



IN-SPIRE Tips and Techniques
"At a Glance"
Version 2.1

December 18, 2003

At a Glance

Table of Contents

Getting an Overview of a Dataset	3
Strategies for Early Use of Spatial Patterns	4
Read Similar Documents Together	5
See More Detail on Important Document Clumps	6
Find Relevant Information (a).....	7
Find Relevant Information (b).....	8
Keep Track of What You Find	9
Inspecting Non-hits Among the Hits	10
Seeing a Time Pattern.....	11
Reuse Queries	12
Manage Evidence and Hypotheses.....	13
Removing Unhelpful Themes	14
Finding Out Who's Saying What.....	15
Finding More Documents Like This One	16
Finding a Document I Know About.....	17
Handling "Squished" Datasets	18
My Data Has Disappeared	19

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 Galaxy and ThemeView show the large and small concentrations of documents; both have the same spatial arrangement.

Step by Step

- 1 Find major and minor themes from ThemeView hills. The higher hills mean more documents and stronger themes.
- 2 The labels of the ThemeView list some strong themes in that region. To show more labels, pull down the Settings menu and select Peak Labels.
- 3 Find related themes from proximity. Hills close together are likely related. Isolated hills show outliers.
- 4 The Galaxy clouds and peak labels correspond to the ThemeView hills and labels. To turn them on and off, use the Settings menu.
- 5 To show strong themes in a location that is not labeled, click on the Probe tool icon and then click on a location in either the ThemeView or Galaxy.
- 6 To see the most common words in a clump of documents, select the documents and open the Gist tool.

ThemeView - EP3

Zoom: Click to zoom in, Ctrl-Click to zoom out, Drag a box to zoom to the region, Ctrl-Shift-Click to reset to the original.

IN-SPIRE - Time Articles 1963

File Galaxy ThemeView Windows Settings Help

Probe

Word	Contribution 1-
building	
clinton	
sen	
exp	
fbi	
rescue	
truck	
blood	
list	
panetta	

Outlier Terms

Gist - Time Articles 1963

Total Words: 16783
Total Docs: 281

All fields Computation fields

Word	Docs
time	281
page	281
source	281
week	248
government	155
party	125
minister	111

Add Words Clear

week

Click to append selected words to the list in the copy window.

Galaxy

Click to select one doc; Ctrl-Click to add one doc; Ctrl-Shift-Click to remove a doc; Drag to group of docs; Ctrl-Drag to add docs; Ctrl-Shift-Drag to remove docs; Shift-Click to clear

Strategies for Early Use of Spatial Patterns

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

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?

- 1 Open the ThemeView and rotate 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.
- 3 Look at the Galaxy. Isolated clumps of documents may show outliers – documents that are very different from the bulk of the set.

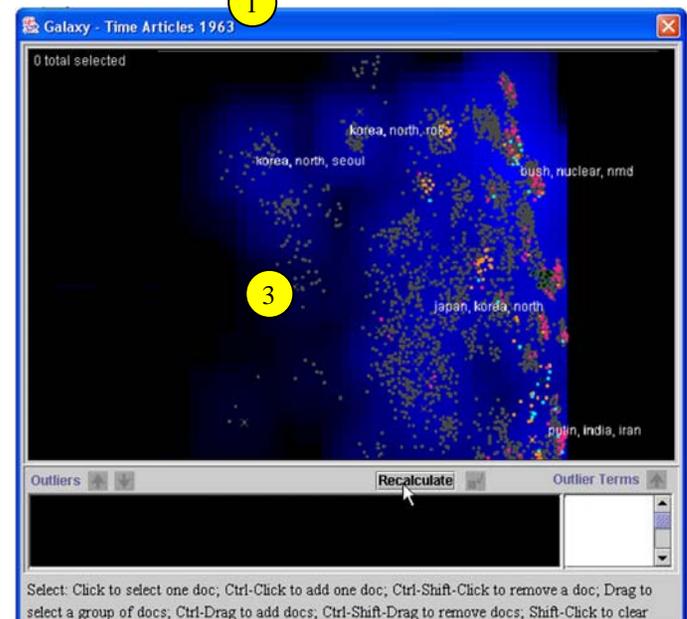
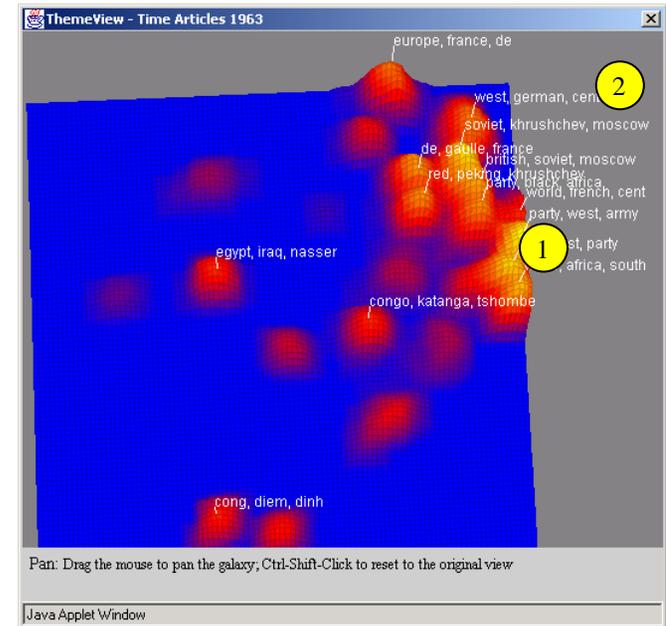
What Questions Can the Spatial Patterns Raise?

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.

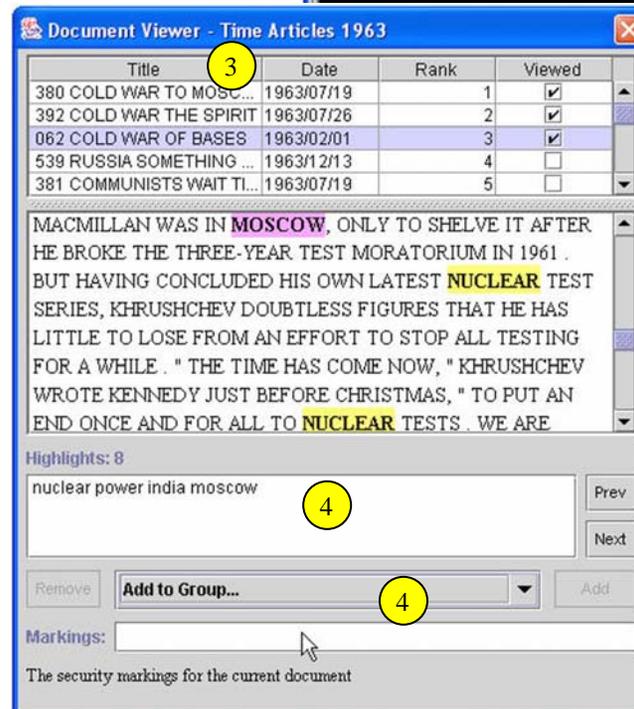
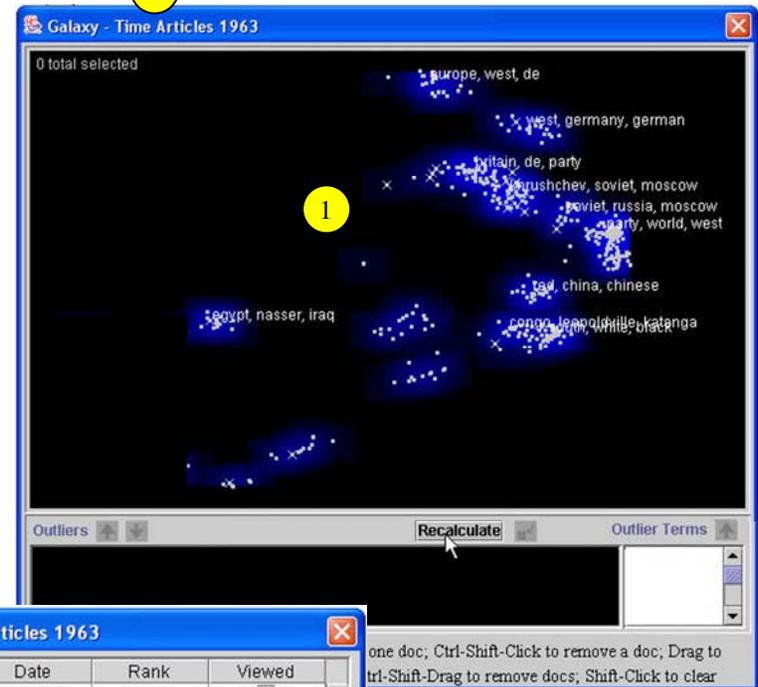


Read Similar Documents Together

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

- 1 Use the labels and the probe tool to get a sense of the various clumps in the Galaxy.
- 2 Select a clump of documents and open the Document Viewer.
- 3 Click on a column heading to sort the documents by title or date. When you click a title, the document will display.
- 4 To highlight particular words, type them in the Highlights pane and hit Enter. To save useful documents, add them to an existing group or make a new one.

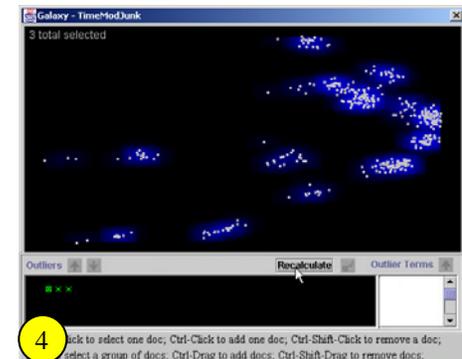
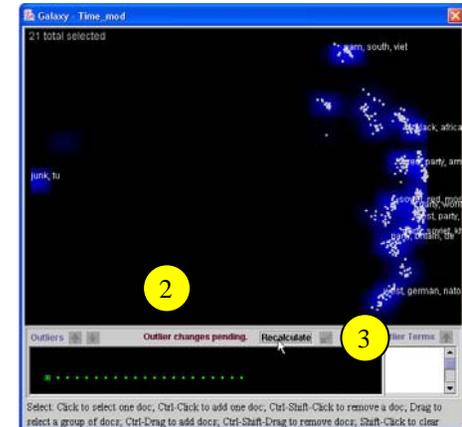
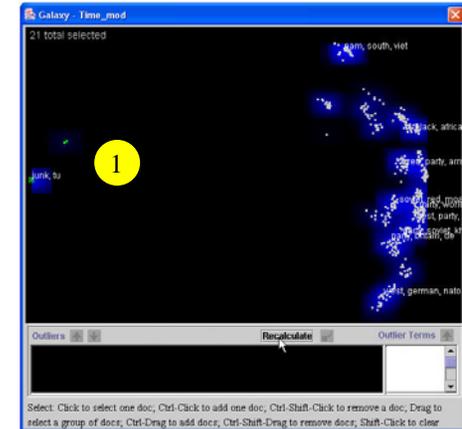


See More Detail on Important Document Clumps

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.

Step by Step

- 1 Select a clump of documents that you want to push aside. Click the down arrow.
- 2 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 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:

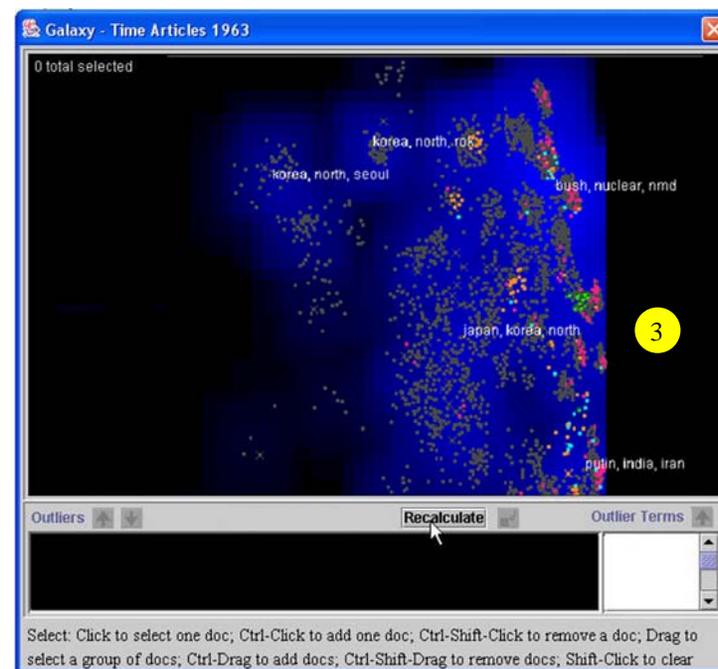
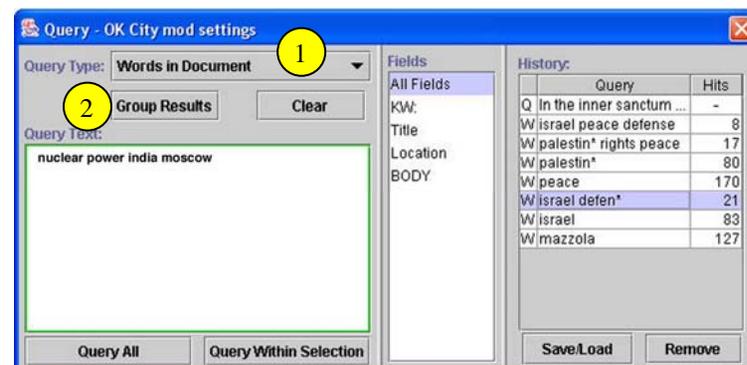
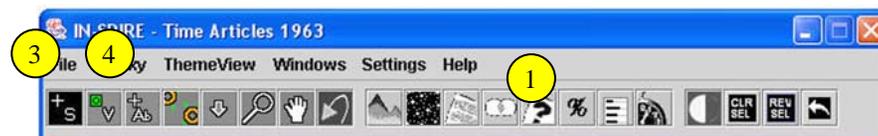
- a query based on word presence or absence in a document
- a query based on thematic similarity
- the clusters based on overall document similarity.

Used together, these form a powerful set.

Step by Step

- 1 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., telephon* will match telephone or telephonic
 - Use * in a phrase, e.g., "walk * dog" will match "walk dog" or "walk the dog"
 - Words with spaces between are interpreted as a Boolean AND.
- 2 Often there are several hits close together in one or more large clumps, with others scattered in smaller ones. Reading the hits by clump lets you read similar documents together. Make a group from the query results.
- 3 Use the Galaxy layout and the pattern of query hits to decide which results to look at first. To select just the portion of the set that you want to see first, click the Select tool. Then hold down the Alt key and drag the mouse over the clump you want. Those documents will stay selected, and the others will show in the group color but will no longer be selected. Look at the documents in the document viewer. Repeat with other portions of the selection set.
- 4 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.



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
- the clusters based on overall document similarity.

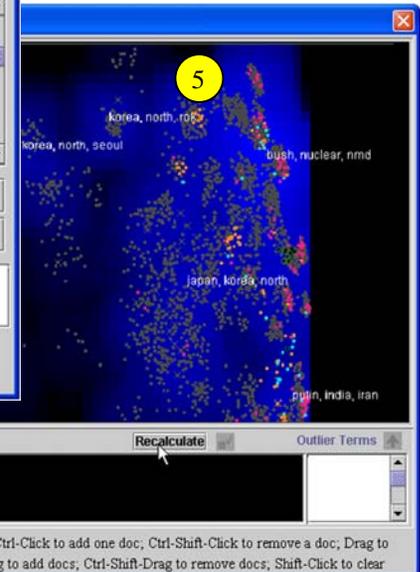
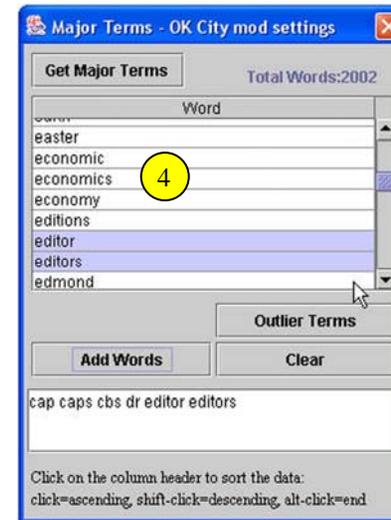
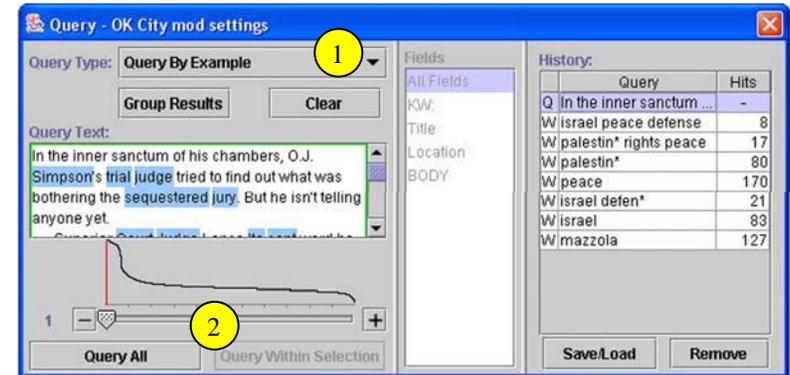
Used together, these form a powerful set.

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.
- 2 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, scroll to the bottom of the list in the Document Viewer 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.
- 4 If an important word is not highlighted, you can insert other similar words. Use the Major Terms window for suggestions.
- 5 The Word and Query by Example usually return somewhat different results (like the orange vs. red groups in the figure).



3

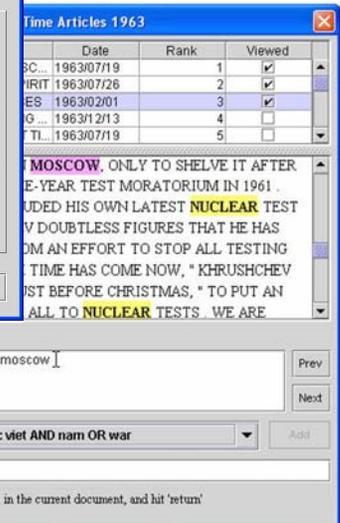
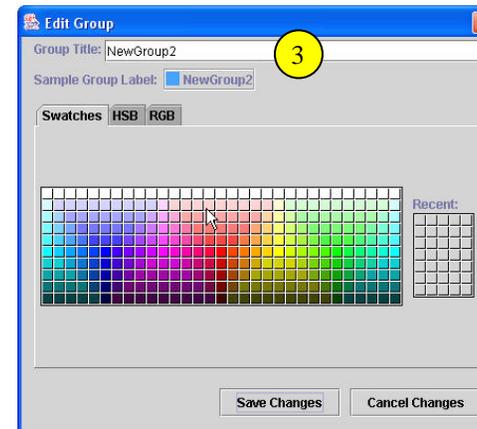
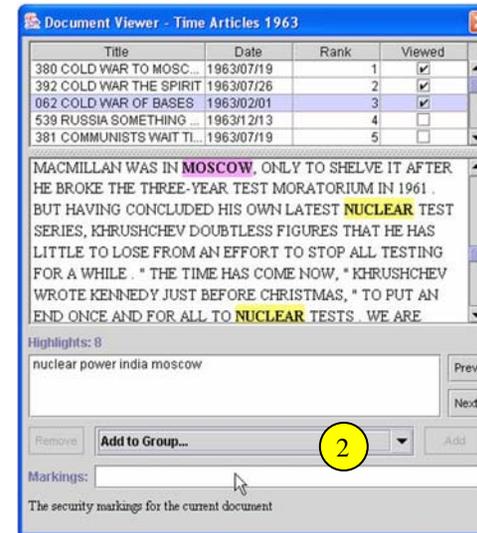


Keep Track of What You Find

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

- 1 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*.
- 3 To start a new group with this document, select New Group from the pull down and click the *Add* button. The Edit Group window opens. Enter a name for the new group (you can also change the group color here) and click *Save Changes*.
- 4 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.

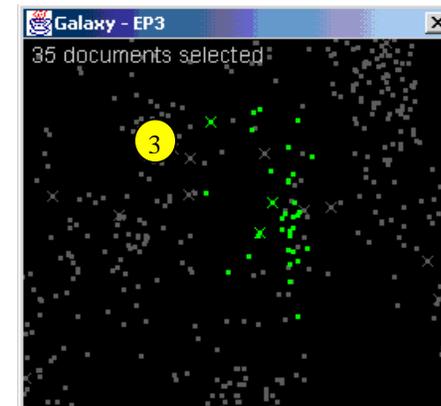
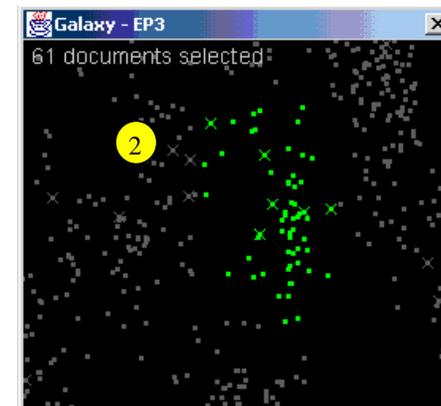
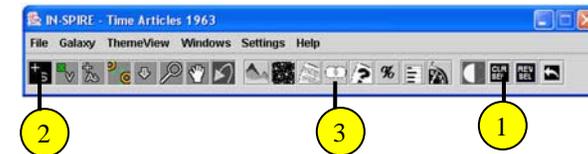
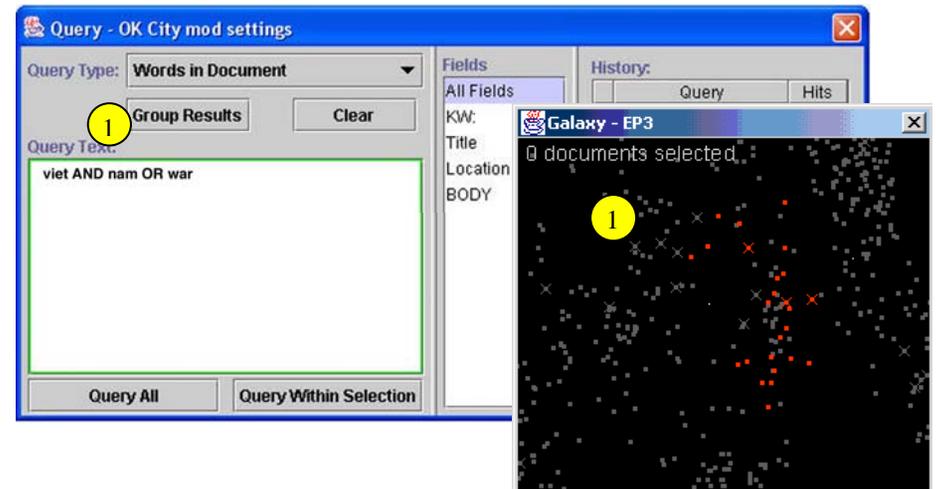


Inspecting Non-hits Among the Hits

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 non-hits, follow these suggestions.

Step by Step

- 1 Make a group from the Query result. Press Clear Selections to see your group in the Galaxy.
- 2 Use the Selection tool to drag over and select all documents in the area of interest.
- 3 Open the Group tool and select the 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.
- 4 Now you can read them to see what all is there.

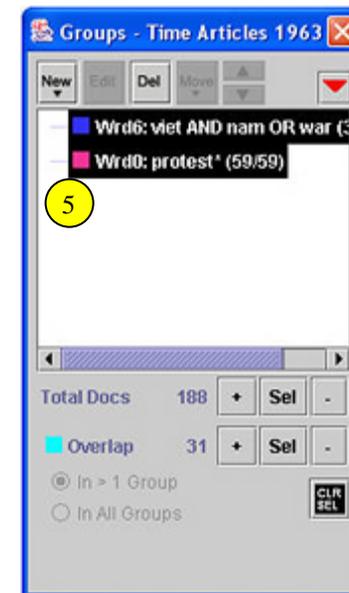
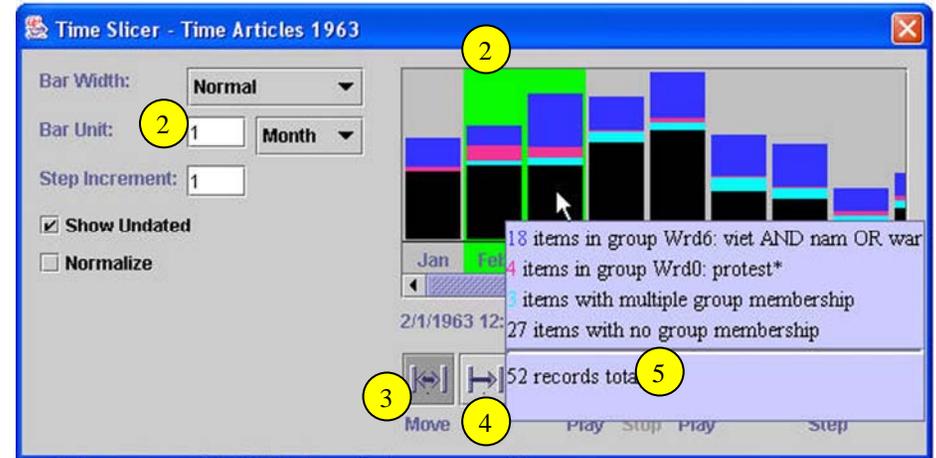
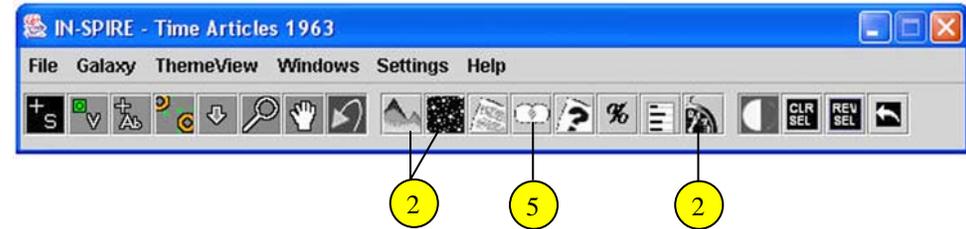


Seeing a Time Pattern

The Time Slicer lets you see relations between themes and time. Documents stay in the same spatial location, but are shown or hidden according to the time period of interest.

Step by Step

- 1 Open the Time Slicer and the ThemeView or Galaxy.
- 2 Change the Unit to the granularity of interest. The green highlighted bars are the ones reflected in the ThemeView and Galaxy. Change the width of the highlight area by dragging the edge.
- 3 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. The histogram shows number of documents per 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.
- 4 Click on the first time slice. Click the Grow button. Step through again, watching the cumulative patterns.
- 5 To see groups over time, open the Group tool. Highlight the groups you're interested in, using SHIFT-click or CTRL-click. The group color in the histogram bars shows the number of members per time slice. An additional color (defaults to aqua) shows 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?



Reuse Queries

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

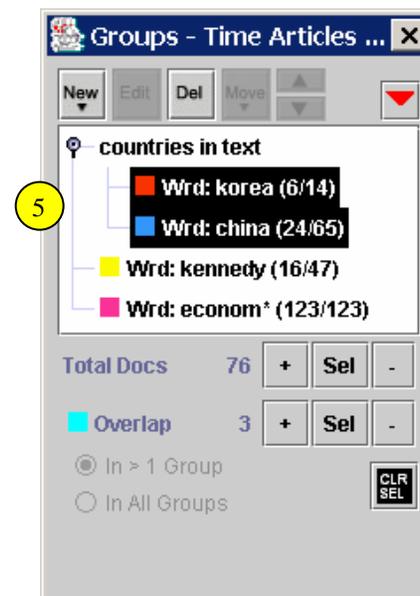
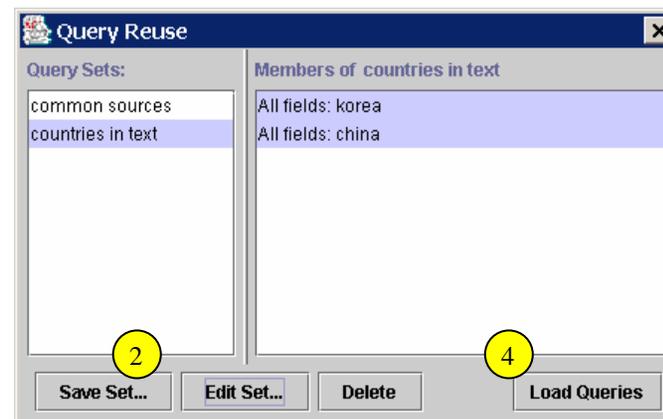
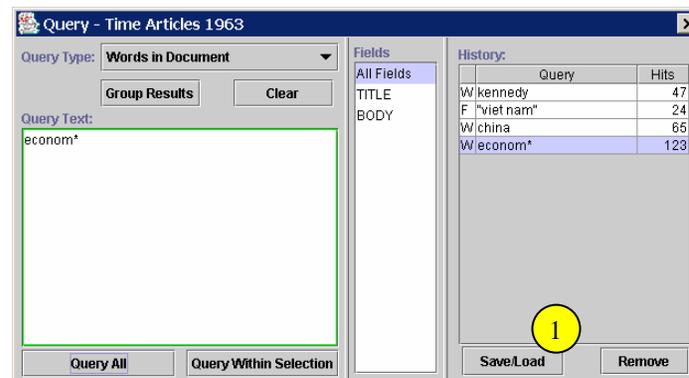
Step by Step

To Save Queries

- 1 On the query tool, click Save/Load.
- 2 A dialog opens Click Save Set... to make a set of saved queries.
- 3 Type in the name for the query set. The dialog will show the Word and Field queries you have made on the current datasets. Select the queries you want to save for use in other datasets. Click the arrow to include them in the list to save. Click the Save key.

To Load Saved Queries

- 4 One the query tool, click Save/Load. Select the set you want. You may highlight some or all of the queries it contains. Click Load Queries.
- 5 IN-SPIRE will make a group for each of the highlighted queries and put them in a group with the same name as the saved query set.

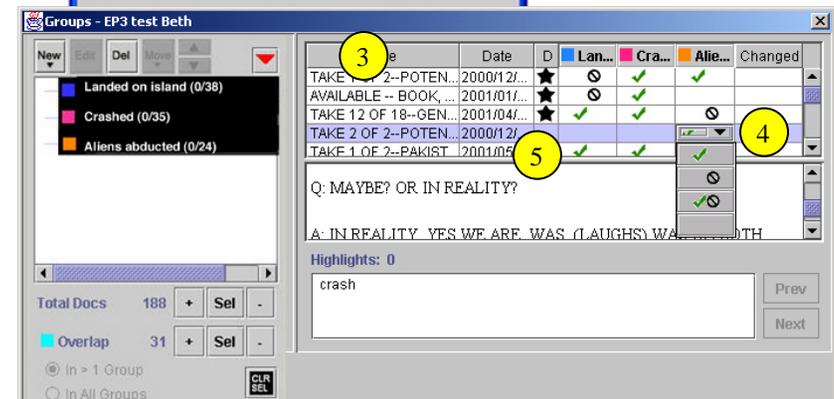
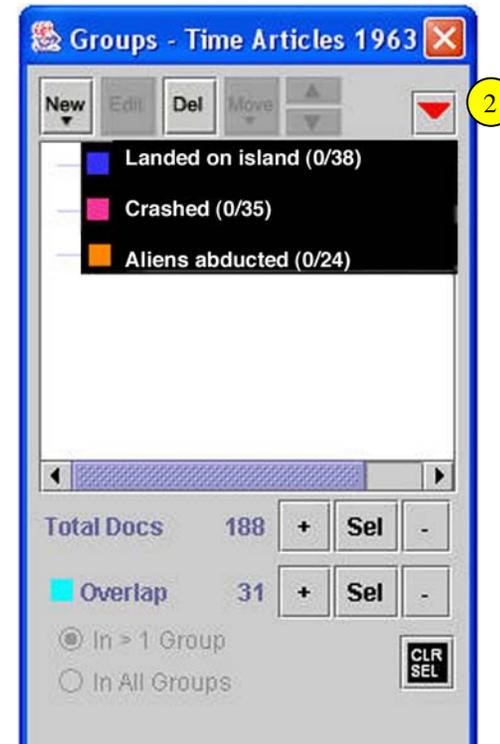
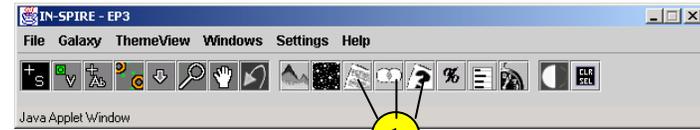


Manage Evidence and Hypotheses

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

- 1 As you explore the dataset, make a group for each hypothesis you are investigating. For example, you may have three hypotheses about the cause of a plane disappearance. 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. To review a document, click its title.
- 4 If you find a document that is inconsistent with a hypothesis, click on the check mark to change it to the symbol for contradictory evidence.
- 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.



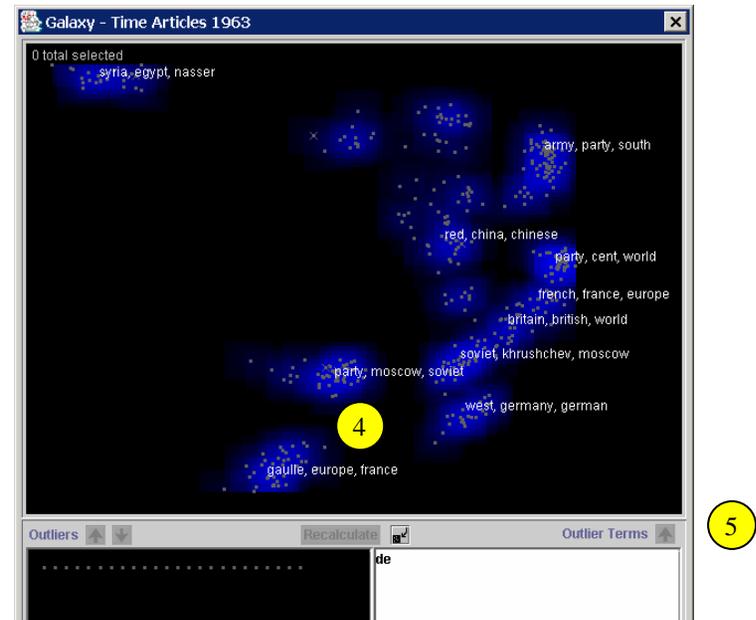
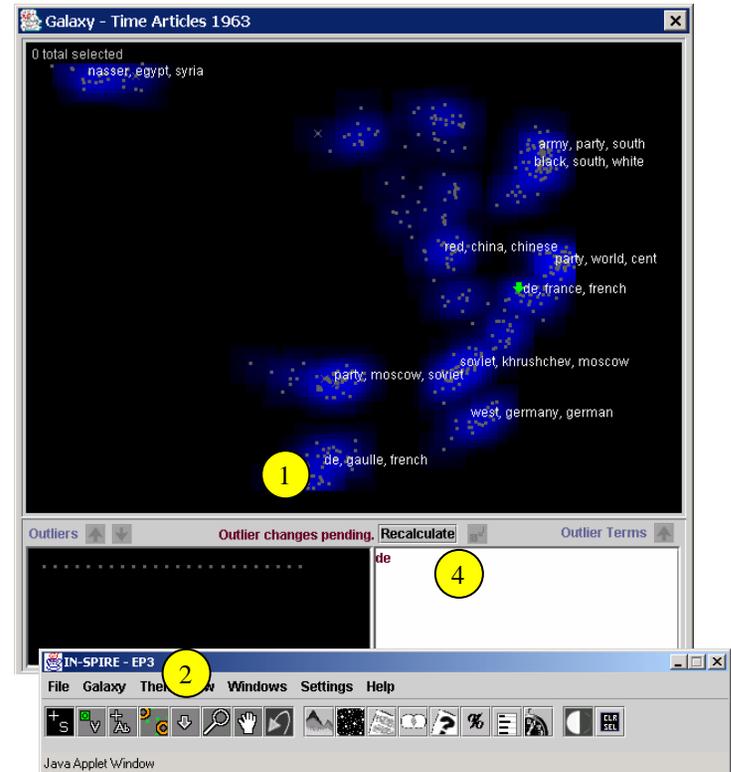
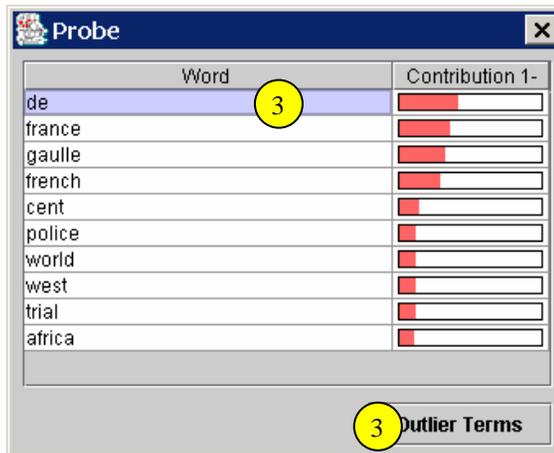
Removing Unhelpful Themes

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 be need finance-related documents, a theme of "finance" is still useful if it helps you identify less interesting clumps.

Step by Step

- 1 Get a sense of what's where in the Galaxy, and the clumps that are likely relevant or not to your interests and task. Identify the themes that are not helpful in making this distinction.
- 2 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.
- 4 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.
- 5 You can put Outlier Terms back by selecting them and clicking the up arrow.

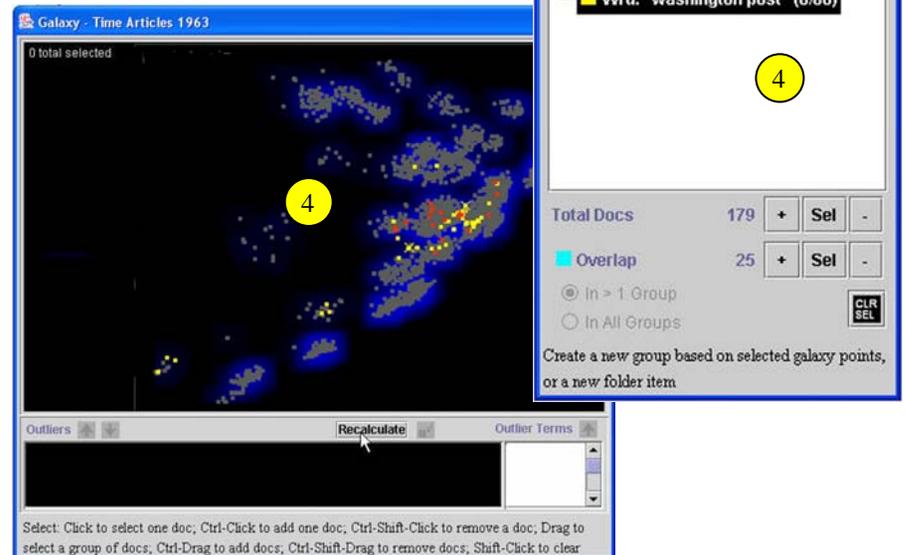
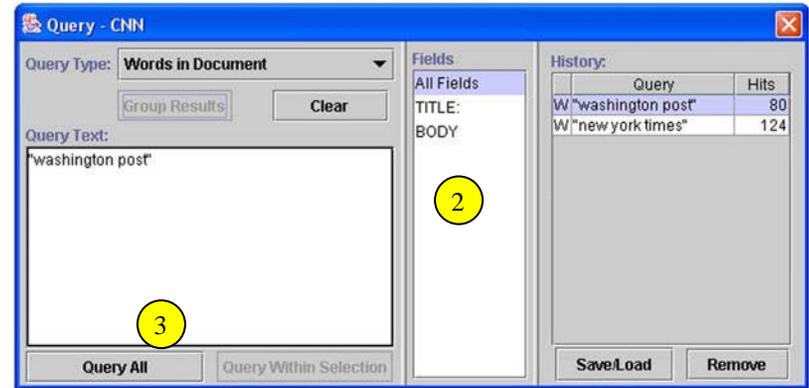
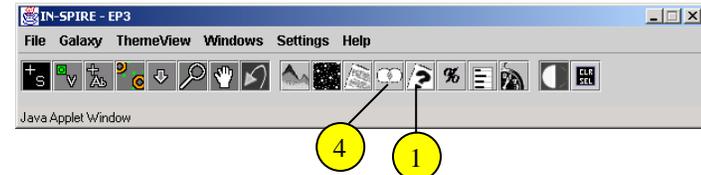


Find Out Who's Saying What

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

Step by Step

- 1 Open the query tool and make sure that Words in Document is selected.
- 2 The central pane lists the queryable fields for this dataset. Select the source field. Type in the query for the source you want.
 - AND, OR, 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, and then click Group Results. Continue to query and make groups for other sources of interest.
- 4 Open the Group tool. Clear 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 that source has different documents from other sources.



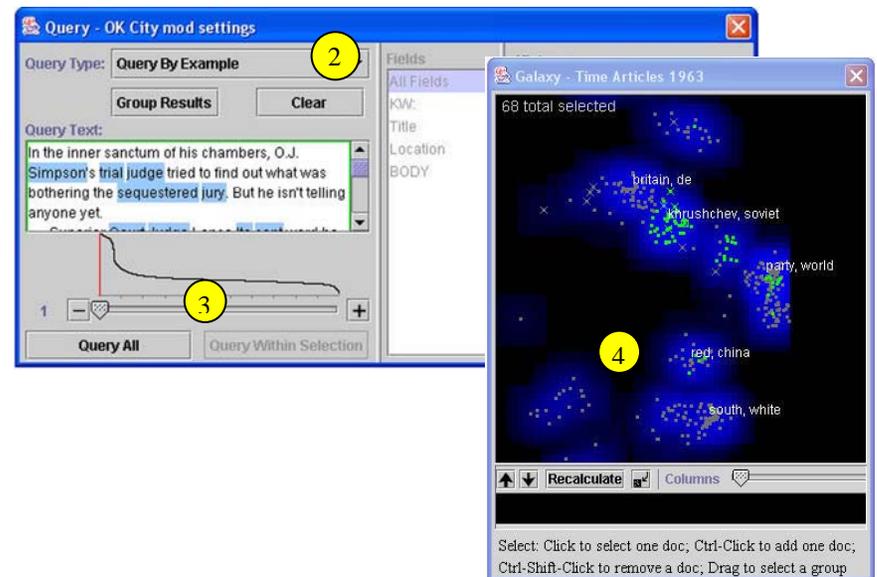
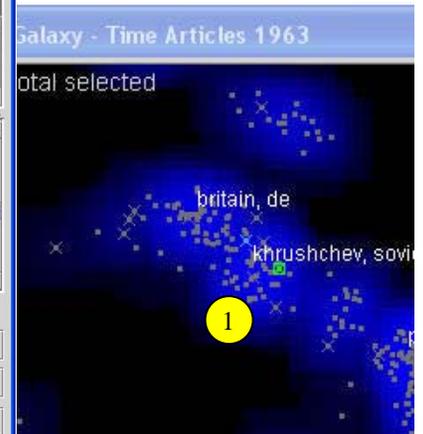
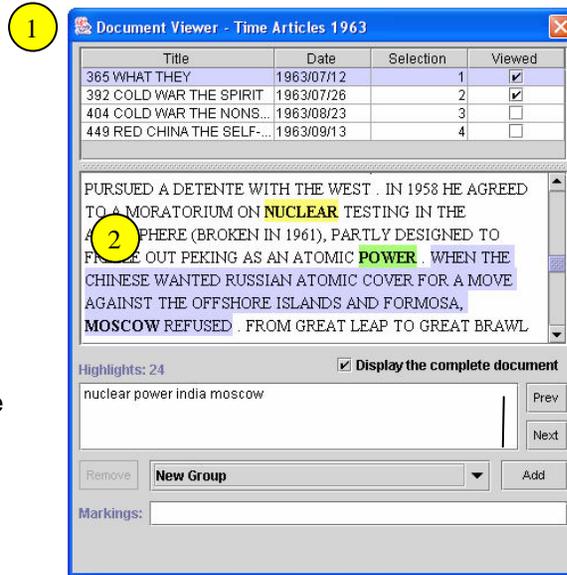
Finding More Documents Like This One

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's 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 Documents close by should be similar, so may also be useful.
 - 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.
- 2 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.
- 3 Move the slider to select progressively 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.
- 4 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.



Finding a Document I Know About

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

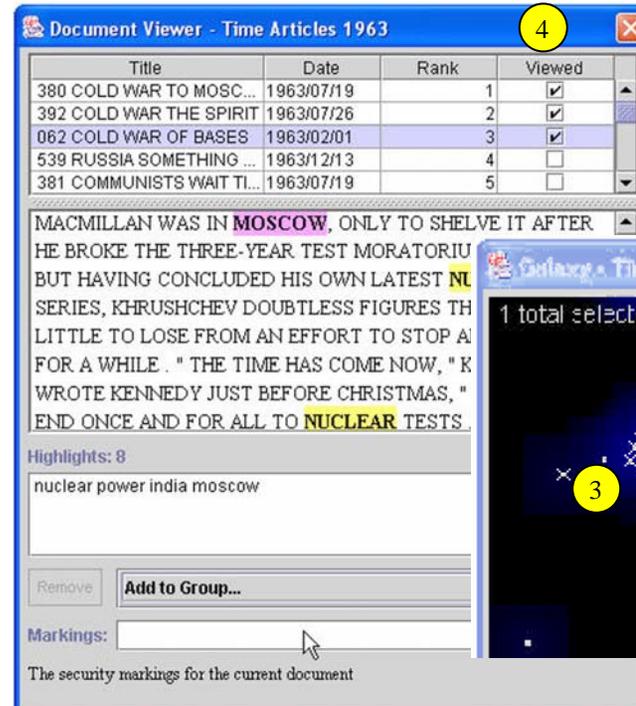
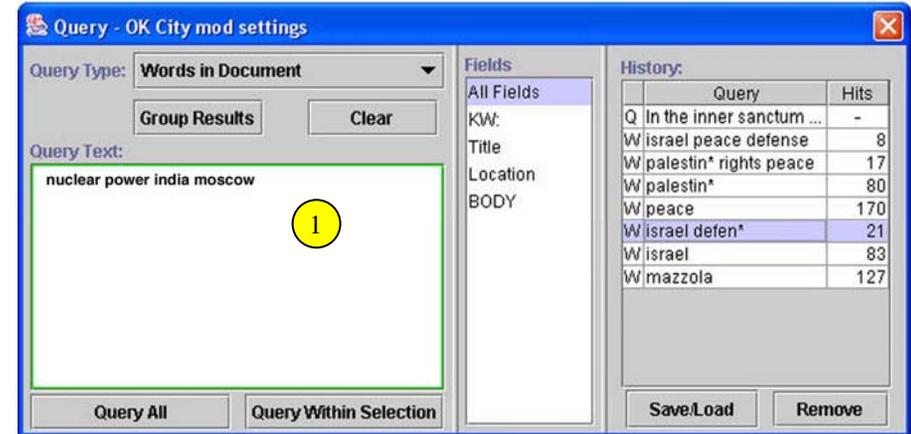
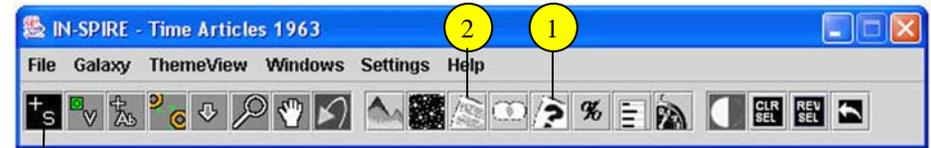
You can use a relatively unique term or field value to find it. If you've read the document that can also help you find it quickly.

Step by Step

By Field or Text

- 1 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.
- 2 Open the Document Viewer. You can sort by title, date, or Viewed status by clicking on the column heading.
- 3 Documents you've 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.
- 4 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.

If You've Read the Document

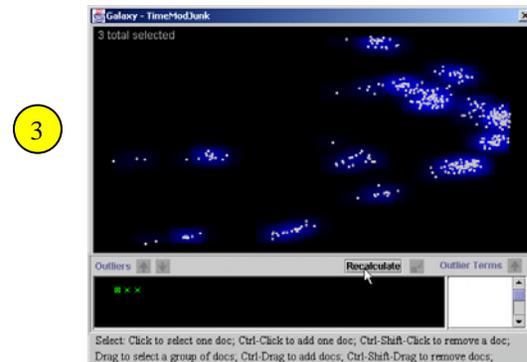
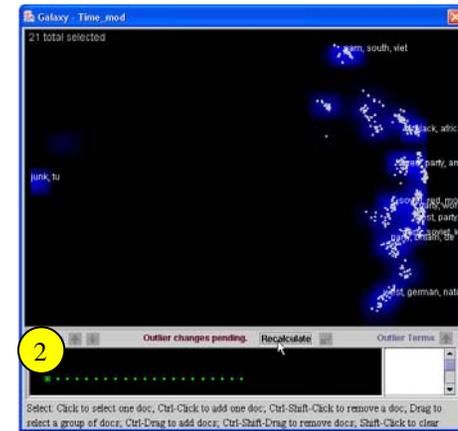
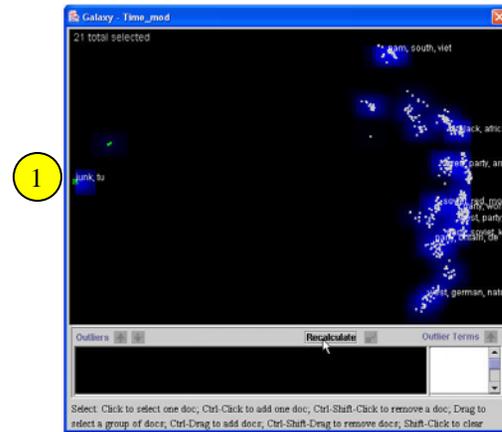


Handling “Squished” Datasets

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 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 pane. Click the Recalculate button.
- 3 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

- 1 Time Slicer**
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.
- 2 Zoom or Pan**
Click the Zoom button, then Ctl-Shift-Click in the Galaxy window to reset the view to its default.
- 3 Reset**
A quick way to get back to where you started on this dataset is to click the Reset button. This will reset all display changes, such as zoom or pan, and will close all windows except the Galaxy window.

