A1	20061130	EP 1888785	A4	20100203	
		US 20070037186	A1	20070215	US 20070037186	A1	20070215	
		EP 1888785	A2	20080220	WO 2006127537	A2	20061130	
		MX 2007014618	A	20080422	WO 2006127537	A3	20090416	
		JP 2008545400	A	20081218				
		WO 2006127537	A3	20090416				
		CN 101501214	A	20090805				
		EP 1888785	A4	20100203				
		BR 200610012	A2	20100518				
21	GQPAT Nucleotides	US20070037186		20070215	EP1888785			CN101501214
		BRPI0610012			US20070037186		20070215	
		CA2609214			WO2006127537			
		CN101501214						
		EP1888785						
		JP2008545400						
		MX2007014618						
		WO2006127537						
22	GQPAT Nucleotides	JP2008545400		20081218	EP1888785			CN101501214
		BRPI0610012			US20070037186			
		CA2609214			WO2006127537			
		CN101501214						
		EP1888785						
		MX2007014618						
		US20070037186						
		WO2006127537						
23	GQPAT Nucleotides	EP1888785		20080220	EP1888785		20080220	CN101501214
		BRPI0610012			US20070037186			
		CA2609214			WO2006127537			
		CN101501214						
		JP2008545400						
		MX2007014618						
		US20070037186						
		WO2006127537						

Software for
Business Intelligence

BizInt Smart Charts

Filtering the Patent Family by Authorities [Patents PT-3]

How did we create this?

SEARCHES:

- Sequence and patent data from GenomeQuest (GeneSeq and GQPAT) and FAMPAT on Orbit.com

BIZINT SMART CHARTS FOR PATENTS (VERSION 3.5.6 OR HIGHER)

1. Import your results into BizInt Smart Charts to create a .chp file.

2. Use the **File | Combine** feature to create a single combined chart. Run the Identify Common Patent Families tool and sort by the Common Patent Family field. Display the Patent Family column.

3. Use **File | Export** to export the chart as Vantage Point - Smart Charts Edition (creates a .vpxi file.)

VANTAGEPOINT - SMART CHARTS EDITION

4. If the file does not launch automatically in VantagePoint - Smart Charts Edition (VP-SCE), then import the .vpxi file into VP-SCE using **Import from BizInt Smart Charts**, under the Home tab.

5. **Filter the Patent Family by Authorities:** On the Summary panel, double-click on Patent Family.

6. Under the **Home** tab, click on **Find** to bring up the Find panel. Click the Clear Selection button to clear any existing selections.

7. In the Find panel, check the Use regular expression box. Type the regular expression "^(US|EP|WO)" into the Find what: box and click the Select All button. All of the US, EP, and WO patent numbers should be selected in the Patent Family window.

8. Click the Add to Group button and type "US/EP/WO" in the group field and click OK. A new column with check marks next to the selected subrows will be created. Click Close to close the Find panel.

9. Under the **Refine** tab, click on **Create Field From | Group Items**. Select the US/EP/WO item under Patent Family, and click OK. *This is your filtered patent family.*

10. **Filter Patent Numbers by Authorities:** If you want a column in your report with patent numbers from selected authorities, return to the Summary view and right-click on the Patent sub-field under the Patent Family field. Choose **Copy Field...** and enter a new field name (such as "All Patent Numbers").

11. Under the **Home** tab, click on **Find** to bring up the Find panel. Click the Clear Selection button to clear any existing selections.

12. In the Find panel, check the Use regular expression box if unchecked. Type the regular expression "^CN" into the Find what: box and click the Select All button. All of the CN patent numbers should be selected in the All Patent Numbers window.

13. Click the Add to Group button and type "CN" in the group field. A new column with check marks next to the selected subrows will be created. Click Close to close the Find panel.

14. Under the **Refine** tab, click on **Create Field From | Group Items**. Select the CN item under All Patent Numbers, enter "Chinese Patents" as the New Field Name and click OK.

15. **Export new fields to BizInt Smart Charts:** Under the **Report** tab click on **Export to BizInt Smart Charts.** In the Export Wizard, click Deselect All and then click the checkboxes to select your new filtered fields. Click OK to export.
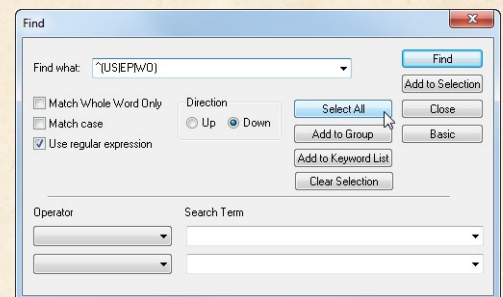
BIZINT SMART CHARTS FOR PATENTS

16. When BizInt Smart Charts for Patents comes to the foreground, select the original chart. It will redisplay with your new filtered patent family and Chinese patent numbers as new columns.

Technique: Using List Cleanup

VP-SCE offers multiple ways to clean up data. The primary tool is List Cleanup found under the Refine tab. The majority of the recipes in this Cookbook include a List Cleanup step. Here we show some options for cleaning company names. The primary difference when cleaning up different fields is your choice of matching ruleset.

Technique: Using List Cleanup

How did we create this?

SEARCHES:

- List Cleanup can be used with any type of data. Here we clean up company names from a drug pipeline search. The same techniques can be used to cleanup assignees (or other fields) in patent data.

BIZINT SMART CHARTS DRUG DEVELOPMENT SUITE (OR PATENTS)

1. Open your chart file if it is not already open.
2. Use **View | Columns** to display the company name field you want to clean up.
3. Use **File | Export** to export as VantagePoint - Smart Charts Edition. VP-SCE should open and import the file.

VANTAGEPOINT - BIZINT SMART CHARTS EDITION

4. When import completes you will see the Summary sheet. Double-click on the company field you wish to clean.
5. **Quick clean up:** Under the **Refine** tab, select **Further Processing | Read Until First | Paren**. This removes parenthetical information to allow List Cleanup to perform better.
6. Click on **List Cleanup**. *You will be prompted to select a matching ruleset to automatically match terms and then go to the Cleanup Confirmation panel to complete the cleanup.*
7. **Choose a matching ruleset:** In this case we'll choose Organization Names (Pharma) and click **OK**.
8. **Display options:** There are different display options available from the drop-down for each of the dual views. The initial display shows **All Items** including what the matching ruleset has combined. Items that are collapsed initially indicate a stronger match, whereas expanded items don't match as closely and should be reviewed. **Combined Items** shows items that have been combined and **Custom Set of Items** is only used with the Find Close Matches feature.
9. **Dual views:** The left hand panel is sorted by record count and the right is sorted alphabetically. You can filter either list by typing into the filter box. Changes made to either of the two views will update both.
10. **Review combined items:** As you review the combined items if you see a group where none of the items should be grouped together, click on the top level item and click the Delete key. *This doesn't delete the individual terms*. To remove a single item from a group, select it and click the Delete key. *The term still exists in the data, but won't be renamed in the cleaned field*.
11. **Drag and drop:** Combine groups by dragging one group onto another. Or drag a term from one group to another.
12. **Rename a term:** Right-click on a term and choose **Rename Item**. Or, if you prefer to use a different sub term as the group name, right-click on your preferred term and choose **Make This Item the Set Name**.
13. **Finish Cleanup and Save:** When you have completed cleanup you can click Accept to create the new, cleaned field. *You may save the cleanup as a thesaurus to use with other datasets if you wish*.
14. **Repeat List Cleanup:** You can continue to clean the field if you wish using **List Cleanup** a second time and choosing a different matching ruleset.
15. **Export back to BizInt Smart Charts:** Under the **Report** tab click on **Export to BizInt Smart Charts**. In the Export Wizard, click to Deselect All and check the box for the final cleaned companies field. Click Export.

BIZINT SMART CHARTS DRUG DEVELOPMENT SUITE (OR PATENTS)

16. When BizInt Smart Charts Drug Development Suite (or Patents) comes to the foreground, select the original chart. The cleaned company field will be added as a new column at the end of your chart.

Software for
Business Intelligence

BizInt Smart Charts

Technique: Extracting Year

This Technique shows how easy it is to extract just the year from a date field and export back to BizInt Smart Charts as a new column. In this example we use VP-SCE to add a Priority Year column to our BizInt Smart Charts table.

The same technique can be used with any chart that includes a date column, such as publication date or clinical start or end date.

	Title	Database	Common Family	Probable Assignee	Patent Family			Priority Date	Priority Year
					Patent	Kind	Date		
1	COMPOSITIONS AND METHODS FOR SITE-DIRECTED DNA NICKING AND CLEAVING	GQPAT Gold+ Proteins	WO2016007604		WO2016007604		20160114	2014-07-09	2014
2	COMPOSITIONS AND METHODS FOR SITE-DIRECTED DNA NICKING AND CLEAVING	PatBase	WO2016007604	GEN9 INC	WO 16007604	A1	2016-01-14	2014-07-09	2014
3	COMPOSITIONS AND METHODS FOR SITE-DIRECTED DNA NICKING AND CLEAVING	GQPAT Gold+ Proteins	WO2016007604		WO2016007604		20160114	2014-07-09	2014
4	COMPOSITIONS AND METHODS FOR SITE-DIRECTED DNA NICKING AND CLEAVING	GQPAT Gold+ Proteins	WO2016007604		WO2016007604		20160114	2014-07-09	2014
5	COMPOSITIONS AND METHODS FOR PRODUCING PLANTS RESISTANT TO GLYPHOSATE HERBICIDE	GQPAT Gold+ Proteins	WO2016007347		WO2016007347		20160114	2014-07-11	2014
6	COMPOSITIONS AND METHODS FOR PRODUCING PLANTS RESISTANT TO GLYPHOSATE HERBICIDE	PatBase	WO2016007347	PIONEER HI BRED INT	WO 16007347	A1	2016-01-14	2014-07-11	2014
7	REGULATED GENE EXPRESSION FROM VIRAL VECTORS	GQPAT Gold+ Proteins	WO2016004010		WO2016004010		20160107	2014-07-01	2014
8	REGULATED GENE EXPRESSION FROM VIRAL VECTORS	GQPAT Gold+ Proteins	WO2016004010		WO2016004010		20160107	2014-07-01	2014

Software for
Business Intelligence

BizInt Smart Charts

Technique: Extracting Year

How did we create this?

SEARCHES:

- You can use any export from BizInt Smart Charts for Patents or the Drug Development Suite that includes a date column. Here we will use Priority Date For this particular example, we used the CAS9 chart file with combined data from GenomeQuest, Patbase, and Derwent WPI used in our PT-12 recipe earlier in this Cookbook.

BIZINT SMART CHARTS FOR PATENTS

1. Open your chart file if it is not already open.
2. If the "Priority Date" column is not displayed in your chart, use View | Columns to add it.
3. Use File | Export to export the chart as VantagePoint - Smart Charts Edition (creates a .vpxi file.)

VANTAGEPOINT - SMART CHARTS EDITION

4. If the file does not launch automatically in VantagePoint - Smart Charts Edition (VP-SCE), then import the .vpxi file into VP-SCE using Import from BizInt Smart Charts under the Home tab.
5. Right-click on the Priority Date field and apply Further Processing | Dates | Extract Years.
6. **Export back to BizInt Smart Charts:** Under the Report tab click on Export to BizInt Smart Charts. Click Export.

BIZINT SMART CHARTS FOR PATENTS

7. When BizInt Smart Charts for Patents comes to the foreground, select your original chart. The new year field will be added as a new column at the end of your chart.
8. Double-click on the title of the new column to go to Column Properties to rename it "Priority Year."

Software for
Business Intelligence

BizInt Smart Charts

Technique: Working with Thesauri

VP-SCE offers multiple ways to create and work with thesauri. In many recipes you will notice that when you clean a field with List Cleanup, you have the option to save as a thesaurus. You can also edit and create thesauri directly with the Thesaurus Editor. Here we will show how it can be used to edit a Drug Development Thesaurus to provide a custom sort in Excel or in VP-SCE visualizations.
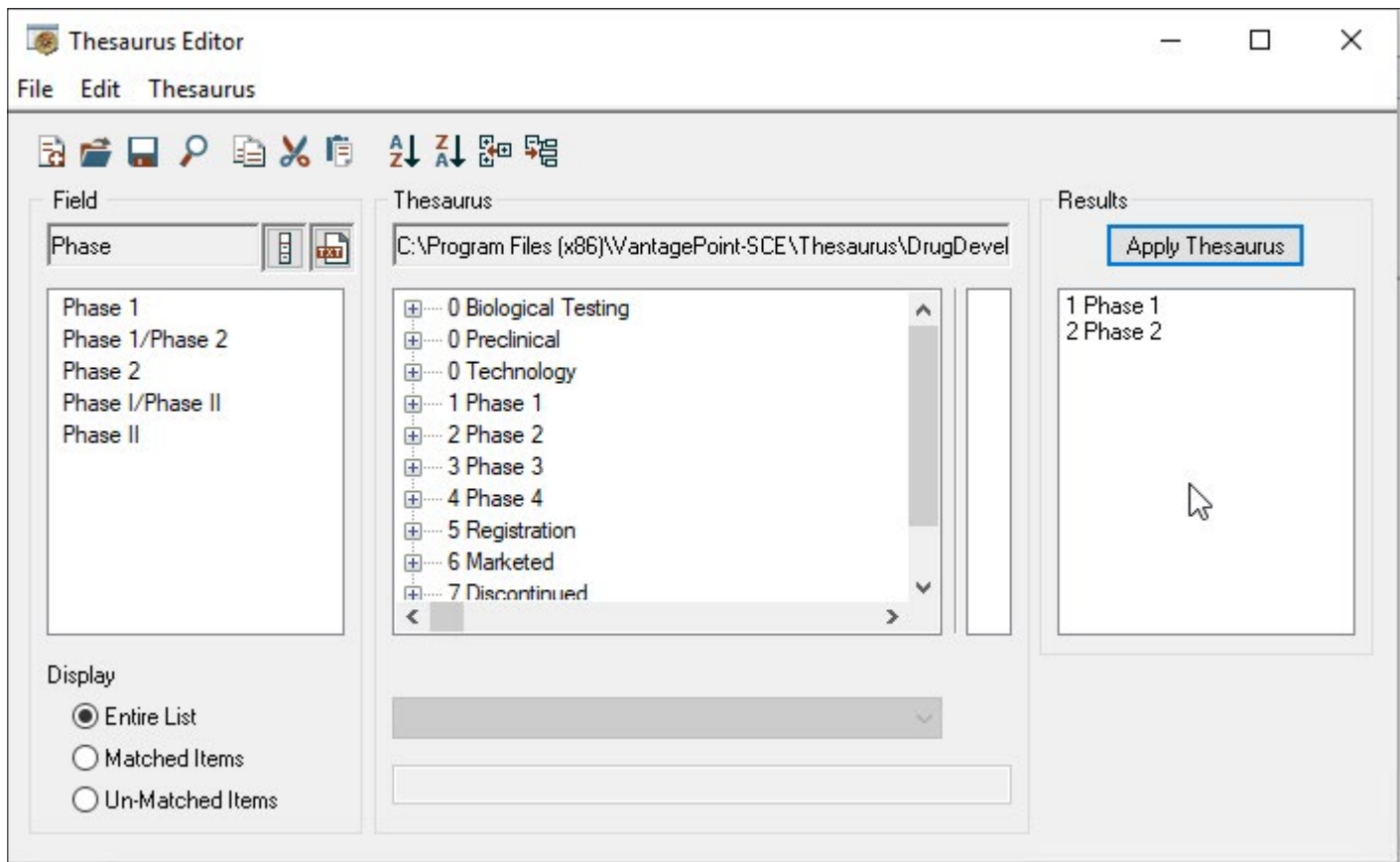
Technique: Working with Thesauri

How did we create this?

SEARCHES:

- We used the DrugDevelopmentPhases_ExcelSort.the as a starting point to create a thesaurus that will sort the way we want. We could also create a thesaurus from scratch, or edit a thesaurus created as part of List Cleanup.

VANTAGEPOINT - BIZINT SMART CHARTS EDITION

1. Go to the **Editors** tab and click **Thesaurus Editor** to open the Thesaurus Editor in a new window.

2. Use **File | Open Thesaurus** and select the thesaurus you want to work with. We'll use DrugDevelopmentPhases_ExcelSort.the.

3. Before making changes, save under a new name. From the **File** menu select **Save as**.

4. There are two levels in a VP-SCE thesaurus, the Set and Item. The center pane of the Thesaurus Editor shows the structure of the current thesaurus. When you open a thesaurus only the Set names are shown. These are the terms that will replace anything matching the associated Items when the thesaurus is used. To expand an item and view the Items associated with a Set you can click the plus sign.

5. **Combine Phases:** Drag one Set term under another one to consolidate them. Here we drag "0 Biological Testing" onto "0 Preclinical" so that records with either preclinical or biological testing as the phase will be treated as preclinical in our custom thesaurus.

6. **Change Phase Sorting:** Right-click on the Top Level Item for the phase you want to change and select **Rename Item**. Then change the numbering to reflect your preferred sort order. We will right-click on "7 Discontinued" and change the 7 to a 0, moving the Discontinued items from the bottom to the top of the list.

7. You can use outline numbering convention of 1, 1.1, 1.1.1, and so on to have phases sort between the existing numbers. Because of the text descriptions, phases will sort alphanumerically, so standard numbering 10 and above will not sort as expected.

8. **Add a New Phase:** To add a new phase, from the **Thesaurus** menu select **Create New Set** and enter the preferred term that should replace the existing text when the thesaurus is run. For this thesaurus, we use a number followed by a text description for the added phase. Here we want to add Phase 1/2 as a separate phase, "1 Phase 1/2".

9. **Check the sort:** Right-click on your new Top Level Item and select **Re-Sort Thesaurus | Ascending**. To change the sort, change the number.

10. **Move the Phase 1/2 terms:** Use ctrl-click to select all the Phase 1/2 Items from the "1 Phase 1" Set and drag and drop them to your new "1 Phase 1/2" Set.

11. You can also add Items to the new phase by right-clicking on the new Set and selecting **Insert New Item**. The drop-down box controls whether you want to match exactly or search within the existing term "contains."

12. Items use regular expressions and can be quite complex, but can also be literal matches. For literal matches begin the Item with a carat (^) and end with a dollar sign ($).

13. **Check your thesaurus:** Use **File | Load Field** and select a field with matching terms, here choose the Phase field and click the Apply Thesaurus button.

14. You'll see the results of applying the thesaurus in the right-hand column, You can also change the radio button on the left-hand side to show only Un-Matched or Matched terms from the field you loaded. This allows you to see what you want to add to your thesaurus.

15. **Save your thesaurus:** Go to **File | Save Thesaurus**. *You can now use your edited thesaurus with recipes in this Cookbook instead of the standard phase thesaurus.*

Software for
Business Intelligence

BizInt Smart Charts

Technique: Extracting Keywords

You may find yourself wanting to extract a set of keywords from a field. For example, you might want to visualize Endpoint Classifications in search results from Citeline TrialTrove. TrialTrove provides this information as part of the "Study Design" column, in one of two different formats. This Technique shows how to extract a set of keywords from a field in VP-SCE using the My Keywords feature.
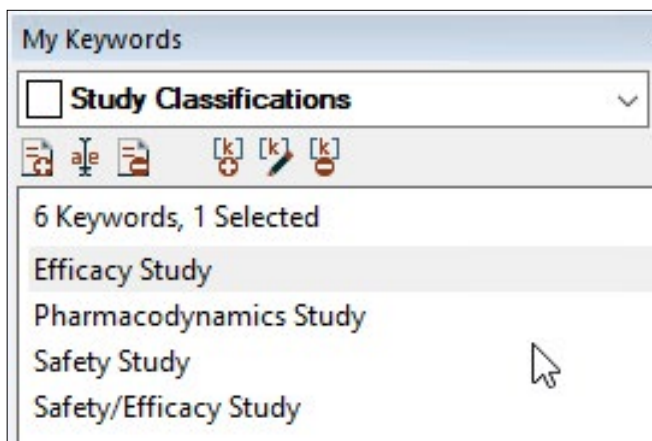
FIGURE 1. THE MY KEYWORDS PANEL

FIGURE 2. KEYWORDS EXTRACTED
USING SIMPLE KEYWORDS

FIGURE 3. KEYWORDS EXTRACTED USING
REGULAR EXPRESSIONS

Technique: Extracting Keywords

How did we create this?

SEARCHES:

- For this particular example, we used an export from Citeline TrialTrove. If you would like to follow along, you can import any results from TrialTrove into BizInt Smart Charts Drug Development Suite, make the Study Design column visible, and then export to VP-SCE.

BIZINT SMART CHARTS DRUG DEVELOPMENT SUITE
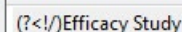
1. Import the search results into BizInt Smart Charts to create .cht files.
2. Display the "Study Design" column.
3. Use **File | Export** to export the chart as VantagePoint - Smart Charts Edition (creates a .vpxi file.)

VANTAGEPOINT - SMART CHARTS EDITION

4. If the file does not launch automatically in VantagePoint - Smart Charts Edition (VP-SCE), then import the .vpxi file into VP-SCE using **Import from BizInt Smart Charts** under the **Home** tab.
5. Double-click on the Study Design field.
6. Make sure the My Keywords panel is visible. Look on the **View** tab and make sure **My Keywords** is selected (in blue). The My Keywords panel is shown in Figure 1 (on the previous page.)
7. **Create your keywords list:** Click on the blue-red gear icon and choose Create New Keyword List. Click the +kw button or press Insert on your keyboard to insert a new keyword. Type the phrase you would like to extract from the raw data. In the example in Figure 1, you can see that we have entered some of the Endpoint Classifications that we see in our data.
8. **Extract keywords:** To extract your keywords from a field, bring up the Manage Keywords menu (under the blue-red gear icon), and hold your mouse over **Extract My Keywords From | Study Design**. (Move your mouse over Extract My Keywords From and then move the mouse to the field name that you want to extract from). This will create a new field with only the extracted keywords.
9. **Export the extracted keywords back to BizInt Smart Charts:** Under the **Report** tab click on **Export to BizInt Smart Charts**. In the Export Wizard, the Study Design : Study Classifications field should already be checked. Click Export.
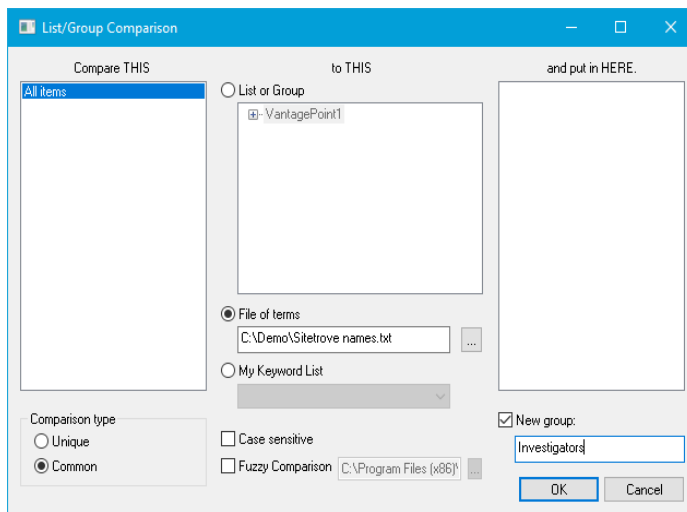
GOING FURTHER

10. Keywords can be case sensitive or not. You can control whether a set of keywords is case sensitive on the Manage Keywords menu (under the blue-red gear icon), and then follow to the **Keyword Options** sub-menu. If you look at the Figure 2 (on the previous page), the first row lists "safety study" in all lower case, but our standard keyword is used in the extracted column.
11. My Keywords can do more than simply match phrases. For example, if you look more carefully at the example in Figure 2, you will see that we are extracting both "Safety/Efficacy Study" and "Efficacy Study" from records where the original text contains "Safety/Efficacy Study". We can use a regular expression to ensure that we only match the exact term "Efficacy Study".
12. To use a regular expression, make sure that there is a check mark next to **Regular Expressions** on the **Keyword Options** sub-menu. Then modify your keyword to include the necessary regular expression code. In this case, we want to only match when the character before "Efficacy" is not a "/". You can see the regular expression we used in the keyword on the right

    `(?<!/)Efficacy Study`

13. When we Extract My Keywords from the Study Design field again, you see in Figure 3 that we now only extract "Efficacy Study" when that exact text is present, and not when the classification is "Safety/Efficacy Study".
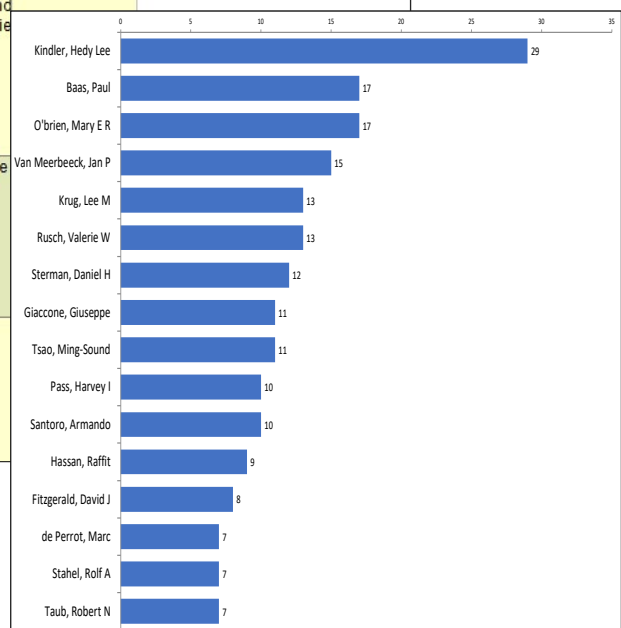
Software for
Business Intelligence

BizInt Smart Charts

Technique: List Comparison

You can use VP-SCE to identify items that appear on two different lists. One list needs to be a field in your current dataset. The other can be a field in any open dataset or a text file. Here we show an example useful for identifying Key Opinion Leaders (KOLs). We will compare authors from a literature search in PubMed to investigator names from relevant clinical trials in Sitetrove. This technique can also be used with any list, such as a speaker list from a conference.

Software for
Business Intelligence

BizInt Smart Charts

Technique: List Comparison

How did we create this?

SEARCHES:

- PubMed: Search for mesothelioma limited to clinical trials. *You can use any supported source with a column for individual names.*
- Names for comparison: You will need a list of names saved as a text file. *Any source that includes individual names can be used, such as a list of board members or names found searching another database. In this example we use a list of names from a search of Sitetrove. Feel free to contact us for help with formatting.*

BIZINT SMART CHARTS DRUG DEVELOPMENT SUITE (OR PATENTS)

1. **Create your literature chart:** Import the PubMed (or other supported database) search results into BizInt Smart Charts.
2. Ensure that the Author and Publication Date columns are visible in your chart. Use **View | Columns** to add if needed.
3. Use **File | Export** to export the chart as VantagePoint - Smart Charts Edition (creates a .vpxi file.)

VANTAGEPOINT - SMART CHARTS EDITION

4. If the file does not launch automatically in VantagePoint - Smart Charts Edition (VP-SCE), then import the .vpxi file into VP-SCE using **Import from BizInt Smart Charts** under the **Home** tab.
5. Double-click on the Author field to view the list. *You will probably note some author name variations. This should help with the comparison. We will use List Cleanup later.*
6. **Identify authors also found on another list**: Go to the **Refine** tab and choose **List Comparison**. See the facing page for the options used in this example. In the center panel, select the File of terms radio button and click the "..." box to browse to your text file.
7. Enter "Investigators" as the New group name and click OK. *You can repeat steps 6-7 with another list of names and a new group name.*
8. **Cleanup author names:** *After running the comparison we no longer need name variations.* Under the **Refine** tab click on **List Cleanup**. For Matching rulest, use Person Names and click OK. *See the Techniques: List Cleanup recipe for additional tips.*
9. **Extract Publication Year:** Return to the Summary Sheet, right-click on Publication Date and choose **Further Processing | Dates | Extract Years**.
10. **Create Author Year Matrix:** Under the **Analyze** tab, click on Matrix. Use Author (Cleaned) for the Rows and Publication Date: Dates/Extract Years for the Columns.
11. **Sort the matrix:** Double-click on Publication Date: Dates/ Extract Years to sort the matrix by years.
12. **Transfer the data to a spreadsheet:** Select rows to export, say authors with 15 or more publications. Copy and paste into a blank Excel sheet.
13. Switch back to the Authors (Cleaned) list and copy the same rows and paste into the same sheet so that author names line up. *You can now work with the data in Excel: delete unneeded columns, add a sparkline using the year data, and use Conditional Formatting on the columns created by groups in VP-SCE, such as Investigators in this example.*
14. **Export back to BizInt Smart Charts:** Under the **Report** tab click on **Export to BizInt Smart Charts**. In the Export Wizard, click to Deselect All and check the box for the Author (Cleaned) field. Click Export.

BIZINT SMART CHARTS DRUG DEVELOPMENT SUITE (OR PATENTS)

15. When BizInt Smart Charts Drug Development Suite (or Patents) comes to the foreground, select the PubMed chart. The cleaned author field will be added as a new column at the end of your chart.

About This Cookbook

This "Cookbook" illustrates how to create some popular visualizations and reports from patent, clinical trials, drug pipeline, and gene sequence data using tools in the BizInt Smart Charts family.

To use these "recipes", you'll need the following software:

- BizInt Smart Charts for Patents 5.0 or higher
- BizInt Smart Charts Reference Rows 2.0 or higher
- VantagePoint - Smart Charts Edition 11.0 or higher

BIZINT SMART CHARTS TRIAL SOFTWARE

You can download BizInt Smart Charts trial software at: **http://www.bizint.com/try**

You can request a trial of VantagePoint - Smart Charts Edition at **http://www.bizint.com/vpsce** or contact us (email: support@bizint.com) .

BIZINT SMART CHARTS PRODUCT DOCUMENTATION

You'll find more information on how to use BizInt Smart Charts software tools at:
http://www.bizint.com/tips

Note: These recipes are constantly being tested and revised based on user feedback. We suggest that you check the website to see if there is a more recent version of your recipe. If you find any errors or need help, please contact us at "support@bizint.com."

October 2020

Software for
Business Intelligence

BizInt Smart Charts

Download the Software

- **Customer downloads:** *http://www.bizint.com/downloads/*
  You will need to enter the userid and password supplied to your company.
- **Trial downloads:** *http://www.bizint.com/try*
  We encourage you to evaluate BizInt Smart Charts using your own data!

Keep up with the Latest Releases

- **Latest release information web page**: provides details on all releases.
  http://www.bizint.com/product/patents/latest_release.php
  http://www.bizint.com/product/drugdevsuite/latest_release.php
- **Subscribe to release announcements** by email or RSS feeds.
  http://www.bizint.com/support/SubscribeReleases.php

Using the Software

- **Creating Reports from Databases and Hosts web page**: follow the links on the left side of the web page
  for step by step instructions for each database/host.
  http://www.bizint.com/support/create/

Go to **www.bizint.com/tips** for links to these useful documents —

- **Mini Guide:** provides a step-by-step overview of the key features of each BizInt Smart Charts product.
- **Combining and Updating Reports**: a detailed discussion of features and options for creating combined
  and updated reports.
- **Distributing Reports**: a guide to choosing and using the export formats available with BizInt Smart Charts,
  including HTML, Word, Acrobat, and Excel.
- **BizInt Smart Charts Reference Rows Mini Guide:** a guide to using the Reference Rows utility for creating
  single rows combining information from related records.
- **Gene Sequence Reports:** describes the latest features for using BizInt Smart Charts with IP sequence
  databases.
- **Cookbook of Reports and Visualizations:** *www.bizint.com/Cookbook*

Contact Us for Help

- **Email us** at **support@bizint.com**
- **Call us** at +1 (714) 289 1000 (office hours are 8 am - 5 pm Pacific)

**Software for
Business Intelligence**

BizInt Smart Charts

Cookbook of Reports & Visualizations

www.bizint.com/Cookbook

The *Cookbook of Reports & Visualizations* illustrates how to create some popular visualizations and reports from drug pipeline, clinical trials, and patent data using tools in the BizInt Smart Charts family.

The Cookbook provides step-by-step recipes showing how you can create these reports and visualizations with your own data.