

IN-SPIRE Tips and Techniques "At a Glance" Version 3.0

November 21, 2005

At a Glance

Table of Contents

Getting an Overview of a Dataset	
Strategies for Early Use of Spatial Patterns	4
Read Similar Documents Together	
Remove Less Interesting Clumps	
Find Relevant Information (a)	7
Find Relevant Information (b)	
Keep Track of What You Find	9
Inspecting Non-hits Among the Hits	
Focus in on What You've Found	11
Seeing a Time Pattern	12
Seeing Themes or Groups over Time	13
Seeing Correlation among Groups (updated for 3.0)	14
Reuse Queries	
Reusable Triage Strategies (new for 3.0)	
Manage Evidence and Hypotheses (updated for 3.0)	17
Summarize Evidence (new for 3.0)	
Removing Unhelpful Themes	19
Finding Out Who's Saying What	
Finding More Documents Like This One	21
Finding a Document I Know About	
Handling "Squished" Datasets	23
My Data Has Disappeared	24

Getting an Overview of a Dataset

IN-SPIRE visual tools can help give you a quick overview of a dataset. This process helps jumpstart an analysis.

Documents that are close together should have similar content. Both the Galaxy and ThemeView show the large and small concentrations of documents; both have the same spatial arrangement.

Step by Step

 $\begin{pmatrix} 1 \end{pmatrix}$

 $\left(2\right)$

 $\left(3\right)$

(4)

(5)

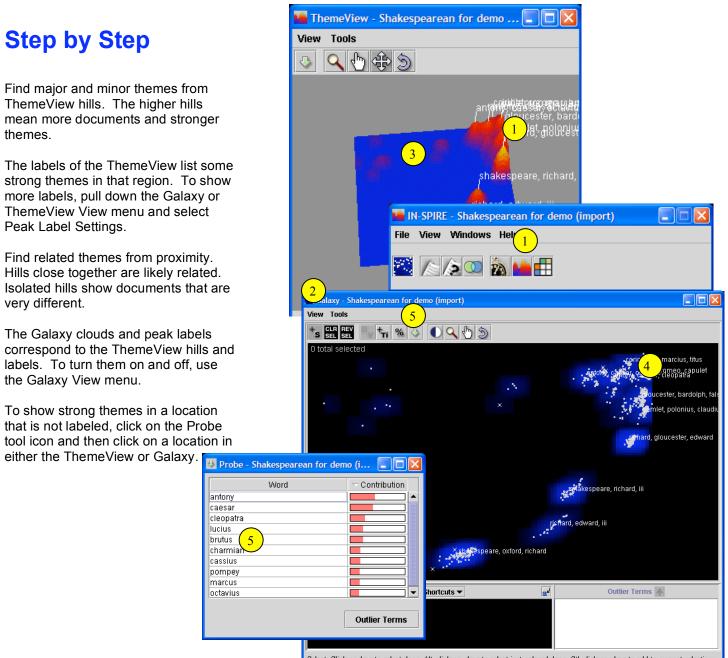
Find major and minor themes from ThemeView hills. The higher hills mean more documents and stronger themes.

The labels of the ThemeView list some strong themes in that region. To show more labels, pull down the Galaxy or ThemeView View menu and select Peak Label Settings.

Find related themes from proximity. Hills close together are likely related. Isolated hills show documents that are very different.

The Galaxy clouds and peak labels correspond to the ThemeView hills and labels. To turn them on and off, use the Galaxy View menu.

To show strong themes in a location that is not labeled, click on the Probe tool icon and then click on a location in



Select: Click or drag to select docs; Alt-click or -drag to select just colored docs; Ctl-click or -drag to add to current selection; Ctl-shift-click or -drag to remove from selection.

Strategies for Early Use of Spatial Patterns

The spatial layout and groupings in the ThemeView and Galaxy can show useful information up front.

(3)

Analysts have also found that the spatial patterns can raise useful questions that lead to profitable exploration.

Step by Step

What Can the Spatial Patterns Tell Me?

Open the ThemeView and rotate the theme to the best angle. The highest peaks show areas with lots of relatively similar documents.

2 Peaks close together suggest themes that are related. Lower hills show groups with fewer members or less overall content. The labels show the strongest discriminating themes in these areas.

Look at the Galaxy. Isolated clumps of documents may show outliers – documents that are very different from the bulk of the set.

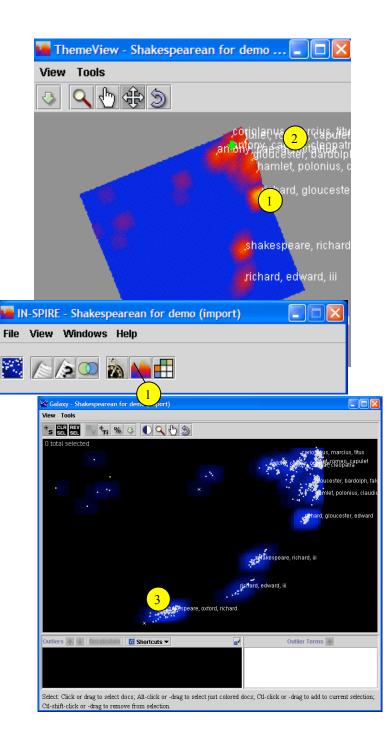
What Questions Can the Spatial Patterns Raise?

Do the patterns of document concentration and outlier match expectations? If not, why?

Are there themes grouped in the same or neighboring peaks that make an unusual combination?

As you get a sense of the main groupings and their themes, look for what's missing. Is there a major theme you expected but don't see? Do a query to find if it's in the set but dispersed.

Does the same theme appear in widely separated labels? This could mean a difference in the related or secondary themes.



4

Read **Similar Documents Together**

An easy way to get started using IN-SPIRE is to use the clumps of similar documents in IN-SPIRE to help organize your reading strategy.

Step by Step

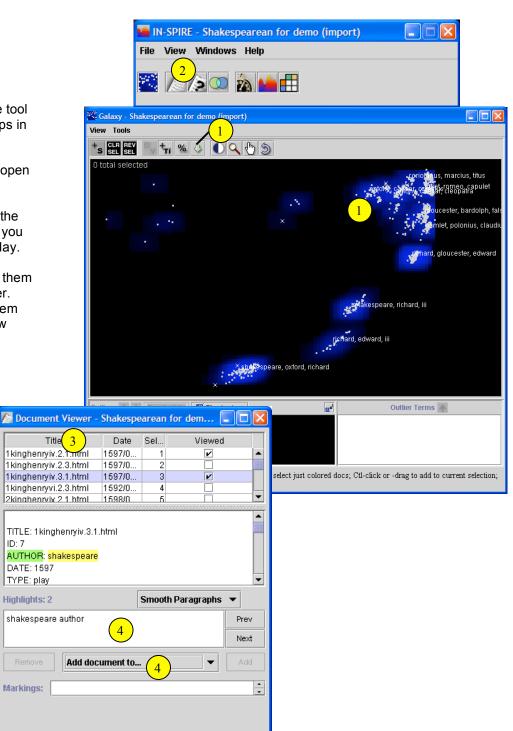
Use the peak labels and the probe tool to get a sense of the various clumps in the Galaxy.

Select a clump of documents and open 2 the Document Viewer.

Click on a column heading to sort the 3 documents by title or date. When you click a title, the document will display.

To highlight particular words, type them 4 in the Highlights pane and hit Enter. To save useful documents, add them to an existing group or make a new one.

ID: 7



Remove Less Interesting Clumps

IN-SPIRE shows the differences and similarities among documents in the dataset. The initial clumps help separate out less relevant documents. Once these are identified, you may want to push them aside to see differences among the remaining documents. You can save Shortcuts to these different views.

Step by Step

2

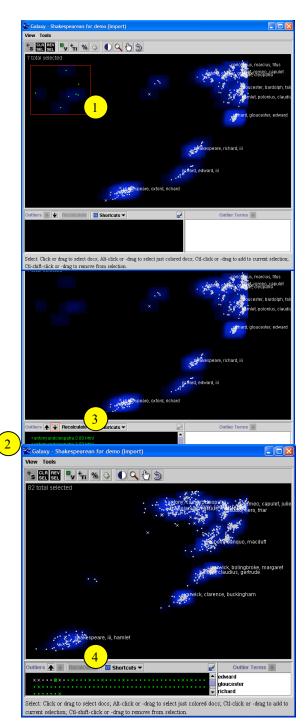
Select a clump of documents that you want to push aside. Click the Outliers down arrow.

The documents will move to the lower pane. Move other clumps down as desired.

3 Click the Recalculate button. IN-SPIRE will rearrange the Galaxy to reflect the similarities and differences among the remaining documents. Usually new themes will show up in the labels.

4 The documents in the lower window can still be selected and read and will still respond to queries. To move some back up to the Galaxy, select them and click the Outliers up arrow.

• You can drag the dividing line to give more space to either the Galaxy or the lower pane.



Find Relevant Information (a)

IN-SPIRE provides several capabilities for finding useful information:

- a query based on word presence or absence in a document
- a query based on thematic similarity
- clusters based on overall document similarity.

Used together, these form a powerful set.

Step by Step

(1)

(2)

 $\left(\begin{array}{c} 3 \end{array} \right)$

 $\left(4\right)$

Open the Query tool. To find documents containing specific words, use the Words in Document query.

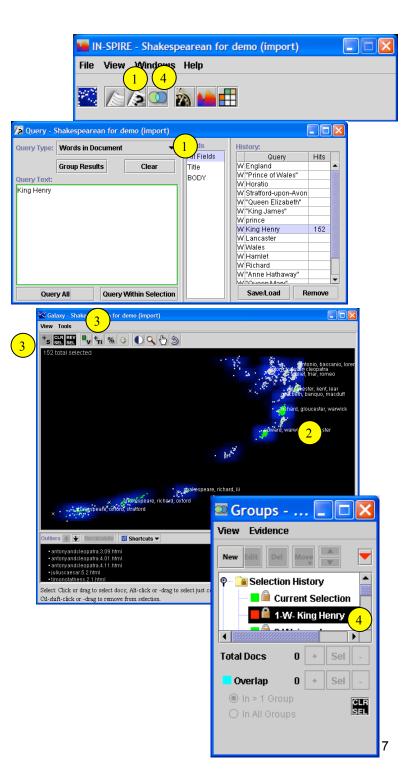
- Word queries support Boolean AND, OR, NOT and allow parentheses.
- The query is case-independent.
- Use quote marks to search for an exact phrase like "white house"
- Use * as a trailing wild card, e.g., chair* will match chair or chairs
- Use * in a phrase, e.g., "walk * dog" will match "walk dog" or "walk the dog". Also "walk *3 dog" will find a phrase with up to 3 words between "walk" and "dog"
- Words with spaces between are interpreted as a Boolean AND.

Usually the hits fall in various clumps, representing different contexts.

- A larger concentration near related labels may be documents most related to the search terms.
- Scattered hits near seemingly unrelated labels may mean a chance mention of the search terms.

Read the hits by clump to read similar documents together. Click the Select tool. Then hold down the Alt key and drag the mouse over the clump you want. Only those documents will stay selected. Look at the documents in the document viewer. To focus the Doc Viewer on a particular selected dot, click the View Selected tool. Center the cursor box over the dot and click. The corresponding document will show in the document viewer and the selection set will not change.

To see the full query result again, open the group tool. The Selection History folder has up to 3 recent queries. Click the 1st one and the hits show in the Galaxy.



Finding Relevant Information (b)

IN-SPIRE provides several capabilities for finding useful information:

- a query based on word presence or absence in a document
- a query based on thematic similarity
- clusters based on overall document similarity.

Used together, these form a powerful set.

Step by Step

(1

To find documents related to themes, use Query by Example. Type words in the query window or paste portions from other documents (use ctrl-c to copy and ctrl-v to paste). The words highlighted in blue will be used to calculate similarity.

- To make a word have a stronger influence on the result, repeat it multiple times.
- Query by Example uses theme relationships determined by the dataset you are using.
- Once the query is run, move the slider to select progressively more documents. The graph shows roughly how the similarity drops off. To decide the cutoff, open the Document Viewer and sort by rank (click the column if needed). Then scroll to the bottom of the list and decide how relevant the last few are. After finding a good cutoff, make a group from the set.

3 Use the Alt-select capability to explore pockets of hits as above.

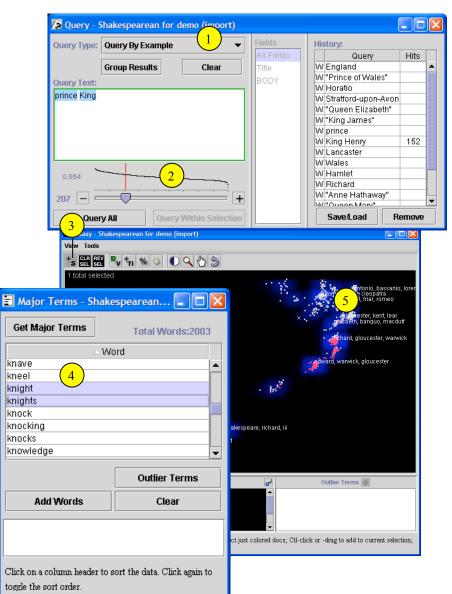
If an important word is not highlighted, you can insert other similar words. Use the Major Terms window for suggestions. It's under the main toolbar Windows menu.

5 The Word and Query by Example usually return somewhat different results. Use them together to find additional useful documents.

 IN-SPIRE - Shake
 arean for demo (import)

 File
 View

 Windows
 Help



Keep Track of What You Find

As you read a document in the document viewer, you may decide you need to keep this document for use later. Here's how.

Step by Step

The document viewer lists the documents currently selected. To read the text of a document, click on its title.

2 To add the document you are reading to an existing group, click the "Add to Group" pull-down, and select the group you want. Click the Add button.

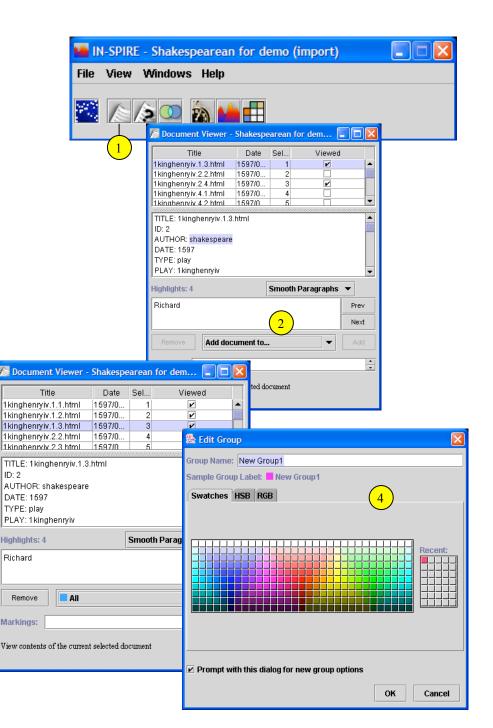
• The pull down will stay on the same group until you change it. As you read more documents, you can add them to that group by just clicking *Add*.

If the document you are reading is already a member of the group shown on the pull down, then the Add button is grayed out. You can remove the document from that group by clicking Remove.

4

To start a new group with this document, select New Group from the pull down and click the *Add* button. Depending on the Settings you have chosen, you may be asked for group options. If so, enter a name for the new group (you can also change the group color here) and click *OK*. If not, your group will use default values.

3



Inspecting **Non-hits** Among the **Hits**

The documents close to your query hits may also be useful, since they should be similar. To see both hits and non-hits together, just select the region. To single out the nearby nonhits, follow these suggestions.

Step by Step

 $\begin{pmatrix} 1 \end{pmatrix}$

4

Make a group from the Query result. Press Clear Selections to see your group in the Galaxy.

Use the Selection tool to drag over and 2 select all documents in the area of interest.

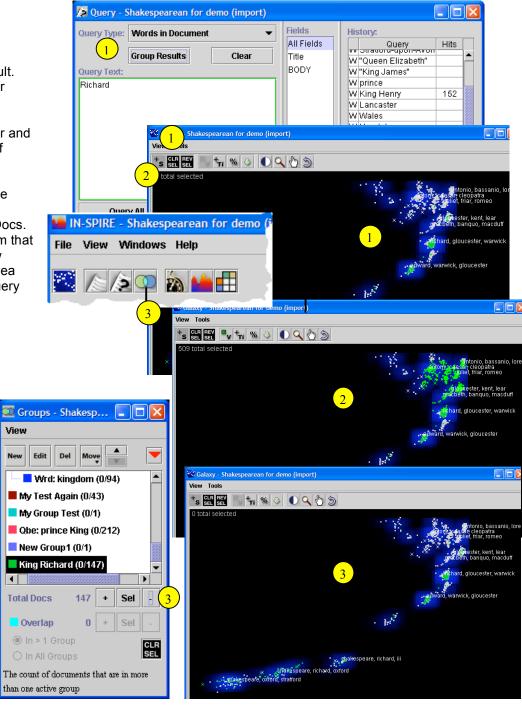
Open the Group tool and select the 3 name of the group you just made. Press the "-" to the right of Total Docs. This deselects the documents from that group. Your selection set will now have only the documents in the area you selected that are not in the query results.

New

.€ [

(3)

Now you can read the selected documents.



Focus in on What You've Found

You can converge to examine a portion of your document collection in greater thematic detail. then bring the rest back into view for a larger context.

Step by Step

To converge on a selection set

- Select the portion you want to focus on, 1 for example, by directly selecting the documents in the Galaxy, or by using the Query tool.
- On the Outlier Shortcut control, choose 2 "Only Selected Documents in Galaxy." All of the other documents will move down to the Outlier area and their influence will be reduced. Your selected documents will show in the Galaxy and the clumps and labels will better reflect those specific documents. To save the view for later use, use the Save As option on the Shortcuts pull-down.
- When you are done, use the Outlier 3 Shortcut control to select either "All Documents in Galaxy" or a previously saved view.

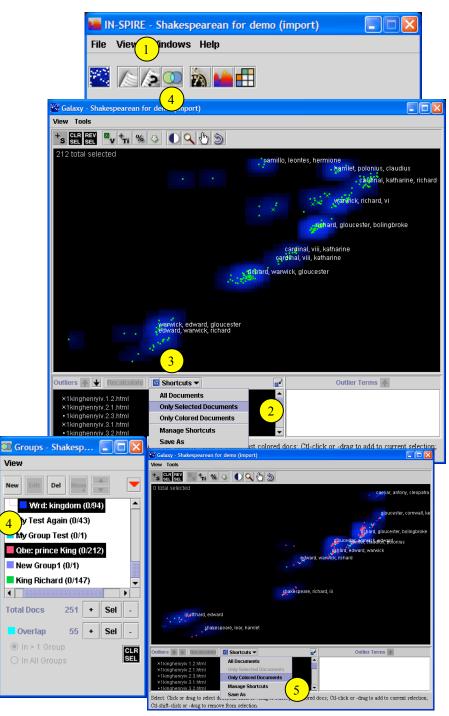
To converge on groups

You can also use this capability to 4 compare documents in two or more groups. First, clear selections. Then open the Group tool and highlight the groups you want to compare. Their colors should show in the Galaxy.

New

Δ

On the Outlier Shortcut control, choose 5 "Only Colored Documents in the Galaxy." The documents in your highlighted groups will show in the Galaxy and the rest will move down to the Outliers area. The clumps and labels should show thematic similarities and differences among the groups.



Seeing a Time Pattern

The Time Slicer lets you see how your document collection changes over time. Documents stay in the same spatial location, but are shown or hidden according to the time period of interest.

Step by Step

1

2

3

4

5

Open the Time Slicer and the ThemeView or Galaxy.

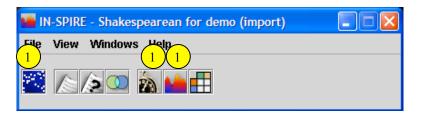
Change the Unit to the granularity of interest. The bars show the number of documents in each time increment. If you hold the cursor over a bar, a pop-up gives more detail. To change the width of the bars, drag one edge of any histogram bar.

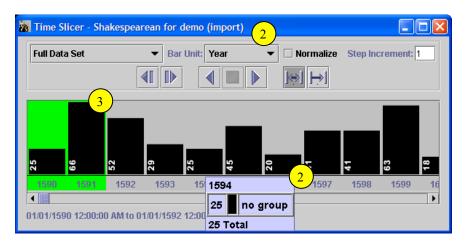
The green highlighted area shows the time period reflected in the ThemeView and Galaxy. To change the width of the green highlight area, drag the edge of the highlight.

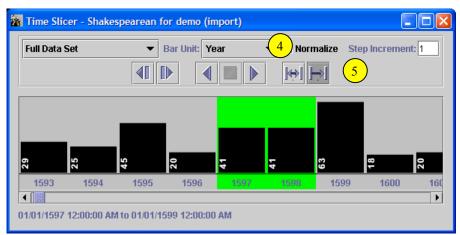
Push the Move button. Step through the time slices, noting the changes in the peaks and labels as they are recalculated to reflect each time slice.

- You can also use the keyboard arrows keys to step forward and back.
- It may help to toggle the forward and back Step buttons a couple of times to examine a change in detail.
- To see a gradual evolution, set the Step Increment to a smaller size than the number of bars in the highlight area.

Click on the first time slice. Click the Grow button. Step through again, watching the cumulative patterns.







Seeing Themes or Groups over Time

In addition to general changes, the time slicer lets you see how individual themes or groups map over time. You can see both the change in amounts, and the change in context.

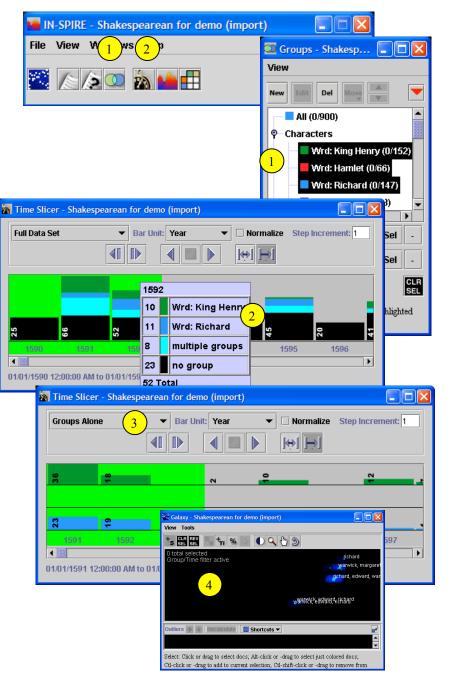
Step by Step

Highlight the groups of interest in the group tool, using SHIFT-click or CTRL-click. Their colors will show in the Galaxy. Documents belonging to more than one group will show as the overlap (aqua) color.

- 2 The time slicer will show the group colors in the histogram bars indicating the number of members per time slice. An additional color (defaults to aqua) shows the number of overlap documents per time slice. Hover the mouse over a bar to see exact counts.
 - It may be helpful to find a trend that stops suddenly. For example, why is there information about this topic for three months, and then no more?

3 To see each group separately, change the pull-down to Groups Alone. Separate histogram sets will appear for each highlighted group, making it easier to see individual patterns. The bars are scaled to show relative proportions among the groups. You can resize the bar width or section heighth by dragging, and they will stay in proportion.

4 The Galaxy will reflect the state of the Time Slicer, showing either the Full Data Set or just the highlighted Groups Alone respectively. This status is also noted in the upper left corner of the Galaxy.



Seeing Correlation Among Groups

You can explore correlations among themes, sources, and time with the Correlation tool. Use it to see overlaps between one set of themes and another set or to see how themes distribute over a list of fields, such as source, type, or author.

Step by Step

(3)

6

Make groups of things you want to explore. In the example, we've made groups from queries on the themes of "economy" and "war" and the sources of London and Paris. Open the Group tool and select the main groups you want to explore; these will be your rows.

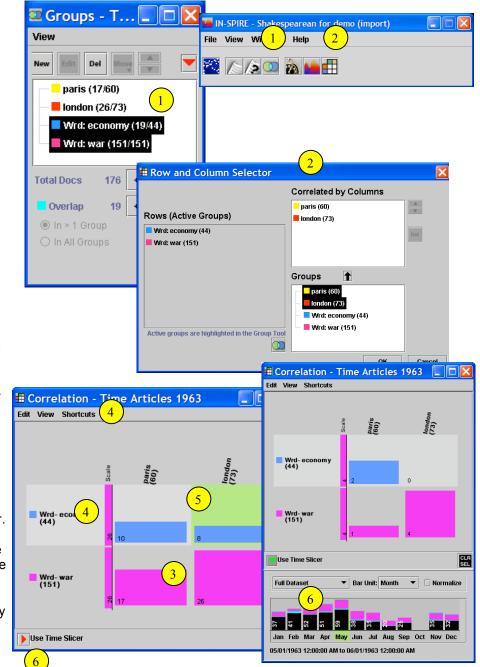
2 Open the Correlation tool. Select Edit/Choose Rows and Columns and set the columns you want to compare to the rows. Click OK.

The Correlation tool will show how the main groups distribute over the columns. Totals are shown for each individual row and column and for the overlaps.

The rows reflect the groups highlighted in the Group tool. If you make a new group from a query, it is automatically highlighted and will show as a new row in the Correlation tool. To save a particular combination of rows and columns, use the Shortcuts menu to Save one.

5 To select any particular overlap set, click the corresponding cell(s). You can read them in the Document Viewer.

To see correlations in time, click on the arrow at the bottom left. This opens the Time Slicer and connects it to the Correlation tool. As you step through time, the correlations shown reflect only the current time window.



Reuse Queries

You may have created certain queries you want to use on more than one dataset. You can save and reuse these queries. Examples might include sources, titles, or topics you often find interesting.

Step by Step

To Save Queries

(1)

On the query tool, click Save/Load.

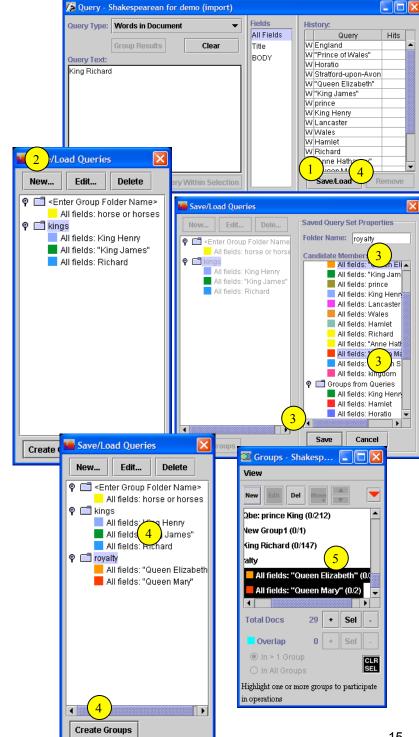
2 A dialog opens. Click New... to make a set of saved queries. The dialog will expand.

3 Type in the name for the query set. The dialog will show the Word queries and Groups you have made on the current dataset. Select the ones you want to save for use in other datasets. Click the Save button.

To Load Saved Queries

On the query tool, click Save/Load. Open the sets you want. You can highlight the set as a whole or only some of the queries it contains. You can also highlight queries from more than one set. Click Create Groups.

IN-SPIRE will make a group for each of the highlighted queries and put them in a folder with the same name as the saved query set. The group colors will be consistent across datasets.

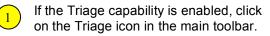


Reusable Triage Strategies

You can set up a network of gueries and query combinations to use over and over again, or to apply to dynamic data automatically. These provide more power than saved queries and let you explore results at different levels of abstraction. In Version 3.0, it is an optional feature that your IN-SPIRE administrator may turn on.

Step by Step

4



2 On the Triage tool, click Manage Available Networks. Select one or more of the networks listed, or type a new name and click New. Click Done.

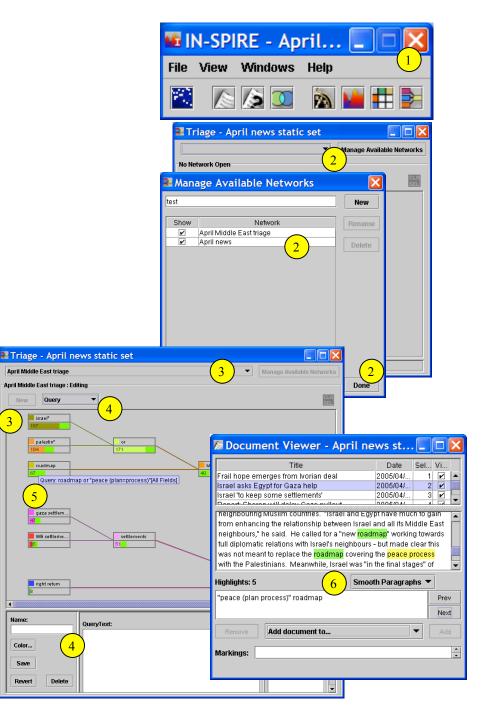
3 The pulldown in the Triage tool now lists the networks you have chosen. Select the one you want, and the query strategy will automatically run, making groups of the results. The colored bar shows how many hits were found in this dataset for each group.

To add new nodes to the strategy, select the type of node you want and click New. The most commonly used are Query and Set Combination.

- Query nodes have the same options as a Word query.
- Set Combination nodes allow group operations as in the Group tool.

5 In this example, basic queries (such as "Israel") are in the left column. Set Combinations are defined to group two or more query results (such as "Israel" or "Palestine"), making different levels of abstraction. To compare results, click any node. The green portions show the overlap between the selected group and others in the network.

6 When a node is selected, the group color shows in the Galaxy, and the group members show in the document viewer, with the query terms highlighted.



Manage Evidence and Hypotheses

You can progressively build up groups of documents relevant to multiple hypotheses, then use the expanded group tool to track your assessments of consistency and diagnostics.

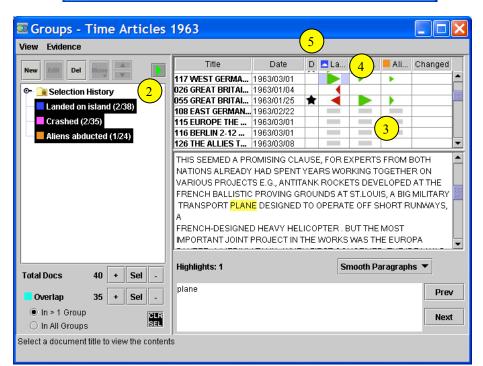
Step by Step

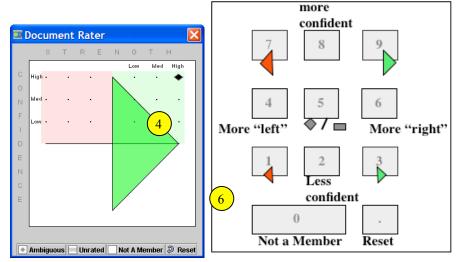
As you explore the dataset, make a group for each hypothesis you are investigating. You can make new groups from the query tool, from the Group tool, and from the Document Viewer.

2 To organize the evidence you have found, open the Group tool. Select the groups corresponding to your hypotheses. Click the arrow in the upper right.

- 3 The expanded tool shows you the documents for the selected groups. A gray rectangle shows which groups contain which documents. To review a document, click its title.
- To rate a document's relevance to a particular hypothesis, click the cell corresponding to that document and that group. The document rater will open. Use the rater to mark this document as consistent (green arrow facing right) or inconsistent (red arrow facing left) with the hypothesis, as ambiguous (gray diamond), as unrated, or even as not relevant (not a member). You can also rate the strength (longer arrows) and confidence (taller arrows) of your judgment.
- 5 The documents that support one hypothesis and contradict another are diagnostic. These show a star in the "D" column. Click on the D to bring these documents to the top.
- 6 To make evidence rating quicker, several hot keys have been defined. Up and down arrows change between documents; right and left arrows change column focus. On an extended keypad, the Num Lock key activates hotkeys on the numeric keypad.

Main-SPIRE - Shakespearean for demo (import)	
File View Windows Help	





Summarize Evidence

IN-SPIRE provides a summary of the ratings you've assigned to the documents in one or more groups. This summary includes the strength and confidence values along with the valence (positive/negative value).

Step by Step

3

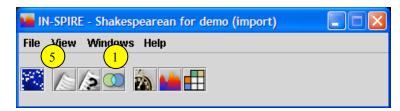
Open the Group tool. Select the groups corresponding to your hypotheses. Click the arrow in the upper right to expand the Evidence pane.

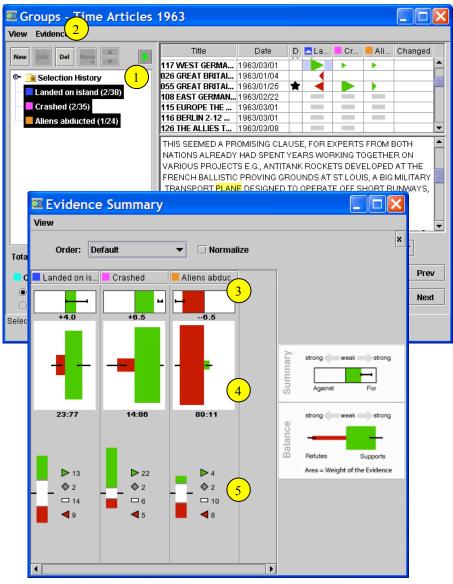
2 To get a summary overview of your ratings, pull down the Evidence menu and select Show Evidence Summary.

The Summary window provides three levels of detail. At the top is a view showing the predominant argument weight of the evidence For or Against each hypothesis, taking into account both strength and confidence ratings. The larger the colored block, the greater the predominance of the evidence.

4 Directly below is a view showing the relative balance of the For and Against evidence, including strength, confidence, and number of documents.

5 The histogram at the bottom shows simple counts of documents assigned to the different evidence values: For, Ambiguous, Unrated, and Against. Click on a portion of the histogram or on the corresponding evidence symbol to select those documents. Open the Document Viewer to review the text.





Removing Unhelpful Themes

Clump labels should help you assess what's in the clump and whether a clump is likely to be relevant or not for your task. The labels also show themes that have influence over the clumping. Sometimes the themes automatically selected don't match well with your interests. You can demote unhelpful themes to remove their influence.

Note that even though you may need finance-related documents, a theme of "finance" is still useful if it helps you identify less interesting clumps.

Step by Step

(1)

 $\left(2 \right)$

 $\left(4\right)$

 $\left(\begin{array}{c} 5 \end{array} \right)$

Get a sense of what's where in the Galaxy and the clumps that are likely relevant or not relevant to your interests and tasks. Identify the themes that are not helpful in making this distinction.

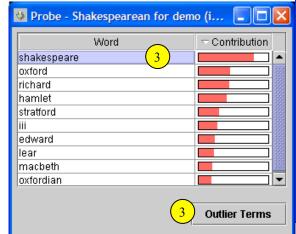
Select the Probe tool and click on a clump with unhelpful themes.

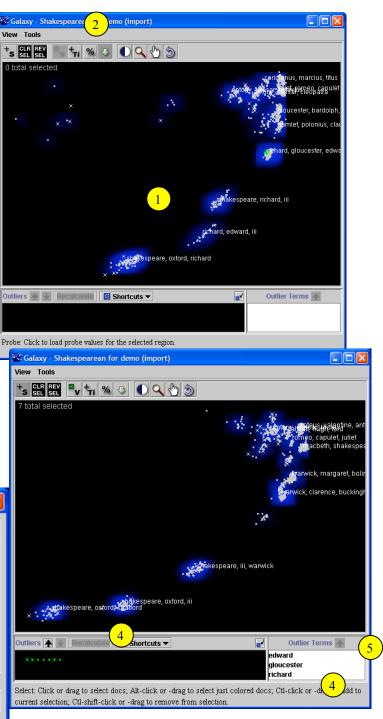
3 On the list of themes, click the ones you want to demote. Click the Outlier Terms button.

The selected themes will show in the Outlier Terms box. Click Recalculate to remove their influence. The Galaxy will change, and the themes will no longer have an influence nor show up as labels.

• If the removed themes were critical to making a clump, that clump may redistribute. If additional themes also strongly influenced the clump, you may only see new labels.

You can put Outlier Terms back by selecting them and clicking the up arrow.





Find Out Who's Saying What

Seeing how documents from different sources are separated or intermingled within the Galaxy can reveal differences in topics or in how they are discussed.

Step by Step

Open the query tool and make sure Words in Document is selected in the drop-down menu.

2 The central pane lists the fields available for querying in this dataset. Select the source field. Type in the guery for the source you want.

> AND, OR, and NOT capabilities can still be used. For example, to find documents from either the Washington Post or Washington Times, type "washington post" or "washington times".

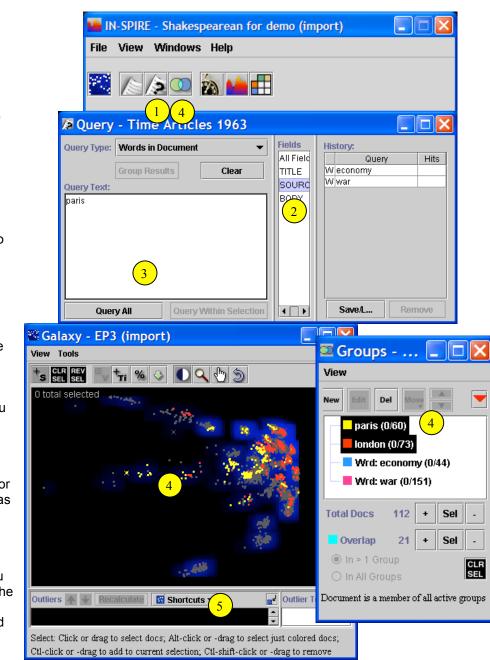
3 Click Query All and then click Group Results. Continue to query and make groups for other sources of interest.

Open the Group tool. Clear any selections and click on the groups you just made. The colors in the Galaxy show which groups are where.

 A clump with many colors means similar documents from multiple sources. A clump with only 1 color indicates that particular source has different documents from other sources.

To focus just on the groups you've made, use the Galaxy Shorcuts menu to show Only Colored Documents. The Galaxy will reflect just those documents and the rest will be moved down to the outliers area.

(5)



Finding More **Documents** Like this One

Once you find a really useful document, there are several ways to look for similar ones.

TIP> You may want to save the useful document so it will be easy to come back to. Make a group from the document, or Copy the full text and save it to a text file in Notepad.

Step by Step

1

 $\left(2 \right)$

4

Documents close by should be similar, and so may also be useful to you.

- Look at the useful document in the document viewer.
- Find where it is in the Galaxy by finding the dot surrounded by the small green square.
- Use the Selection tool and drag the mouse over the nearby documents.

Use the Query by Example on key portions of the Useful document

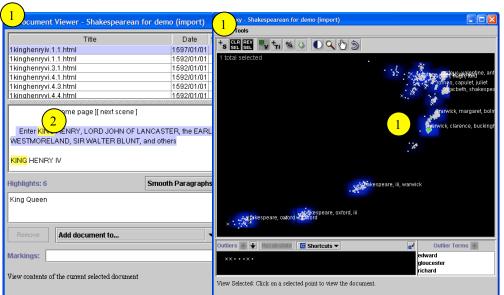
- Use ctl-C to copy key sentences or paragraphs from the useful document
- Open the Query tool and select Query by Example.
- Use ctl-V to paste the text into the • query window. The words that show with a blue background are ones that will be used to calculate similarity.
- Press the Query button. IN-SPIRE will find similar documents and rank them.

Query Text:

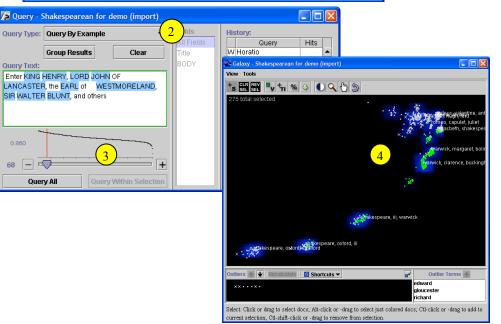
0.860

Move the slider to select progressively 3 more documents: the more that are selected, the less similar they will be to the Useful document. To help decide what cutoff to use, scroll to the bottom of the list in the Document Viewer and decide how relevant the last few are.

Watch the patterns in the Galaxy as the selections grow. There should be pockets of selections. You can make a group from the current set, and then explore each pocket separately.



Maile Selfer - Shakespearean for demo (import) File View Windows Help 2



Finding a Document I Know About

If you know where a favorite document is in the Galaxy space, you can use it as a landmark to find new favorites that may be nearby.

You can use a relatively unique term or field value to find it. If you have read the document that can also help you find it quickly.

Step by Step

By Field or Text

 $\begin{pmatrix} 1 \end{pmatrix}$

2

4

Open the Query tool and select Words in Document. Select the field(s) that you want to search. Type in one or more words that should be relatively unique for the document you want. You do not need to put anything between the words; IN-SPIRE interprets a blank as an AND. Run the query.

• If you want to search entire document contents, use the default of All Fields.

Open the Document Viewer. You can sort by title, date, or Viewed status by clicking on the column heading.

If You've Read the Document

3 Documents you have read in this dataset are marked with an X in the Galaxy. If you remember the general area where the document was, you may be able to see the corresponding X.

If there are a few X's close by, select the documents in that area, open the document viewer, and sort by the Viewed column.

File 2 1 ndows Help		
🔀 Query - Shakespearean for demo (import))	
Query Type: Words in Document	 Fields 	History:
Group Results Clear Query Text: Shakespeare Hamlet	All Fields Title BODY	Query Hits w Other Adaptave W "Queen Mary" W "William Shakesp W "William Shakesp W Innover King Q Prince King Q Richard
		W Prince and king W queen and hamlet W shakespeare 168 W England W queen and king W "Prince of Wales"
Query All Query Within Selection	DN	Q Enter KING HENR W Shakespeare Ha 45 - Save/Load Remove
Document Viewer - Shakespearean for demo (i	mport)	
	iction Viewed	
brown.hamlet.htm 199: TITLE: hamlet.1.2.html ID: 267 AUTHOR: shakespeare DATE: 1601 TYPE: play PLAY: hamlet ACT: 1 Fichelichter 16	3 ×	ontenaista ella valentine, ar pomeo, capulet, juliet macbeth, shakespe
fighlights: 46 Shakespeare Hamlet		warwick, margaret, bol

Click on a column header to sort the data. Click again to toggle the sort order.

Handling "Squished" Datasets

The Galaxy shows similarities and differences among documents. Sometimes a very few documents are so unique that these differences outweigh the other relationships. The result may be a Galaxy with most all the documents "squished" into a relatively small part of the screen.

Step by Step

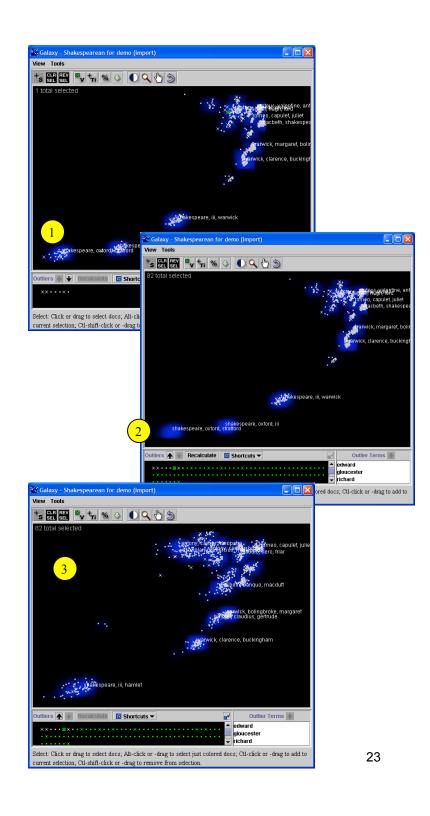
1

3

Select the few documents that are off to themselves and open the Document Viewer to see why they are so different. For example, there may be almost no text, or it may be in a different language.

2 Click the down arrow on the Outlier tool to move the oddball documents to the lower Outliers pane. Click the Recalculate button.

The Galaxy will recluster and rearrange the remaining documents. If you want to keep track of the documents you moved down, open the Group tool and make a new group out of them while they are still selected.



My Data Has Disappeared

Sometimes you may realize that the visualization looks unexpectedly sparse and doesn't match what you remember for this dataset. There are a few interaction capabilities that might be the cause. Follow these steps to get the view "back to normal."

Step by Step

Time Slicer

 $\left(1\right)$

2

Look in the upper left corner of the Galaxy window to see if the Time Slicer is active. If so, click the Time Slicer button to bring it to the front and close it.

Zoom or Pan

You may have zoomed or panned the data off screen. Use the Reset button on either the Galaxy or ThemeView to reset all display changes back to their original defaults.

📽 Galaxy - Shakespearean for demo (import)	
View Tools	
+s Sel Sel ♥v +ti % ○ ● ♀ ⊕	۵
82 total selected Time filter active	
1	demetriu្ល់ដែមរ៉េអៀមតែមន្តា) X
🖬 IN-SPIRE - Shakespearean for demo (import)	
File View Window 1 lp	

