# Table of Contents

1 Adding translated text to documents ................................................................. 19
   1.1 Recipe overview .......................................................................................... 19
   1.2 Requirements .............................................................................................. 19
   1.3 Directions ................................................................................................... 19

2 Adding documents to an existing Assisted Review project ................................. 20
   2.1 Recipe Overview ......................................................................................... 20
   2.2 Requirements .............................................................................................. 20
   2.3 Directions ................................................................................................... 20

3 Cluster Visualization QC Workflow ..................................................................... 21
   3.1 Requirements .............................................................................................. 21
   3.2 Directions ................................................................................................... 21

4 Coding the first item in a family group with the skip function ............................. 26
   4.1 Recipe overview ......................................................................................... 26
   4.2 Requirements .............................................................................................. 26
   4.3 Directions ................................................................................................... 27

5 Creating a read-only layout to lock produced documents .................................... 28
   5.1 Recipe overview ......................................................................................... 28
   5.2 Requirements .............................................................................................. 28
   5.3 Directions ................................................................................................... 28
   5.4 Considerations ............................................................................................ 30

6 Cluster Visualization QC Workflow ..................................................................... 30
   6.1 Requirements .............................................................................................. 30
   6.2 Directions ................................................................................................... 31

7 Creating an application for managing attorney lists .......................................... 35
   7.1 Recipe overview ......................................................................................... 35
7.2 Requirements ........................................................................................................35
7.3 Directions .............................................................................................................35

8 Creating an optimized Relativity Analytics index (8.2 and higher) ................37
  8.1 Requirements .....................................................................................................37
  8.2 Directions ...........................................................................................................37
    8.2.1 Important Points .......................................................................................38

9 Creating commonly used Pivots ........................................................................38
  9.1 Recipe overview ...............................................................................................38
  9.2 Requirements ...................................................................................................38
  9.3 Directions .........................................................................................................38

10 Creating secured saved search folders for multiple groups .....................39
  10.1 Recipe overview ............................................................................................39
  10.2 Requirements ..................................................................................................39
  10.3 Directions .......................................................................................................39

11 Creating static searches using multiple-choice fields ..................................40
  11.1 Recipe overview ............................................................................................40
  11.2 Requirements ..................................................................................................40
  11.3 Directions .......................................................................................................40

12 Creating training resources tabs ......................................................................41
  12.1 Creating training resources tabs ..................................................................41
    12.1.1 Requirements .........................................................................................41
    12.1.2 Directions ...............................................................................................41
  12.2 Adding training resources to a workspace ......................................................42
    12.2.1 Requirements .........................................................................................42
    12.2.2 Directions ...............................................................................................42

13 Displaying family groups in a view .................................................................43
  13.1 Recipe overview .............................................................................................43
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.3.2</td>
<td>Offline coding form</td>
<td>49</td>
</tr>
<tr>
<td>18</td>
<td>Filtering to find empty fields</td>
<td>50</td>
</tr>
<tr>
<td>18.1</td>
<td>Recipe overview</td>
<td>50</td>
</tr>
<tr>
<td>18.2</td>
<td>Requirements</td>
<td>50</td>
</tr>
<tr>
<td>18.3</td>
<td>Directions</td>
<td>50</td>
</tr>
<tr>
<td>19</td>
<td>Finding duplicates or near duplicates in a new set that match up to a previous set</td>
<td>51</td>
</tr>
<tr>
<td>19.1</td>
<td>Requirements</td>
<td>51</td>
</tr>
<tr>
<td>19.2</td>
<td>Directions</td>
<td>51</td>
</tr>
<tr>
<td>19.3</td>
<td>References</td>
<td>51</td>
</tr>
<tr>
<td>20</td>
<td>Finding emails exchanged within the same domain</td>
<td>51</td>
</tr>
<tr>
<td>20.1</td>
<td>Recipe overview</td>
<td>52</td>
</tr>
<tr>
<td>20.2</td>
<td>Requirements</td>
<td>52</td>
</tr>
<tr>
<td>20.3</td>
<td>Directions</td>
<td>52</td>
</tr>
<tr>
<td>21</td>
<td>How to calculate Precision and Recall without a control set</td>
<td>54</td>
</tr>
<tr>
<td>21.1</td>
<td>Recipe overview</td>
<td>54</td>
</tr>
<tr>
<td>21.2</td>
<td>Requirements</td>
<td>55</td>
</tr>
<tr>
<td>21.3</td>
<td>Precision and Recall Defined</td>
<td>55</td>
</tr>
<tr>
<td>21.3.1</td>
<td>Initial setup</td>
<td>56</td>
</tr>
<tr>
<td>21.3.2</td>
<td>View setup</td>
<td>57</td>
</tr>
<tr>
<td>21.3.3</td>
<td>Calculating Precision</td>
<td>57</td>
</tr>
<tr>
<td>21.3.4</td>
<td>Calculating Recall</td>
<td>59</td>
</tr>
<tr>
<td>21.3.5</td>
<td>Additional considerations</td>
<td>61</td>
</tr>
<tr>
<td>22</td>
<td>Identifying emails between two or more specific individuals</td>
<td>61</td>
</tr>
<tr>
<td>22.1</td>
<td>Requirements</td>
<td>61</td>
</tr>
<tr>
<td>22.2</td>
<td>Directions</td>
<td>61</td>
</tr>
<tr>
<td>23</td>
<td>Identifying standalone emails</td>
<td>66</td>
</tr>
<tr>
<td>23.1</td>
<td>Requirements</td>
<td>67</td>
</tr>
</tbody>
</table>
23.2 Directions .................................................................................................................. 67

24 Importing productions in Relativity with natives .......................................................... 69
  24.1 Overview .................................................................................................................... 69
  24.2 Requirements ........................................................................................................... 69
  24.3 Directions .................................................................................................................. 69

25 Legal Hold checklist ....................................................................................................... 71
  25.1 Requirements ........................................................................................................... 71
  25.2 Directions .................................................................................................................. 72
  25.3 References ................................................................................................................ 73

26 Leveraging processing metadata for quality control ...................................................... 73
  26.1 Overview .................................................................................................................... 73
    26.1.1 Requirements ....................................................................................................... 73
  26.2 Directions .................................................................................................................. 73
    26.2.1 Important considerations ...................................................................................... 79

27 Loading external productions into Relativity ................................................................. 80
  27.1 Requirements ........................................................................................................... 80
  27.2 Instructions ............................................................................................................... 80
  27.3 Notes .......................................................................................................................... 80

28 Maintaining previous Bates numbers between production sets .................................. 81
  28.1 Recipe overview ....................................................................................................... 81
  28.2 Requirements ........................................................................................................... 81
  28.3 Directions .................................................................................................................. 81
    28.3.1 Option 1 .............................................................................................................. 81
    28.3.2 Option 2 .............................................................................................................. 82

29 Mapping Processing Fields ............................................................................................. 83
  29.1 Requirements ........................................................................................................... 83
  29.2 Directions .................................................................................................................. 83
30 Mass importing into a file field ........................................................................86
30.1 Overview ........................................................................................................86
30.2 Requirements .................................................................................................86
30.3 Directions .......................................................................................................86
31 Migrating cases from Summation ....................................................................88
31.1 Recipe overview .............................................................................................88
31.2 Requirements .................................................................................................88
31.3 Directions .......................................................................................................88
31.3.1 Migrating images ......................................................................................88
31.3.2 Migrating data ............................................................................................88
31.3.3 Migrating transcripts ................................................................................89
31.3.4 Migrating productions ...............................................................................89
31.3.5 Migrating redactions .................................................................................90
32 Moving relevant choices to the top ..................................................................90
32.1 Overview ........................................................................................................90
32.2 Requirements .................................................................................................91
32.3 Directions .......................................................................................................91
33 Multi-matter workspace setup in Relativity ......................................................91
33.1 Recipe overview .............................................................................................91
33.2 Requirements .................................................................................................92
33.3 Directions .......................................................................................................92
33.3.1 User Groups/Security ..............................................................................92
33.3.2 Folder Setup ...............................................................................................93
33.3.3 Field and choice setup ..............................................................................93
33.3.4 Views Setup ...............................................................................................93
33.3.5 Layout Setup ..............................................................................................94
33.3.6 Search indexes ...........................................................................................94
37.4.3 Update Duplicate Status script ................................................................. 117
37.5 Viewing the results ......................................................................................... 118
37.5.1 All Custodians script .................................................................................... 118
37.5.2 All Source Locations script ........................................................................ 119
37.5.3 Update Duplicate Status script .................................................................... 120

38 Performing QC of Productions ......................................................................... 121
38.1 Recipe overview ............................................................................................. 121
38.2 Requirements .................................................................................................. 121
38.3 Directions ........................................................................................................ 122
   38.3.1 Conversation before Production ................................................................. 122
   38.3.2 Items to check in a production set .............................................................. 122
   38.3.3 QC of production data ............................................................................... 123

39 RAR checklist .................................................................................................... 123
   39.1 Checklist ......................................................................................................... 124

40 Relativity Assisted Review Reviewer Protocol ..................................................... 125
   40.1 Overview .......................................................................................................... 125
   40.2 Requirements ................................................................................................... 125
   40.3 What is Assisted Review? ................................................................................ 125
   40.4 Selecting good example documents ............................................................... 125
      40.4.1 Sufficient text ......................................................................................... 125
      40.4.2 Email headers and repeated content ...................................................... 126
      40.4.3 Images ..................................................................................................... 126
      40.4.4 Numbers .................................................................................................. 127
      40.4.5 Families and the "Four Corners" test ......................................................... 129
   40.5 Handling good language / bad example exceptions ...................................... 130
   40.6 Applying the Use as Example field ............................................................... 131
   40.7 Using the Excerpt text box ............................................................................. 131
46.2 Annotations ............................................................................................................. 143
46.3 Security and authentication ...................................................................................... 143
46.4 Emailing documents ................................................................................................. 144

47 Relativity Structured Analytics – Best practices and workflows ................................. 144
47.1 Requirements ........................................................................................................... 144
47.2 Directions ................................................................................................................. 144

48 Relativity Structured Analytics - Email Threading ...................................................... 148
48.1 Requirements ........................................................................................................... 148
48.2 Directions ................................................................................................................. 148
    48.2.1 Create saved searches (Relativity 9.5.89.76 and lower) ...................................... 148
    48.2.2 Create or configure an Analytics profile .............................................................. 149
    48.2.3 Create a Structured Analytics set ....................................................................... 150
    48.2.4 Email threading results - reports ...................................................................... 151
    48.2.5 Setting up an Email Threading View .................................................................. 152
    48.2.6 Email Threading Display ................................................................................... 154
    48.2.7 Thread groups in the related items pane ............................................................ 155
    48.2.8 Working with email threading results ................................................................. 156
48.3 Special considerations .............................................................................................. 157
48.4 References ............................................................................................................... 158

49 Removing documents from a RAR example set ......................................................... 158
49.1 Requirements ........................................................................................................... 158
49.2 Directions ................................................................................................................. 158
49.3 References ............................................................................................................... 159

50 Resolving common processing errors ........................................................................ 159
50.1 Recipe overview ....................................................................................................... 159
50.2 Requirements ........................................................................................................... 159
50.3 Directions ............................................................................................................... 159
66 Transfer tags from other document review applications .......................................... 197
66.1 Recipe overview .................................................................................................. 197
66.2 Directions ............................................................................................................ 198

67 Threading review setup ......................................................................................... 198
67.1 Requirements ...................................................................................................... 198
67.2 Directions ............................................................................................................ 198
  67.2.1 Sample view for Inclusive email with family .............................................. 198
  67.2.2 Sample view for Inclusive and non-Duplicate email with family ............ 199
  67.2.3 Sample view of the Related Items pane for an Email Thread Group .... 199
  67.2.4 Sample view for Near Dupe Documents .................................................... 200
  67.2.5 Sample view of the Related Items pane for a Near Dupe Group .......... 200

68 Understanding the statistical sampling formula ..................................................... 202
68.1 Recipe overview .................................................................................................. 202
68.2 Requirements ...................................................................................................... 202
68.3 Formula breakdown .............................................................................................. 202

69 Use case for Relativity Analytics - Clustering & Categorization ......................... 202
69.1 Common scenarios ............................................................................................... 203
69.2 Requirements ...................................................................................................... 203
69.3 Directions ............................................................................................................ 203
  69.3.1 Scenario 1 ...................................................................................................... 203
  69.3.2 Scenario 2 ...................................................................................................... 203

70 Using Analytics with CJK characters and non-English languages ....................... 204
70.1 Requirements ...................................................................................................... 204
70.2 Recipe overview .................................................................................................. 204
70.3 Considerations for using Analytics with CJK characters ............................... 205
70.3.1 Considerations for creating Analytics indexes containing non-English languages ... 205
70.3.2 Considerations for using Structured Analytics with non-English languages ........... 207
70.3.3 Checklist for using Analytics with non-English language and multiple language workspaces .............................................................. 210

71 Using categorization to find privileged documents .................................................. 210
71.1 Requirements ........................................................................................................ 211
71.2 Directions ............................................................................................................. 211

72 Using categorization to identify hot documents ...................................................... 211
72.1 Requirements ........................................................................................................ 211
72.2 Directions ............................................................................................................. 212

73 Using Cluster Visualization to jump start your RAR project .................................. 212
73.1 Requirements ........................................................................................................ 212
73.2 Directions ............................................................................................................. 212

74 Using custom objects to secure coding choices ...................................................... 216
74.1 Recipe overview .................................................................................................... 216
74.2 Requirements ........................................................................................................ 217
74.3 Directions ............................................................................................................. 217

75 Using dtSearch to identify documents in other languages .................................... 217
75.1 Recipe overview .................................................................................................... 217
75.2 Requirements ........................................................................................................ 218
75.3 Directions ............................................................................................................. 218

76 Using Near Duplicate analysis in review ............................................................... 218
76.1 Requirements ........................................................................................................ 219
76.2 Overview ............................................................................................................... 219
76.3 Directions ............................................................................................................. 219
76.4 Workflow considerations ...................................................................................... 220

77 Using redactions from other applications ............................................................... 221
83.2.1 Publishing or loading additional documents .......................................................... 234

84 Validating RegEx filters in Structured Analytics .......................................................... 235
84.1 Directions .................................................................................................................. 235

85 Verifying Analytics filters ............................................................................................ 238
85.1 Requirements ........................................................................................................... 239
85.2 Directions .................................................................................................................. 239
1 Adding translated text to documents

If your documents require translation, this recipe describes one way of including both the original text and the translated text in your Relativity workspace after you receive the translated documents from the translation service.

1.1 Recipe overview

This recipe explains how to add translated text to your workspace, then search, view the translated text alongside the original text, and potentially produce the translated text.

1.2 Requirements

- Relativity 6.9 or higher
- Relativity security access with the ability to edit/add fields
- Relativity Desktop Client

1.3 Directions

1. Create a field to hold the translated text using the following settings:
   - Field type - Long Text
   - Available in Viewer - Yes
   - Included in Text index - Yes (not required, but helpful when searching with the keyword index)

2. After the documents have been translated and are ready for import, use the Relativity Desktop Client to overlay the information into the newly created field. The translated text must exist in the load file and not as separate text files.

3. Index the translated text. The newly added text may not be immediately searchable. The following three options exist for indexing the text:
   - Build a new dtSearch index with only the new translated text.
   - Add this field to an existing dtSearch index, and then rebuild.
   - Set the Include in Text index attribute to Yes on the translated text field to add the content to the keyword index.

4. Configure your workspace to display both the translated and original text:
   a. Launch the standalone viewer.
   b. Change the Extracted text drop-down to the translated text field you created in step 1.
   c. Click the Sync button at the top of the standalone viewer to continuously update this window as you navigate through documents.

5. If there is an agreement to share the translated text, export the content as you would any other field.

- Creating and editing fields
- Importing through the RDC
- dtSearch
- Viewer
2 Adding documents to an existing Assisted Review project

2.1 Recipe Overview

Relativity allows you to add new documents to an Assisted Review project that is already in progress. This recipe discusses the suggested steps to add documents to a live project.

2.2 Requirements

- Relativity Assisted Review
- Analytics Core
- Analytics

2.3 Directions

To apply a categorized value to new documents, you must complete a round. It is easier to get immediate results if you have a round in process when beginning these steps. When you finish the round, the new documents get categorized from the examples already in the system. Generally it will be a QC round in progress, but any round type except Control Set will work. If you are not in the middle of a round, begin one and code at least one document so a round can be finished.

1. Load the new documents into the workspace.
2. Ensure the new documents are included in your saved searches for the Analytics index.
3. Navigate to the Indexing & Analytics tab, select Search Indexes and then the Analytics index.
4. From the Analytics Index console, rebuild the Analytics index to include the new documents. A full rebuild of the Analytics index may or may not be needed. Consider the following:
   - When adding new concepts or types of documents, a full rebuild is recommended.
   - If the new documents are more of the existing types of documents and the system does not need to learn anything new, an incremental population will probably suffice.
5. In the Assisted Review console, click the View Project Settings link. Ensure the saved search used for Documents to be categorized includes the new documents.
6. Once manual review is completed for the current round and the index is finished rebuilding, click Finish Round. The new documents will categorize alongside the old ones.
7. Return to the Project Home and verify that the total in the Docs in project column reflects the additional documents.
8. Continue the Assisted Review project as normal.

Note: Any previously established control set in the project will no longer be in sync with the document set. To correct this, make a new control set round to replace the previous one after the existing round is completed.

- Relativity Assisted Review
- Analytics
3 Cluster Visualization QC Workflow

Improve the efficiency and effectiveness of your quality control workflow by using cluster visualization to help identify potential coding discrepancies. This recipe describes how you can use cluster visualization to identify patterns and visually compare groups of conceptually similar documents to uncover potential coding inconsistencies.

3.1 Requirements

- Relativity 9.2 or higher
- An active Analytics index
- An existing cluster

3.2 Directions

This recipe involves a scenario where the case team has performed a first-level review for responsiveness. Cluster visualization can help QC the responsiveness coding. The approach described here could also be used to QC privilege coding or issue coding.
1. Ensure that the reviewed and coded documents are included in an existing cluster set.
2. From the Documents tab, select the cluster set in the cluster browser and click Visualize Cluster.

![Visualize Cluster](image)

3. In the search panel, click Add Condition and select the responsiveness coding field. Create a condition that will return documents coded as responsive. Click Apply, then click Run Search.
4. The system displays a heat map overlay on the cluster visualization that indicates the percentage of documents coded as responsive in each cluster. The darker the shading, the higher the concentration of responsive documents in the cluster.

5. Throughout this scenario, you can focus in on the documents in a cluster by right clicking the cluster and clicking Select, then Apply. You can select multiple clusters by doing this for each cluster you want to focus on.

6. The main cluster visualization view (the circle pack view) provides the following visual clues to aid in your QC review:
a. Darker clusters – The darker shading indicates that a cluster contains a high percentage of documents coded as responsive, meaning that the documents in the cluster have been coded consistently. Since documents within a cluster are conceptually similar, we would generally expect them to be treated similarly, and this is what the darker shading tells us.

b. Lighter clusters – The lighter shading indicates that only a small percentage of documents in the cluster have been coded as responsive. The documents in these lighter shaded clusters would warrant further investigation to determine why only a small portion of conceptually similar documents were coded as responsive.
7. Continue your QC review by right clicking a darker shaded cluster (one with a high percentage of responsive documents) and selecting View Nearby Clusters from the right click menu.

8. The Nearby Clusters visualization places the selected cluster in the middle of the screen and shows you other clusters that are conceptually similar to it. The closer a cluster is to the center cluster, the more conceptually similar it is.

9. The Nearby Clusters visualization reveals the following additional insights to help guide your QC review:
   a. Darker clusters – We would expect clusters that are nearby, or conceptually similar to, the center cluster to also have a darker shading, indicating that they too contain a high percentage of documents coded as responsive.
   b. Lighter clusters - Clusters with a lighter shading would probably warrant additional investigation. The system is telling us that these clusters are conceptually similar to the dark cluster in the middle (which contains a high percentage of responsive documents). However, only a small
portion of the documents in these lighter clusters have been coded as responsive; we might want to investigate why this is. A cluster that is completely white contains no responsive documents, yet is conceptually similar our center cluster, and should also be investigated.

Note that with the Responsiveness filter set and active, only documents coded as responsive are returned. If you would like to access all the documents in a selected cluster, simply clear the box on the filter card, click Run Search, and all documents within a cluster will be returned.

4 Coding the first item in a family group with the skip function

When a document in a family group is marked as responsive, the rest of the group is typically grouped as responsive. If the other family items are going to a second pass team for review, you can quickly prioritize which groups you send to the second pass team to determine privilege and importance.

4.1 Recipe overview

This recipe explains how to use a saved search, a view, and the document skip function to determine the next document available for review. By determining whether at least one item in a family has been coded as responsive, you can remove any other family items from the review queue based on that one coding decision, which is not propagated to the family group.

4.2 Requirements

This recipe is applicable to all versions of Relativity.
4.3 Directions

1. Create a new saved search and name it **Responsiveness is Responsive and Family**.
2. Choose the following search information and conditions:
   - Set **Includes** to **Include Family**.
   - Set **Scope** to **Entire Workspace**.
   - Set **Field** to **Responsiveness**, **Operator** to **any of these**, and **Value** to **Responsive**.

3. Create a new view:
   - a. Name the view.
   - b. Give the view any order number.
   - c. Set **Visible in Dropdown** to **Yes**.
   - d. Set **Object Type** to **Document**.
   - e. Include the fields to display.
4. Select the following conditions:
   - a. For the first condition, set **Field** to **Responsiveness** and **Operator** to **is not set**.
   - b. Select **AND**.
   - c. For the second condition, set **Field** to **(Saved Search)**, **Operator** to **Document is not in**, and **Value** to **Responsiveness is Responsive and family**.
5. Use the Skip function with this view. After the first item in a family group is coded, even without propagation set, the rest of the family group is removed from the review set without being coded. This way you can manually code each document without the need for a first level review.

- Views
- Document skip
- Saved search

### 5 Creating a read-only layout to lock produced documents

Rolling productions can potentially reverse the document coding that has already been produced. If you still need to review documents, you can't lock down these coding fields. However, you can make reviewed documents read-only to prevent changing coding on produced documents.

#### 5.1 Recipe overview

This recipe describes how to add an object rule to your Document object that allows documents to be viewed in a read-only layout.

#### 5.2 Requirements

- Applicable to all Relativity versions
- Relativity security access
  - Fields: Add
  - Object rules: Add
  - Ability to build layouts

#### 5.3 Directions

Perform the following steps to create a read-only layout:
1. From the Layouts tab, create a new layout with the following settings to display the documents that have been coded:
   a. Object type: Document
   b. Name: Previously Produced
   c. Order: 10

   Alternatively, you can copy the original coding form and change only the fields that should not be altered after production to read-only.

2. From the Fields tab, create a new field with the following settings to indicate which documents need to be locked down:
   a. Object type: Document
   b. Name: Produced and locked
   c. Field type: Single choice
   d. Available in field tree: Yes

3. From the Choices tab, add a single choice field with the following settings:
   a. Field: Produced and Locked (Document)
   b. Name: Yes
4. From the Administration tab, select Object type and select the Document object.
5. In the Rules section, click New and create a new rule with the following settings:
   a. Rule type: Default layout
   b. Name: Production lock documents
   c. Field: Produced and locked
   d. Value: Yes
   e. Action: Previously produced
   f. Leave User can select another layout unchecked.
6. Once you've completed steps 1-5, a system admin must switch the produced documents choice in the Produced and Locked field to Yes.
7. Examine the records. In the layout drop-down, there should only be the read-only version.

5.4 Considerations

- Make sure your reviewers don't have rights to edit layouts, as they can then switch read-only to No.
- This doesn't work on Mass Edit operations, and you must limit that right for users.

- Layouts
- Object rules

6 Cluster Visualization QC Workflow

Improve the efficiency and effectiveness of your quality control workflow by using cluster visualization to help identify potential coding discrepancies. This recipe describes how you can use cluster visualization to identify patterns and visually compare groups of conceptually similar documents to uncover potential coding inconsistencies.

6.1 Requirements

- Relativity 9.2 or higher
- An active Analytics index
- An existing cluster
6.2 Directions

This recipe involves a scenario where the case team has performed a first-level review for responsiveness. Cluster visualization can help QC the responsiveness coding. The approach described here could also be used to QC privilege coding or issue coding.

1. Ensure that the reviewed and coded documents are included in an existing cluster set.
2. From the Documents tab, select the cluster set in the cluster browser and click **Visualize Cluster**.

![Visualize Cluster](image)

3. In the search panel, click **Add Condition** and select the responsiveness coding field. Create a condition that will return documents coded as responsive. Click **Apply**, then click **Run Search**.
4. The system displays a heat map overlay on the cluster visualization that indicates the percentage of documents coded as responsive in each cluster. The darker the shading, the higher the concentration of responsive documents in the cluster.

5. Throughout this scenario, you can focus in on the documents in a cluster by right clicking the cluster and clicking Select, then Apply. You can select multiple clusters by doing this for each cluster you want to focus on.

6. The main cluster visualization view (the circle pack view) provides the following visual clues to aid in your QC review:
a. Darker clusters – The darker shading indicates that a cluster contains a high percentage of documents coded as responsive, meaning that the documents in the cluster have been coded consistently. Since documents within a cluster are conceptually similar, we would generally expect them to be treated similarly, and this is what the darker shading tells us.

b. Lighter clusters – The lighter shading indicates that only a small percentage of documents in the cluster have been coded as responsive. The documents in these lighter shaded clusters would warrant further investigation to determine why only a small portion of conceptually similar documents were coded as responsive.
7. Continue your QC review by right clicking a darker shaded cluster (one with a high percentage of responsive documents) and selecting View Nearby Clusters from the right click menu.

8. The Nearby Clusters visualization places the selected cluster in the middle of the screen and shows you other clusters that are conceptually similar to it. The closer a cluster is to the center cluster, the more conceptually similar it is.

9. The Nearby Clusters visualization reveals the following additional insights to help guide your QC review:
   a. Darker clusters – We would expect clusters that are nearby, or conceptually similar to, the center cluster to also have a darker shading, indicating that they too contain a high percentage of documents coded as responsive.
   b. Lighter clusters - Clusters with a lighter shading would probably warrant additional investigation. The system is telling us that these clusters are conceptually similar to the dark cluster in the middle (which contains a high percentage of responsive documents). However, only a small
portion of the documents in these lighter clusters have been coded as responsive; we might want to investigate why this is. A cluster that is completely white contains no responsive documents, yet is conceptually similar our center cluster, and should also be investigated.

Note that with the Responsiveness filter set and active, only documents coded as responsive are returned. If you would like to access all the documents in a selected cluster, simply clear the box on the filter card, click Run Search, and all documents within a cluster will be returned.

7 Creating an application for managing attorney lists

At the start of a project, it’s typical for reviewers to receive a list of attorneys associated with a case. To manage these lists as the review progresses, create a master list within Relativity, which can be easily updated and maintained.

7.1 Recipe overview

This recipe shows you how to create a single object with multiple fields on it, allowing you to import a CSV attorney list into your project.

7.2 Requirements

- Applicable to all versions of Relativity
- Relativity Desktop Client (RDC)
- Microsoft Excel attorney list (this can be saved as a CSV file)

7.3 Directions

1. Create a new object in your workspace called Attorney. You may wish to place this new tab under an Attorney List parent tab.
2. A Name field will automatically be created on the Attorney object. Change the Name field to Last Name, First Name.
3. Click **New Field** to create the following single-choice fields on the **Attorney** object:
   a. **Alphabet**  
      
      *Note: This field is optional.*
   
   b. **Law Firm or Employer**
   c. **Client**

4. Edit the **Attorney** tab view and layout to include the new fields.

5. In Excel, create or modify the attorney list you wish to import.
   a. Name the columns in the spreadsheet the same as the fields created above.
   b. The column that corresponds with the Alphabet field should be populated with the first letter of each attorney’s last name. This will facilitate sorting and quick access to the list for your reviewers.

   ![Excel Spreadsheet](image)


7. Launch and log in to the RDC.

8. Choose your workspace, and change the dropdown from the **Document** object to the **Attorney** object.

9. Choose your CSV file and import your list.

10. Enable filters to search on any combination of the newly created single-choice fields.

11. Add new attorneys as needed during review.

12. If necessary, add other fields and columns to match your workflow.

   - **Applications**
   - **Relativity Objects**
   - **Fields**
   - **Layouts**
   - **Relativity Desktop Client (RDC)
8 Creating an optimized Relativity Analytics index (8.2 and higher)

Published October 20, 2016

Relativity Analytics uses only the documents you provide to make a search index. Because no outside word lists are used, you must create Saved Searches to dictate which documents are used to build the index. Generally, one index is built for each workspace. However, if you wish to limit search results to certain document groups or have more than one language in the document set, multiple indexes might give you better results.

8.1 Requirements

- Relativity Analytics 8.2 or higher
- Existing decimal-type field

8.2 Directions

1. The key search is based on text size. Your processing tool might provide the text size. If not, run the Set Extracted Text Size script. This updates a decimal field with the text size in KB.
2. Create a saved search of the data and (name it Analytics Data Search or something similar).
   - Bring back all text less than 30 MB which translates to “Extracted Text Size is less than 30,000 and Extracted Text Size is greater than 0. (This will return documents that have had this script run and nothing too large that will encumber the system.)
   - Ensure that the Extracted Text field is the only one returned.
3. Run Structured Analytics > Repeated Content Identification on your Analytics Data Search.
   - Set the Minimum number of occurrences to 0.5% of the total population. Leave the rest of the settings to the default.
   - Scan the results of repeated content identification to ensure that the repeated text blocks correspond to unwanted (non-authored) content.
4. Create a new Analytics profile:
   - Set the Email header filter to Yes.
   - Set OCR and Go Words to No.
   - If you intend to run email threading, you can match up fields for Structured Analytics email threading.
   - Save the Analytics profile and link any repeated content filters which comprise non-authored content.
5. Create your Analytics index
   - Use the Analytics Data Search created above for both training and searchable (unless it is for a very large workspace).
   - Select the Analytics profile created above.
   - Set Optimize Training Set and Auto Remove Signatures to Yes.
6. Build your index.
The following are explanations of the various settings selected above.

- **Email header filter** - removes email headers throughout email-formatted text.
- **Repeated Content** - removes blocks of text, such as email footer information. Instances of repeated content may be detected and converted into filters automatically by a structured analytics set.
- **Optimize Training Set** - this option automatically excludes poor quality index training documents based on an analysis of their text. Commonly excluded content includes extremely large documents and files predominated by tables of numbers, long garbage strings of characters, and a preponderance of symbols (rather than words).
- **Auto-removal of email signatures and footers** - this option automatically excludes email signatures and footers when it finds them (although it is intentionally conservative to avoid over-exclusion. This feature works for English signatures and footers only.

### 8.2.1 Important Points

- Single word queries tend to not return valid results, except with keyword expansion.
- Analytics indexes are not available for searching when they are in the build phase of an incremental update or if they are completely repopulated. A second inactive index can serve as a backup and will allow for continued access during updates. But remember—this secondary index will take up server space.
- Prior to population, remember to link the desired repeated content filters to the Analytics profile to be used with your index.

### 9 Creating commonly used Pivots

Relativity Pivot is a powerful feature that can help you visualize your data set in a digestible way. Through interactive charts and graphs, Pivot allows you to gain insight into trends and patterns in your case. The following suggestions describe how to make the most of Pivot.

#### 9.1 Recipe overview

This recipe describes a number of Pivot charts that may be useful for analyzing data throughout your project.

#### 9.2 Requirements

Relativity 7.0 or higher

#### 9.3 Directions

Set the following fields as suggested to create a number of useful Pivot charts.
<table>
<thead>
<tr>
<th>Group by</th>
<th>Pivot on</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Custodian</td>
<td>Sent date</td>
<td>Review sent dates of emails for each custodian to verify a date range is not</td>
</tr>
<tr>
<td></td>
<td>File type</td>
<td>Verify all documents were collected, including emails and loose files.</td>
</tr>
<tr>
<td>Custodian</td>
<td>STR-Keywords</td>
<td>Examine the count of keywords that occur by custodian.</td>
</tr>
<tr>
<td>Sent date</td>
<td>Issues</td>
<td>Examine the occurrence of issues over a time period.</td>
</tr>
<tr>
<td>Batch::Assigned</td>
<td>Responsiveness</td>
<td>View reviewer progress of responsive items across batches.</td>
</tr>
<tr>
<td>Cluster</td>
<td>STR-Keywords</td>
<td>Find out which clusters or document groups contain the most interesting</td>
</tr>
<tr>
<td>Authordomains</td>
<td>Recipient domains</td>
<td>Dive into internal emails, potentially privileged documents, or find</td>
</tr>
<tr>
<td>Authordomains</td>
<td>Issues or STR-Keywords</td>
<td>Determine the most important domains where issues or keywords appear.</td>
</tr>
</tbody>
</table>

10 Creating secured saved search folders for multiple groups

When two groups need to add saved searches to your workspace, without sharing them with each other, you can secure the individual saved searches.

10.1 Recipe overview

This recipe describes how to create two secure folders to prevent two groups from seeing each other’s saved searches.

10.2 Requirements

Applicable to all versions of Relativity.

10.3 Directions

Set workspace security as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>Security Setting</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &amp; 2</td>
<td>Search: View</td>
<td>This ensures users in these groups can access the saved search browser, but can't save searches outside of their assigned folders.</td>
</tr>
<tr>
<td>1 &amp; 2</td>
<td>Search Folder: View</td>
<td></td>
</tr>
<tr>
<td>1 &amp; 2</td>
<td>Saved Search</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Browser Access</td>
<td></td>
</tr>
</tbody>
</table>
1. Create two folders in the saved search browser, one for each group.
2. Secure the first search folder to remove Group 2, and then set permissions for Group 1 as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>Security Setting</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Search: Edit and Add</td>
<td>This allows the user to save searches only in this folder. You can substitute Delete for the Edit permission if you want to allow the users to delete their own saved searches or sub-folders.</td>
</tr>
<tr>
<td>1</td>
<td>Search Folder: Edit and Add</td>
<td></td>
</tr>
</tbody>
</table>

3. Repeat step 2 for the search folder for Group 2.

**Note:** If you need to make any modifications to either the folders or saved searches within the secured folders, you must address these changes individually at the folder level and not in workspace permissions.

- [Saved search](#)

### 11 Creating static searches using multiple-choice fields

As databases update with additional documents, or as edits are made, saved searches provide different results that include new documents. To maintain the same search results from before a database update, you can create a field to store the results.

#### 11.1 Recipe overview

This recipe describes how to create a multiple-choice field and populate it with records in a static search. It also explains how to make this new field visible in the field tree to manage and review your static searches.

#### 11.2 Requirements

- Applicable to all Relativity versions
- Mass updates
- Relativity security access with the following permissions:
  - **Fields**: edit/add
  - **Choices**: edit/add
  - **Document**: edit

#### 11.3 Directions

1. Create a new multiple choice field with the following settings to hold the items for each search.
   - **Object Type** - Document
   - **Name** - Static Searches
Field Type: Multiple Choice
Available In Field Tree: Yes

2. Add a choice to this field, and then name it after the appropriate search.
3. Build, and then run the desired search.
4. Mass update the Static Searches field with the choice you created in step 2.
5. Review the Field Tree and the Static Searches field to view your new search choices and results.

Fields
Saved search

12 Creating training resources tabs

While in Relativity, users can access various resources by adding tabs that display training resources so that users don't have to leave the Relativity environment and can quickly jump back into their workspaces once they’ve found the information they need.

The first recipe shows you how to create tabs in your workspace that contain documentation, webinars, and recipes. The second recipe helps you import a training resources application, which creates a tab containing all of Relativity’s training and reference materials within the Relativity interface.

12.1 Creating training resources tabs

12.1.1 Requirements

This recipe is applicable to all Relativity versions.

12.1.2 Directions

1. From the home screen, click Admin Workspace Configuration > Tabs.
2. Click New Tab and create the Training Resources parent tab:
   a. Name the tab Training Resources.
   b. Set Order to 10.
   c. Set Link Type to Parent.
3. Click New Tab and create the On-Demand Videos tab:
   a. Name the tab On-Demand Videos.
   b. Set Order to 10.
   c. Set Link Type to External.
   d. Set Parent to Training Resources.
   e. Set Link to https://www.relativity.com/relativity/ediscovery-resources/training/.
4. Click New Tab and create the Documentation tab:
   a. Name the tab Documentation.
   b. Set Order to 20.
   c. Set Link Type to External.
   d. Set Parent to Training Resources.
   e. Set Link to the appropriate URL based on your Relativity version:
5. Click **New Tab** and create the Recipes tab:
   a. Name the tab **Recipes**.
   b. Set Order to **30**.
   c. Set Link Type to **External**.
   d. Set Parent to **Training Resources**.
   e. Set Link to the appropriate URL based on your Relativity version:
      - 9.5: https://help.relativity.com/9.5/Content/Site_Resources/RecipesHome.htm
      - For earlier versions, locate the appropriate link on the documentation archives page:

### 12.2 Adding training resources to a workspace

#### 12.2.1 Requirements
Applicable to all versions of Relativity.

#### 12.2.2 Directions

1. Visit the **Relativity Community** and search for **RA_Training**.
2. Download the training application.
3. From the Applications tab in either the workspace or admin mode, install the application to your desired workspace(s).
4. Once the application installs successfully, a new tab, **Training Resources**, appears in the workspace(s) you selected. You will also see the sub-tabs for **Tutorials**, **Videos**, and **Documentation**. Here, you can navigate to these tabs to access training resources without leaving Relativity.

- **Tabs**
- **Relativity applications**
13  Displaying family groups in a view

During review, it may be helpful to quickly see documents as they exist in their family groups. Building a view to organize documents by family can support this workflow.

13.1 Recipe overview

This recipe describes how to create a view to display documents organized by family.

13.2 Requirements

- Workspace access with permission to the following tabs:
  - Administration | Views

13.3 Directions

Create a view with the following settings:

13.3.1 Information

View information:

- Object Type: Document
- Name: Documents
- Owner: Public
- Order: 10

Other:

- Group Definition: Family Group (Family related items field)
- Relativity Applications: none
- Query Hint: leave blank
- Visible In Dropdown: Yes

13.3.2 Fields

Suggested fields - Include Edit, File Icon, Control Number, Group Identifier, File Name (or Name or Unified Title), Record Type, and any additional fields you want to place on the view

13.3.3 Conditions

Set conditions - None

13.3.4 Sort

It is necessary to sort based on the related item field. It might be Group Identifier or Family Group.
Then how the families are displayed is optional.

- Control Number
- Family Date - Use the Propagate Sent Date to Family Documents script to set values for this field.

Save the View.

Final View:

<table>
<thead>
<tr>
<th>Number</th>
<th>Edit</th>
<th>Control Number</th>
<th>Family Date</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Edit</td>
<td>JGRIFFITHD0000090007</td>
<td>NEW WEATHER SWAPS ON THE INTERCONTINENTAL EXCHANGE</td>
<td>11/16/2001 3:52 PM</td>
</tr>
<tr>
<td>11</td>
<td>Edit</td>
<td>JGRIFFITHD0000090007:0001</td>
<td>Enron Announcements/Corp/Enron@ENRON</td>
<td>9/14/2001 5:27 AM</td>
</tr>
<tr>
<td>12</td>
<td>Edit</td>
<td>JGRIFFITHD0000090009:0002</td>
<td>Issler Paulo</td>
<td>10/29/2001 1:27 AM</td>
</tr>
<tr>
<td>14</td>
<td>Edit</td>
<td>JGRIFFITHD0000090009:0005</td>
<td>Mandatory Harassment Avoidance Training</td>
<td>9/14/2001 5:27 AM</td>
</tr>
<tr>
<td>15</td>
<td>Edit</td>
<td>JGRIFFITHD0000090009:0006</td>
<td>Historical Correlation Calculator</td>
<td>10/29/2001 1:27 AM</td>
</tr>
<tr>
<td>16</td>
<td>Edit</td>
<td>JGRIFFITHD0000090010</td>
<td>FW: Presentation Announcement</td>
<td>10/29/2001 1:27 AM</td>
</tr>
<tr>
<td>17</td>
<td>Edit</td>
<td>JGRIFFITHD0000090011</td>
<td>Prices for Kroger</td>
<td>10/29/2001 1:27 AM</td>
</tr>
<tr>
<td>18</td>
<td>Edit</td>
<td>JGRIFFITHD0000090011:0001</td>
<td>Credit Watch List - Week of 10/29/01</td>
<td>10/30/2001 1:34 AM</td>
</tr>
<tr>
<td>19</td>
<td>Edit</td>
<td>JGRIFFITHD0000090012:0001</td>
<td>Enron Announcements/Corp/Enron@ENRON</td>
<td>10/30/2001 1:34 AM</td>
</tr>
<tr>
<td>20</td>
<td>Edit</td>
<td>JGRIFFITHD0000090013:0001</td>
<td>Lexis-Nexis Training: Houston &amp; Worldwide / Dow Jones Training</td>
<td>10/30/2001 1:34 AM</td>
</tr>
<tr>
<td>21</td>
<td>Edit</td>
<td>JGRIFFITHD0000090014</td>
<td>Florida Spread Option</td>
<td>10/30/2001 1:34 AM</td>
</tr>
<tr>
<td>22</td>
<td>Edit</td>
<td>JGRIFFITHD0000090014:0001</td>
<td>10/29/2001 11:42 PM</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The screen shot is not intended to show as indented (with dots), but to illustrate the view you see when you use the indented list method to group the families together. Notice in the view that each family is now divided by bolded lines.

- **Views**
- **Indented lists**

### 14 Email threading and near dupe - workflow alternatives

Relativity workspaces often have a need for email threading and document near dupe detection. System admins need to decide whether both functions should be addressed together or independently.
14.1 Recipe overview

This recipe assists you in making an informed decision that is right for your workflow. This includes suggested workflows for Email Threading and Near Dupe detection with examples of how to utilize applicable Relativity functions.

14.2 Requirements

- Applicable to Relativity 8.0 or higher
- Workspace & system admin rights

14.3 Directions

The following sections contain three workflow alternatives to help system admins who decide to run a Structured Analytics Set (SAS) involving Email threading and/or Near Dupe detection on a data set.

14.3.1 Running the Email Threading & the Near Dupe in a single SAS

Used this workflow to identify email duplicate spares independent of the duplicate loose documents. In other words, use this where the case demands that you handle the workflow for emails and loose files separately. If both options are selected on a single set, only run near duplicate identification against non-emails (attachments and loose files).

Once you run the SAS, you can create a document view or a batch of documents that contain the documents flagged inclusive (Inclusive Email = Yes) and marked as non duplicate email (Email Duplicate Space = No) for email documents.

You can then identify the loose documents as the documents flagged as Text Duplicate Principle (Text Near Duplicate Principal = Yes).

14.3.2 Running only Email Threading in an SAS.

This workflow works well if you want to perform email threading and email duplicate detection on an email only document set. For example:

- A case where emails are the only important documents. The case team wants to see only the most inclusive, non-duplicate emails. Running a SAS with just email threading would be best.

14.3.3 Running only Near Dupe in an SAS.

Use this workflow to identify duplicates whether they are emails or loose files. For example:

- Two parties in discovery produce documents. There is an agreement that each is to review documents that were outside the set provided to the opposite side. In this case, the workflow requires you treat all documents as loose documents and identify all near duplicate groups of single documents (just a Near Duplicate Principal and no other documents in the group). You can then isolate all such documents via
a saved search, placed into a singular view or into review batches.
- A case team must identify duplicates across two or more groups of produced documents. Running the Near Dupe process would be ideal. This allows the team to review only the necessary documents—the Near Duplicates—while the other documents in the group are the duplicates.

Analytics

15 Estimating disk space utilization for databases and Analytics indexes

Published July 19, 2013

When working with Relativity workspaces, system admins are often asked to estimate the disk space and the cost associated with a Relativity database and Analytics index. This recipe provides some workflows and associated scripts that can help you calculate this for your workspaces.

15.1 Requirements

- Relativity 7.0 or higher
- Workspace with system admin-level access

15.2 Directions

The rule of thumb measurements and methods below help system admins estimate the size of their Relativity database and associated Analytics indexes.

It’s noteworthy that these metrics have a high standard deviation and tend to scale up or down linearly. Every data set is different, so please note that this is only an average. Files will vary in size considerably.

15.2.1 Workspace sizes

On average, every 100,000 records consume about 10 GB of database space. Database disk space usage is roughly one third of the total file size, including natives and images, for a workspace.

In addition to the above approximation, system admins can execute the Billing Statistics – Case Roll-up script that is preinstalled in every Relativity environment. This script helps you calculate accurate numbers on the size of your SQL database and the size of the files associated with the workspace.

Note: This script can only be run in Home mode.

15.2.2 Analytics index sizes

With default Analytics profile settings for an Analytics index, disk space usage is approximately 20 percent of the size of the corresponding SQL database’s MDF file.
To get an accurate measurement, system admins can use the Relativity Analytics Billable Estimate script to determine the precise size of an Analytics index. This script runs on a workspace level against a saved search of documents submitted to the Analytics index.

Analytics

16 Exporting PDFs using document identifiers as file names

Exporting documents in PDF format is a good way to create a portable subset of documents that you can use when you can't access your Relativity environment, or when you need to share documents with other parties. For ease of identification, you should create a cross-reference between the files and the documents in Relativity, which you can do by naming the PDF files after the Document Identifier field.

16.1 Recipe overview

This recipe describes how to export documents from Relativity in PDF format using the Relativity Desktop Client and a saved search.

16.2 Requirements

- Workspace access with the following permissions:
  - Search edit/add
  - Corresponding search browser
- Relativity Desktop Client

16.3 Directions

1. Create a saved search with the documents to export.
2. Confirm all documents in the saved search have images created.
3. Use the Relativity Desktop Client to export the saved search.
   a. Select Tools | Export | Saved Search.
   b. Select the location for the export.
   c. Under the Image section, select the following settings:
      - Data File Format - Select either Opticon or IPRO
      - File Type - PDF
   d. Under the Metadata section, select the following settings:
      - Data File Format - Select the format that suits your needs.
      - Data File Encoding - Select the encoding that suits your needs.

The export consists of two load files and PDF files named by the Document Identifier.
17 Exporting to HTML for expert witness review

This recipe describes how to share documents and their metadata with an expert witness without giving the witness access to Relativity. Using the HTML export feature and a CSV file, you can review documents and their metadata, as well as enter comments to later overlay into Relativity.

17.1 Recipe overview

This recipe walks you through the process of exporting documents to an HTML format that include PDF, TIFF, or Native files.

17.2 Requirements

- Applicable to all versions of Relativity
- Workspace access with the following permissions:
  - Saved search
  - Print to TIFF
  - Export to File mass action
- Relativity Desktop Client (RDC)

17.3 Directions

1. Create a saved search containing the documents you would like to review.
2. Image the documents.

   **Note:** To create PDF files—which may be easier to review—make sure you have TIFF images available.

3. In the RDC, select to export a saved search.
   a. Select the search you’ve created for this project, or click the ellipsis to use production images.

      **Note:** Fields are automatically brought over from the saved search.

   b. Populate your destination settings according to your preferences.
      i. Change the Metadata: Data File Format to HTML (.html).
      ii. For file type, you can select PDF, TIFF, and/or Natives.
17.3.1 Results

The result is an HTML document that includes all of your selected fields and hyperlinks to the image files.

17.3.2 Offline coding form

If your expert witnesses need to make any comments while reviewing these records, create a CSV file export from the saved search.

1. Edit your search to show the control number.
2. Using mass operations, export this list of control numbers to a CSV file.
3. Track their comments in this file.
4. Overlay this information into a new document field.

- Exporting with the RDC
- Saved search
18 Filtering to find empty fields

Occasionally, you may need to locate documents that are missing field content. You can use advanced filtering options to find text fields that don't contain data.

18.1 Recipe overview

This recipe describes how to use a custom filter type to identify empty fields.

18.2 Requirements

This recipe is applicable to all versions of Relativity.

18.3 Directions

1. Under the Administration tab, select Fields.
2. Click Edit for the desired text field.
3. Change the Filter Type to CustomOnly.
4. Click Show Filters from the view.
5. Select (Advanced...) from the field filter’s drop-down to activate the CustomOnly filter.
6. Select is not set from the value drop-down and leave the text query field blank.
7. Click Apply to see the fields that do not contain data.
19 Finding duplicates or near duplicates in a new set that match up to a previous set

When a new set of documents arrives, it can be important to know how the documents match up to the current set that have been reviewed. This recipe can be useful for the following scenarios:

- Finding the overlap between one document set and another.
- Finding the documents that match up to the coded items in your initial set.
- Comparing items to determine the coding process.

19.1 Requirements

- Structured Analytics
- Saved searches

19.2 Directions

1. Define the old and new datasets by marking a field or using the prefix of the ID.
2. Create a Structured Analytics set of Near Duplicate items across old and new data.
3. Create a saved search of the tagged items in the old document set and include the condition that Textual Near Duplicate Group is set. Add the include Near Duplicate Items from the Relational Items drop down to see both the old information and new data that contains near duplicates of the original data.
4. Create a second search with the first condition as \( \text{(Saved Search)} = \text{First Search} \). The second condition is to return all the new documents.
5. This search returns documents that are in the new set which match up as duplicates or near duplicates of the requested items in the first set. You can include the Near Duplicates to this search to see both the original and new set items.

19.3 References

- Structured Analytics
- Searching

20 Finding emails exchanged within the same domain

During the course of developing your case strategy, you might need to identify the amount of email traffic within the same company or between two companies. This can help identify some of the main players in the case along with communication patterns. Using Relativity, you can identify email communications at the company domain level and analyze those communications using sender and recipient information.
20.1 Recipe overview

This recipe describes how to use a transform set to parse email domain information from your data and use Pivot to analyze the volume of communications between or among companies.

20.2 Requirements

- Relativity security access:
  - Field tab with add/edit permission
  - Object type tab with add/edit permission
  - Transform set tab with add/edit permission
  - Transform add/edit permission
  - Use Pivot/Chart permission

20.3 Directions

1. From the Administration tab, click the Object Type sub-tab.
2. Click New Object Type and create a new object type with the following settings:
   a. Name: Domain
   b. Leave all other fields as default
3. Create a new field with the following settings to hold the domain information:
   - Object type: Document
   - Name: Recipient Domains
   - Field type: Multiple Object
   - Associative Object Type: Domain
   - Allow Group By: Yes
   - Allow Pivot: Yes
   - Leave all other fields as default

   **Note:** The object field type must be Multiple Object for the transform set to run.

4. Create a new field with the name Author Domains.
5. Create a saved search that serves as the data source for your transform set.
   a. From the Documents tab, click the Saved Searches browser icon to create a new search with the following settings:
      - Name: Domains Saved Search
      - Conditions: Set conditions so that the saved search only returns documents for which your Email From (the field that stores the author email address) or Email To (the field that stores recipient email addresses) fields are populated. This allows the transform set to run as efficiently as possible.
        - Email From::is set OR
        - Email To::is set
      - Fields: Select only the Email From and Email To fields.
6. From the Transform Sets tab, click New Transform Set and create a transform set with the following settings:
- Name: Domains
- Handler: Domain Parsing
- Data Source: Domains Saved Search (the saved search you set up in the previous step).

7. Click New next to the transform and enter the following:
   - Transform set: Domains
   - Name: Author Domains
   - Source field: Field that stores the author email address (e.g., Email From or Email Author)
   - Destination field: Author Domains field created above (where Object Type = Document)

8. Click New next to the transform again and enter the following:
   - Transform set: Domains
   - Name: Recipient Domains
   - Source field: Field that stores the Recipient email address (e.g., Email To or Recipients)
   - Destination field: Recipients Domains field created above (where Object Type = Document)

   You can repeat this process with the CC and BCC field domains in the Recipient Domains field. You just have to create a transform to do so. For this example, we concentrate on just the two fields we already set up.

9. Click Start Transform: Full on the Transform Set Console. This runs through the workspace and extracts domains from the email address field. When this process is finished, the status changes to Completed.

   Now you can use Pivot to determine the overlap between the two fields containing the author and recipient domains.

10. From the Documents tab, click the Pivot icon.
    a. Turn on Grid and List.
    b. Group by Author domains and Pivot on Recipient domains.

11. Click Go.
If you scroll to the author domain enron.com and then go across to the recipient domain enron.com, you can see in our example that 68 items are from and to people within the Enron domain. You can also review those 68 documents from the list view beneath the grid.

### 21 How to calculate Precision and Recall without a control set

#### 21.1 Recipe overview

Relativity will automatically calculate Precision and Recall for Assisted Review projects that have a control set. However, the control set is an optional round type, and not recommended for all projects, specifically (though not limited to) situations where documents continue to be added or removed throughout the process.

There are situations, however, where you might wish to calculate these values without a control set, or alternatively other instances where you feel that your control set is not a reliable model of your document universe.

Although this calculation is not performed for you automatically under these conditions, the process is very simple, requires just a few fields on a view, some filtering, and very basic math.
21.2 Requirements

- Relativity Assisted Review
- Analytics
- View / Edit permission

21.3 Precision and Recall Defined

Precision is a measurement of system accuracy. Ask yourself: Of everything the system thought was Responsive, how much of it truly was Responsive according to the human reviewers?

The equation:

\[
\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}
\]

Recall is the measurement of completeness. Ask yourself: Of everything Responsive that the system was supposed to find, how much did it actually find?

The equation:

\[
\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}
\]

The following table illustrates the meaning of terms like “True Positive”:

<table>
<thead>
<tr>
<th></th>
<th>Responsive (System)</th>
<th>Non-Responsive (Human)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responsive (System)</strong></td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td><strong>Non-Responsive (System)</strong></td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

Consider the illustration below:

- The entire rectangle represents all of the documents of a control set, or for the purposes of this exercise, a QC round which is being used as a control set.
- The circle contains all documents which the system has categorized as Responsive.
- Blue documents, regardless of their location, are truly Responsive, as decided by the human reviewer.
  - The blue documents inside the circle are True Positives.
  - The blue documents outside the circle are False Negatives.
- Orange documents, regardless of the location, are truly Non-Responsive, as decided by the human reviewer.
  - The orange documents inside the circle are False Positives.
  - The orange documents outside the circle are True Negatives.
Under different circumstances, some of the documents outside the circle might also be uncategorized. This workflow focuses solely on categorized documents, however.

21.3.1 Initial setup

The purpose of a control set is to serve as a miniature model of your document universe where the documents in the set are not used as seeds which influence categorization. They can however, be categorized by seed documents from other rounds. This lack of direct influence allows the system admin to gauge the accuracy and success of the project through the control set categorization results.

However, it is possible to use a QC Round as an ad hoc control set. In order to do so, the following conditions are required:

- **The QC Round must be a random sample.** This is true of all QC rounds, and this recipe workflow is no exception. Remember, judgmental sampling is typically not appropriate for QC rounds.
- **Only sample from** [Documents to be categorized] - Categorized, where [Documents to be categorized] is the saved search in your project settings. This saved search is comprised of all your categorized documents for both categories (such as Responsive and Not Responsive).
- **Coding for the round should be complete, but the round itself should not be finished, and post-round categorization should not be triggered.** Once you finish the current round, the documents in it become examples, and are no longer viable for this ad-hoc control set.
21.3.2 View setup

- Control Number
- RAR Sample Set
- Designation (this name may vary but should be your human reviewer designation field)
- Categories - XXX RAR Designation Cat. Set (XXX represents the project prefix)

![View setup table]

21.3.3 Calculating Precision

There are two ways to calculate Precision.

**Method 1:**

The first method is extremely simple because Precision is the mathematical complement of your Responsive Overturn Rate. In other words: 100% - Responsive Overturn % = Precision %.

To calculate Precision using this method:

1. Navigate to the Overturn Summary Report and scroll down to the section called Overturns by Designation Category.
2. Identify the Responsive overturned docs percentage for the current round.
3. Subtract this value from 100% to calculate your Precision.

![Overturns by Designation Category table]

So, for the above example, 100% - 26.54% (Responsive Overturn) = 73.46% (Precision).

**Method 2:**

This method involves filters on the view which was set up earlier.
To calculate Precision using this method:

1. Navigate to your Assisted Review document view; ensure before you start that all filters have been cleared.
2. Filter on the **RAR Sample Set** field for the current round. Again, coding should be complete but the round should not have been Finished/Categorized.
3. **Ask yourself the question:** Of everything the system thought was Responsive, how much of it truly was Responsive according to the human reviewers?
4. Filter for Responsive on the **Categories - XXX RAR Designation Cat. Set** field. This is everything the system thought was Responsive. Make a note of the number of documents returned. This will be the denominator for your calculation.
5. Filter for Responsive on the **Designation** field. Make a note of the number of documents returned. This will be the numerator for your calculation.
6. Divide the numerator by the denominator. This is your Precision.

Example, Precision denominator (True Positive + False Positive): 456 documents:

![Image of a screen with RAR Sample Set, Designation, and Categories fields]  

Example, Precision numerator (True Positive): 335 documents:
Example, Precision calculation: $335 \div 456 = 73.46\%$.

21.3.4 Calculating Recall

This method involves filters on the view which was set up earlier. To calculate Recall using this method:

1. Navigate to your Assisted Review document view; ensure before you start that all filters have been cleared.
2. Filter on the RAR Sample Set field for the current round. Again, coding should be complete but the round should not have been Finished/Categorized.
3. Ask yourself the question: Of everything Responsive that the system was supposed to find, how much did it actually find?
4. Filter for Responsive on the Designation field. This is everything the reviewers determined to be Responsive this round. Make a note of the number of documents returned. This will be the denominator for your calculation.
5. Filter for Responsive on the Categories - XXX RAR Designation Cat. Set field. Make a note of the number of documents returned. This will be the numerator for your calculation.
6. Divide the numerator by the denominator. This is your Recall.

Example, Recall denominator (True Positive + False Negative): 630 Documents:
Example, Recall numerator (True Positive): 335 documents:

Example, Recall calculation: $335 \div 630 = 53.17\%$. 
21.3.5 Additional considerations

- Ensure that filters are applied in the proper order as failing to do so will give incorrect results.
- Note that numerator is the same (True Positive) for both equations.
- It is also important to understand that these calculations apply only to your categorized population, and do not include your uncategorized documents. These documents are often removed from projects and/or reviewed separately, essentially negating the need to determine their Precision and Recall values. Please note, however, this differs from the automatic control set calculation, which does account for uncategorized documents.
- It is also possible to calculate Precision from a Responsive-only QC Round, as all of the required information will still be available. Please note that a Recall calculation will not be possible under these circumstances, however.

References

- Assisted Review
- Analytics
- Views
- Filters
- Views
- Filters
- Assisted Review
- Analytics

22 Identifying emails between two or more specific individuals

This recipe describes how you can use dtSearch and search terms reports to identify emails in which two or more individuals were involved in the correspondence. This includes emails between the specific individuals, as well as instances in which the individuals are only in the recipient fields.

22.1 Requirements

- Relativity 8.2 or higher

22.2 Directions

This recipe involves a scenario where the case team has a list of specific individuals and only wants to review emails in which two or more of these individuals were involved in the email correspondence.
1. Create a custom index containing only the correspondence fields (To, From, CC, BCC):
   a. Create a saved search that returns only the correspondence fields (To, From, CC and BCC).
   ![Saved Search Image]
   b. Create a custom dtSearch index using this new saved search as your searchable set.
   ![Custom Index Image]

2. Create a new Search Terms Report (STR).
   a. Select the custom dtSearch Index that contains only the correspondence fields.
   b. Select Report and tag for Type.
   c. Ensure that Calculate Unique Hits is set to Yes.
d. Click **Save**.

![Search Terms Report](image)

e. Click **Modify Terms**, and then paste in the email addresses for the individuals you are trying to identify.
f. Click **Add Terms**, and then click **Done**.

g. Click **Run All Terms**.

3. Create a saved search based on the two STR fields that were created (setting Calculate Unique Hits to Yes creates a second STR field) when the STR completes. In the following example, the conditions are as follows:

   a. STR – Email Addresses_v1 – Unique “not these conditions” STR – Email Addresses_v1 Unique – is set.
      
   AND

   b. STR – Email Addresses_v1 “these conditions” STR – Email Addresses_v1 – is set.

   c. Ensure the STR fields are selected as fields returned in your saved search. You may also want to
include the To, From, CC, and BCC fields for reference.

You can now set up a workflow for reviewing these emails in which two or more individuals were identified.
23 Identifying standalone emails

This recipe describes how you can use saved searches after email threading is complete to identify emails that have not been replied to or forwarded. You can then batch these documents together in order to expedite the review process.
23.1 Requirements

- Relativity 8.2 or higher
- Existing email threading set

23.2 Directions

This recipe involves a scenario in which the case team has performed email threading and would now like to identify any standalone emails.

1. Ensure email threading is complete.
2. To find all emails that have not been replied to or forwarded, you need to return all emails that have Indentation equal to 1, but where there is no Indentation greater than 1 within the thread group. This requires you creating three saved searches:
   a. Saved search 1 contains documents in which Indentation is equal to 1 and thread group is included.
   b. Saved search 2 contains documents in which Indentation is greater than 1 and thread group is included.
<table>
<thead>
<tr>
<th>Information</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name:</td>
<td>Indentation &gt; 1 + Thread Group</td>
</tr>
<tr>
<td>Owner:</td>
<td>Public</td>
</tr>
<tr>
<td>Scope:</td>
<td>Entire Workspace</td>
</tr>
<tr>
<td>Notes:</td>
<td>Currently searching entire workspace</td>
</tr>
</tbody>
</table>

**Conditions**

1. Indentation
   - Is greater than 1

**Save All Conditions**

---

c. Saved search 3 gives you the final results and is a nested search of saved searches 1 and 2. The conditions for this search should be as follows:

- Document is in saved search “Indentation is 1 + Thread Group”
- Document is not in saved search “Indentation > 1 + Thread Group”
- Sort the results by Group Identifier to keep families together

<table>
<thead>
<tr>
<th>Information</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name:</td>
<td>Standalone Emails</td>
</tr>
<tr>
<td>Owner:</td>
<td>Public</td>
</tr>
<tr>
<td>Scope:</td>
<td>Entire Workspace</td>
</tr>
<tr>
<td>Notes:</td>
<td>Currently searching no folders</td>
</tr>
</tbody>
</table>

**Conditions**

1 AND 2

1. *(Saved Search)*
   - Document is in Indentation is 1 + Thread Group

2. *(Saved Search)*
   - Document is not in Indentation > 1 + Thread Group

**Clear All Conditions**
The result of saved search 3 are the standalone emails which have not been replied to or forwarded.

- **Saved Search**
- **Email Threading**
- **Analytics**

# 24 Importing productions in Relativity with natives

When working with productions in Relativity, system admins may need to make documents available as images and natives to expert witnesses. To do this, system admins will import the production back into Relativity as images (along with the natives) when natives are available. This creates duplicate files, which use more storage space.

## 24.1 Overview

This recipe shows you how to set up your workspaces to support the import of productions with natives when natives are available. This approach conserves disk space by storing only one copy of the images and the natives.

## 24.2 Requirements

- Workspace access
  - Documents: Add, Edit, Delete
  - Images: Add, Delete
- Relativity Desktop Client

## 24.3 Directions

1. When setting up and importing into the workspace, be sure that the files are available in a shared repository accessible from your Relativity environment, and import the data using the links to the repository.
2. Create a saved search for the production set.
3. Export the saved search. Make sure to:
   - Include all fields necessary for your project.
   - Clear the **Copy Files from Repository** box when exporting from the RDC.

4. Export your production set. Make sure to:
   - Include all fields necessary for your project.
   - Clear the **Copy Files from Repository** checkbox.
5. Create a new folder in your workspace. Secure the folder so only your expert witness group can access it.
6. Import the production set you used in step 4 into the new Productions folder. Be sure to select the following File Repository Preferences setting:
7. Overlay the natives using the load file and the export from the saved search in step 3. Be sure to select the following File Repository Preferences setting:

Note: You must perform the imports (Steps 6 and 7) on a machine that has access to the repository to avoid import errors. The import of the natives (step 7) executes successfully only when natives have been imported into the workspace (step 1) using the files that already reside in a valid location that Relativity can access.

Importing through the Relativity Desktop Client

25 Legal Hold checklist

When working with Relativity Legal hold, it's helpful to have a checklist of items you can address to expedite the setup time in Relativity. This information helps you establish an efficient, repeatable process.

25.1 Requirements

- Workspace access
- Relativity Legal Hold
25.2 Directions

As you set up holds using Relativity Legal Hold, consider the following:

1. Determine whether you'll use a single workspace for all holds or separate workspaces for holds.
   - Relativity Legal Hold supports having multiple legal holds in a single workspace. However, it's beneficial as a system admin, to set up workspaces based on organizational groups (associated with the hold), or similarity of workflow. This helps simplify security on holds.
   - For example, you could have a workspace that contains all holds concerning HR-related matters, and set up additional workspaces based on hold types. Alternatively, you could set up holds in a workspace per organization or department.

2. Determine hold owners and other participants who need access to data.
   - Create a Hold Owners group that has permissions to set up and configure a hold.
   - Create a group for other participants who expect to receive (or have access to) hold reports.

3. Identify your custodians and the potential need for alert groups.
   - For example, you can set up groups for IT, HR, Executives, Marketing, etc.
   - A sample use case would be:
     - Setting up an alert for the IT group when someone responds "Yes" to the question "Do you use your home desktop for work purposes?" in a questionnaire.
     - You can set up custom reminder notices for the Executives group.

4. Define your hold workflow with:
   a. Notification & reminder templates (including criteria for defensible hold notifications):
      - Define the overview of the matter.
      - Indicate explicitly (with examples) the types of information you want to retain.
      - Specify data sources.
      - Collect additional information on other possible sources via questionnaires.
      - Specify date ranges.
      - Clarify the legal obligations to preserve information.
   b. Communication & reminder frequencies:
      - Does one size fit all, or do I set them up depending on the group (for example, Executives get a reminder less frequently)?
      - Do I need to set up a Global reminder (for example, a blanket email every 3 months)?
      - Does one questionnaire fit the need for all groups, or do I need to craft separate questionnaires for each group of recipients?

Note: Relativity maintains an audit log of all the actions associated with the legal hold; however, we strongly recommend that you document all your activities leading up to the legal hold.

In addition, keep the following in mind as you establish a repeatable process to execute a legal hold:

- Factor in employee mobility and turnover.
- Consider checking for non-traditional data sources (for example, data in Cloud Services, mobile devices).
- Define a process for immediate collection (to account for potential equipment failures).
- Record all activities and materials that leverage past preservation efforts (to be able to learn from those efforts and replicate or improve).

Overall, remember that legal holds are an iterative process. To be in compliance, keep the following best practices in mind:

- Issue timely written legal hold directives.
- Ensure custodians are clear about the whats and hows of compliance.
- Schedule timely follow-ups for confirmation and compliance.
- Provide periodic updates and reminders.
- Document every action and the basis for the decisions executed on a legal hold (for example, releasing a custodian from a hold).

### 25.3 References

**Relativity Legal Hold**

### 26 Leveraging processing metadata for quality control

At some point, unprocessable files will likely appear in your processing projects. To correctly address these files in your workspace, you will need to review the files and determine whether they require further attention.

#### 26.1 Overview

Files that are designated unprocessable don't actually have an error associated with them. This is because these files weren't able to be processed in the first place, so they weren't able to register a processing error. This recipe walks you through the process of identifying and reviewing unprocessable files from processing.

#### 26.1.1 Requirements

- Relativity 8.2 or higher
- Relativity processing
- Processing repository access

#### 26.2 Directions

To identify and review unprocessable files, perform the following steps:

1. Create a view in your workspace called Unprocessable Documents on the document object with the following properties:
2. In the view, include the processed metadata fields that provide additional descriptive information about the files, such as the Document Extension, File Type, File Name (which includes the original document extension), and File Size (especially useful for empty files) fields.
3. Next, add a condition to the view. The Unprocessable field uses a Yes/No field type. The yes/no value indicates whether a file was able to be processed. If the file couldn’t be processed, this field is set to Yes.

**Note:** The Unprocessable field is set to Yes on any file for which Relativity doesn’t have an Invariant plugin that is capable of extracting text or imaging/OCR’ing that document type. For example, it’s not set for a corrupt file for which Relativity can’t extract text, such as a corrupt Word document that logs an error during data extraction. Unprocessable documents don’t have errors associated with them because they never reach a point at which they can register a processing error.
4. Finally, set your Sort to sort by Document Extension and File Type. This allows you to complete your review while addressing similar file types together.

5. Once your view is created, you can create a dashboard. Set the Group By value to Document Extension and the Pivot On value to File Type to examine the different file types associated with the document extensions.
You can also reference your list view to identify records individually using the view.
Once you've segregated the unprocessable files, you can begin your review.
26.2.1 Important considerations

- While these files are single records, you will want to watch for unsupported containers that could potentially contain additional compressed files. Proprietary archives, like self-decrypting archives in .exe format are becoming more prevalent.
- In the example shown here, there are also files created using a Macintosh machine. If data was collected and processed from a non-Windows device, special attention should be made to files without text extraction.
- Special considerations may be considered for file types not supported by the viewer. Installing the appropriate software may be required on the reviewer workstations.
- If your processing profile did not include the Unprocessable field, you can reference the two reports available on the Relativity Processing | Reports tab. These reports provide information about the file types, by extension, with a section listing the unprocessable files.
  - Discovered Files by Custodian - This report provides information on the file types discovered during processing for the custodians associated with the selected processing sets. This report identifies the total processable and unprocessable file types discovered and categorizes the totals by custodian. You can run this report on discovered or published processing sets.
  - Discovered Files by File Type - This report provides information on the file types discovered during processing for the custodians associated with the selected processing sets. This report identifies the total processable and unprocessable file types discovered and categorizes the totals by file type. You can run this report on discovered or published processing sets. See Supported file types for processing for a list of file types and extensions supported by Relativity for processing.

Using the extensions in the report, you can create saved searches to identify the unprocessable files using the document extensions:
27 Loading external productions into Relativity

You can import external productions into Relativity through the Relativity Desktop Client. This recipe illustrates the basic workflow to set up a new production set and then load the load files (usually one for metadata and one for the images) into the production set.

This is often useful when the case team asks that, after a production goes out the door, they can access the documents just as they were in the production and that they be available for searching. This recipe also helps when you receive a production from the other side to help organize the received data in the workspace.

27.1 Requirements

- Applicable to all versions of Relativity
- Relativity Desktop Client (RDC)

27.2 Instructions

1. Build a new production set. This is an empty 'staging' production to load the received production into.
2. Open the RDC.
3. Click Tools > Import > Production Load File... to:
   a. Load the image load file (opt file) into the ‘staging’ production.
   b. Use Append Only to ensure only new documents are created.
4. Click Tools > Import > Import Document Load File... to:
   a. Load the .dat metadata file.
   b. Map the appropriate metadata files and use the Overlay Only option on the Overwrite behavior to the production documents that you want to load.

27.3 Notes

- Loading documents into a 'staging' production marks the production set as produced.
- When loading a production set into Relativity, the systems loads the newly created documents into the root folder. It may be prudent to move these documents to a new folder to ensure proper foldering and security of the documents.
- Using Append/Overlay during the metadata import creates new documents that did not have images, but these documents aren't associated with the production. To associate these files with a production set, create and load an image load file pointing to placeholder images and then overlay the metadata. Using the Overlay Only option shows an error on any document not already loaded, and you can use the error file as a list to identify documents that may not have had images.

Importing through the RDC
28 Maintaining previous Bates numbers between production sets

In some cases, you may need to produce documents with a second Bates number—while maintaining their original Bates numbering—on the production image.

28.1 Recipe overview

This recipe shows you how to build production images that maintain their original Bates numbering.

28.2 Requirements

- Workspace access
  - Saved searches
  - Production sets
- Relativity Desktop Client
- Relativity 7 or higher

28.3 Directions

There are two options for creating a new set of production images that retain the Bates numbering from a previous production. Both options generate the same results.

28.3.1 Option 1

Load the documents to re-produced back into Relativity as new documents.

1. Create a saved search of the documents to produce again.
2. Export the documents’ images from the saved search.
   Ensure you select the production version of the images in the Production Precedence section of the Relativity Desktop Client.
3. Import the documents back into your original workspace as new documents. Use the Bates number as the control number.

4. Create a production set with the documents you just loaded back into Relativity. The images contain the Bates numbers from the original production, and from the new production.

28.3.2 Option 2

Produce the documents in a “staging” workspace.

1. Export the documents to re-produce from the original workspace.

   **Note:** Ensure you select the correct production version in the *Production Precedence* section of the Relativity Desktop Client.

2. Create a new “staging” workspace. Load the first version of the produced images into the staging workspace.

3. Create a production set in the staging workspace, and then produce the documents. The images contain the Bates numbers from the original production, and from the new production.

4. Export the re-produced documents from the staging workspace.

   **Note:** Ensure you select the new production version in the *Production Precedence* section of the Relativity Desktop Client.

5. Create a production set in the original workspace.
6. Overlay the re-produced documents as a production in the original workspace. This gives you a production set containing the original documents in the original workspace—which contains images with both sets of Bates numbers.

- [ ] Saved search
- [ ] Production sets

29 Mapping Processing Fields

This application loads the fields necessary for processing. These fields also store metadata in your workspace. The application runs a script that maps identically named Relativity processing fields to the newly loaded fields.

29.1 Requirements

You need to install the Relativity Processing application in the workspace before you install the Mapping Workflow application:

- [ ] Fields
- [ ] Scripts

29.2 Directions

**Note:** Verify that Relativity Processing is already be installed in the workspace.

1. On the Relativity Applications tab, click **New Relativity Application**.
2. Select **Import from File** and click **Browse**.
3. Select the Mapping Workflow application and click **Open**.

![Application Type](image)

**Application Type**
- Select from Application Library
- Import from File

**Application Information**
- Application Name: Create Processing Fields
- Version: 0.0.0.2
- File Name: RA_Create_Processing_Fields_20140228182805.xml

**Application Artifacts**
- Create Processing Fields

**Map Fields**
- **Select Object:** Document

**Application Fields**
- Attachment Document IDs
- Author
- Comments
- Conversation
- Conversation Family
- Conversation Index
- Date Created
- Date Last Modified
- Date Last Printed
- Date Received
- Date Sent
- Document Subject
- Email DCC
- Email CC
- Email Categories
- Email From
- Email Subject
- Email To
- File Name

**Workspace Fields**

4. Click **Import**.

**Note:** Relativity displays an error message after installation. This error is expected. Select **Unlock Processing Application** and click **Retry Import**.

![Import Status](image)

**Import Status**
- Your application failed to import. There are 1 errors that need your attention. Please resolve all errors and try your import again.
- **Export Error File**

![Locked Conflicts - Processing](image)

**Locked Conflicts - Processing**
- This view can't be imported. The locked Relativity application Processing in the target workspace shares the Duplicates Documents view with the application you're trying to import.
- **Unlock Processing Application**

When the installation completes, Relativity displays a message indicating that import was successful.

5. Navigate to the Script Library.
6. Execute the **Map Processing Fields** script provided with the application. After the script completes, you can view the processing fields mapped to their respective Relativity fields.
<table>
<thead>
<tr>
<th>Processing Field Name</th>
<th>Relativity Field</th>
<th>Field Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edit Attachment Document IDs</td>
<td>Attachment Document IDs</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Author</td>
<td>Author</td>
<td>Fixed-Length Text</td>
</tr>
<tr>
<td>Edit Comments</td>
<td>Comments</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Conversation</td>
<td>Conversation</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Conversation Family</td>
<td>Conversation Family</td>
<td>Fixed-Length Text</td>
</tr>
<tr>
<td>Edit Conversation Index</td>
<td>Conversation Index</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Date Created</td>
<td>Date Created</td>
<td>Date</td>
</tr>
<tr>
<td>Edit Date Last Modified</td>
<td>Date Last Modified</td>
<td>Date</td>
</tr>
<tr>
<td>Edit Date Last Printed</td>
<td>Date Last Printed</td>
<td>Date</td>
</tr>
<tr>
<td>Edit Date Received</td>
<td>Date Received</td>
<td>Date</td>
</tr>
<tr>
<td>Edit Date Sent</td>
<td>Date Sent</td>
<td>Date</td>
</tr>
<tr>
<td>Edit Document Extension</td>
<td>Document Extension</td>
<td>Fixed-Length Text</td>
</tr>
<tr>
<td>Edit Document Subject</td>
<td>Document Subject</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Domains (Email BCC)</td>
<td>Domains (Email BCC)</td>
<td>Multiple Object</td>
</tr>
<tr>
<td>Edit Domains (Email CC)</td>
<td>Domains (Email CC)</td>
<td>Multiple Object</td>
</tr>
<tr>
<td>Edit Domains (Email From)</td>
<td>Domains (Email From)</td>
<td>Multiple Object</td>
</tr>
<tr>
<td>Edit Domains (Email To)</td>
<td>Domains (Email To)</td>
<td>Multiple Object</td>
</tr>
<tr>
<td>Edit Email BCC</td>
<td>Email BCC</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Email Categories</td>
<td>Email Categories</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Email CC</td>
<td>Email CC</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Email From</td>
<td>Email From</td>
<td>Fixed-Length Text</td>
</tr>
<tr>
<td>Edit Email Subject</td>
<td>Email Subject</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Email To</td>
<td>Email To</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit FileName</td>
<td>File Name</td>
<td>Fixed-Length Text</td>
</tr>
<tr>
<td>Edit File Size</td>
<td>File Size</td>
<td>Decimal</td>
</tr>
<tr>
<td>Edit FileType</td>
<td>File Type</td>
<td>Fixed-Length Text</td>
</tr>
<tr>
<td>Edit Group Identifier</td>
<td>Group Identifier</td>
<td>Fixed-Length Text</td>
</tr>
<tr>
<td>Edit Has Hidden Data</td>
<td>Has Hidden Data</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Edit Importance</td>
<td>Importance</td>
<td>Single Choice</td>
</tr>
<tr>
<td>Edit Lotus Notes Other Folders</td>
<td>Lotus Notes Other Folders</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit MD5 Hash</td>
<td>MD5 Hash</td>
<td>Fixed-Length Text</td>
</tr>
<tr>
<td>Edit Message Type</td>
<td>Message Type</td>
<td>Single Choice</td>
</tr>
<tr>
<td>Edit Number of Attachments</td>
<td>Number of Attachments</td>
<td>Whole Number</td>
</tr>
<tr>
<td>Edit Other Props</td>
<td>Other Props</td>
<td>Long Text</td>
</tr>
<tr>
<td>Edit Password Protected</td>
<td>Password Protected</td>
<td>Single Choice</td>
</tr>
<tr>
<td>Edit SHA1 Hash</td>
<td>SHA1 Hash</td>
<td>Fixed-Length Text</td>
</tr>
<tr>
<td>Edit SHA256 Hash</td>
<td>SHA256 Hash</td>
<td>Fixed-Length Text</td>
</tr>
</tbody>
</table>

- Applications
- Scripts
30 Mass importing into a file field

A file type field allows users to link an outside document to a record. This field is only available in custom objects. File fields offer the flexibility to store documents outside the scope of discovery in your workspace, such as CVs for your expert witnesses.

30.1 Overview

This recipe shows you how to upload multiple documents into a file field stored on a custom object.

30.2 Requirements

- Applicable to all Relativity versions
- Dynamic object with a file field
- Workspace access
  - Relativity Desktop Client

30.3 Directions

After you create your custom object and add a new file field, you can use the RDC to import into this object. This example uses an Expert Witness object.

1. Open the Relativity Desktop Client and your workspace.
2. Click the drop-down for the **Documents** object and select the object that stores the file field you’re using.
3. Click **Tools > Import**, and then choose **Expert Witness Load File**.
4. Browse to your load file with the file field path.
5. Change **Encoding and Characters** to match the load file, and then click to the **Field Maps** tab.

   **Note:** When you look at the list of available fields in your workspace, your file field does not appear. However, the Native File Behavior is enabled.

6. Ensure the Load Native Files checkbox is selected.
7. Fill in the Native File Path Contained in Column field with the appropriate field from your load file.

8. Click **Import > Import File**.
9. Navigate back to your workspace. In the file field, you should see your records with a hyperlink to the file.
31 Migrating cases from Summation

At times, you may need to migrate case data from other software into Relativity. While a button for transferring data from one database to another doesn't exist, there are some ways you can help mitigate the process.

31.1 Recipe overview

This recipe highlights some best practices for migrating case information from Summation to Relativity.

31.2 Requirements

- Applicable to all Relativity versions
- Workspace access
- Relativity Desktop Client

31.3 Directions

31.3.1 Migrating images

- Clients sometimes find it easier to use the image volumes and load files on their server rather than exporting images from Summation. You can account for all the images by comparing the total images in Summation with the total number of images contained in the volumes on the server.
- Use a conversion tool like iConvert to convert .dii files to .opt files.

   **Note:** Use a text editor like TextPad to combine multiple .dii files into one. Doing so saves time during the conversion and import processes.

31.3.2 Migrating data

- Often, not all data stored in Summation is contained in the original volume load files on the server. You should perform an export from Summation that contains all metadata and work product fields.
- Fields that use lookup tables in Summation—such as Names, Issues, and Doctype—should be created in Relativity as either single- or multiple-choice fields to match the previous fields.
- Summation folder information is stored in the Folder field. You can use that information in two ways:
  - If you load the Folder field into a multiple-choice field, you see the folder structure in the tag tree. Ensure that, when loading the data, you delimited to maintain the nested value to the “\” character.
If you want to reproduce the Summation folder structure in your Relativity folders, you need to export the Folder field and import it into Relativity. During import, set the Folder Information Column accordingly.

31.3.3 Migrating transcripts

- It may be easier to work with the original files received from the court reporting agency and import them directly into Relativity. Once they’re in your workspace, use the Process Transcript mass action to build the word index.
- Create a separate folder for your transcripts. You can build out a folder tree with each deponent’s name as a sub-folder, or place them all in the root of the new transcript folder.

31.3.4 Migrating productions

You can export production images from Summation and imported into Relativity as a production set. You need to create a production set in Relativity and import documents into the set through the Relativity Desktop Client.
31.3.5 Migrating redactions

To migrate redactions from Summation to Relativity, you need to burn the redactions into the images and then import those images into Relativity.

- Relativity Desktop Client
- Saved search

32 Moving relevant choices to the top

When working in Relativity, reviewers or system admins may request some of their most frequently used choices to be moved to the top of the Choice field list for coding convenience.

32.1 Overview

This recipe demonstrates how to use Relativity's flexibility to move the choices associated with a single choice or multi-choice field.
32.2 Requirements

- Workspace access
- Edit and delete permissions on fields

32.3 Directions

You can use the following workflow on all single and multi-choice fields on any object in Relativity:

1. Click the **Choices** tab, and then filter for the field of your choice.

2. Click **Sort Top Level Choices** and use the pop-up to sort the order of the choices.
   
   After you sort the choices to the order you want, Relativity reorders the choices to that order on a user's layout.

33 Multi-matter workspace setup in Relativity

When working with a single workspace supporting multiple review matters, there are several ways system admins can expedite setup and improve performance.

33.1 Recipe overview

This recipe highlights the items that system admins can reference while managing workspaces that support multi-matter ESI for processing, review, and production.
33.2 Requirements

- System admin access
  - Add new users and groups permission
- Workspace access
  - Edit, Delete permission for all items under the Admin tab
- Applicable to all versions of Relativity.

33.3 Directions

In this illustration, we’ll reference a single workspace that contains three different matters: A, B, and C. Each matter is actively being reviewed by a combination of reviewers, some common to all three matters and some who are unique to certain matters.

System admins should use the following best practices as they configure this type of workspace.

33.3.1 User Groups/Security

Set up groups with reviewers:

- Common to the three matters, there are four potential combinations:
  - A,B,C
  - A,B
  - B,C
  - A,C

- Unique to matters, there are three groups:
  - A
  - B
  - C

The following table represents the different user groups and how they are distributed:

<table>
<thead>
<tr>
<th>User Groups</th>
<th>Matter - A</th>
<th>Matter - B</th>
<th>Matter - C</th>
<th>Matter - A</th>
</tr>
</thead>
<tbody>
<tr>
<td>UG - 1</td>
<td></td>
<td></td>
<td></td>
<td>UG - 7</td>
</tr>
<tr>
<td>UG - 2</td>
<td>UG - 3</td>
<td>UG - 4</td>
<td>UG - 5</td>
<td>UG - 6</td>
</tr>
</tbody>
</table>

As illustrated, you’ll have a minimum of seven unique user groups for the three matters in this workspace. The minimum number of unique user groups a workspace requires (ignoring tiered permissions within a user group, which will be described in the following example) is determined by the following formula:

\[(2^m) – 1, \text{ where } m \text{ is the number of matters being set up in the workspace}\]

System admins must analyze which users belong in which groups and consider the permission levels required for each matter and each user. For example, if we add a user to UG-4 and UG-3, and the permission levels for the matters are identical, that particular user may be for UG-7.
However, if the permission levels at UG-5 and UG-3 are different—perhaps the user has permissions to code in Matter A and Matter B, but not the documents in Matter C—then tiered user group access is required. In that case, UG-4 and UG-3A — which, unlike UG-3B, do not include permissions to code documents — may be more suitable.

A best practice for creating folders is to continue to set up the documents in a workspace by custodian. There are two advantages to this:

- For end users, the workspace will look similar to the typical single matter workspaces, and navigation will be familiar.
- If some custodians are not part of a certain matter, those custodian folders can be locked at the folder level for the appropriate user groups, thus restricting document access to those users.

### 33.3.2 Folder Setup

### 33.3.3 Field and choice setup

For quick identification of documents in a multi-matter workspace, system admins can use a required multi-choice field to tag documents appropriately. For example, it could be titled Matter ID and include the matter names as choices. Use this field to identify documents not tagged for any of the matters, as well as for creating batches.

Add fields and choices to the document object. However, if you anticipate adding more fields — such as for review or production— outside of the standard Designation and Issue fields, there are two ways to approach this.

- Continue to add the system admin fields as a part of the document object.
- We recommend adding the system admin fields to a separate object, with one object set for each matter. Refer to the Building Custom Objects Workbook.

Adding additional fields will affect performance as the database grows, but you can make the database more scalable for new fields as your document count or the needs of each of the matter grow by adding them on a separate object.

With the second option, the new fields created on a separate object will need to be set up so that they will be available in the views on the Document object. This can be accomplished by opening the field to association, as highlighted by the red box below.

### 33.3.4 Views Setup

If fields are set up on a separate object, those fields and their associated choices are available in a view under the Documents object, as described in the Field & Choice Setup section of the Building Custom Objects Workbook. Once the fields are available in a view, you can use the view to filter for documents. Below is a view with fields from the Document object (boxed in blue) and the Matter-1 object (boxed in green).  

Relativity | 9.2 Recipes - 93
33.3.5 Layout Setup

Fields set up on a separate object can be added to the document layout with an Associative Object List, highlighted in the red box below. Once the object has been included, it will appear in the coding pane — shown in the blue box — along with views that have associative object fields and their choices.

Once the object has been included in the coding pane, the field will appear in the layout as shown below. Note that the New, Link, and Unlink buttons can be turned off depending on the permission level for the user group to the object.

You can link multiple associative objects to the layout being built if you set up separate objects for each matter.

33.3.6 Search indexes

Depending on the needs of your workspace, you can create separate dtSearch indexes for documents in each of the matters.

This may not be recommended in all multi-matter workspace setups, but in such cases it is recommended to have at least two tiered indexes. Refer to the following table for more details.

<table>
<thead>
<tr>
<th>Search Index Tiers</th>
<th>Use</th>
<th>Drawback</th>
</tr>
</thead>
</table>
| Across all documents (and all matters) in the workspace | - For linear review and batching  
- Supports review QC for responsiveness to matters. | - Slower Index Builds and slower query response times |
| Across a subset of documents responsive to each matter | - Supports quicker index builds and faster querying  
- Securable to user groups | |

Keyword search automatically indexes all documents within the workspace, so if your search access is matter-based, access to keyword search needs to be disabled for appropriate user groups.

33.3.7 Batches

The first step is to identify the documents included in the different matters within the workspace. Batching can be done as part of the linear review workflow, or a Relativity Analytics-based categorization workflow. In a categorization workflow, begin by sampling documents the same way you would perform issue coding using categorization.

As described in the Fields Setup section of the Building Custom Objects workbook, batching can leverage the Matter ID field for further review or to tag the documents during initial review.
All other system admin setup functions — such as production sets, markup sets, fields, persistent highlight sets, and choices on the document object — are managed the same way they would be in a single matter workspace. However, we recommend the following best practices:

- Use intuitive naming conventions. For example:
  - SEC productions on Matter A could be MA – Production SEC 20110229
  - EPA markup sets on Matter B could be MB – EPA Priv Set
- System admins need to be particularly aware of permissions provided to each user group. We recommend setting up tiered access for each user group combination based on who needs access to the views, layouts, markup sets, production sets, etc., as described above.
- Since batching is based on saved searches, further review, batching, and dtSearch indexes can be based on search results responsive to the matters in question.

**Workspaces**

### 34 Native Imaging - Dithering Algorithm Options

Relativity native imaging offers multiple dithering algorithms you can apply towards imaging your documents.

#### 34.1 Recipe overview

This recipe presents the various dithering algorithms available within Relativity native imaging. This recipe provides you with additional information and use cases for each algorithm, including the size and clarity of images you create.

#### 34.2 Requirements

- Relativity 8.0 or higher
- Workspace and system admin access

#### 34.3 Directions

System admins can select from one of eight different dithering algorithms when they choose to image with native imaging:

- Clustered 6x6
- Clustered 8x8
- Clustered 16x16
- Dispersed 4x4
- Dispersed 8x8
- Dispersed 16x16
- Floyd & Steinberg
- Threshold
### 34.3.1 Dithering results for text-rich documents

Below are the testing results with the different algorithms for text-rich documents. All results are 300 DPI B&W only.

<table>
<thead>
<tr>
<th>Image Dithering Type</th>
<th>Size</th>
<th>Image Sample Magnified (150% scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustered 6x6</td>
<td>27 KB</td>
<td>Transwestern's average deliveries to California were 1,074 MMMBtu/d (99%), with San Juan lateral throughput 390 MMMBtu/d.</td>
</tr>
<tr>
<td>Clustered 8x8</td>
<td>25 KB</td>
<td>El Paso's average deliveries to California were 2,319 MMMBtu/d (78%): - PKR-ETop, capacity 1,140 MMMBtu/d, deliveries of 670 MMMBtu/d (59%) - SoCalETop, capacity 1,299 MMMBtu/d, deliveries of 1,147 MMMBtu/d (88%) - SoCalTop, capacity 340 MMMBtu/d, deliveries of 502 MMMBtu/d (93%)</td>
</tr>
</tbody>
</table>

Transwestern's average deliveries to California were 1,074 MMMBtu/d (99%), with San Juan lateral throughput 390 MMMBtu/d. El Paso's average deliveries to California were 2,319 MMMBtu/d (78%): - PKR-ETop, capacity 1,140 MMMBtu/d, deliveries of 670 MMMBtu/d (59%) - SoCalETop, capacity 1,299 MMMBtu/d, deliveries of 1,147 MMMBtu/d (88%) - SoCalTop, capacity 340 MMMBtu/d, deliveries of 502 MMMBtu/d (93%)
<table>
<thead>
<tr>
<th>Image Dithering Type</th>
<th>Size</th>
<th>Image Sample Magnified (150% scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustered 16x16</td>
<td>23 KB</td>
<td>Transwestern's average deliveries to California were 1074 MMBtu/d (99%), with San Juan lateral throughput 390 MMBtu/d.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>El Paso's average deliveries to California were 2319 MMBtu/d (78%):</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- PO&amp;E, capacity of 1140 MMBtu/d, deliveries of 670 MMBtu/d (59%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal, capacity 1299 MMBtu/d, deliveries of 1147 MMBtu/d (88%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCalTop, capacity 540 MMBtu/d, deliveries of 502 MMBtu/d (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Friday's posted Gas Daily prices:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SoCal gas, large plgs 6.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PG&amp;E, large plgs 4.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW San Juan 2.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW Permian 3.405</td>
</tr>
<tr>
<td>Floyd &amp; Steinberg</td>
<td>24 KB</td>
<td>Transwestern's average deliveries to California were 1074 MMBtu/d (99%), with San Juan lateral throughput 390 MMBtu/d.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>El Paso's average deliveries to California were 2319 MMBtu/d (78%):</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- PO&amp;E, capacity of 1140 MMBtu/d, deliveries of 670 MMBtu/d (59%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal, capacity 1299 MMBtu/d, deliveries of 1147 MMBtu/d (88%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCalTop, capacity 540 MMBtu/d, deliveries of 502 MMBtu/d (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Friday's posted Gas Daily prices:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SoCal gas, large plgs 6.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PG&amp;E, large plgs 4.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW San Juan 2.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW Permian 3.405</td>
</tr>
<tr>
<td>Dispersed 4x4</td>
<td>24 KB</td>
<td>Transwestern's average deliveries to California were 1074 MMBtu/d (99%), with San Juan lateral throughput 390 MMBtu/d.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>El Paso's average deliveries to California were 2319 MMBtu/d (78%):</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- PO&amp;E, capacity of 1140 MMBtu/d, deliveries of 670 MMBtu/d (59%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal, capacity 1299 MMBtu/d, deliveries of 1147 MMBtu/d (88%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCalTop, capacity 540 MMBtu/d, deliveries of 502 MMBtu/d (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Friday's posted Gas Daily prices:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SoCal gas, large plgs 6.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PG&amp;E, large plgs 4.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW San Juan 2.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW Permian 3.405</td>
</tr>
<tr>
<td>Dispersed 8x8</td>
<td>24 KB</td>
<td>Transwestern's average deliveries to California were 1074 MMBtu/d (99%), with San Juan lateral throughput 390 MMBtu/d.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>El Paso's average deliveries to California were 2319 MMBtu/d (78%):</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- PO&amp;E, capacity of 1140 MMBtu/d, deliveries of 670 MMBtu/d (59%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal, capacity 1299 MMBtu/d, deliveries of 1147 MMBtu/d (88%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCalTop, capacity 540 MMBtu/d, deliveries of 502 MMBtu/d (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Friday's posted Gas Daily prices:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SoCal gas, large plgs 6.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PG&amp;E, large plgs 4.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW San Juan 2.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW Permian 3.405</td>
</tr>
<tr>
<td>Dispersed 16x16</td>
<td>24 KB</td>
<td>Transwestern's average deliveries to California were 1074 MMBtu/d (99%), with San Juan lateral throughput 390 MMBtu/d.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>El Paso's average deliveries to California were 2319 MMBtu/d (78%):</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- PO&amp;E, capacity of 1140 MMBtu/d, deliveries of 670 MMBtu/d (59%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal, capacity 1299 MMBtu/d, deliveries of 1147 MMBtu/d (88%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCalTop, capacity 540 MMBtu/d, deliveries of 502 MMBtu/d (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Friday's posted Gas Daily prices:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SoCal gas, large plgs 6.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PG&amp;E, large plgs 4.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW San Juan 2.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW Permian 3.405</td>
</tr>
<tr>
<td>Image Dithering Type</td>
<td>Size</td>
<td>Image Sample Magnified (150% scale)</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Threshold 0</td>
<td>2 KB</td>
<td></td>
</tr>
<tr>
<td>Threshold 127</td>
<td>22 KB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transwestern’s average deliveries to California were 1074 MMBtu/d (99%), with San Juan lateral throughput 390 MMBtu/d.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>El Paso’s average deliveries to California were 2319 MMBtu/d (78%) :</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- PG&amp;E Top, capacity 1140 MMBtu/d, deliveries of 670 MMBtu/d (39%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal, capacity 1299 MMBtu/d, deliveries of 1147 MMBtu/d (88%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal Top, capacity 540 MMBtu/d, deliveries of 502 MMBtu/d (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Friday’s posted Gas Daily prices:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SoCal gas, large plgs 6.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PG&amp;E large plgs 4.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW San Juan 2.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW Permian 3.495</td>
</tr>
<tr>
<td>Threshold 215</td>
<td>22 KB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transwestern’s average deliveries to California were 1074 MMBtu/d (99%), with San Juan lateral throughput 390 MMBtu/d.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>El Paso’s average deliveries to California were 2319 MMBtu/d (78%) :</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- PG&amp;E Top, capacity 1140 MMBtu/d, deliveries of 670 MMBtu/d (39%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal, capacity 1299 MMBtu/d, deliveries of 1147 MMBtu/d (88%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal Top, capacity 540 MMBtu/d, deliveries of 502 MMBtu/d (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Friday’s posted Gas Daily prices:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SoCal gas, large plgs 6.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PG&amp;E large plgs 4.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW San Juan 2.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW Permian 3.495</td>
</tr>
<tr>
<td>Threshold 255</td>
<td>22 KB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transwestern’s average deliveries to California were 1074 MMBtu/d (99%), with San Juan lateral throughput 390 MMBtu/d.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>El Paso’s average deliveries to California were 2319 MMBtu/d (78%) :</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- PG&amp;E Top, capacity 1140 MMBtu/d, deliveries of 670 MMBtu/d (39%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal, capacity 1299 MMBtu/d, deliveries of 1147 MMBtu/d (88%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SoCal Top, capacity 540 MMBtu/d, deliveries of 502 MMBtu/d (93%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Friday’s posted Gas Daily prices:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SoCal gas, large plgs 6.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PG&amp;E large plgs 4.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW San Juan 2.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TW Permian 3.495</td>
</tr>
</tbody>
</table>

34.3.2 Dithering results for image-rich documents

Below are the testing results with the different algorithms for image-rich documents. All results 300DPI B&W only.
<table>
<thead>
<tr>
<th>Image Dithering Type</th>
<th>Size</th>
<th>Image Sample Magnified (150% scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustered 6x6</td>
<td>45 KB</td>
<td><img src="image1.png" alt="Image Sample Magnified" /></td>
</tr>
<tr>
<td>Clustered 8x8</td>
<td>39 KB</td>
<td><img src="image2.png" alt="Image Sample Magnified" /></td>
</tr>
<tr>
<td>Image Dithering Type</td>
<td>Size</td>
<td>Image Sample Magnified (150% scale)</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>Clustered 16x16</td>
<td>33 KB</td>
<td><img src="image.png" alt="Image Sample" /></td>
</tr>
<tr>
<td>Image Dithering Type</td>
<td>Size</td>
<td>Image Sample Magnified (150% scale)</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Floyd &amp; Steinberg</td>
<td>70 KB</td>
<td><img src="image.png" alt="Image Sample" /></td>
</tr>
<tr>
<td>Image Dithering Type</td>
<td>Size</td>
<td>Image Sample Magnified (150% scale)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Dispersed 4x4</td>
<td>64 KB</td>
<td></td>
</tr>
<tr>
<td>Image Dithering Type</td>
<td>Size</td>
<td>Image Sample Magnified (150% scale)</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Dispersed 8x8</td>
<td>52 KB</td>
<td></td>
</tr>
<tr>
<td>Image Dithering Type</td>
<td>Size</td>
<td>Image Sample Magnified (150% scale)</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>Dispersed 16x16</td>
<td>69 KB</td>
<td><img src="image_url" alt="Image Sample Magnified" /></td>
</tr>
<tr>
<td>Image Dithering Type</td>
<td>Size</td>
<td>Image Sample Magnified (150% scale)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Threshold 0</td>
<td>2 KB</td>
<td></td>
</tr>
<tr>
<td>Image Dithering Type</td>
<td>Size</td>
<td>Image Sample Magnified (150% scale)</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>Threshold 127</td>
<td>79 KB</td>
<td><img src="image.png" alt="Image Sample" /></td>
</tr>
</tbody>
</table>
35 OCR Redacted Production Documents & Export Text

Redactions are a ubiquitous aspect of document review. The challenge is to prepare redacted documents for production in a timely and efficient manner. This recipe assists in completing OCR on redacted documents, and exporting text, while minimizing unnecessary work.
35.1 Requirements

- Workspace access
  - OCR Profile, OCR Set, Production Sets
  - Create access on Saved Searches
  - Create Field access
- Relativity Desktop Client

35.2 Directions

First, OCR a production. You should OCR your production after you run the production, but before you perform the final export. Perform the following steps to OCR documents with redactions:

1. Create an OCR profile with your preferred settings. You must have an OCR profile before you can save an OCR Set.
2. Create a long text field to hold your OCR results. You can't OCR directly into the Extracted Text field.
3. Create an OCR Set.
4. Point the OCR Set to completed Production. If you only want to OCR redacted documents for the production, select Yes next to Only OCR Production Documents Containing Redactions.
5. Click Save on the OCR Set. The OCR Set Console appears.
6. Click OCR Documents from the OCR Set Console to kick off the OCR job.

When the job completes, the text becomes ready to export.

Now you can export a production using text precedence. To only produce non-redacted text when exporting a production, you must use the Text Precedence feature. When you export a production through the RDC, don't add the Extracted Text field to the exported field list. This results in producing redacted text. In order to preserve the integrity of your redactions, don't add the Extracted Text field to the selected columns tab.
When you switch to the Destination Files tab, you can set Text Precedence in the Metadata section. To set Text Precedence:

1. Set the File Format, Encoding, and Export Files options to meet your needs.
2. Click the ellipsis next to Text Precedence. The Pick Text Precedence window appears.
   - Based on the screenshot below, the RDC first checks to see if there is OCR’d text available in the OCR Destination Field. If this field isn't set, the RDC exports the Extracted Text.
3. Export your production set documents.
36 Performing overturn analysis for Relativity Assisted Review

During a computer-assisted review, a case team moves through several rounds of coding to train the system on the document collection and validate the computer’s results. A meta round isn’t a formal round, but a between-rounds workflow used to make any necessary adjustments to the project to prepare for the next round. A meta round usually consists of identifying, analyzing, and correcting (re-coding) documents which have a significant and adverse effect on project results.
We recommend that you make these adjustments prior to finishing a round and categorizing documents. This allows the system to make use of the corrections performed during the meta round and apply them to the next true round.

### 36.1 Recipe overview

This recipe teaches Relativity project managers to identify high-ranking overturns and highly influential seed documents that are responsible for a large number of overturns.

### 36.2 Requirements

- Admin rights or access to the console
- Analytics index (active with queries enabled)

### 36.3 Directions

#### 36.3.1 Correcting coding inconsistencies between true or conceptual duplicates

- Each seed-overturn pair has a rank (or score) which indicates the degree of conceptual similarity they share. The maximum possible score is 100, which means the two documents are conceptual duplicates. Conceptual duplicates are documents which may or may not have identical text, but do contain the same conceptual content according to the Analytics index. While it is possible that conceptual duplicates may also be exact textual duplicates (i.e., documents with the same MD5 hash value), this should not be assumed from a score of 100.
- We recommend that you use the Overturn Documents report to locate these documents by filtering on the round and sorting by descending rank. A good best practice is to re-evaluate each seed-overturn pair having a rank of 95 and higher to see which document was coded correctly, as well as whether each is a suitable example.

#### 36.3.2 Identifying and correcting the most influential seed documents

- When viewing overturn reports at the end of a round, the same few documents can be responsible for many overturns. If those seed documents were incorrectly coded, they can greatly inflate the overturn rate for the entire project. Finding and correcting these situations is an essential component of meta round protocol.
- The quickest way to find the most influential documents is by using Pivot on the Overturned Documents report. Simply choose Seed document in the Group by drop-down and leave the <Total Only> drop-down as is.

#### 36.3.3 Using the Overturn Analysis related items pane

- Once a document has been targeted for re-evaluation during a meta round, you can navigate directly to it using the hyperlinks in the Overturned Documents report. Once you reach the core reviewer interface, open the Overturn Analysis related items pane by clicking the RAR icon in the bottom right corner.
Clicking the file icon next to the document's control number opens the target document in a separate viewer window. This allows a reviewer to compare the two documents side by side to assist in the decision-making process.

Analytics

37 Processing duplication workflow

The Processing Duplication Workflow is a Relativity application that identifies master and duplicate documents, all custodians, and all source files for documents. It also provides capabilities to identify unique, master and duplicate files based on a relational field.

You can download the Processing Duplication Workflow solution files from the Relativity Community.

This page contains the following sections:

- Before you begin below
- Deploying and configuring the solution on the next page
- Preparing the workspace on the next page
- Running the solution on the next page
- Viewing the results on page 118

37.1 Before you begin

37.1.1 Supported versions

This solution is supported in Relativity 9.0 – 9.5 and RelativityOne.

Note: Some versions of this application may not be eligible for support by Relativity Client Services. For more information, see the Version support policy on the Relativity documentation site.

<table>
<thead>
<tr>
<th>Solution version</th>
<th>Supported Relativity version</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>9.4 - 9.5</td>
</tr>
<tr>
<td>6.0</td>
<td>9.0 - 9.3</td>
</tr>
</tbody>
</table>

37.1.2 Components

This custom solution consists of the following components:

- Relativity application
- Relativity scripts that run at the workspace level within a script group

37.1.3 Considerations

- This script should only be run by a system admin. If you are not a system admin, we recommend you do not run this script.
The solution may require you to create fields in your environment. Please see Update Duplicate Status script for detailed information about the fields that need to be created.

This solution is especially useful when you’re using Relativity's ECA and Investigation feature, which was introduced in Relativity 9.4. For more information, see ECA and Investigation.

37.2 Deploying and configuring the solution

To deploy and configure the solution, add it to the Application Library as a Relativity application by following these steps:

1. Log in to Relativity.
2. Click the user drop-down menu in the upper-right corner of Relativity, and then click Home.
3. Click on the Applications & Scripts tab, and then click the Application Library tab.
4. Click on the Upload Application button.
5. Next to Application File, click Choose File.
6. Navigate to and select the Processing Duplication Workflow.rap file included in the solution, and then click Open.
7. Click the Save button.

37.3 Preparing the workspace

After you add the solution to the Application Library, you’re ready to install and configure it to a workspace by following these steps:

1. Click the user drop-down menu in the upper-right corner of Relativity, and then click Home.
2. Click the Applications and Scripts tab, and then click the Application Library tab.
3. Click the name of the Processing Duplication Workflow application.
4. Next to Workspaces Installed, click Install.
5. Next to Workspaces, click ⋱.
6. Select the check box next to the workspace that you want to install the solution in. To install the solution in more than one workspace, select the check box next to each additional workspace that you want.
7. Click OK.
8. Click Save.
9. After refreshing the page, the Workspaces Installed category displays the progress of the installation.

37.4 Running the solution

After you install and configure the solution to a workspace, a Processing Duplication Workflow tab is added to the workspace. Within the Processing Duplication Workflow tab is a tab called Processing Scripts which contains three Relativity scripts within a script group:

1. All Custodians script
2. All Source Locations script
3. Update Duplicate Status script
37.4.1 All Custodians script

The All Custodians script populates a field with the names of all custodians who own a document. To run the All Custodians script:

1. Click the Processing Duplication Workflow tab, and then click the Processing scripts tab.
2. The Processing script page appears. Select the radio button next to All Custodians.

3. Complete the following fields:
   - **Saved Search** - Required. Select the saved search which will contain the group of documents the script will be executed against.
   - **Relational Identifier** - Required. Select the relational field which defines groups of duplicate documents.
   - **Custodian** - Required. Select the Custodian object that contains your custodian information.
   - **Custodian Field** - Required. Select the field that contains the custodian name.
   - **Destination (output)** - Required. Select the long text field to store the semi-colon delimited list of custodians.

4. Click Run.
37.4.2 All Source Locations script

The All Source Locations script identifies all source locations for a duplicate document. To run the All Source Locations script:

1. Click the Processing Duplication Workflow tab, and then click the Processing scripts tab.
2. The Processing script page appears. Select the radio button next to All Source Locations.

3. Complete the following fields:
   - **Saved Search** - Required. Select the saved search which will contain the group of documents the script will be executed against.
   - **Relational Identifier** - Required. Select the relational field which defines groups of duplicate documents.
   - **Source Field** - Required. Select the long text field that contains the source location for documents.
   - **Destination Field (output)** - Required. Select the long text field to store the semi-colon delimited list of source locations.
4. Click Run.
37.4.3 Update Duplicate Status script

Before running the Update Duplicate Status script, you need to create fields in your environment by doing the following:

1. Navigate to the workspace in which you are running the solution
2. Create a new single choice field on the Document object with three choices (names must be exact, but are not case-sensitive): Unique, Master, or Duplicate.
3. Make sure a whole number field is created for the Custodian object if utilizing the order of the custodians associated with the documents.
4. Make sure each custodian has a unique order number. Do not use numbers less than 1.

5. Verify that a Custodian single-object field exists on the Document object. Create a Custodian field if one does not already exist.

To run the Update Duplicate Status script:

1. Click the Processing Duplication Workflow tab, and then click the Processing scripts tab.
2. The Processing script page appears. Select the radio button next to Update Duplicate Status.
3. Complete the following fields:

- **Saved Search** - Required. Select the saved search which will contain the group of documents the script will be executed against.
- **Duplicate Status Field** - Required. Select the single choice field that indicates what kind of duplicate the document is.
- **Relational Identifier** - Required. Select the relational field which defines groups of duplicate documents.
- **Duplicate Sort Order** - Not Required. Select the whole number field that you want the duplicates to be sorted by. See Viewing the Results for more information on how this field operates.

### 37.5 Viewing the results

The results of running depend on which script was executed.

#### 37.5.1 All Custodians script

After you run the All Custodians script, the results appear as a report on the script page.
The following table lists and describes the columns in the report.

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Custodian</td>
<td>The value of the Custodian associated with the document.</td>
</tr>
<tr>
<td>Relational Identifier</td>
<td>The value of the relational field chosen to define groups of related documents.</td>
</tr>
<tr>
<td>All Custodians per Relational Group</td>
<td>The semi-colon delimited list of all custodians that are associated with the document in the relational group.</td>
</tr>
</tbody>
</table>

37.5.2 All Source Locations script

After you run the All Source Locations script, the results appear as a report on the script page.
The following table lists and describes the columns in the report:

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Folder Path</td>
<td>The value of the source field file location associated with the document.</td>
</tr>
<tr>
<td>Relational Identifier</td>
<td>The value of the relational field chosen to define groups of related documents.</td>
</tr>
<tr>
<td>All Source Locations</td>
<td>The semi-colon delimited list of all source locations associated with the document.</td>
</tr>
</tbody>
</table>

37.5.3 Update Duplicate Status script

When the script is executed, the selected Duplicate Status Field is cleared for all documents in the workspace. Once the field is cleared, the Duplicate Status Field is populated with the following values for any documents included in the selected saved search:

- **Unique** – Documents where there is only one document in the saved search with the same relational identifier
- **Master** – If Duplicate Sort Order is selected, documents where there is more than one document in the saved search with the same relational identifier and the lowest order of the associated custodian. If multiple documents in the same group shared the same custodian, then the lowest document artifact ID. If Duplicate Sort Order is not selected, documents where there is more than one document in the saved search with the same relational identifier which have the lowest document artifact ID in the relational group.
- **Duplicate** – If Duplicate Sort Order is selected, documents where there is more than one document in the saved search with the same relational identifier, which do not have the lowest ordered custodian in the relational group. If Duplicate Sort Order is not selected, documents where there is more than one document in the saved search with the same relational identifier which do not have the lowest artifact ID in the relational group.
- **Not Set** — Documents will remain "Not set" in the selected Duplicate Status Field if the selected Relational Identifier is not set.

Any documents not included in the selected saved search will be excluded from the logic to calculate duplicate status and will not be populated in the selected Duplicate Status Field.

When the script is complete, you will receive an "Update Complete" message on the script page. Documents in the saved search are updated.

<table>
<thead>
<tr>
<th>Artifact ID</th>
<th>Control Number</th>
<th>MDS Hash</th>
<th>Custodian</th>
<th>Duplicate Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1039576</td>
<td>JGRIFFITH_0000284</td>
<td>6DBC9FF8489363C8C62E5F4D9AB9186</td>
<td>Griffith John</td>
<td>Master</td>
</tr>
<tr>
<td>1039577</td>
<td>JGRIFFITH_0000285</td>
<td>A6E37E912E79BE9AF112D4F0AF175F0D</td>
<td>Griffith John</td>
<td>Master</td>
</tr>
<tr>
<td>1039582</td>
<td>JGRIFFITH_0000286</td>
<td>6DBC9FF8489363C8C62E5F4D9AB9186</td>
<td>Griffith John</td>
<td>Duplicate</td>
</tr>
<tr>
<td>1039579</td>
<td>JGRIFFITH_0000287</td>
<td>A6E37E912E79BE9AF112D4F0AF175F0D</td>
<td>Griffith John</td>
<td>Duplicate</td>
</tr>
<tr>
<td>1039580</td>
<td>JGRIFFITH_0000288</td>
<td>9CCF63F5E9EF0A4CABC5ABC2EA9EF41</td>
<td>Griffith John</td>
<td>Master</td>
</tr>
<tr>
<td>1039581</td>
<td>JGRIFFITH_0000289</td>
<td>9CCF63F5E9EF0A4CABC5ABC2EA9EF41</td>
<td>Griffith John</td>
<td>Master</td>
</tr>
<tr>
<td>1039582</td>
<td>JGRIFFITH_0000290</td>
<td>7656E009DD8D624AE32374FCDCE11F3</td>
<td>Griffith John</td>
<td>Master</td>
</tr>
<tr>
<td>1039583</td>
<td>JGRIFFITH_0000291</td>
<td>9CCF63F5E9EF0A4CABC5ABC2EA9EF41</td>
<td>Griffith John</td>
<td>Duplicate</td>
</tr>
</tbody>
</table>

#### 38 Performing QC of Productions

Creating and exporting a production successfully can be complicated. It's important to perform a QC of your productions to ensure accuracy.

#### 38.1 Recipe overview

This recipe points out all possible areas of conflict and errors to consider when creating and exporting a production.

#### 38.2 Requirements

- Workspace Access
  - Saved Searches
  - Production sets
- Relativity Desktop Client
- This recipe is applicable to all versions of Relativity.
38.3 Directions

When you're preparing to create a production, it's beneficial to verify some standard checklist items to prevent incorrect items mistakenly included in a production. Many of these items are intended to check for conflicts between items in a group and to show where unwanted items are brought in while segmenting desired items and including a group.

This recipe assist you in preparing a plan on how to move forward when locating unwanted data. This recipe also assist you in identifying who reviews this data and when a review of the data should occur.

See [Using saved searches to complete conflict checks](#) for conflict searching when considering the following in the production set:

### 38.3.1 Conversation before Production

Consider the following questions before you create a production:

- What field are reviewers using to code responsiveness of documents?
- What field are reviewers using to denote privilege?
- What markup set does the production use?
- Which fields do you want to produce?
- Which fields do you want to produce if you're producing redactions?
- Is there a production set tag?
- What order are you producing items in?
- What is the bates prefix?
- How many digits are in the bates number?
- What language do you want to use on the placeholders? (if applicable)
- What type of production is it? (images, natives, both)
- Do you want to export tiffs or PDFs?

### 38.3.2 Items to check in a production set

Review the production set for the following document codings:

- Not responsive
- Privileged
- Redacted, not tagged redact (make sure to search on proper markup set)
- Tagged redact, not redacted (make sure to search on proper markup set)
- Previously produced
- Privilege screen terms, not marked privileged
- Imaged and to be produced natively
- To be imaged, not imaged
- Corrupt/unprocessable in production set
- Password protected documents in production set
- Duplicate/email thread/near duplicate of privilege document
- Not reviewed/question/not sure for responsiveness or privilege
- Outside of relevant date range
38.3.3 QC of production data

Review load files for:

- Ensure the number of rows in DAT file is correct.
- Ensure the fields in the DAT file are correct.
- Ensure the native and text paths are correct.
- Verify the number of images in image load files.
- Verify the number of natively produced documents.
- Ensure the bates and sort order are correct.
- Ensure the header/field names in DAT file are correct.

Sample and Check Review images for:

- The correct first bates number.
- The correct confidentiality endorsement.
- Redactions are in the appropriate location.
- Any other special endorsements.
- The correct number of images from image load files.
- The image type (black and white, color).
- No thumbs.db file is present.

Review text files for:

- Ensure the text of the first document matches image of the first document.
- The OCR text for redacted documents.
- Ensure empty text files only exist for documents with no text in database.
- Ensure no thumbs.db file is present.
- Ensure file names contain the bates number.

Sample and Review native files for:

- The correct number of native files.
- The proper extensions/document types produced natively.
- No thumbs.db file is present.
- Ensure file names contain the bates number.

39 RAR checklist

Before you begin a RAR project, think about whether the project is a good fit for assisted review. Assisted review centers on the concept of training the system, so it learns how to interpret uncategorized documents. The system learns best from documents that are good examples. To be good examples, documents should have rich text with lots of concepts, not just numbers.

Also consider what constitutes a responsive document. If, for instance, responsiveness hinges on a name or a date, that’s not enough for RAR, because there are no concepts to learn, only absolutes. Successfully completing a RAR project requires you to spend a little time at the beginning of the process to determine if RAR is the best way to proceed.
Every RAR project has specific needs, goals, and deliverables. Customize the following checklist to fit the needs of each project. The checklist is also useful as a guide to the most commonly required items for RAR.

### 39.1 Checklist

1. Ensure the document set you plan to use is a good population for Assisted Review. Ensure the population contains the following:
   - Minimum 50k records with text
   - Concept rich files (not primarily numbers)
   - Issue or privilege coding is in a separate field or not part of RAR workflow
2. Make sure you set your timeline and goals. The stakeholders should discuss goals and timelines prior to beginning a RAR project. This helps establish clear deliverables. Decide and determine the following:
   - Determine level of Precision, Recall, and F1
   - Decide on a manual review plan (i.e., all docs categorized as Responsive; privilege screen only)
   - Put a production plan in place
3. Ensure your administrative prep is complete. Use the following to ensure you complete all administrative tasks before reviewers look at documents. This decreases the chance of delays during the project.
   - Install the Assisted Review application
   - Create a layout and fields for the project
   - Build a layout
   - Create a RAR designation field
   - Create choices
   - Create the use as example field
   - Create a text excerpt field
   - Build a Reviewers’ view
   - Complete the index build
   - Eliminate documents without enough text
   - Run the extracted text size script
   - Create saved searches (training set and searchable set)
     a. Optional – Use the following if the project requires categorizing for issues:
        - Key issues field (single choice)
        - Key issues excerpt field (long text)
        - Cluster all documents in RAR universe
        - Pull UNCLUSTERED documents from project
     b. Optional - Use the following to further cull documents:
        - Email threading
        - Key word culling
        - STR
4. Ensure you prepare your reviewers. Reviewer preparation is key to success. A RAR project isn’t like other document coding. Use all the tools available to train all reviewers in RAR protocols.
   - Show the RAR for End Users webinar to reviewers
   - Distribute and discuss Assisted Review Reviewer Protocol
5. Complete the RAR project. Planning in advance ensures a successful wrap up. Ensuring all tasks are complete is important for the client’s satisfaction as well as defensibility.
   - Meet project goals
   - Precision/recall
   - Achieve stabilization
   - Manual review under way
   - Complete production

40 Relativity Assisted Review Reviewer Protocol

40.1 Overview

This recipe is intended as a reference for reviewers in a Relativity Assisted Review project. We’ve included some best practices and special considerations that are unique to a computer-assisted review.

40.2 Requirements

Analytics Server

40.3 What is Assisted Review?

Computer-assisted review is a workflow that captures and analyzes reviewers’ coding decisions and amplifies them across a data set. Based on user input, the system learns how to categorize conceptually similar documents in the document universe.

Relativity Assisted Review utilizes a specific type of text analytics called latent semantic indexing (LSI). It is important to remember that LSI uses only text in its decision-making; numbers, symbols, and images are not considered during the machine learning process.

40.4 Selecting good example documents

Because all machine learning is derived from text, it is important to note that some documents may be highly responsive but undesirable as example documents for an Assisted Review project.

40.4.1 Sufficient text

In order for a document be considered a good example for machine learning, it must contain a sufficient quantity of text to train the system. Assisted Review’s text analytics engine learns from concepts, rather than individual words or short phrases. It’s best to think in terms of sentences or paragraphs rather than a few words when deciding if a document has sufficient conceptual language. For that reason, very short documents that contain only a few words or phrases are typically not good examples.
40.4.2 Email headers and repeated content

Email headers, confidentiality footers, and other types of repeated content are typically filtered out prior to review. They should not be considered when determining whether a document is a good example for a computer-assisted review.

Consider the following document:

![Email header example]

In this example, the system will only learn from the text that is not framed by a red box. Even if the subject line and short sentence fragment that remain are responsive, there is not enough text to warrant this document’s inclusion as an example.

40.4.3 Images

Consider the following example from the Enron data set. This document, which appears to contain useful text, is actually a JPEG image of a paper document:
A reviewer might read the above language and find the document to be responsive. However, when we switch to the extracted text of the document we see the following:

Because the system only works with a document’s extracted text, all of the responsive text located in the image will be unavailable for machine learning. Consequently, this document, while highly responsive, turns out to be a poor example document.

40.4.4 Numbers

As mentioned above, numbers are not considered in the machine learning process. It follows that spreadsheets consisting largely of numbers, while potentially responsive, do not make good example documents.

Consider the two examples below. Both are spreadsheets, but only the second would make a good example document.
<table>
<thead>
<tr>
<th>REGION</th>
<th>REGION</th>
<th>Sum of DEALS</th>
<th>% of TOTAL DEAL COUNT</th>
<th>Sum of VOLUME</th>
<th>% of TOTAL VOLUME</th>
<th>Sum of VALUE</th>
<th>% of TOTAL NOTIONAL VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CENTRAL</td>
<td>EOL</td>
<td>174,414</td>
<td>86</td>
<td>13,498,922,762</td>
<td>62</td>
<td>16,782,521,064</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>OTC</td>
<td>27,792</td>
<td>14</td>
<td>3,058,917,028</td>
<td>38</td>
<td>20,650,852,684</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>202,206</td>
<td></td>
<td>16,557,839,788</td>
<td>98</td>
<td>37,433,373,748</td>
<td>100</td>
</tr>
<tr>
<td>EAST</td>
<td>EOL</td>
<td>152,618</td>
<td>77</td>
<td>11,613,432,735</td>
<td>48</td>
<td>21,859,684,338</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>OTC</td>
<td>40,780</td>
<td>22</td>
<td>12,603,937,390</td>
<td>52</td>
<td>20,934,173,185</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>193,398</td>
<td></td>
<td>24,217,369,125</td>
<td>78</td>
<td>42,793,857,523</td>
<td>80</td>
</tr>
<tr>
<td>ECC-CANADA WEST</td>
<td>EOL</td>
<td>70,777</td>
<td>30</td>
<td>7,267,166,377</td>
<td>47</td>
<td>21,152,634,317</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>OTC</td>
<td>30,614</td>
<td>12</td>
<td>6,208,359,206</td>
<td>39</td>
<td>20,332,985,566</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>101,391</td>
<td></td>
<td>13,475,525,583</td>
<td>53</td>
<td>41,485,620,883</td>
<td>29</td>
</tr>
<tr>
<td>CNA-CANADA EAST</td>
<td>EOL</td>
<td>12,569</td>
<td>21</td>
<td>1,124,971,796</td>
<td>61</td>
<td>1,111,172,440</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>OTC</td>
<td>3,322</td>
<td>21</td>
<td>713,294,347</td>
<td>29</td>
<td>2,632,316,915</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>15,891</td>
<td></td>
<td>1,838,266,143</td>
<td>60</td>
<td>3,743,489,355</td>
<td>30</td>
</tr>
<tr>
<td>DL-DAILY-EST</td>
<td>EOL</td>
<td>30,798</td>
<td>86</td>
<td>6,633,131,653</td>
<td>76</td>
<td>31,239,679,833</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>OTC</td>
<td>5,942</td>
<td>14</td>
<td>2,890,471,797</td>
<td>24</td>
<td>2,058,365,358</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>36,740</td>
<td></td>
<td>9,523,603,450</td>
<td>100</td>
<td>51,398,045,191</td>
<td>100</td>
</tr>
<tr>
<td>NO-PRICE</td>
<td>EOL</td>
<td>154,006</td>
<td>71</td>
<td>61,443,109,017</td>
<td>40</td>
<td>387,076,402,780</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>OTC</td>
<td>60,376</td>
<td>29</td>
<td>92,294,711,102</td>
<td>60</td>
<td>408,640,744,946</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>214,382</td>
<td></td>
<td>153,737,820,119</td>
<td>79</td>
<td>795,717,147,727</td>
<td>80</td>
</tr>
<tr>
<td>TEXAS</td>
<td>EOL</td>
<td>21,867</td>
<td>56</td>
<td>3,637,944,922</td>
<td>31</td>
<td>1,759,836,154</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>OTC</td>
<td>10,951</td>
<td>44</td>
<td>8,059,565,596</td>
<td>56</td>
<td>13,421,562,534</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>32,818</td>
<td></td>
<td>11,697,510,518</td>
<td>77</td>
<td>25,181,398,688</td>
<td>70</td>
</tr>
<tr>
<td>WEST</td>
<td>EOL</td>
<td>126,096</td>
<td>79</td>
<td>16,541,615,390</td>
<td>51</td>
<td>21,223,379,220</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>OTC</td>
<td>32,088</td>
<td>21</td>
<td>15,798,806,153</td>
<td>49</td>
<td>20,083,286,542</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>158,184</td>
<td></td>
<td>32,340,421,543</td>
<td>78</td>
<td>41,306,665,762</td>
<td>74</td>
</tr>
<tr>
<td>TOTAL</td>
<td>EOL</td>
<td>731,875</td>
<td>77</td>
<td>121,793,163,892</td>
<td>45</td>
<td>431,184,230,311</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>OTC</td>
<td>221,006</td>
<td>23</td>
<td>148,168,879,818</td>
<td>55</td>
<td>533,432,373,748</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>952,881</td>
<td></td>
<td>269,962,043,710</td>
<td>100</td>
<td>964,616,593,939</td>
<td>100</td>
</tr>
</tbody>
</table>

Spreadsheet – Bad Example
40.4.5 Families and the "Four Corners" test

A document is only a good example if there is text on the document’s face—within the four corners of the document—that makes it responsive.

The following scenarios violate the Four Corners Test, and will not offer good example documents:

- The document is a family member of another document which is responsive.
- The document comes from a custodian whose documents are presumptively responsive.
- The document was created within a date range which is presumptively responsive.
- The document comes from a location or repository where documents are typically responsive.

This issue is especially prevalent with regard to document families, so additional emphasis is warranted:

A reviewer should never include a document as an example based on the content of a family member.
For example, consider the following email. Note that it mentions an attachment. If the attachment is responsive, the reviewer might be tempted to include the email as a responsive example, too. Doing so would violate the Four Corners Test. Again, if there is no sufficient, responsive language on the document’s face, it should not be used as an example.

<table>
<thead>
<tr>
<th>From:</th>
<th>Jack Boatman</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sent:</td>
<td>Friday, January 12, 2001 5:29 PM</td>
</tr>
<tr>
<td>To:</td>
<td>John Hodge; Raimund Grube</td>
</tr>
<tr>
<td>Subject:</td>
<td>Draft Letter</td>
</tr>
<tr>
<td>Attach:</td>
<td><a href="#">Southeast Florida 011201.doc</a></td>
</tr>
</tbody>
</table>

Initial draft for your review and comments.

**********
EDRM Enron Email Data Set has been produced in EML, PST and NSF format by ZL Technologies, Inc. This Data Set is licensed under a Creative Commons Attribution 3.0 United States License [http://creativecommons.org/licenses/by/3.0/us/](http://creativecommons.org/licenses/by/3.0/us/). To provide attribution, please cite to “ZL Technologies, Inc. ([http://www.zli.com](http://www.zli.com)).”
**********

### 40.5 Handling good language / bad example exceptions

Quite frequently, a reviewer may encounter a document containing highly responsive language which should be coded as non-responsive for an external reason, such as a carve-out agreement between the litigating parties. For example, consider the draft contract fragment below:
Let’s assume that the parties have agreed to produce only final, executed contracts. This contract is a draft, which would make it non-responsive. However, this document contains a great deal of extremely responsive language, and to submit it as a non-responsive example would teach the system incorrectly. Consequently, the correct action here is to mark it as non-responsive, but not submit it as an example.

**40.6 Applying the Use as Example field**

Users select or reject documents as examples for the system via the Use as Example field. During categorization, Assisted Review will check the Use as Example field checkbox for all documents that are part of a project sample set.

If a document appears to be a good example for machine learning, simply code the designation field as you normally would and leave the **Use as Example** field checked.

If a document is known to be responsive but is not a good example, code the document as responsive and uncheck the Use as Example button.

**40.7 Using the Excerpt text box**

When a document is mostly non-responsive but contains examples of responsive text, use the Excerpt Text box. Highlight the responsive text, right click, and choose **Add to Excerpt Text**. Code
Relativity pastes the selection into the Excerpt Text field on the Assisted Review coding layout. You can perform this action multiple times per document. Each time you select text and click Add to Excerpt Text, the text is automatically appended to the Excerpt Text field.
However, it is important to remember that good example language is measured in concepts (sentences and paragraphs) and not single words or short phrases. Do not use the Excerpt Text option for short keywords or phrases.

### 40.8 Coding Tips

- Consistency is crucial.
  - Consult fellow reviewers on difficult coding decisions to ensure unanimity.

- Don’t touch.
  - Never add choices to the Designation tag.
  - If you are unsure about a document or have a technical difficulty, consult with the project manager to identify workflow solutions.

- Double check.
  - Always check the extracted text of a document to be sure it matches the content in other views. Whenever possible, review from the Extracted Text viewer.

- When in doubt, ask.
  - If there is an aspect of the Assisted Review workflow that is confusing, do not guess. Ask a system admin or project manager about the proper course of action.

Optimal use of Assisted Review mandates careful adherence to the recommended workflow. Always consult with a system admin or review manager when confusion or problems arise. For additional support for Assisted Review, please contact solutions@relativity.com.

- **Assisted Review**
41 Recording page-level information for a document

Occasionally, documents may not be paginated correctly or may require changes to pagination for use as exhibits. This workflow allows you to note these page breaks.

41.1 Recipe overview

This recipe uses an object to store information about a document to call out page-level information.

41.2 Requirements

- Workspace access
- Object creation permission

41.3 Directions

Perform the following steps:

1. Create a new object type called Work Orders.
2. Create fields in the Work Orders type for coding values, such as Page Number, Bates Number, or Description.
3. Add the newly created fields to a document coding layout.
4. Set the Work Orders object as an associative object on the layout. The Work Order tab can hold all the individual Bates Numbers of the work orders.
5. You can also create a document view to identify the documents with work orders.

41 Dynamic Objects

42 Regular expression searching - SSN and EIN

During a review, you may want to search documents for social security numbers and/or employer identification numbers. Customizing a dtSearch index and using regular expressions can assist in locating these numbers.

42.1 Recipe overview

This recipe shows you how to customize a dtSearch index and use a regular expression to return a set of documents containing social security numbers or employer identification numbers.

42.2 Requirements

- Relativity 7.2 or higher
- Workspace access
  - Search index – Edit/Add and corresponding tab
42.3 Directions

1. Create a dtSearch index.
2. Change the alphabet file to include the dash as a searchable character.
   a. Enter the following value under [Letters] // Original letter, lower case, upper case, unaccented:
      i. [Space] [-][Space] [-][Space] [-][Space] [-]

         *Note: You must have a leading space. You can't have a trailing space.*

   ii. Place a second instance of the string for lower case.

   b. Delete the definition of [Hyphens].
3. Perform a full build of the dtSearch index.

4. Run a search using the following regular expression for social security numbers:

```
"##((?!000)(?!666)([0-8]d(2)))\-(\(?!00)\d(2))\-(\(?!0000)\d(4))"
```

5. Run a search using the following regular expression for employer identification numbers:

```
"##\d(2)\-\d(7)"
```

**Note:** The quotes and # characters are necessary for the dtSearch index to recognize the regular expression. Quotes can't be Smart Quotes; they must be ASCII quotes.

### 42.4 General considerations for all regular expression searches

Each regular expression in a search request must match a single word in the index. This results in the following limitations:

- A regular expression can't contain a space or any character that's indexed as causing a word break.
- It can only reference searchable letters.
- The terms are subject to the maximum length of a word, which is 32 by default. You can change this limit using the Options.MaxWordLength configuration table value.
- The case should match the case in the index (lower case unless the index is case-sensitive).

**Note:** dtSearch uses TR1 regular expressions as implemented in Visual C++, which [Microsoft's Developer Network](https://docs.microsoft.com/en-us/visualcpp/regex) documents.

- [dtSearch](https://www.dtsys.com/)

---

Relativity | 9.2 Recipes - 136
43 Searching for symbols

This recipe demonstrates how to make symbols searchable in a dtSearch index. It also demonstrates how to use regular expressions to search for certain symbols reserved as search operators, such as the % sign.

43.1 Requirements

- Relativity 7.2 or higher
- Workspace access
- Search index – Edit/Add and corresponding tab

43.2 Directions

1. Create a dtSearch index.
2. Change the alphabet file to include the % sign as a searchable character:
   a. Enter the following under [Letters] // Original letter, lower case, upper case, unaccented:
      i.  [Space] [%] [Space] [%] [Space] [%] [Space] [%]
   Note: You must have a leading space. You can’t have a trailing space.
      
      **Alphabet**

      ![](image)

      ii. Place a second instance of the % % % % string in the lower [Letters] section immediately
after __ __:

<table>
<thead>
<tr>
<th>Alphabet</th>
</tr>
</thead>
<tbody>
<tr>
<td>X x X X</td>
</tr>
<tr>
<td>Y y Y Y</td>
</tr>
<tr>
<td>Z z Z Z</td>
</tr>
<tr>
<td>% % % %</td>
</tr>
<tr>
<td>a a A a</td>
</tr>
<tr>
<td>b b B b</td>
</tr>
<tr>
<td>c c C c</td>
</tr>
<tr>
<td>d d D d</td>
</tr>
</tbody>
</table>

b. Delete the % sign from the [Ignore] section.

c. Repeat these steps for any other symbols that you need to make searchable. Delete the appropriate symbols from the [Spaces] section.

3. Perform a full build of the dtSearch index.

You can now search for terms containing the % sign using a regular expression. For example, if you need to search for documents that contain the term 75%, you would enter the following in your search box (ensuring you select the proper dtSearch index):

![Search Terms](image)

W is the regular expression metacharacter that stands for any symbol. A dtSearch for "#75\W" hits on 75%, as well as 75# and 75! if you had made # and ! searchable. "##" signals to Relativity to treat the string as a regular expression. Let’s say you need to search for documents that contain any variety of percentage values, where the strings are least 2 numbers and at most 3, such as: 100%, 37%, 214%, 95%. You could use the following regular expression:
The quotes and # characters are necessary for the dtSearch index to recognize the regular expression. Quotes can’t be Smart Quotes; they must be ASCII quotes.

Note: For most symbols, once you have made them searchable in the alphabet file, you can type them directly into the dtSearch box without using regular expressions. Thus to search for 30!, enter 30! into the dtSearch box after you have made ! a searchable character. You only need to employ \W for certain symbols such as the % sign.

43.3 References

- Searching with Regular Expressions
- Searching with Regular Expressions
- Regular Expressions for Beginners webinar

44 Relativity Binders best practices

44.1 Introduction

Relativity Binders is an iPad app for partner-level attorneys who need access to their most important case documents from anywhere. This recipe will walk you through some basic best practices for using and administering the app.

44.2 Requirements

- Relativity Binders Application in Relativity
- iPad with Relativity Binders

44.3 Best practices

44.3.1 Internet connection

- While Binders can be used via a 4G data connection, Wi-Fi is always the preferred data connection, especially when a user is downloading multiple binders.
- If you have a poor data connection and need to download binders, avoid using the Download All button. Instead, download one binder at a time in order of importance.
44.3.2 Creating binders

All documents are converted to PDFs.

A binder can be created from:

- A saved search
- A saved search and overlapping production set

**Note**: If a document exists in both the saved search and the production set, Relativity will use the production image for the binder.

44.3.3 Binder size considerations

- While Binders can manage having 100+ documents, a binder with 30-50 documents is optimal for navigation and reading purposes.
- The ideal number of documents for your binders—and the number of binders that should be stored on your device—also depends on your iPad’s available storage space.
  - Repeated warnings will appear on your iPad once you start running out of space.
- For reference, 400 pages of transcripts take up just 5 MB of space—less than a single song on iTunes. Conversely, presentation decks can be only 10 slides but, depending on content, may require 17 MB in space.
- You can delete a binder to free up additional space on your iPad.
  - Annotations will be backed up on the Relativity server.
  - If you download the binder again, annotations will be preserved.

44.3.4 Binders and system performance

- Relativity syncs with Binders when a Relativity account has been linked. It will download all workspaces and binders associated with each account.
- You will only see binders that are linked to your Relativity account. The more binders you have, the more syncing is required. This may impact iPad performance.
- On the iPad, the best practice is a maximum of 10 binders per workspace. Within Relativity, you may have as many binders as necessary to support users.
- In Relativity, the number of workspaces with binders linked to users has an impact on system performance.
  - This metric affects performance because the API has to call and pull back information for every workspace that has linked binders. The more items to sync, the more resources it takes from the iPad.
  - Once a case or workspace is no longer needed, unlink the users from any binders in that workspace. This will prevent Relativity from syncing with Binders.

[Binders Admin]
45 Relativity Binders field preparation

45.1 Recipe overview

Though Relativity Binders offers an easy way to bring your database with you, there’s difficulty in displaying the valuable data in the iPad’s abbreviated format. This recipe will walk you through how to combine fields to get the most out of metadata fields.

45.2 Requirements

You want to get the most out of each record when looking at your document list from Binders, but space is limited.

Because of this space limitation you can get creative with the fields. Creating a new field that combines two other metadata fields will provide a more informative document summary.

Examples:

- **File Name**
  - Combine DocID and File Name field (e.g., ABC000001 - Enron North America.doc)
  - Combine DocId and Custodian (e.g., ABC000001 – Kenneth)

- **Date**
  - Make a combined field of both Date Sent for emails and Date Created for other files
  - Propagate the Sent Date to the attachments

- **Excerpt**
  - Attorney notes
  - Move From and To information into a new field to use as excerpt
  - Move Issues to a new text field

**Binders Admin**
46 Relativity Binders FAQs

The Relativity Binders FAQ list provides answers to common questions regarding implementing, configuring, securing, and using Binders.

46.1 General

The following frequently asked questions apply to general Relativity Binders requirements and app usage.

Is there an additional license requirement to use Binders?
No, you just need a Relativity license and login to use Binders.

Can I use existing images in Relativity with Binders?
No, Binders creates images through Native Imaging for all documents including production images.

Can someone search within a binder?
No, searching is currently available only document-wide.

Are binders static or dynamic?
Binders are static. You must actively update a binder in Relativity through an incremental or full build for them to show up on the iPad. If you tag new documents as hot and want them in your binder, you must add them to the binder via an incremental build to have them on the iPad. If you tag documents that are already in a binder, the user must have the field map set to show the field you tagged, and he or she must delete and re-download the binder to see the new field map.

For more information regarding updating binders, mapping fields, and configuring user profiles, see Setting up binders for app users.

Do you need to tag searches or production sets as binders, or are all searches and production sets automatically made into binders?
You will need to create a binder from the saved search and/or production set. They are not automatically made into binders.

For steps to create a binder with a saved search, see Setting up binders for app users.

Can binders be accessed by groups or are they user-specific?
Binders are accessible by users added to the Binders ADS app. These users must first be part of a group that has permissions to the workspace.

For steps to grant users access to a binder, see Setting up binders for app users.

If you have a large binder, but a slow connection speed, can you look at a portion of the binder or do you have to wait until it loads fully?
You have to wait for a binder to completely download before viewing the documents inside it.

**Do binders require a great deal of storage on the iPad?**

This depends on the native document type, size, and your imaging and OCR profiles.

### 46.2 Annotations

The following frequently asked questions apply to using and storing annotations with Relativity Binders.

**Is redaction an option?**

Relativity Binders does not have coding functionality. It’s currently intended for partner-level attorneys and has functionality that includes making notes, highlighting, drawing, organizing annotations, and instantly syncing/downloading documents from Relativity.

**Are notes and annotations created in Binders viewable in Relativity?**

Notes and annotations created in Binders are not viewable in Relativity. They are only backed up for iPad syncing purposes.

**Are notes and changes made on the iPad searchable in Relativity?**

Notes, highlights, and drawings are not visibly added to the Relativity database and cannot be searched or indexed. Currently, these annotations are saved in Relativity for backup purposes only.

### 46.3 Security and authentication

The following frequently asked questions address security concerns and authentication types.

**Is the app integrated with RSA authentication?**

Relativity Binders currently does not have RSA authentication integration available.

**If an iPad is lost, can someone get to your documents without going through the app?**

If an iPad is lost and someone trying to access the device does not have the iPad unlock code, the documents in Binders are encrypted and cannot be exported as unencrypted. If Binders is installed and used on a jail broken iPad, Relativity does not ensure encryption on this device.

**Is it securable so a private search is a private binder?**

Private searches do not create private binders. You will have to item level secure binders from groups.

**Does Relativity Binders support Good for Enterprise?**

No, currently the Binders app does not support any Good Technology mobile products.
We use Active Directory authentication for our users. How will that affect the setup? Would users use their domain login?

Users cannot use their domain login. Binders uses Relativity login accessibility. Users must log in with their Relativity login credentials. If a user does not know his or her username or password, the user must contact their system admin to obtain their login or reset their password.

46.4 Emailing documents

The following question applies to emailing documents from Binders.

Is it possible to email multiple documents at one time?

No, you can only email one page of a document or one document at a time.

47 Relativity Structured Analytics – Best practices and workflows

Published September 30, 2013

This recipe provides some best practices around the setup and use of Structured Analytics.

47.1 Requirements

- Relativity 8.0 or higher
- Workspace access and system admin rights
- Relativity Structured Analytics application

47.2 Directions

Consider the following guidelines when using Structured Analytics in Relativity:

- There are advantages to running each of the Structured Analytics Set operations individually. For example, if you need to make changes to a particular operation, it is best to make the changes and rerun the set for that one operation, rather than re-running a set with multiple operations selected.
- The ideal groups of Structured Analytics operations to perform as separate sets are:
  - Email threading and Near Duplicate Identification
  - Language Identification
  - Repeated Content Identification
- When setting up email threading, the Identify emails in document set search items should also be in the Select document set to analyze search, but the documents to analyze should also include attachments.
When running Near Duplicate Identification, system admins have two options:

- **Run with email threading** – the Near Duplicate Identification process identifies all textual duplicates within your document set which are not emails, while the email threading operation identifies all the email duplicates.
- **Run it without email threading** (either by itself or with another Structured Analytics operation) – the Near Duplicate Identification process includes all email and non-email textual duplicates.

When performing Repeated Content Identification, you should change the default settings if you’re working with large data sets. See [Sampling for repeated content](#) for more information.

Once you run the Structured Analytics Set, the following workflows can assist with an efficient review process.

- System admins can set up a review to include only the inclusive email threads. Listed below are some settings for a Saved Search that can be used to batch documents:
  - Search name and settings (you can choose whether to include family members)

  ![Search settings](image1)

  ![Search conditions](image2)

- Search conditions
Similar to identifying email threads, you can set up a review for all near duplicate items by setting up a saved search to identify such documents.

- Search name and settings (you can choose whether to include Near Dupe Groups)
As a system admin, you can continue to manually identify repeated content items and include them as a part of the Repeated Content Filters tab. Alternatively, you can use the Repeated Content Identification feature and link to an Analytics profile.

To facilitate review in multiple languages, system admins can use the Language Identification feature to batch out documents based on the primary language identified, or use the Pivot feature to identify and filter on the languages the system identified.

Analytics
48 Relativity Structured Analytics - Email Threading

Published January 22, 2015

Relativity Analytics Email Threading reduces the time of email review by arranging entire email conversations in sequence. Structured Analytics identifies inclusive emails, the most complete message in a thread. This allows reviewers to bypass other emails in a thread that contain only part of the inclusive content. Reviewing non-inclusive emails is unnecessary.

This recipe provides the necessary setup instructions to properly view and utilize Email Threading results and reports.

48.1 Requirements

- Relativity 8.0 or higher
- Workspace access and system admin rights
- Relativity Structured Analytics application

48.2 Directions

48.2.1 Create saved searches (Relativity 9.5.89.76 and lower)

First, create the following saved searches when setting up a Structured Analytics Email Threading set.

- **Saved Search #1** = Parent emails in document set: Create a saved search that returns only the parent emails. The search conditions you set will depend on the metadata fields available in your workspace.
  - "Email From" field is set
  - Parent Document ID is not set

- **Saved Search #2** = Parent Emails and Family: Create a saved search that returns all documents across which you intend to run email threading. You can set the saved search so that it returns all the documents in your workspace or limit the results according to any particular conditions you may want to set. Ensure that you include both the parent emails and any attachments in your search results. You can do this by setting the drop-down to Include Family.
  - (Saved Search) : Document is in [Parent Emails]
  - Include Family

**Note:** Relativity version 9.5 and above, documents with more than 30 MB of extracted text will automatically be excluded from the structured analytics set and you no longer need to manually exclude them using an additional search.

- **Saved Search #3** = Document set to analyze: Using search #2 above, exclude documents based on extracted text size. Create a third saved search with the following conditions:
48.2.2 Create or configure an Analytics profile

Next, create or configure an Analytics Profile for your Email Threading set. Perform the following steps:

1. From the Indexing and Analytics tab, select Analytics Profiles. A list of existing Analytics profiles displays.
2. To create a new profile, click New Analytics Profile. Or, you can edit an existing profile. The Analytics Profile layout appears.

   **Note:** Rather than modify the Default profile, you should always create a new profile or edit an existing profile you have created.

3. **Filter Configuration** - these settings affect the building of the conceptual Analytics Index, but will not have any effect on your Structured Analytics Email Threading set. Thus, you may leave the default settings as is in this section. (filter configuration moved to the Analytics index starting in 9.5.196.102).

4. **Email Threading Settings** - map the email threading fields on the Analytics Profile Layout to the appropriate workspace fields containing the following specified metadata.

   - **Email Header Fields:**
     - **Email from field** - sender’s email address.
     - **Email to field** - recipients’ email addresses.
     - **Email cc field** - carbon copied (Cc) recipients’ email addresses.
     - **Email bcc field** - blind carbon copied (Bcc) recipients email addresses.
     - **Email subject field** - email subject line.
     - **Email date sent field** - email sent date. Typically a date field, this field needs to contain both the date and time.

   **Note:** When setting up an Email Threading Structured Analytics set, if you choose to not use email header fields, then you will not need to map any of the above email header fields. In this scenario, email threading relies on extracted text, and the Parent Document ID and Attachment Name fields.

   - **Email Metadata Fields**
     **Note:** The following fields are all optional.

     - **Attachment name field** - file name of the document attachment. This only needs to be populated for child documents and not the parent email. This field is used to display an attachment’s file name and icon in the Email Threading Display visualization field. Relativity Processing creates the Attachment name field. You may use a File name field instead.
     - **Group ID field** - relational family group field which is used to group emails with their attachments.
5. Advanced Options - these settings affect the building of the conceptual Analytics Index, but will not have any effect on your Structured Analytics Email Threading set. Thus, you may leave the default settings as is in this section.

48.2.3 Create a Structured Analytics set

You are now ready to create a Structured Analytics Email Threading Set.

1. From the Indexing and Analytics tab, select Structured Analytics Set. A list of existing Structured Analytics sets displays.
2. To create a new set, click New Structured Analytics Set. The Structured Analytics set layout appears.
3. Set Structured Analytics Set Information:
   - Enter a name for your Structured Analytics set, and then select Email threading as the operation to perform.
   - In Select document set to analyze, select the saved search you created that will return all documents across which you wish to run email threading.
   - In Analytics server selection, select Automatic.

Enter your email address in the "email addresses" text box if you wish to receive a notification when the email threading analysis is completed.

4. Set Email Threading Settings:
   - In Identify parent emails in document set, select the saved search you created that will return only parent emails from the document set across which you wish to run email threading.
In **Select profile for field mappings**, select the Analytics profile that you had set up earlier.

In **Use email header fields**, select Yes in order to use any email header fields in your workspace for the threading analysis. Setting this option to No will not send any email header fields for analysis, forcing Structured Analytics to rely only on the extracted text field of an email for threading. If you choose No, you will not need to map the email header fields in your Analytics profile.

Click **Save**. The Structured Analytics Set Console appears.

5. **Run Email Threading Analysis**:

- Click **Run Full Analysis** to begin running your Email Threading analysis.

You can monitor the progress of the email threading analysis in a separate window by clicking on the pop-out icon in the upper right corner of the progress pane.

### 48.2.4 Email threading results - reports

After running an email threading operation, we recommend reviewing the available reports. On the Structured Analytics Set console, click **View Email Threading Summary** to open the report. This report contains a graphical summary and tables that list a breakdown of all the emails analyzed. See **Viewing the Email Threading Summary**.
The Email Threading Summary pie chart provides a graphical representation of the percentages of inclusive emails, inclusive duplicate emails, and non-inclusive emails.

48.2.5 Setting up an Email Threading View

To view the results of the email threading operation, we recommend creating an Email Threading view in the Documents tab.

For more information on creating views, see Views.
Navigate to the Documents tab in your workspace.
Click to create a view.

1. **Email Threading View - Basic Information**
   - **View Information:**
     - Fill out the appropriate information for Name, Order, Visible in Dropdown.
   - **Advanced Settings:**
     Select the following Advanced Settings options when creating the Email Threading view:
     - Object Type - Document
     - Visualization Type - Indented List
     - Group Definition - Email Thread Group
     - Indentation Method - Numeric
     - Indentation Definition Field - Indentation
     - Display Field - Email Threading Display (Choose Control Number if you wish to see indentation dots in the view.)

2. **Email Threading View - Set Fields**
   Add the following output fields to your Email Threading view:
   - Email Threading Display
   - Email Thread Group
   - Email Threading ID
   - Inclusive Email
   - Inclusive Reason
   - Email Duplicate Spare
   - Control Number

3. **Email Threading View - Set Conditions**
   To only return email thread groups (or possibly documents that have been part of an email threading set) set the following condition:
   - Email Thread Group : is set

4. **Email Threading View - View Sorting**
   Configure each field below to sort in ascending order:
   - **Email Thread Group** - initial nine characters of the Email Threading ID. This value is the same for all members of an Email Thread Group which are all of the replies, forwards, and attachments following an initial sent email. This field is relational.
   - **Indentation** - a whole number field indicating the node level of the document.
   - **Email Threading ID** - ID value that indicates which conversation an email belongs to and where in that conversation it occurred. The first nine characters indicate the Email Thread Group, (all emails with a common root), and the subsequent five-character components show each level in the thread leading up to an email’s top segment. For example, an email sent from A to B, and then replied to by B, has the initial nine characters for the message from A to B, plus one five-character set for the reply.
Your email threading view will appear similar to the following. The blue line between rows separates distinct threads of email messages, and is activated as a result of setting the view's visualization type to Indented list.

**Note:** If you wish to see the latest email in a thread first, set Indentation to sort in descending order.

48.2.6 Email Threading Display

The Email Threading Display field provides the following visual information about email threading:

1. **Indentation Bubbles**
   
   The indentation bubbles indicate each email message's indentation level within the thread. For example, the first email in the chain would be "1," an email responding to the first email would be "2," and an email responding to the second email would have a "3." Email indentation bubbles go up to 99. For messages with an indentation level over 99, the bubbles display "99+.

   The color of the indentation bubble indicates inclusiveness. Inclusive email messages contain the most prior message content and have a "Yes" value in the Inclusive Email field.

   Emails that are both inclusive and non-duplicative appear blue. Non-inclusive or duplicate spare messages appear gray. In the screen shot below, Emails #1 and #2 are non-inclusive. Email #3 is inclusive and not a duplicate spare.

---

<table>
<thead>
<tr>
<th>#</th>
<th>Control Number</th>
<th>Email Threading Display</th>
<th>Email Thread Group</th>
<th>Email Threading ID</th>
<th>Inclusive Email</th>
<th>Inclusive Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EN446520</td>
<td>[Robert Newport]</td>
<td>T001</td>
<td>T001</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>EN446521</td>
<td>[Bill Deshkart]</td>
<td>T001</td>
<td>T001</td>
<td>Yes</td>
<td>ATTACHMENT</td>
</tr>
<tr>
<td>3</td>
<td>EN446524</td>
<td>[Q3 2014. earnings.pd]</td>
<td>T001</td>
<td>T001</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>EN446522</td>
<td>[Robert Newport]</td>
<td>T001</td>
<td>T001</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>EN446523</td>
<td>[Bill Deshkart]</td>
<td>T001</td>
<td>T001+</td>
<td>Yes</td>
<td>MESSAGE</td>
</tr>
<tr>
<td>6</td>
<td>EN446525</td>
<td>[Craig Middlebrooks]</td>
<td>T002</td>
<td>T002</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>EN446526</td>
<td>[Jennifer Barkley]</td>
<td>T002</td>
<td>T002</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>EN446527</td>
<td>[Craig Middlebrooks]</td>
<td>T002</td>
<td>T002</td>
<td>Yes</td>
<td>MESSAGE</td>
</tr>
</tbody>
</table>
2. **Message and file type icons**

   The Email Threading Display field includes the following file type icons:
   - **Reply** - original file name begins with RE:
   - **Forward** - original file name begins with FW:
   - **Other** - original file name does NOT begin with either RE: or FW:
   - **Unknown** - file type cannot be found.
   - **Email contains attachments** - file is an email containing attachments. Attachments are documents included in your emails in document set saved search. The following conditions apply to the icon display in the Relativity workspace:
     - The document type determines the attachment’s file type icon.
     - The attachment’s file name that appears is based on the value in the attachment name field.

### 48.2.7 Thread groups in the related items pane

To improve the review efficiency, the Email Threading Display field automatically appears when viewing the Thread Group in the related items pane. You can also add the Email Threading Display field to any other view in Relativity.
In the related items pane, click the Thread Group icon to display all messages in the same thread group as the selected document. You can also click the Email Duplicates icon to show all of the messages identified as duplicates of the selected document, or click the Near Dupe Groups icon to show all items that are textual near duplicates of the selected document.

After deploying the Analytics application, you can create fields that propagate coding choices to these Structured Analytics related groups. See Applying propagation to documents for more information.

48.2.8 Working with email threading results

After running an email threading operation and setting up your email threading view, use the following sample workflow to narrow down your documents to review.
48.2.8.1 Identifying unique documents to review

By reviewing only inclusive emails and skipping duplicates, your review process would be most efficient. The Analytics engine derives the email threads and determines which subset of each conversation constitutes the minimal inclusive set. Non-inclusive emails are redundant because all non-inclusive content is also contained in inclusive emails. The inclusiveness analysis ensures that even small changes in content will not be missed by reviewers.

To avoid redundant review, you should look at the Email Duplicate Spare field in conjunction with the Inclusive Email field. The Email Duplicate Spare field will be set to Yes for all but one duplicate in each group of duplicative emails.

A suggested workflow is to review the inclusives and their attachments, while skipping the duplicate spares and their attachments. These fields may be used in Saved Searches and views to easily locate the desired documents, as follows.

To identify and view a list of only unique documents using your email threading results and email threading view, perform the following steps:

1. Click the Documents tab in your workspace.
2. Select your new email threading view from the drop-down menu on the view bar.
3. In the drop-down to the right of In This Folder & Subfolders, select Include Family.
4. Click located next to the Search... drop-down menu to toggle on search conditions.
5. Add the following search conditions with an AND operator:
   - Field: Email Duplicate Spare
     Operator: isValue: No
   - Field: Inclusive Email
     Operator: isValue: Yes

6. Click Search.

To save this search, click the Save as search icon. See Saving searches on the Documents tab for additional information.

48.3 Special considerations

Consider the following guidelines when using Structured Analytics in Relativity:

- There are advantages to running each of the Structured Analytics Set operations individually. For example, if you need to make changes to a particular operation, it is best to make the changes and rerun the set for that one operation, rather than rerunning a set with multiple operations selected.
The ideal groups of Structured Analytics operations to perform as separate sets are:
  - Email threading and near duplicate identification
  - Language Identification
  - Repeated Content Identification

48.4 References

Email Threading

49 Removing documents from a RAR example set

When working in Relativity, you may need to remove certain documents from a RAR project if they're not good examples for training the RAR project. This recipe explains how to remove the set of example documents provided to a RAR project at any point during the RAR project.

49.1 Requirements

- Relativity access
- An active RAR project

49.2 Directions

This workflow assumes that you are between rounds in a RAR project (i.e., you've finished the current round and are ready to begin the subsequent round). Perform the following steps to remove documents from a RAR project after the start of the project.

1. Identify the documents that are not good example documents for the RAR project per the Reviewer Protocol specifications.
   1. Create a saved search against your RAR project universe with the following conditions:
      - (RAR Use As Example = Yes) AND
      - ((Extracted Text Size <= 0.75KB) OR
      - ((Extracted Text Size >= 2048KB) AND (RAR Excerpt Text = Not Set)) OR
      - (File Type = compressed files, RAR, ZIP, system files, image files, CAD drawings, maps, and calendar items, number-rich documents))

2. Once you've identified bad example documents, flag the documents by setting the RAR Use as Example field to No.

3. Proceed with the regular RAR workflow. Create new training or QC rounds per the needs of your project/reports (you can create a round with a single document to force categorization).

Once you complete the round in step 3, all the documents that were flagged RAR Use as example = No (in step 2) will no longer be used by the categorization engine to make suggestions on the documents in the RAR project universe.
49.3 References

- Analytics
- Relativity Assisted Review

50 Resolving common processing errors

Password-protected and corrupt files may appear in your processing projects. To add these files to your workspace, you need to remove the password restriction or repair the corrupt files.

50.1 Recipe overview

This recipe describes how to repair files and remove password protection from them. You can then add the files to your workspace during processing.

50.2 Requirements

- Relativity 8.0 or higher
- Relativity Processing
- Processing repository access

50.3 Directions

1. Select the Processing > Processing Sets tab.
2. Select the processing set with the errors that you want to resolve.
3. Click View Errors on the console for the current processing set. You can now view the errors in the processing set on the Processing > Errors tab.
4. Click on the name of the file with errors that you want to resolve. You can now view the error details, which lists two different file paths:

- **Document file location** - This path includes the actual file ending with the .doc file extension. You can click this link to open the document.
- **Document folder location** - This path indicates the path to the folder that contains the document with the error. You can click this link to open the folder.

5. Perform one of the following tasks to resolve the specific type of error:

- **Password protection error** - If password protection caused the error, use the password bank to resolve the error.
- **Corrupt file** - If a corrupt file is corrupt caused the error, click the **Document folder location** path. Locate the file in the folder. Next, open it with a tool used to repair corrupt files. After repairing the file, save the new file in the same file location. Ensure that the new file has the same file name as the corrupt one.

**Note:** When you repair the file, you modify it, therefore changing the meta data. Consider making a copy of the file before making any changes to preserve the original meta data.
1. Retry the discovery phase after you resolve the file errors:
   a. Use the mass operations window in the lower-left corner to check the files you fixed.
   b. Select Checked and Retry.
   c. Click Go.

2. Click Ok to retry the errors.

3. Monitor the retry progress via the progress window.
   - Processing

51 Reviewer QC workflow ideas

51.1 Overview

The quality and accuracy of the review team’s work is a very important component of any project. This recipe outlines some suggested workflows to perform Quality Control measures on first level reviewers’ work. The requirements for QC and the extent of QC conducted is always dictated by the needs of each project, but some form of QC will almost always be necessary. This recipe
discusses building and using a separate QC layout, setting up a QC sample set to gain insight into the team’s accuracy, selecting and applying key substantive and privileged terms to QC measures, and checking for tagging conflicts.

51.2 Requirements

- Relativity 8.1 or higher
- Workspace access
  - Use Sampling, Pivot, Tally/Sum/Average
  - Delete and Add Fields, Choices, Layouts, Views
  - Delete and Add all Search functions

51.3 Directions

General QC workflow

1. Create the QC Layout
2. Batch documents for the QC team to review
3. Identify total documents put through QC and resulting overturns

51.3.1 Create a Quality Control layout

Once the first level team completes its review on a given set of documents, the QC team can get started. The best workflow is to be sure first level review is complete on the set of documents before placing the documents in any QC queue. We suggest setting up a separate layout for the QC team’s exclusive use.

To set up the layout:

1. Include all fields the first level reviewers coded, such as any Responsiveness determination fields, Privilege, Confidentiality and Issue Designation.
2. Add a “Second Level QC” field to track the number of documents put through QC.
3. Add choices for the QC team to make its decisions. The QC team may record any overturns on second level as well. Name the choices appropriately for your matter.
4. Secure the layout so that only the QC team can view and code it.

Below are two examples of QC layouts. Both allow users to QC the first level review and make any necessary coding changes.
Or, configure the QC Layout to track the QC team's overturns. If the QC team disagrees with the first level reviewer’s decision, they can mark the changes by category, e.g. “Responsiveness Change” or “Privilege Change” as in the following example.

<table>
<thead>
<tr>
<th>Quality Control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Number</strong></td>
<td>CTRL.0000000004</td>
</tr>
<tr>
<td><strong>Responsive Designation</strong></td>
<td></td>
</tr>
<tr>
<td>○ Responsive</td>
<td>○ Non-Responsive</td>
</tr>
<tr>
<td>Add</td>
<td></td>
</tr>
<tr>
<td><strong>Privileged</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>○ Privileged</td>
</tr>
<tr>
<td>Add</td>
<td></td>
</tr>
<tr>
<td><strong>Confidential Designation</strong></td>
<td></td>
</tr>
<tr>
<td>○ Confidential</td>
<td>○ Highly Confidential</td>
</tr>
<tr>
<td>Add</td>
<td></td>
</tr>
<tr>
<td><strong>Issue Designation</strong></td>
<td></td>
</tr>
<tr>
<td>○ Hot</td>
<td>○ Issue A</td>
</tr>
<tr>
<td>Add</td>
<td></td>
</tr>
<tr>
<td><strong>Attorney Review Comments</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Second Level QC</strong></td>
<td></td>
</tr>
<tr>
<td>○ QC Complete</td>
<td>Add</td>
</tr>
</tbody>
</table>

Replicate the first level layout
51.3.2 Batch reviewed documents for QC review

1. Create a saved search of all documents that have been through first level review (Responsive Designation is set).
2. Create a QC Batch Set using the reviewed documents search as the Batch Data Source. Enable Auto Batch to continuously populate new batches for the QC team.
51.3.3 Identify total documents put through QC and resulting turnovers

After the QC team finishes its second level review, use a saved search and Pivot to calculate turnovers and team accuracy statistics.

- The saved search conditions should include:
  - Second Level QC → is set and
  - Second Level QC Overturns → is set (Include any specific first level review batches or folders as well)
- Run the search
- Pivot on Second Level QC Overturns and view the results in a chart:

To include the statistics in your reporting, save results by exporting to a file from the Pivot view:
Use Statistical Sampling

Create a QC Sample set of all reviewed documents

The Relativity sampling features can be used to determine overall accuracy rates for individual reviewers, groups, or the entire review team. Here are the steps to create the sample set:

1. Set the view to all documents.
2. Filter on documents where Responsiveness is set.
3. Set up a sample set:
   - Activate the sampling menu.
   - Choose either the Fixed Size, Percentage or Statistical sampling type.
4. Save as a list to keep the set static.
5. Save the list as a search.

6. Create review batches using the sample set search as the batch data source.

7. Assign to the QC team for review using the procedure described above.

Note: To learn more about using sampling and how best to apply it to your review, see Sampling and Understanding the Statistical Sampling Formula.

51.4 Focus QC on Substantive Issues or Privilege

51.4.1 Conduct targeted substantive QC

A suggested method to make sure the reviewers are catching all responsive documents is to conduct some QC based on the substantive issues in the litigation. Targeted QC review focuses only on documents that contain relevant words or phrases, but where the first level reviewer marked the document Non-Responsive. Use a list of relevant terms to define search parameters. You may be able to obtain a list from the attorneys or case team handling the litigation. Or, use the terms that were used to set up the search terms reports.

Following are examples of how to set up targeted QC:

- Use key terms from a Search Terms Report:
  - On the Field Tree, choose an STR with Unique hits.
  - Filter to show only those documents marked Non-Responsive.
  - Create a saved search of the returned documents.
  - Create review batches for the QC team or create a sample set for review.

OR

- Create a saved search to find documents where Responsiveness may have been missed.
  - Enter the relevant text as the search terms.
  - Set the search conditions to return only documents marked Non-Responsive.
51.4.2 Conduct Privilege QC

If privilege is an issue in the review, some measure of QC focused on either privileged terms, attorneys’ names, or others who might create privilege is likely needed. Use the methods outlined above for substantive QC to conduct the Privilege QC.

For targeted Privilege QC

1. Enter the list of names or firms who may create privilege as the search terms.
2. Set the search conditions as Privilege is not set.

51.4.3 Check for family tagging conflicts

- Often, you will need to check for inconsistent coding in family groups. We recommend constructing a set of saved searches to ensure that all members of each family are coded consistently for Responsiveness and for Privilege.
- The Advice team's recipe Searching Workflows includes a step-by-step process to accomplish a tagging conflicts check.
- Sampling

52 Sampling for repeated content

Relativity Analytics Indexes benefit greatly from targeted removal of boilerplate text, especially email confidentiality footers. Structured Analytics offers a Repeated Content Identification operation, but on large collections it can be slow and resource intensive.

This recipe provides information on using random samples to achieve repeated content identification in a much quicker fashion and with significantly lower resource utilization than running on the full collection - all without sacrificing the quality of the results.

52.1 Requirements

- Relativity 8.1 or higher
- Relativity Structured Analytics application
- Document set larger than 100,000 documents (no need to sample if you're already smaller than that)

52.2 Directions

52.2.1 Creating the sample

1. First, create your random sample of the target documents.

Note: When creating STRs, choose “Report and Tag” to automatically list the tags and search terms in the Field Tree.
2. Navigate to the Documents tab and restrict yourself to the documents you want to focus on. It might be everything in the workspace, the searchable documents from your index, or a set limited by file type, email inclusives, or some other subset.

3. Once you're looking at the document set you want to analyze, create a random sample by clicking ![sample icon].

4. In the Sampling dialog, set the following:
   - Type: Fixed
   - Size: 100,000

5. Click Sample. You should now see only 100,000 documents listed on the page.

### 52.2.2 Save sample as list and list as saved search

At this point, you're ready to create the list. From there you'll create the saved search that you'll reference in your Repeated Content Identification run.

**Note:** In Relativity 8.1-9.2, you create lists with an icon rather than a mass action. Other than that minor workflow change, this recipe works for releases prior to 9.3. See [lists](#) for more information. See the Lists topic in the Admin guide for more information.

To create the list as a mass operation:

1. From the bottom of the page, click **all 100,000** and then click **Save as list**.
2. When prompted, name the list "100k random sample".
3. Navigate to the **Lists** tab.
4. Click **100k random sample**, then click **Create Search from List**. You should now have the **100k random sample** search under the List Searches folder in your saved search browser.

### 52.2.3 Create and run repeated content identification as Structured Analytics set

To create a Structured Analytics Repeated Content Identification Set:

1. From the **Indexing and Analytics** tab, select **Structured Analytics Set**. A list of existing Structured Analytics sets appears.
2. To create a new set, click **New Structured Analytics Set**. The Structured Analytics Set layout appears.
3. In the Structured Analytics Set Information layout, set the following:
   - **Structured analytics set name**: Enter a name for your Structured Analytics set, such as "Repeated Content Identification on sample".
   - **Select operation**: **Repeated content identification**
   - **Select document set to analyze**: select your saved search, **100k random sample**.
   - **Analytics server selection**: select **Automatic**.
   - **Enter your email address**: enter an email address if you wish to receive a notification when the email threading analysis completes.
4. In the Repeated Content Identification layout, consider the following:
   - You should modify the Repeated content settings slightly to suit your needs. We recommend the defaults with the exception of the Minimum number of occurrences. If you are using a random sample, we recommend you change this setting to 0.004 (i.e. 0.40%) multiplied by the
number of documents you are submitting. For example, with 100,000 documents, set the Minimum number of occurrences to 400. If you are not using a sample (for instance, if you have a small collection with fewer than 100,000 documents), then we advise setting it to 0.005 times the number of documents you’re submitting. See Special considerations below for more information.

Note: You may over time decide that you prefer setting this value higher or lower, but the above recommendation is consistent with our experience with most customers. Contact solutions@relativity.com if you would like to discuss in greater detail.

5. Click Save. The Structured Analytics Set Console appears.
6. To run repeated content identification analysis, click Run Full Analysis. You can monitor the progress of the operation in a separate window by clicking the export icon in the upper right corner of the progress pane.

52.2.4 Review results

After the operation completes, review the resultant filters to ensure that they are indeed boilerplate and not authored content. Accomplish this task by using filters along with the Ready to index field.

52.2.5 Special considerations

As mentioned above, the minimum number of occurrences setting should be configured in proportion to the number of documents. While 400 is generally appropriate for 100,000 document samples, larger or smaller sets necessitate proportional modification of that number.

To retrieve more filters, the minimum number of occurrences can be reduced. However, we don’t recommend setting it lower than 100 on a 100,000 document sample, as the results can become more subject to sampling error. Consider running the operation across a judgmental sample instead. For example, just parent emails, or just Word and PDF documents.

Analytics

53 Searching for custom object information on the Document object

To search for custom object information on the Document object create a multiple object field that links both objects together.

53.1 Recipe overview

You can search for the custodians' information from the document object using a custom custodian object. The workspace should also contain documents associated with certain custodians.
53.2 Requirements

This recipe is applicable to all Relativity versions.

53.3 Directions

To search for custom object information on the Document object, perform the following steps:

1. Create a multiple object that links the Document object to the custom Custodian object. In this example, Custodian Obj Linker is the name of the multiple object.

2. Set Open to Associations to Yes on the Custodian object fields you want to make searchable on the Document object (in this example, the Start Date and End Date).
When creating a new Saved Search, or toggling search conditions on, you can search across your Custodian object Start Date and End Date fields. Look for the field name that starts with the name of your multiple object. In this example, it is Custodian Obj Linker.

You can also filter on the Custodian object Start Date and End Date fields by adding them to a view on your Document object.
You can now filter on these date fields on this Document view.

- Relativity dynamic objects
- Fields

54 Searching for dates in Relativity

This recipe provides steps to search for dates within Relativity using a filter, a saved search, or a dtSearch.

54.1 Requirements

Applicable to all versions of Relativity.

54.2 Types of date searches

You can search for dates in Relativity using three different methods:

- Field filters
  - Search for single date or single date range
  - Requires metadata
- Saved search
  - Search for multiple date or multiple date ranges
  - Requires metadata
- dtSearch auto-recognize
  - Search for dates within document text and in multiple formats
  - Does not require metadata
54.3 Directions

54.3.1 Field filters

To search for a date using a field filter, perform the following steps:

1. Ensure the date field you want to search for is in the current view. Add it to the view if necessary.
2. Turn on filters, and then enter a valid search string in the text box filter.

The following table lists examples of valid date and number searches, as well as the expected result set.

<table>
<thead>
<tr>
<th>Valid search strings</th>
<th>Return items where...</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;= 7/24/2008</td>
<td>[FIELD VALUE] &gt;= '7/24/2008'</td>
</tr>
<tr>
<td>&lt;= 7/24/2008</td>
<td>[FIELD VALUE] &lt;= '7/25/2008'</td>
</tr>
<tr>
<td>= 07/27/2008 1:23 pm</td>
<td>[FIELD VALUE] = '07/27/2008 1:23 PM'</td>
</tr>
<tr>
<td>&gt;= 100</td>
<td>[FIELD VALUE] &gt;= '100'</td>
</tr>
<tr>
<td>&lt;= 100</td>
<td>[FIELD VALUE] &lt;= '100'</td>
</tr>
<tr>
<td>= 100</td>
<td>[FIELD VALUE] = '100'</td>
</tr>
</tbody>
</table>

The following table includes examples of invalid data and number search strings.

<table>
<thead>
<tr>
<th>Invalid search strings</th>
<th>Invalid reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 7/24/2008</td>
<td>You must use the equal sign with the greater than operator (as in &gt;=).</td>
</tr>
<tr>
<td>&lt; 7/24/2008</td>
<td>You must use the equal sign with the less than operator (as in &lt;=).</td>
</tr>
<tr>
<td>&gt;= 0/24/2008</td>
<td>The search string includes the value 0 for the month.</td>
</tr>
<tr>
<td>= 0/24/2008</td>
<td>The search string includes the value 0 for the month.</td>
</tr>
<tr>
<td>0/24/2008 BETWEEN 8/24/2008</td>
<td>The search string includes the value 0 for the month in the starting date.</td>
</tr>
<tr>
<td>7/24/2008 BETWEEN 0/24/2008</td>
<td>The search string includes the value 0 for the month in the ending date.</td>
</tr>
</tbody>
</table>

54.3.2 Saved search

To search for a date in Relativity using a saved search, perform the following steps:

1. Create a saved search.
2. Select the date field you want to search for in the Conditions section.

You can set additional conditions to search for non-consecutive dates, including non-consecutive date ranges.
54.3.3 Index search

Some cases require you to search for dates contained within the text of a document. Because dates can appear in various formats, the auto-recognize feature in dtSearch is useful.

To use this feature, build a new dtSearch index with **Auto-recognize date, email, and credit card numbers** set to Yes. Once the index is complete and active, you can search for dates within the text of a document.

The auto-recognize feature searches for strings that appear to be dates. It uses English-language months, including common abbreviations, and numerical formats. For example, dtSearch recognizes the following date formats:

- January 15, 2006
- 15 Jan 06
- 2006/01/15
- 1/15/06
- 1-15-06
- The fifteenth of January, two thousand six

It doesn't recognize DD/MM/YYYY format.

Note the following date and date range search strings:

- To search for a date, enter a date expression between the parentheses in the string "date()"; for example, "date(jan 10 2006)"
- To search for range of dates, enter a date range between the parentheses in the string "date()"; for example, "date(jan 10 2006 to jan 20 2006)"
- To search for a range of dates near the word "apple," enter "date(jan 10 2006 to jan 20 2006) w/10 apple"
- dtSearch doesn't support unterminated date. To search for any date after or before a particular date, enter a bounded range with a maximum or minimum value for the bounds. The maximum value for a year is 2900, and the minimum value is 1000. For example, "DateField contains date(jan 10 2006 to jan 1 2900)"

54.4 Notes

- By default, date fields use the filter type of text box. However, you can change this to custom in the field information.
- If you need more complexity than the text box filter type can provide, you can set up a saved search.
55 Searching for documents with incorrectly loaded text

If data is loaded with the wrong encoding selected, you must locate those documents and ensure they load into Relativity properly.

55.1 Recipe overview

This recipe shows you how to search for documents that loaded into Relativity with encoding set incorrectly.

55.2 Requirements

Applicable to all versions of Relativity.

55.3 Directions

1. Create a new dtSearch index, and then remove all stop words.
2. Create a search terms report with all of the stop words using that index.
3. Create a saved search with the following conditions:
   - the STR field is not set AND
   - the Extracted Text field is set

The results yield all documents with extracted text, but don't contain any stop words. Documents that fall into this category are the documents with the wrong encoding.

Note: Documents in other languages may also appear in your results.

- Relativity Desktop Client
- Importing documents
- Searching

56 Searching for handwritten documents

This recipe provides a workflow that can help identify handwritten documents.

56.1 Overview

Relativity doesn't automatically identify handwritten documents. This recipe demonstrates how you can use custom search indexes and search terms reports to execute searches that can identify
potential handwritten documents or documents that contain poor quality OCR.

56.2 Requirements

- Workspace access
- Edit and delete permissions on search indexes
- Edit and delete permissions on searches

56.3 Directions

**Note:** This recipe assumes that the documents contain available text or were previously OCR-ed to capture all possible text.

To search for handwritten documents, perform the following steps:

1. Create a dtSearch index called **dtSearch-NoNoiseWords**.
   a. Set the Searchable Set to the subset of documents you want to search for handwriting, and ensure the extracted text or OCR field returns in the search.
   b. Copy and paste the list of noise words into a Notepad file, and then remove all the noise words from this index.
   c. Save the index, and then perform a full build.

2. Create a search terms report (STR) with the following settings:
   - **Name**: STR-NoiseWord
   - **Index**: dtSearch-NoNoiseWords
   - **Searchable Set**: Same set you previously identified
   - **Type**: Report and Tag
   - **Calculate Unique Hits**: No
   b. Click **Modify terms**, and then enter the list of noise words that you previously copied and pasted.
c. Click Run All Terms.
3. Create a search with the following settings to identify documents that may contain handwritten text once the STR completes.
   a. Name - Potential Handwritten Docs
   b. Conditions:
      ■ Field- STR-NoiseWord
      ■ Operator- Not These Conditions
      ■ Value- STR-NoiseWord is Set

The results of the search narrow down the universe of documents to the documents that potentially contain handwritten text or poor quality OCR.

Searching

57 Searching for single standalone emails

To assist in organizing document review, this recipe shows you how to separate standalone emails. Standalone emails are emails with no further conversation other than the original email.

57.1 Requirements

■ Applicable to Relativity 8.0 and above
■ Structured Analytics - Email Threading
■ Saved Search
57.2 Directions

1. Set up a saved search to return multiple emails with multiple threads.
   a. Use the Email Threading ID field. This field indicates how many threads are in a conversation. The search in the below example asks for items that contain second level emails.
   b. Include the Thread group to pull in the rest of the conversation.

2. Create another saved search to return all records where Email Threading is set, but excludes the search in step 1.
Email threading
Saved search

58 Searching for terms while excluding email footers

While searching for a term, you may receive false positive hits because that term appears in standard email confidentiality footers. For example, if you search for documents that contain the term “Confidential,” the term returns extraneous results. These hits can interfere with relevant results.

With some simple search syntax, you can search for a term, and exclude it only when it appears in a standard email footer.

58.1 Recipe overview

This recipe shows you how to use the NOT operator with a proximity search to search for a specific term. By using this search you can exclude hits in a repeated email disclaimer.

58.2 Requirements

- Applicable to all Relativity versions
- dtSearch index
58.3 Directions

Using your dtSearch index, construct a search for your term using the NOT and w/n operators to exclude the term only when it's part of a repeated email disclaimer. For example, to find documents that contain the term “Confidential” use the following syntax:

**Example 1 for one footer:**

Confidential NOT w/1 "The message, together with any attachment, may contain confidential and privileged information"

**Example 2 for more than one footer:**

Confidential NOT w/1 "The message, together with any attachment, may contain confidential and privileged information" or "Privileged and Confidential information provided in this message is not for use"

Using this syntax returns documents that contain the term “Confidential” and excludes documents where “Confidential” only appears in the email footer. If a document contains “Confidential” in the body of the email and in the footer, it returns as a hit.

Your proximity is equal to the number of words that appear between the first word in the disclaimer and your search term. In this example, “Confidential” is one word from the start of the email footer string.

- dtSearch

59 Searching for unique identifiers

You can run searches for a list of unique identifiers, like control numbers or Bates numbers, within Relativity.

59.1 Overview

This recipe shows you how to set up custom search indexes and then use simple search operators to identify a list of documents.

59.2 Requirements

- Workspace access
  - Edit and delete permissions on search indexes
  - Edit and delete permissions on searches
59.3 Directions

This recipe assumes that the documents you’re attempting to identify contain a unique identifier in a fixed-length text or long text field.

The examples in this recipe show you how to search for the control number of the document. However, you can use the same workflow process for other unique identifiers like Bates numbers.

59.3.1 Option A

Use this option when you have fewer than a few hundred documents to identify.

Create a search with the following conditions to identify documents with the control number:

- **Name**: Control Number List
- **Conditions**
  - **Field**: Control Number
  - **Operator**: Is
  - **Value**
    - ABC001
    - ABC002
    - ABC003

Ensure that you separate the unique identifier values by a line break. This may require you to reformat the string of control numbers using Notepad++ or Textpad.

59.3.2 Option B

Use this option when you have more than a few hundred documents to identify.

1. Create a saved search called **ControlNumber - for Index** with the subset of documents you want to search against.
2. Ensure that the only field the saved search returns is the Control Number.
3. Create a new dtSearch index called **dtSearch-ControlNumber**.
4. Perform a full build on this new index using the saved search from step 1.
5. Create a search with the following settings to identify documents with the control number once the index is complete:
Ensure that you separate the unique identifiers with an “OR” operator. This may require you to reformat the string of control numbers using Notepad++ or Textpad).

The results of the search help identify the documents by their control numbers.

You can also paste this list of identifiers into a search terms report, which provides an OR search for all of the identifiers. Select the Report and Tag option to write this information to a field.

The following table helps you determine the use case for each workflow option.

<table>
<thead>
<tr>
<th>Option A</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal when searching for fewer than a few hundred documents.</td>
<td>Ideal when searching for more than a few hundred documents.</td>
</tr>
<tr>
<td>Keyword indexes any length of text, but limits the search for control numbers to 4000 characters in length (overall length of all characters you search).</td>
<td>dtSearch doesn't index (by design) control numbers longer than 32 characters.</td>
</tr>
<tr>
<td>Ensure you use the IS operator when searching. You can also use the CONTAINS operator if, for example, you are trying to bring in all family members that begin with the same prefix.</td>
<td>If your control number contains periods (.) or hyphens (-), you must index those characters prior to building the index, or enclose the control number in quotes.</td>
</tr>
</tbody>
</table>

**Searching**

**60 Searching workflows**

Relativity provides great flexibility to help you identify documents through searches so that you can ready them for further review and analysis. This recipe highlights setting up multiple search indexes, applying various search techniques, and using filters properly when executing searches.

**60.1 Requirements**

- Workspace access
  - Edit, Delete access on Search Indexes
  - Edit, Delete access on Searches
60.2 Directions

Relativity offers two types of search indexes, Basic Keyword Search and Specialized dtSearch.

The Basic Keyword Search Index populates automatically, is available out of the box, and supports basic Boolean operations. These operations include AND, OR, NOT, and the wildcard (*) operator.

The Specialized dtSearch Index is custom built and must be set up with an index to query against. In addition to basic Boolean operations, it also supports the following functionality:

- Wildcard operator for a single digit (=)
- Stemming operations (~)
- Fuzzy search operation (%)
- Range Searching on Numeric fields (~~)
- Customize data set to be indexed via custom saved search
- Proximity Searching
- Customized Noise Word List
- Auto Recognition of email addresses, credit card numbers, and dates.

The following includes best practices for the use of advanced operators and workflow options:

60.2.1 Proximity search

Proximity search uses operators to search certain terms in proximity to other terms in a document. Use the following recommendations when employing proximity searching:

- To use directional proximity searching use "pre /x" operators. To use non-directional proximity searching use "w /x" operators.
- To determine the beginning and end of a document, use reserved word with either ("xfirstword") or ("xlastword"). Use these operators to search for metadata like email addresses or footers within a document.
- To determine proximity. The distance between terms is important so you know whether you receive true hits or false positives. The following image illustrates the proper syntax for proximity searches:

<table>
<thead>
<tr>
<th>Searching for</th>
<th>Blue Orange Yellow Red Green Brown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>1 2 3 4 5</td>
</tr>
</tbody>
</table>

Blue w/1 Orange w/1 Yellow w/1 Red w/1 Green w/1 Brown \[\text{INCORRECT}\]

((((Blue w/1 Orange w/2 Yellow) w/3 Red) w/4 Green) w/5 Brown) \[\text{CORRECT}\]

The following image shows all correct and incorrect combinations of proximity searches that result in successful hits when using proximity operators:
Sometimes additional factors affect the distance between terms. The following image demonstrates a search for a string of words in proximity to another word. The search in the following image takes the following into account:

- Noise words count as words when calculating proximity.
- Punctuation counts as whitespace when using default settings.
- Relativity treats line breaks and consecutive space characters as single spaces.
- dtSearch default stop words and connector words like AND, OR, BETWEEN, and NOT count as words when calculating proximity.

The message together with any attachment, may contain confidential and/or privileged information.

In this case, we continue to calculate the distance of the string ("confidential and/or privileged information:) from the word ("message") as shown above. However, when using connector words, system admins should create an index that removes stop words from the noise word list.
Relativity reserves the following noise words and characters, which continue to behave as operators, as well as being noise words: and, or, not, to, contains, xfirstword, xlastword, ",", ( ), *, ?, %, @, ~, #, &,
, ;, =.

Once that index is available for query you can either place the search string in quotes or apply stemming to the connector words to override their function as a connector.

60.2.2 Auto-Recognition
Auto-Recognition identifies email addresses, dates, and credit card numbers in the data set you want to index. You can turn this feature on and off when you build your index. The search returns the results regardless of the data format. Auto-Recognition adds some time to your index build, but, depending on your case, the benefits can be significant.

<table>
<thead>
<tr>
<th>Type</th>
<th>Syntax</th>
<th>Results</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email</td>
<td>mail[&quot;* <a href="mailto:doe@kcura.com">doe@kcura.com</a> ]</td>
<td><a href="mailto:jdoe@kcura.com">jdoe@kcura.com</a></td>
<td>Treats email addresses as a single term regardless of alphabet settings for &quot;@&quot; and &quot;.&quot; characters. Can be used with wildcards.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:john.doe@kcura.com">john.doe@kcura.com</a></td>
<td></td>
</tr>
<tr>
<td>Credit Card Numbers</td>
<td>Creditcard(1234678 13246578)</td>
<td>1234 5678-1234 5678 1234 5678 1234567812345678</td>
<td>Returns sets of numbers that satisfy the criteria for valid credit card numbers. Can be used with wildcards.</td>
</tr>
<tr>
<td>Date</td>
<td>date(1/15/2006)</td>
<td>January 15, 2006; 15 Jan 06; 2006/01/15; 1/15/06; 1-15-06; The fifteenth of January, two thousand six</td>
<td>Returns date regardless of format. Dates mentioned in body of documents and not in metadata.</td>
</tr>
</tbody>
</table>

60.2.3 Searching for times
If you want to search for times in the body of documents, perform a full-text search with your dtSearch index for a specific time. Keep in mind that some characters cause a word break, such as the colon and period. Searching for 12:15 p.m. results in searching for four words: 12, 15, p, and m.

60.2.4 Filters
In addition to the Search Indexes, you can also use filters to search on metadata fields and narrow down the review set of documents.
The following list includes metadata fields, their corresponding field types, and the filter type available in Relativity:

- **File Type** - Single-Choice, List Filter
- **Custodian** - Single-Choice, Pop-Up Filter
- **Date Sent** - Date, Textbox Filter
- **Email To** - Long Text, Textbox Filter
- **Email Subject** - Fixed Length Text, Custom Filter
- **Designation** - Single Choice, Multi-Choice List

The following table shows different types of filters you can set up with available fields in Relativity.

<table>
<thead>
<tr>
<th>Type</th>
<th>Text Box</th>
<th>Single Choice List</th>
<th>Multi-Choice List</th>
<th>Pop-Up</th>
<th>Custom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Length Text</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>Long Text</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>✔️</td>
<td></td>
<td></td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Number (Whole, Decimal, Currency)</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes/No</td>
<td></td>
<td></td>
<td></td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Choice (Single, Multiple)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Object (Single, Multiple)</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td>✔️</td>
</tr>
</tbody>
</table>

**60.2.5 Troubleshooting workflow**

To troubleshoot and test searches, use a white board approach to map out searches before you run them. As part of the process, try to take into account all possible variations, so you can see which documents Relativity returns and to gain a better understanding of the search. For instance, if you want to find email family groups with inconsistent coding, numerous possibilities for searches exist. The following illustration shows a mapping of nine email family groups, each with an email and two attachments.

Construct searches (1 and 2) and mark the corresponding documents, and their family members, with hits for each of the searches. At this point, when you look at the board, you can see that the four email family groups have one thing in common. The documents are responsive to both searches. So, to close it out, construct a third search that pulls back documents common to both searches.
61 Setting up CJK document workspaces in Relativity

When working with documents that contain Asian character sets, specific settings help expedite your setup time in Relativity. This information can help you build a clean workspace with more accurate fields to support your Asian language document sets.

61.1 Recipe overview

This recipe highlights items to be aware of as you work with workspaces that contain CJK character sets.

61.2 Requirements

- Applicable to all versions of Relativity
- Workspace access with the following permissions:
  - Field: edit/delete
  - Searches: edit/delete
- Relativity Desktop Client
61.3 Directions

1. In your workspace, set the **Unicode** field to **Yes**.

2. Set up fields such as Custodian, email fields, File Name, and Extracted Text as Unicode-compliant to capture CJK characters.

3. Your processing vendor should have provided you with data files that are Unicode/CJK character-compliant. If this is the case, import the data in a Unicode-compliant format. Be sure to include any extracted text.
Setting up the fields as Unicode-compliant, and leaving the default settings in the dtSearch index (dtSearch index alphabet list: CJKRanges = 2e80-ac00 ac00-d7af f900-faff fe30-fe4f), allows you to index characters in CJK languages where each character is treated as a separate word.

### 61.3.1 Searching considerations

Consider the following with SQL, dtSearch, and CJK documents. Essentially each character is its own word or token so proximity searching works.

#### 61.3.1.1 Keyword Search

- Within a Relativity workspace, a system admin can select the language to use in the SQL Full Text Language. SQL Full Text Language determines the correct stemming and word-break characters used in the full text index.

- Once you configure your SQL full-text language to the correct language (Japanese in this example), you can perform keyword searching and filtering in that language.

#### 61.3.1.2 dtSearch

**Index**

Once you set the fields to be Unicode compliant, leaving the defaults in the dtSearch index setup (dtSearch index alphabet list: CJKRanges = 2e80-ac00 ac00-d7af f900-faff fe30-fe4f) allows you to index characters in CJK languages where each character is treated as a separate word, enabling word breaking with CJK.
Word/Character Search

As explained previously, you can store or convert Chinese, Japanese, and Korean text to Unicode so that you can use dtSearch to search for words in these languages just as you search for words in other languages. However, while dtSearch can search for literal word matches (or wildcard or fuzzy matches), there are some limitations on the support in dtSearch for Chinese, Japanese, and Korean text. Those limitations include:

- Some documents store text in a way that does not separate the words with spaces. Instead, all of the text in a document runs together, and a language-specific dictionary is needed to find word breaks. dtSearch does not have the ability to identify word breaks in these documents because it doesn’t include any language-specific dictionaries. To make this type of text searchable, enable an option in dtSearch to automatically insert word breaks around Chinese, Japanese, and Korean characters. Once you enable this option, Relativity treats each character as a single word for indexing and searching purposes.
- This feature is turned on by default in Relativity dtSearch. Because each term is searchable, we recommend searching for multiple characters to assist in retrieving more accurate results. When looking for multiple characters, use the proximity connectors to assist in finding desired results.
- The same text can be presented in different ways depending on the context. dtSearch searches for a word as it is provided in the search request and does not generate additional grammatical or script variations for words in Chinese, Japanese, and Korean.
- The dtSearch Engine has an API that you can use to integrate with dictionary-based language analyzers from companies such as Basis Technologies. But while the dtSearch standalone desktop does allow for integration, the instance within the Relativity environment does not support integration.

For non-Western languages (such as Chinese, Japanese, Hindi, and so forth) there are additional considerations and workarounds that may provide assistance in locating search hits.

61.3.1.3 Mandarin considerations

- Mandarin characters are treated as words. Therefore the Mandarin characters do not need extra space because each character can be compared against a list. If there is a match, then it’s highlighted.
- Numbers are not treated like Mandarin characters. For example, when you use the number 54, you need everything until there is a space. Add a space to the subject so the term is 54 所 [space], then the rest of the line, it will highlight.

<table>
<thead>
<tr>
<th>54 所 (other characters)</th>
<th>54 所</th>
<th>No Match</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>所</td>
<td>Match</td>
</tr>
<tr>
<td>54 所 (other characters)</td>
<td>Match</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>54 所 space (other characters)</th>
<th>54 所</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>所</td>
<td>Match</td>
<td></td>
</tr>
<tr>
<td>54 所 (other characters)</td>
<td>No Match</td>
<td></td>
</tr>
<tr>
<td>(other characters)</td>
<td>Match</td>
<td></td>
</tr>
</tbody>
</table>

- The term 54 所 does not match the subject. The subject has the additional characters, hence no match.
The Mandarin characters do not need the space because each character, such as 所, is treated uniquely, therefore, it highlights.

**Keyword Search** — Once you configure your SQL full-text language to the correct language (Japanese in this example), you can perform keyword searching and filtering in that language.

**dtSearch** — Once you set up the fields to be Unicode compliant, leaving the defaults in the dtSearch Index setup (dtSearch index alphabet list: CJKRanges = 2e80-ac00 ac00-d7af f900-faff fe30-fe4f), you can index characters in CJK languages where each character is treated as a separate word, enabling word breaking with CJK.

### 61.3.2 Analytics considerations

- Use the structured data analytics language identification operation (or PLI language technique) to identify the languages of the documents in your workspace.
- Training your Analytics index in one language as opposed to multiple languages generally produces a better quality index. However, if many of your documents contain multiple languages, this can change.
- Be sure to set up your stop words in the same language as your index.

**Note:** Refer to the Using Analytics with CJK characters recipe for more detailed considerations.

**Relativity Desktop Client**

### 62 Sorting documents and keeping email families together

You may want to sort your document collection by date; however, it's difficult to keep emails and their attachments together. This is because family items can contain different dates or date values in different fields. To sort documents chronologically and keep email families together, you need a date field that is consistent for emails and their attachments.

### 62.1 Recipe overview

This recipe shows you how to create an additional date field. It also shows you how to populate it with the sent date from the email to the attachments using a script. This creates a consistent date field for emails and their attachments, which makes the sorting process easier.

### 62.2 Requirements

- Applicable to all versions of Relativity.
- Relational field identifying email families.
- Relativity security access
  - Field- edit/Add
  - Document- edit
  - Permissions- access to Scripts tab
  - Groups- system Admin and Script Admin (if you didn't add the script to the workspace.)
62.3 Directions

1. Create a Date field to hold the date information. Call this field Parent Date, Family Date, etc.
2. Create a Yes/No field to flag documents when the script affects them. If you add additional data to the workspace at a later date, and need to run the script again, this flag ensures the script runs only on the newly added documents. Call this field Date Propagated.
3. Run the Propagate Sent Date to Family Documents script. If you don’t see this script listed under the Scripts tab, add it to the workspace from the Relativity Script Library. Once you run this script, Relativity presents you with the following inputs:
   a. Date Field – Original - enter the field from which you want the date propagated, such as Sent Date.
   b. Date Field – Sort - enter the field to which you want the date propagated (the field you created in step 1).
   c. Family Identifier - enter the relational field that defines the email families.
   d. Date Propagated - enter the Yes/No field you created in step 2.
   e. Document Sort Field - select the field for which to order documents by instead of using Artifact ID to order documents in each family. This accounts for some document families where the lowest Artifact ID isn’t the parent.

The script takes the date value from the field you select in step 3(a) from the email, and writes it to the field you select in step 3(b) for the email and attachments. Once the script completes, you can use the field you create in step 1 to sort your documents and keep email family groups together.

- Fields
- Relativity Case Script - Propagate Sent Date to Family Documents

63 Starting a RAR project after review has begun

While working in Relativity, you may need to start a RAR project after review has already begun on the documents in your workspace. This recipe explains how to use Assisted Review on an existing data set that has been partly reviewed.

63.1 Requirements

- Relativity 9.0 or higher
- An active Analytics index
- Assisted Review Application

63.2 Directions

This recipe assumes that the following:

- The Assisted Review application is installed.
- An optimized Analytics index has been built and includes the project documents.
- The project is a good fit per the RAR Checklist.
Once you've established these prerequisites, perform the following steps to integrate an Assisted Review workflow midway into review:

1. Create an Assisted Review project. Keep in mind the “Documents to be categorized” saved search must include all the documents on which you’d like the engine to make its coding decisions.
2. Create a saved search to identify the documents that have been already reviewed and coded.
3. Identify the documents from step 2 that were coded per the Reviewer Protocol specifications to ensure they are good training examples. Once identified, tag the documents on the following fields, both of which you specified when you set up the project in step 1:
   a. RAR Designation field
   b. RAR Use as Example field
4. Create a saved search for all documents tagged on the RAR Designation field that were good examples (from step 3).
5. Create a new round in the Assisted Review project for pre-coded seeds to include the documents that have already been coded during manual review that you identified as good examples (use the saved search created in step 4).

   **Note:** The saved search must exclude documents marked No for use as example, as RAR uses all coded documents in a pre-coded seed saved search as training examples.

6. Proceed with the regular RAR workflow. Finish the pre-coded seed round and proceed to subsequent rounds (training or QC rounds per the needs of your project / reports).

### 63.3 References

- Analytics
- Relativity Assisted Review

### 64 Tally all email recipients

This recipe describes a workflow to count the total number of recipients to an email, based on recipient metadata.

#### 64.1 Recipe overview

To achieve this workflow, the user will consolidate the Email To, Email CC, and Email BCC fields into one field, separated by semicolons. Then, the user will run the Delimiter Count By Saved Search script to count the number of semicolons in the consolidated field.

#### 64.2 Requirements

- Correctly populated email metadata fields
- The Delimiter Count By Saved Search script from NSerio, located on their [website](#)
64.3 Directions

To tally all email recipients, use the following steps:

1. Create a field to store the consolidated values of Email To, Email CC, and Email BCC.
   1. **Field Type**: Long Text
   2. **Field Name**: Email All Recipients
2. Create a field to store the number of delimiters.
   1. **Field Type**: Whole Number
   2. **Field Name**: Email Recipient Count
3. Create a saved search of all the documents for which you'd like to tally the number of recipients.
   1. (Optional) You can pull these results in a saved search by marking the Email To, Email CC, or Email BCC as **set**.
      1. Return these fields:
         - Email All Recipients
         - Email Recipient Count
         - Email To
         - Email CC
         - Email BCC
4. Using the filter pane, filter on the **Email To** field for **Is Set**.
5. Use the Replace mass action on all documents returned by the filter.
   1. **Field to Update**: Email All Recipients
   2. **Action**: Replace Entire Field
   3. **Update With**: Field
   4. **Field**: Email To
6. Use the Replace mass action on all documents returned by the same filter as step 4.
   1. **Field to Update**: Email All Recipients
   2. **Action**: Append to End
   3. **Update With**: Text
   4. **Text**: ; (semicolon, no spaces)
7. Clear your previous filter.
8. Using the filter pane, filter on the **Email CC** field for **Is Set**.
9. Use the Replace mass action on all documents returned by the filter.
   1. **Field to Update**: Email All Recipients
   2. **Action**: Append to End
   3. **Update With**: Field
   4. **Field**: Email CC
10. Use the Replace mass action on all documents returned by the same filter as step 4.
    1. **Field to Update**: Email All Recipients
    2. **Action**: Append to End
    3. **Update With**: Text
    4. **Text**: ; (semicolon, no spaces)
11. Clear your previous filter.
12. Using the filter pane, filter on the **Email BCC** field for **Is Set**.
13. Use the Replace mass action on all documents returned by the filter.
   1. Field to Update: Email All Recipients
   2. Action: Append to End
   3. Update With: Field
   4. Field: Email BCC

14. Use the Replace mass action on all documents returned by the same filter as step 4.
   1. **Field to Update**: Email All Recipients
   2. **Action**: Append to End
   3. **Update With**: Text
   4. **Text**: ; (semicolon, no spaces)

15. Install the Delimiter Count by Saved Search script from NSerio, located on their [website](#). Directions and documentation about the script are available from NSerio.

16. Run the Delimiter Count by Saved Search script with the following settings:
   1. **Saved Search**: Search created in step 3.
   2. **Source Field**: Email All Recipients
   3. **Field Delimiter**: ; (semicolon, no spaces)
   4. **Count Destination Field**: Email Recipient Count

17. When the script completes, it updates the Email Recipient Count field for all documents returned in your search to reflect the total number of recipients across their Email To, Email CC, and Email BCC fields.

### 64.4 References

- [Fields](#)
- [Mass replace](#)
- [Saved search](#)
- [Scripts](#)
- [Fields](#)
- [Mass Replace](#)
- [Saved search](#)
- [Scripts](#)

### 65 Tracking categorized documents to their example document

During a categorization project, it’s sometimes important to know which example document was used by the Relativity Analytics index to categorize the rest of your documents. This information can help you improve your example set and yield a more accurate categorization set.

#### 65.1 Recipe overview

This recipe shows you how to create a tab containing a view that displays both your categorized documents and the example document used by the Analytics index to make its categorization
decision.

65.2 Requirements

- Applicable to all Relativity versions
- Relativity security access
  - Permissions: Access to Tabs option under the Administration tab
  - Tab: Edit/Add
  - View: Edit

65.3 Directions

1. For Relativity 7.1 and higher, your categorization set must have a Maximum Categories per Document value greater than 1.
2. Under Tabs, create a new tab to display the Analytics categorization results.
   a. Name the new tab and select an Order.
   b. Set Link Type to Object.
   c. In the Parent field, you may wish to place the new tab under the Analytics tab.
   d. Set Object Type to Analytics Categorization Result.
   e. Leave Is Default as No.
3. Navigate to the newly created tab.
4. Create a new view with the object type of Analytics Categorization Result.
5. Ensure the following fields are added to the view:
   a. Document – [Name of Categorization Set]
   b. Category
   c. Category Example

These fields show you the document that was categorized, the name of the category, and the example document that was used by the Analytics index to make its categorization decision.

- Analytics
- Tabs

66 Transfer tags from other document review applications

Migrating data from other document review applications to Relativity is a straightforward process through the exporting and importing of load files. However, some preparation is necessary to transfer tag information from the applications.

66.1 Recipe overview

This recipe describes how to transfer tags to the field tree in your Relativity workspace.
66.2 Directions

1. In the application database, write the tag(s) information to a field. Use a semicolon to separate the individual tag or folder naming.
2. Export the data from the application, using default delimiters. Ensure that at least the Document Identifier and Tag fields are in the export.
3. Create a field in your Relativity workspace to hold the tag information.
   a. Set the Field type to Multiple choice.
   b. Ensure the field is available in the field tree.
4. Import the tag data into the field created in step 3. Ensure that the multi-value delimiter is set to semicolon in the Relativity Desktop Client.
5. Navigate to the field tree and observe your tags.

67 Threading review setup

This recipe suggests some ideal views to make available to your reviewers to assist with the document review process.

67.1 Requirements

- Relativity 8.0 or higher
- Workspace access
- Relativity Structured Analytics application

67.2 Directions

The following sections contain sample views that you can create along with workflow use cases for each. The use cases provide details about the results of the Email Threading Structured Analytics sets. These views can be very useful for your reviewers. See the table that appears after the following views for an explanation of each of the fields.

67.2.1 Sample view for Inclusive email with family

You can use this view to identify all the inclusive emails that need to be reviewed. This view is particularly useful in the following situations:

- You're running threading analysis on a Custodial level.
- Duplicate emails matter from a production perspective.

This view provides context for review by displaying family members like attachments to an email when going doc to doc.

In conjunction with this view, the Related Items pane for a Thread Group view can help reviewers find context for the email that is being reviewed.
67.2.2 Sample view for Inclusive and non-Duplicate email with family

You can modify the conditions on the preceding view to return only inclusive, non-duplicate emails, plus family. You can use this view to identify all the inclusive emails that need to be reviewed. This view is useful when you're running threading analysis across the entire dataset, and the case team has agreed to produce just the inclusive non-duplicate emails.

This view provides context for reviewing by displaying family members like attachments to an email when going doc to doc.

In conjunction with this view, the Related Items pane for a Thread Group view can help reviewers find context for the email that they're reviewing.

67.2.3 Sample view of the Related Items pane for an Email Thread Group

This default view is available in the Related Items pane. It allows reviewers to identify the context of an email that is being reviewed. You can use this view with both sample views identified previously.
67.2.4 Sample view for Near Dupe Documents

This view displays all near duplicate documents that are grouped together and identifies the percentage of similarity to the Principal documents.

**Note:** From a review standpoint, it’s important for the reviewer/case team to be aware that the document with the most amount of text is identified as the Principal document.

67.2.5 Sample view of the Related Items pane for a Near Dupe Group

This default view is available in the Related Items pane. It allows reviewers to identify the context of an email that is being reviewed. You can use this view in conjunction with the Sample view for Near Dupe Documents.

<table>
<thead>
<tr>
<th>Label</th>
<th>Field(s) / Logical Groups</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Control Number</td>
<td>Unique Identifier for the document</td>
</tr>
<tr>
<td>B</td>
<td>Email Threading Display</td>
<td>Custom (system) field created following Email Thread analysis that provides a visual representation of the Indentation of the email (Label-F), Subject Line, and</td>
</tr>
<tr>
<td>Label</td>
<td>Field(s) / Logical Groups</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>C</td>
<td>Email Thread Group</td>
<td>Attachments to the email.</td>
</tr>
<tr>
<td>D</td>
<td>Inclusive Email / Inclusive Reason</td>
<td>Custom (system) field created following Email Thread analysis that maintains the relationship of all the emails in an email thread.</td>
</tr>
</tbody>
</table>
| D     | Inclusive Email / Inclusive Reason | Custom (system) field created following Email Thread analysis that identifies if the email is Inclusive and the reason for inclusiveness (typically tied to contents of the message and/or presence of an attachment)  
- Inclusive email = Yes: The email contains all the contents of the prior emails.  
- Inclusive Reason = Message: The contents of the email classify the email as Inclusive.  
- Inclusive Reason = Attachment: The email contains an attachment that needs to be reviewed. |
| E     | Email Duplicate Spare / Email Duplicate ID | Custom (system) field created following Email Thread analysis that identifies whether the document is a duplicate email in the collection analyzed and provides a unique identifier to that duplicate email.  
- Email Duplicate Spare = No: Not a Duplicate email |
| F     | Indentation               | This Custom (system) field created following Email Thread analysis defines the depth of the email in the email chain. The example here represents that the email was third in the chain. |
| G     | N/A                       | An example email along with associated attachment. |
| H     | N/A                       | An example Inclusive email (based on the message contents). |
| I     | N/A                       | An example Inclusive, but duplicate email (based on the message contents). |
| J     | N/A                       | An example Inclusive email (based on the message contents & presence of an attachment). |
| K     | Textual Near Dupe Group   | Custom (system) field populated following Near Dupe analysis that maintains the relationship of textually near duplicate documents. |
| L     | Textual Near Dupe Similarity / Textual Near Duplicate Principal | Custom (system) field created following Near Dupe analysis that identifies whether the Principal document (the one with the most text) and the Similarity score provides the score of all the associated Near Duplicate documents in that Near Dupe Group. |
| M     | Email Threading Relational Field Icon | Icon available in the Related Items pane in the document viewer that, when clicked, identifies all the documents related to that Email Thread group. |
| N     | Near Dupe Relational Field Icon | Icon available in the Related Items pane in the document viewer that, when clicked, identifies all the documents related to that Near Dupe group. |
| O     | Relativity Compare Icon   | This text comparison icon, when clicked, allows you to view the differences in text between two documents (useful for comparing near dupe documents.) |

- Views  
- Fields
68 Understanding the statistical sampling formula

Relativity uses human decisions on a sample set of documents to learn the coding parameters of your review project. One way to select your sample set is to pull a random group of documents from your data set. When this option is selected, calculates the number of documents to pull with a statistically valid formula.

68.1 Recipe overview

This recipe breaks down the formula used to calculate the number of documents in a random sample set. You can reference this formula to ensure your random sample set is statistically valid.

68.2 Requirements

This recipe is applicable to all versions of Relativity.

68.3 Formula breakdown

The following formula uses the total number of documents as well as the confidence level and margin of error you select for your project to determine a statistically valid size for your sample set.

\[
ss = \frac{Z^2 \cdot (p) \cdot (1 - p)}{c^2}
\]

- \(Z\) = Z value, which is based on the confidence level (e.g., 1.96 for 95% confidence level)
- \(p\) = reflects the chance of a defect (i.e., overturn), expressed as a decimal; Assisted Review conservatively sets this at .5
- \(c\) = confidence interval — or margin of error — expressed as a decimal (e.g., .04 = ± 4%)

69 Use case for Relativity Analytics - Clustering & Categorization

Published September 27, 2013

Relativity Analytics features — including clustering and categorization — allow users to expedite a review by organizing, identifying, and prioritizing documents. This recipe highlights techniques for expediting review using Relativity Analytics, as well as tips for utilizing the Analytics index to identify other hot documents.
69.1 Common scenarios

Common scenarios for using clustering and categorization include the following:

- You are working with a large number of documents. This data may not be coded at all, or it may contain documents that were coded with a multiple choice field for issues.
- You have limited time to complete the review.
- You have a limited number of reviewers.

69.2 Requirements

- Relativity 7.0 or higher
- Relativity Analytics index
- Workspace access
  - Document – Edit

69.3 Directions

69.3.1 Scenario 1

You have no prior knowledge of the data set. Clustering in Relativity groups conceptually similar documents without the need for example documents or user input.

Perform the following steps:

1. Use clustering to automatically organize documents into groups of related data.
2. Batch and assign documents to reviewers based on the clusters.
3. Bulk review groups of clustered documents.

Once the above operation is complete, you may find clusters of documents that are clearly irrelevant to your case, such as spam emails. Instead of reviewing hundreds or even thousands of junk emails one at a time, reviewers and system admins can eliminate impertinent documents with minimal time, effort, or subject matter expertise.

69.3.2 Scenario 2

You need to find key documents in an opposing production. Relativity’s categorization functionality identifies and groups similar documents together based on a set of example documents manually identified by the user. When you receive documents in an opposing production, and hot documents have been identified in your documents, you can pass these document values to the new production items by making them examples for the Analytics engine.

Perform the following steps:

1. Create a categorization set and use the Issue Designation multiple choice field as the Categories and Examples source.
2. Use the Synchronize feature for your new categorization set.
3. Click **Categorize All** to categorize the opposing counsel's documents with your categorization set and to identify opposing counsel's documents which are similar to your hot documents.

Setting the available Categories and Examples source option to use your multiple choice designation field enables the Synchronize feature for categorization. The Synchronize feature automatically creates categories for all choices associated with the specified field and designates example records for all documents with this field coded. With the example document records identified in your data set, categorization identifies and organizes similar documents in the opposing counsel's data set.

### 70 Using Analytics with CJK characters and non-English languages

*Published January 19, 2016*

Relativity Analytics works with Chinese, Japanese, and Korean characters (otherwise known as CJK characters in relation to computers). CJK languages have writing systems based entirely or partly on Chinese characters. These languages have some special characteristics that need to be handled properly in regards to computing and searching methodologies. Some of these characteristics are also common in other languages, such as Thai or Arabic.

The characteristics include:

- Non-delimiting of words with spaces
- Lack of word wrapping and line breaking
- Existence of thousands of complex characters
- Existence of multiple, conflicting versions of character sets
- Ability to edit characters by authors
- Horizontal and vertical layout frequently used

The system uses a process called tokenization to break a stream of text into a list of tokens, which are words, phrases, symbols, or other meaningful elements distinguishable within the text. The list then becomes input for further processing, such as searching.

### 70.1 Requirements

- Relativity 9.0 or higher
- Workspace access and system admin rights

### 70.2 Recipe overview

This recipe describes how Relativity Analytics handles CJK languages and provides some special considerations for leveraging text Analytics in a workspace with CJK documents to create Analytics indexes and to take advantage of Structured Analytics features.

---

**Note:** This document draws from various Content Analyst (CAAT®) resources with their permission.
70.3 Considerations for using Analytics with CJK characters

The Analytics engine is powered by individual terms, rather than sentences or phrases. Therefore, challenges with various sentence structures are not a concern with the Analytics engine. Regardless of the language, the text must be split into words (i.e., tokenized) in order to be indexed by Relativity Analytics. CJK languages use long strings of characters, and what constitutes a word can be context-specific and subject to debate. Given that words (or "features") are one of the primary building components behind CAAT's concept indexes, a process for splitting CJK text into words is crucial. To support this functionality, a number of algorithms run during document ingestion and filtering which split the CJK text into words. Once these algorithms run, you can build high quality indexes in these languages. This saves time and money by dramatically reducing the effort and expense of reviewing and understanding unstructured data.

Relativity Analytics uses the CAAT® software for conceptual indexing. CAAT® uses the getWordInstance method of the Java breakiterator class to handle all language tokenization, except Chinese. The globalization libraries are provided through the International Components for Unicode (ICU). Services provided by ICU are widely used by companies like Google, Apple, IBM, and others.

**Note:** For more information on the ICU, see [http://site.icu-project.org/](http://site.icu-project.org/).

There are two approaches to word-breaking (tokenization) for all Java languages (other than Chinese):

1. **Rule-based Iteration:** In this approach, spaces and punctuation define word boundaries.
2. **Dictionary-based Iteration:** In this approach, there are ambiguous word boundaries (no space separation, spaces in the middle of words, long compound words, etc.). This approach is used primarily for supporting non-Latin based languages. Java only applies the dictionary-based approach when necessary.

Chinese poses some unique challenges for Analytics. First of all, almost all characters can be one-character words by themselves. Moreover, they can join other characters to form multi-character words. This leads to a large amount of segmentation ambiguities. Second, compounding is the predominant word-formation device in modern Chinese. It is often difficult to tell whether a low-frequency compound is a word or phrase, and the lexicon can never exhaustively collect all low-frequency compounds. Third, the same pool of characters is also used in constructing proper names. Identifying proper names will be a problem, too. Finally, some specific morphological structures such as reduplication and A-not-A construction also need to be taken into consideration.

To handle these characteristics in Chinese, a different approach is taken. CAAT® applies a Maximal Matching algorithm in order to properly tokenize Chinese. This dictionary-based word segmentation alternative approach is taken due to previous limitations with the Java class.

70.3.1 Considerations for creating Analytics indexes containing non-English languages

When creating an Analytics index, it is recommended, if possible, that the index only contains one
language in its documents. Training the index in one language as opposed to multiple languages produces a better quality index. Additionally, if the data set is overwhelmingly one language (e.g., English) with only a small amount of another language (e.g., French), then the French documents would likely be outweighed by the English documents. The conceptual analysis of the French documents would not be as deep, and they might all be clustered together due to sharing so many common terms that are not found in the English documents.

Typically, there is a language that is used for the majority of documents in a workspace (i.e., the Primary language). Other languages are referred to as secondary languages. If a secondary language is less than 1K documents, an Analytics index might not be feasible for such a small data set. In this case, you may either include the secondary language documents in the index, or instead use Structured Analytics to analyze these documents.

If the index contains many multi-language documents, it may be impossible to split the languages. In this case, consider the following Analytics index settings below, especially Concept Stop Words, to ensure that it is a good quality index. You may apply stop words from both languages to this index.

### 70.3.1.1 Analytics index settings

<table>
<thead>
<tr>
<th>Supports non-English</th>
<th>English support only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept stop words</td>
<td>Email header filter</td>
</tr>
<tr>
<td>Regular expression filter</td>
<td>Automatically remove signatures and footers</td>
</tr>
<tr>
<td>Repeated content filter</td>
<td>OCR filter</td>
</tr>
<tr>
<td>Optimize training set</td>
<td>Go Words filter</td>
</tr>
</tbody>
</table>

### 70.3.1.2 Concept stop words considerations

Concept stop words are very common terms that are filtered from the Analytics index in order to improve quality. By default, the Analytics profile contains a Concept Stop Words field of only English stop words. These are customizable by the user in Relativity. The English stop words should be replaced by the appropriate stop words for the language being indexed. Analytics profiles are copied over when a workspace is used as a template; it is recommended that you create multiple profiles in the workspace template for each commonly indexed language.

kCura does not publish stop word lists. The following are three places where you can find stop word lists:

- Ranks NL [http://www.ranks.nl/stopwords](http://www.ranks.nl/stopwords) - 40 languages are available.
- Microsoft SQL Server Stop Lists - 33 languages are available. Perform a query as follows:

```sql
SELECT LCID, Name FROM sys.syslanguages
SELECT * FROM sys.fulltext_system_stopwords WHERE language_id = ###
```

- Take the default English stop words and use a translator tool to translate to the desired language.
70.3.1.3 Regular expression filter considerations

The Regular Expression filtering allows for a high level of customization. Creating these filters requires a working knowledge of regular expression methodology. This filter uses the regular expression syntax supported by the java.util.regex.Pattern Java class, which is very similar to Perl regular expressions.

Due to the English-only support by the Email Header filter, it is recommended to create a regular expression filter for the indexed language to suppress email headers. For example, consider the following German email header:

```plaintext
Von: John Smith
Gesendet: Mittwoch, 26 August 2015 17:05
An: Simon, Jane
Cc: Johnson, Ed
Betreff: Hallo
```

The Email Header filter would not find this text. Instead, use a Regular Expression filter as follows:

```regex
(?i)(Von.*?Gesendet.*?An.*?Cc:.*?Betreff.*?\r\n)
```

**Note:** The regular expression flag (?i) forces a case insensitive match. This is necessary because matching for what to filter out from Structured Analytics analysis is case-sensitive.

70.3.1.4 Repeated content filters considerations

It is highly recommended to use Repeated Content filters on the Analytics index. The **Auto-remove signatures and footers** feature only supports English, so the Repeated Content filter is needed to filter out the unwanted text. Repeated Content identification supports all Java-compatible languages and may be used to identify the filters. As always, it is recommended to evaluate the contents and impact of the filters before applying them to an Analytics profile.

70.3.1.5 Optimize training set considerations

The Optimize training set feature produces a better quality Analytics index by excluding documents from the training set that are poor quality for index training. This function takes into account many different characteristics, such as number of unique terms, punctuation, whitespace, size of text, etc. This feature is applied after the word-breaking has been applied to the documents.

**Note:** The **Optimize training set** feature supports all Java-compatible languages and should be enabled.

70.3.2 Considerations for using Structured Analytics with non-English languages

The following table lists the four Structured Analytics features along with their language support:
<table>
<thead>
<tr>
<th>Function</th>
<th>Supports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language identification</td>
<td>173 languages</td>
</tr>
<tr>
<td>Textual Near Duplicate identification</td>
<td>All Java-compatible languages</td>
</tr>
<tr>
<td>Repeated content identification</td>
<td>All Java-compatible languages</td>
</tr>
<tr>
<td>Email threading</td>
<td>English Email Headers only</td>
</tr>
</tbody>
</table>

**70.3.2.1 Language identification considerations**

Language identification examines the extracted text of each document to determine the primary language and up to two secondary languages present. This allows you to see how many languages are present in your collection, and the percentages of each language by document. You can then easily separate documents by language and batch out files to native speakers for review.

For multi-language documents, it returns the top three languages found and their approximate percentages of the total text bytes (e.g. 80% English and 20% French out of 1000 bytes of text means about 800 bytes of English and 200 bytes of French). Language identification does not use a conceptual index for this operation; a Structured Analytics set is used. The Analytics engine is trained on the prose of each supported language. The operation analyzes each document for the following qualities to determine whether it contains a known language:

- Character set (e.g., Thai and Greek are particularly distinctive)
- Letters and the presence or absence of accent marks
- Spelling of words (e.g., words that end in “-ing” are likely English)

Language identification is a Naïve Bayesian classifier, using one of three different token algorithms:

- For Unicode scripts such as Greek and Thai that map one-to-one to detected languages, the script defines the result.
- For CJK languages, single letters (rather than multi-letter combinations) are scored.
- For all other languages, language identification ignores single letters and instead uses sequences of four letters (quadgrams).

It also ignores punctuation and HTML tags. Language identification is done exclusively on lowercase Unicode letters and marks, after expanding HTML entities &xyz; and after deleting digits, punctuation, and <tags>. For each letter sequence, the scoring uses the 3-6 most likely languages and their quantized log probabilities.

The analysis does not use a word list or dictionary. Instead, the engine examines the writing to determine the language. The training corpus is manually constructed from chosen web pages for each language, then augmented by careful automated scraping of over 100M additional web pages. Though the algorithm is designed to work best on sets of at least 200 characters (about two sentences), testing has shown that it does perform well with small fragments, such as Tweets.

**Note:** Language identification in Relativity 9+ supports 173 languages. Language identification considers all Unicode characters and understands which characters are associated with each of the supported languages. For example, Japanese has many different character sets - Kanji, Katagana, Hirigana, all of which are supported. See Supported languages matrix in the System Guides for the full list of supported languages.
70.3.2.2 Textual Near Duplicate identification considerations

Textual Near Duplicate identification finds documents with highly similar text and places them into relational groups. You may then use near duplicate groups in searching or filtering.

**Note:** This feature supports all Java-compatible languages.

70.3.2.3 Repeated content identification considerations

Repeated content identification finds blocks of text that are repetitive throughout the documents and automatically creates filters. You can choose which filters you’d like to link to the Analytics profile in order to optimize the Analytics index.

**Note:** This supports all Java-compatible languages and is highly recommended to use with non-English cases.

70.3.2.4 Email threading considerations

Email threading is designed to find English email headers.

- When email headers themselves are in English, then the Analytics engine will thread them even if the contents are not in English (e.g., a Japanese email subject).
- When the headers themselves are not in English (this commonly happens when a non-English speaker hits "reply" in his or her email client and the email client then includes the headers in the embedded message), the Analytics engine will not be able to parse them out for email threading.

Processing engines tend to insert English-language headers on top of extracted email body text when they process container files such as PST, OST, or NSF. These headers, such as "To," "From," "Subject," etc., take their contents from specific fields in the container file. But the email body text does not, strictly speaking, contain the headers.

When the Analytics engine parses emails, it looks for English-language cues to determine when it is or isn't in an email header. In particular, it's looking for words like "To, From, CC, Subject" in the case of the traditional headers. And it's looking for format such as "on <date>, <author> wrote:" for single-line replies. There are other variations, but the email header logic is English-centric.

Email threading will be affected as follows by non-English email headers:

- Groups of emails which should be in a single thread will split into multiple threads. This is due to not matching up the foreign-language-surrounded segment with its English-language version (either when the email itself is in the collection, or when the email was replied to by both an English and a non-English email client).
- There will be fewer segments than desired in the **Email Thread Group** of a document which contains foreign email headers.
- If emails contain mixed languages in header fields, for example some field names are in English and some are in a different language, your **Indentation** field is lower than expected because Analytics doesn't identify the emails to be emails.
Note: The Analytics engine does not offer capability to effectively deal with email headers which are in a non-English language.

70.3.3 Checklist for using Analytics with non-English language and multiple language workspaces

The following table can be used as a general ordered checklist for using Analytics successfully in workspaces with a non-English primary language or with multiple languages:

<table>
<thead>
<tr>
<th>Process description</th>
<th>Non-English workspaces</th>
<th>Multi-language workspaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run Textual Near Duplicate identification</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Run language identification</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Determine if there will be separate indexes (i.e., at least 1K documents of a given language)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Find appropriate stop word list</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Create one new Analytics profile with appropriate stop words</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Create two new Analytics profiles with appropriate stop words (if separate indexes)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Create saved searches</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Run Repeated Content identification against the Searchable Set search</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Run two different Repeated Content identification sets against each Searchable Set search (separately)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Choose Repeated Content filters to apply</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Apply Repeated Content filters to the Analytics profile</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Apply Repeated Content filters to each appropriate Analytics profile</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Create index(es) with <strong>Optimize Training Set</strong> set to Yes and <strong>Auto-Remove Signatures</strong> set to No</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Setting up CJK document workspaces in Relativity

Analytics

71 Using categorization to find privileged documents

*Published February 22, 2012*

Relativity’s categorization functionality can help you locate potentially privileged documents that may have been overlooked during review. This recipe describes how to use the **Synchronize** feature
in Relativity Analytics to categorize existing privileged documents and use them as examples to locate similar documents that may also be privileged.

### 71.1 Requirements

- Relativity 7.1 or higher
- An active Analytics index
- Categorization data source (saved search)
- Edit/Add access to Analytics Categorization Set

### 71.2 Directions

1. Create a new Analytics Categorization Set (found under the Analytics tab) with the following settings:
   - a. Set the Documents to Be Categorized and Analytics Index.
   - b. Set Privilege as the Categories and Examples Source.
   - c. Save your Categorization Set.
2. Click Synchronize in the Categorization Set console.
3. Click Categorize All Documents.
4. Once categorization is complete, a field is created that contains the categorization results. It will be named Categories – <your Categorization Set name>. To begin screening for privileged documents, sort on the CSR – <your Categorization Set name>:Category Rank field, and prioritize the higher-ranking documents.

**Note:** Categorization has the potential to return many false-positive results. You can focus on more viable results by raising the Minimum Coherence Score.

### 72 Using categorization to identify hot documents

*Published April 18, 2012*

Using Relativity Analytics and methods for finding initial key documents, you can easily find other potentially hot documents.

This recipe describes how to use Analytics to find hot documents based on records that have already been marked as hot.

### 72.1 Requirements

- Applicable to all Relativity versions
- Active Analytics index
- Analytics categorization set
- Edit/Add access to Analytics categorization set
72.2 Directions

1. Identify hot documents to serve as example documents, using any of these methods:
   a. Use key search terms to find hot documents.
   b. Use key documents from your review to serve as examples against opposing production.
   c. Use the Find Similar Documents functionality to begin a review of main document items.
   d. Cluster and analyze documents from each cluster.
2. Mark your hot documents as examples, using any of these methods:
   a. Use a layout, and then apply values to those records.
   b. Right-click, and then apply the document as an example.
   c. Import a list of document identifiers into the Analytics Example Object.
3. Categorize the examples against the database, and check for more hot documents.

Analytics

73 Using Cluster Visualization to jump start your RAR project

This recipe describes how to using cluster visualization can help you locate training documents for your Assisted Review project. This recipe assumes that the case team has identified and tagged a few responsive or hot documents based on an initial assessment or investigation. Using cluster visualization, additional documents can be found that are conceptually similar to this initial set of documents. These conceptually similar documents will then be used to create a training round in a RAR project.

73.1 Requirements

- Relativity 9.2 or higher
- An active Analytics index
- An existing RAR project
- A cluster set (using default settings) that includes the RAR project documents

73.2 Directions

1. In the Cluster browser, select the cluster set that includes the documents in your RAR project, and then click Visualize Cluster.
2. In the filter panel, create a filter condition that returns the previously identified responsive or hot documents, and then click Run Search.

3. The system applies a heat map overlay to indicate which clusters contain the coded documents.
   a. Set the cluster visualization depth slider to 2 to see the sub-clusters that contain matching documents.
b. Optional: Click **Highlight Matches** in the legend to better visualize the relevant clusters (this feature became available in the 9.3.256.4 product update).
4. Select the clusters that contain matching documents by holding down the control key (command key on Mac), and then clicking on each of the shaded clusters. Click **Apply**.

5. The filter is currently set to return only the previously coded documents. To return all the documents in the selected clusters, disable the filter by clearing the check box in the filter card and clicking **Run Search**.

6. Now that all the documents in the selected clusters are displayed in the document list, save these documents as a search using the Save as Search functionality. Save the search in the search folder for your RAR project so it can be accessed when creating a new round. In the following example, the search is saved in the folder for the Responsive Review RAR project.
Note: You can also use the Save as List feature to save the documents as a list, then create a search from the list and move the search into the RAR project search folder.

7. Next, go to your RAR project and create a new training round:
   a. On the Project Home page, click Start Round.
   b. Select Training for the round type, and in the saved search for sampling field, select the search created above.
   c. Under the Sampling Methodology section, adjust the settings to control the number of documents that you want the system to randomly select from the saved search. For example, you could set the sampling type to fixed sample size, and then enter the number of documents you want in the round.
   d. Click Go.

You have now created a training round that focuses on documents that are conceptually similar to the previously identified important documents.

74 Using custom objects to secure coding choices

You may need to secure coding choices in a field so that you can control which groups are able to see them.

74.1 Recipe overview

This recipe shows you how to create a custom object and connect it to a document object via a field to secure coding values.
74.2 Requirements  
This recipe is applicable to all Relativity versions.

74.3 Directions  
1. Create a new object in your workspace, and name it **Coding Choices**.  
2. A Name field is automatically created on the Coding Choices object. Change the Name field to **Issue Choices**.  
3. Click **New Field** to create a field for the issue type.  
   a. Name the field **Issue Type**.  
   b. Set **Object Type** to **Coding Choices**.  
   c. Set the **Field Type** to **Single Choice**.  
   d. Click **Save and New**.  
4. Create another field to hold the issue selection on the Document object.  
   a. Name the field **Issues**.  
   b. Set the **Object Type** to **Document**.  
   c. Set the **Field Type** to **Multiple Object**.  
   d. Set the **Associative Object Type** to **Coding Choices**.  
5. Click the **Coding Choices** tab.  
6. Edit the view to include the **Issue Type** field.  
7. Import items using the Relativity Desktop Client, or use the New Coding Choices button to add items. This recipe allows you to add choices manually.  
8. Edit the Coding Choices layout to include the **Issue Type** field.  
9. Add issues to the **Issue Types**.  
10. Add items to the **Coding Choices** list and secure them as necessary.  
    - **Relativity dynamic objects**  
    - **Fields**

75 Using dtSearch to identify documents in other languages  
A Relativity Analytics index running Primary Language Identification (PLI) can be very helpful to determine languages. However, you can encounter a language not included in your index, or when a simpler dtSearch workflow is sufficient for identifying documents in other languages.  

75.1 Recipe overview  
This recipe describes how to use stop words as search terms to yield dtSearch results for a specific language.
75.2 Requirements

- Workspace access
  - dtSearch creation rights
  - Saved search creation permissions
- A list of stop words for the desired language

75.3 Directions

The concept behind this process is simple. A document in any language is likely to contain a stop word in that language. Consequently, you can take any language’s stop words and convert them into go words.

1. Create a dtSearch index of your extracted text. You may wish to double-check that your Extracted Text field was enabled for Unicode characters, if applicable.
2. Obtain or create a list of stop words for the language you wish to locate in your data set.
3. Construct a search using the dtSearch index, using the stop words as search terms separated by OR.

76 Using Near Duplicate analysis in review

Relativity can identify textually similar documents to assist in and speed up the review process. Near duplicate analysis is best suited for grouping documents which can then be batched for review based on the similarity, or used to create new document sets for further analysis. The goal is for reviewers to have the ability to see similar documents at the same time based on their textual similarity.
76.1 Requirements

- Relativity 8.0 and up
- Analytics server setup

76.2 Overview

After running a Near Duplicate analysis, system admins should view the Textual Near Duplicate Summary on the set’s Structured Analytics console, which breaks down the number of textual near duplicate groups that have been identified, along with averages of percentage of similarity and of the number of documents per similar document group.

Textual Near Duplicate Identification sorts the documents by size, from largest to smallest. This is the order in which they are processed. The most visible optimization and organizing notion is the principal document. The principal document is the largest document in a similar group and is the document that all others are compared to when determining whether they are near duplicates. If the current document is a close enough match to the principal document, as defined by the Minimum Similarity Percentage, it is placed in that group. If no current groups are matches, the current document becomes a new principal document. When the process is complete, only principal documents that have one or more near duplicates are shown in groups.

When running the process, a Minimum Similarity Percentage is assigned. This parameter indicates how similar a document must be to a principal document to be placed into that principal's group.

76.3 Directions

System admins should create a Textual Near Duplicates view for the review team.

1. In the Near Duplicate Identification view, add the following output fields:
   - **Textual Near Duplicate Principal** - identifies the principal document with a “Yes” value. The principal is the largest document in the duplicate group. It acts as an anchor document to which all other documents in the near duplicate group are compared.
   - **Textual Near Duplicate Similarity** - the percent value of similarity between the near duplicate document and its principal document.
   - **Textual Near Duplicate Group** - the identifier for a given group of textual near duplicate documents.
2. Add a condition to only show documents where the Textual Near Duplicate Group field is set.
3. Set the following sort orders on the Near Duplicate Identification view to list the textual near duplicate principals with the highest percentage of textual near duplicate similarity at the top:
   - **Textual Near Duplicate Group** - Ascending
   - **Textual Near Duplicate Principal** - Descending
   - **Textual Near Duplicate Similarity** - Descending

All documents should be reviewed in this process. Use the grouping and similarity to speed up the review process. The Relativity Compare function can compare two documents to assess their similarities and differences.
Reviewers will be able to view documents that are extremely similar but not identical to each other. For example, the case team may need to ensure a series of very similar reports are coded the same way. Another possible use is to help locate additional privileged documents that might have been missed during first pass review. In situations like these, it is common to use a view that displays textual near duplicates. Prior to the view’s creation, a system admin will run a near duplicate analysis during which documents with similar text patterns are placed together into relational groups. Exact syntax and word order are heavily considered during this analysis.

Here are the two most common fields you’ll likely encounter on a textual near duplicate view:

- **Textual Near Duplicate Principal** - identifies the principal document with a “Yes” value. The principal is the largest document (as measured by amount of text) in the duplicate group. It acts as an anchor document to which all other documents in the near duplicate group are compared.

- **Textual Near Duplicate Similarity** - the percent value of similarity between the near duplicate documents in a given group and their principal document. Each group of textually similar documents will contain a principal, which is typically the document in the group that contains the most text. All documents in a near duplicate group will be assigned a score that indicates how similar each document is to its principal.

Consider the following example. The first document, AZIPPER_0011323, is the group’s principal, as indicated by the “Yes” value in the Textual Near Duplicate Principal field. It also has a score of 100. All principals will have a score of 100, as they are by definition 100% similar to themselves, however, not all documents with a score of 100 are necessarily principals. The documents underneath are part of the principal’s relational group. The second and third documents are identical to the principal. We know this because they are 100% similar to it, as shown in the Textual Near Duplicate Similarity field on the far right of the view. The last three documents are very closely similar to the principal but are not exact duplicates; their scores indicate they are each 99% similar to the principal.

### 76.4 Workflow considerations

Textual near duplicate groups have a relational field that can be used to code several documents at once. Documents contained in the near duplicate group are textually similar, but similarity is usually not enough to treat near-duplicates as identical documents for the purposes of review. As such:

- It is not recommended to propagate coding on near duplicates, unless other analysis or evidence points to their similarity to justify such a step.

- It is not recommended to delete or otherwise shelve near duplicates and opt to focus solely on the principal document of each group.

- Document attachments may also differ on near duplicates, furthering the necessity to review or analyze all documents in the group.

**Textual Near Duplicate Identification**
77 Using redactions from other applications

When transferring data to Relativity from other systems, redactions are often part of the original data. No one wants to lose this work, but there is no way to export and import redactions between databases. The following solution helps retain these redactions.

77.1 Recipe overview

This recipe shows you how to create a production from another application with the redactions burned into the images.

77.2 Requirements

All versions of Relativity

77.3 Directions

1. Export data from the original database.
2. Create a production, burning in redactions.
3. Export the production with single-page TIFFs and an OPT file.
4. Export either native files, multi-page TIFFs, or PDFs from the original application with the path in a DAT file.
5. Import the data into the new database.

   **Note:** Load native files/multi-page image files as native files with metadata.

6. Import the production from your original application as single-page images into Relativity. Use redacted versions of images for production unless changes need to be made. A user can then delete the image and create it again from the clean native version.

   - Markups
   - Mass delete

78 Using reviewed documents to prioritize new documents via Analytics

*Published April 3, 2013*

As a case develops, you may need to add more documents to your workspace several times throughout your review. When this happens, the documents you have already coded can help prioritize your review of the new documents.

This recipe shows you how to use Relativity's near duplicate functionality to leverage previously coded documents to determine the potential values of new documents in your workspace.
78.1 Requirements

- Relativity 9.2 and below
- Structured Analytics for Near Duplicates including new and reviewed documents
- Workspace access
  - Saved searches
  - Structured Analytics

78.2 Directions

1. After new documents are loaded into your workspace, update your Structured Analytics Saved Search and Run Full Analysis.
2. Using a saved search, identify the documents already coded as responsive and add Include Near Dupe Groups. Based on those results, you can create a second saved search to identify potentially responsive documents.
   a. Your first saved search should contain responsive and Near Dupes.
      - **Name**: Responsive and Near Duplicates
      - **Includes**: Include Near Dupe Groups
      - **Conditions**: Designation is Responsive
   b. A second saved search will contain potentially responsive documents.
      - **Name**: Potentially Responsive Documents
      - **Conditions**:
        - Documents are in the Responsive and Near Duplicates search
        - Designation is not set
3. To leverage non-responsive documents, follow the same protocol.
   a. The first saved search will contain non-responsive and Near Duplicate documents.
      - **Name**: Not Responsive and Near Duplicates
      - **Includes**: Include Near Dupe Groups
      - **Condition**: Designation is Not Responsive
   b. The second saved search will contain potentially non-responsive documents.
      - **Name**: Potentially Not Responsive Documents
      - **Conditions**:
        - Documents are in the Not Responsive and Near Duplicates saved search
        - Designation is not set

- Analytics
- Batches
- Saved search

79 Using saved searches to complete conflict checks

When producing documents, check for conflicts between family members in a group and for family members that didn't receive a coding value. You can combine saved searches to perform this conflict check.
79.1 Recipe overview

This recipe shows you how to create a search for a value and include the family, and then create a search for the value only. By creating these searches you can compare the searches to see if coding conflicts exist within family groups.

Note: To read a technical breakdown of how Relativity builds and executes complex searches such as this conflict check, see Searching behind the scenes.

79.2 Requirements

This recipe is applicable to all versions of Relativity.

79.3 Directions

To create saved searches for your conflict check:

1. Create a saved search to find responsive items and their family items.
   a. Set Name to Responsive Designation.
   b. Set Includes to Include Family.
   c. Set Scope to Entire Workspace.
   d. Set Field to Designation, Operator to any of these, and Value to Responsive under Conditions.
   e. Click Save.
2. Create a saved search to find items without coding.
   a. Set Name to Items Not Set.
   b. Set Scope to Entire Workspace.
   c. Set Field to Designation, and Operator to is not set under Conditions.
   d. Click Save.
3. Combine the above saved searches to find items without consistent coding across family groups. The following equation illustrates this: Responsive Designation + documents not in Items Not Set = responsive family groups where not every item is responsive.
   a. Set Name to Family Conflicts.
   b. Set Field to (Saved Search), Operator to Document is in, and Value to Responsive Designation for the first condition.
   c. Select AND.
   d. Set Field to (Saved Search), Operator to Document is in, and Value to Items Not Set for the second condition.

This final search returns all items that aren't consistently coded as responsive in the database across family groups.

- Saved search
80 Using stop words and making some characters searchable in a dtSearch

Relativity ignores words that don't act as meaningful criteria when you create dtSearch and keyword queries. Ignored words are known as stop or noise words. Search indexes automatically include the default list of stop words. However, you can edit this list in the dtSearch list to suit your needs. This recipe includes an overview of stop words and steps to create custom lists.

Using the dtSearch alphabet file, system admins can take originally ignored words and make them searchable. However, dtSearch contains some characters and punctuation characters that you can't make searchable.

80.1 Requirements

- Workspace access
  - Edit, delete access on search indexes
  - Edit, delete access on searches

80.2 Directions

Relativity references the default list of stop words each time you create a new index. System admins can't edit stop words in keyword searches. The default stop word list consists of punctuation marks, single letters and numbers, and the following words:
<table>
<thead>
<tr>
<th>Begins with...</th>
<th>Stop words</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>about, after, all, also, another, any, are, as, at</td>
</tr>
<tr>
<td>B</td>
<td>be, because, been, before, being, between, but, both, by</td>
</tr>
<tr>
<td>C</td>
<td>came, can, come, could</td>
</tr>
<tr>
<td>D</td>
<td>did, do, does</td>
</tr>
<tr>
<td>E</td>
<td>each, else</td>
</tr>
<tr>
<td>F</td>
<td>for, from</td>
</tr>
<tr>
<td>G</td>
<td>get, got</td>
</tr>
<tr>
<td>H</td>
<td>has, had, he, have, her, here, him, himself, his, how</td>
</tr>
<tr>
<td>I</td>
<td>if, in, into, is, it, its</td>
</tr>
<tr>
<td>J</td>
<td>just</td>
</tr>
<tr>
<td>L</td>
<td>like</td>
</tr>
<tr>
<td>M</td>
<td>make, many, me, might, more, most, much, must, my</td>
</tr>
<tr>
<td>N</td>
<td>never, no, now</td>
</tr>
<tr>
<td>O</td>
<td>of, on, only, other, our, out</td>
</tr>
<tr>
<td>S</td>
<td>said, same, see, should, since, so, some, still, such</td>
</tr>
<tr>
<td>T</td>
<td>take, than, that, the, their, them, then, there, these, they, this, those, through, to, too</td>
</tr>
<tr>
<td>U</td>
<td>under, up, use</td>
</tr>
<tr>
<td>V</td>
<td>very</td>
</tr>
<tr>
<td>W</td>
<td>want, was, way, we, well, were, what, when, where, which, while, who, will, with, would</td>
</tr>
<tr>
<td>Y</td>
<td>you, your</td>
</tr>
</tbody>
</table>

**Note:** Relativity ignores stop words. However, Relativity doesn't ignore their position in the search phrase set. So, if you execute the query “apple w/6 pear,” the search returns the phrase “apple tree is far from the pear” even though it contains the stop words “is,” “from” and “the.”

### 80.2.1 dtSearches and stop words

The default list of stop words is the same in a dtSearch as in a keyword search. The primary difference is that you can customize the dtSearch index list. For example, if the word "never" is important to your litigation, remove it from the stop words list, so that your search results always return that word.

To create a custom stop word list, perform the following:

1. Create a new dtSearch index, and then name it "dtSearch - updated stop words."
2. Select your extracted text search for the Searchable set.
3. Delete the word "never" from the Noise Words list.
4. Save the list, and then perform a full build on your new index.

80.2.2 Stop words in languages other than English

You can set up stop words to search documents in other languages. If the workspace primarily contains documents in a different language, see this page for an overview of suggested stop words for use in nineteen additional languages.

Note: The following are restricted from dtSearch in Relativity. These are reserved noise words and characters which will continue to behave as operators as well as being noise words:

Relativity restricts the following from dtSearch. The following noise words and characters continue to behave as operators: and, or, not, to, contains, xfirstword, xlastword, ",, (,), *, ?, %, @, ~, #, &, ;, =.

80.2.3 Searching for Special Characters

80.2.3.1 The Alphabet file

dtSearch defines letters as characters to index. For most characters, this also means they are searchable. However, you can never make the characters discussed in the previous section searchable. By default, this includes all alphabetic characters (a-z and A-Z) and all digits (0-9).
dtSearch is case insensitive. It’s not possible to make dtSearch case-sensitive in Relativity by modifying the Letters section of the Alphabet file.

**80.2.3.2 Spaces**

dtSearch defines a space character as a character that causes a word break. Relativity doesn't index these characters, so you can't search them. By default, dtSearch treats the following characters as spaces:

\09\0a\0c !"#$&'()*+,./:;<=>?\[\5c]^`\

Values listed as \## are Unicode characters. Their definitions are:

- \09 - horizontal tab
- \0a - line feed
- \0c - form feed
- \0d - carriage return
- \5c - backslash (\)

**80.2.3.3 Ignored**

dtSearch defines an ignored character as a character that's ignored when processing text. You can't search these characters. By default, dtSearch ignores the following characters:

\08 %

\08 is the backspace character in Unicode.

**80.2.3.4 Hyphen**

dtSearch defines hyphens as characters that receive special processing in dtSearch. By default, dtSearch only classifies the - character as a hyphen.

**80.2.3.5 Restricted characters**

Either a limitation in dtSearch or the implementation of the dtSearch API restricts characters from being searchable.

**80.2.3.6 Always Restricted**

dtSearch always restricts the following characters, regardless of implementation:

" ( ) * ?

**80.2.3.7 Using reserved characters in the Alphabet list**

dtSearch treats reserved characters as operators regardless of what is set in the Alphabet list. Remember how those operators act when interpreting whether a solution works.

For example, will adding the % sign to the alphabet, and removing it from the Ignore List return "apple%"?

The % is the fuzzy operator, meaning, meaning any one character (or no character) can be in this spot and return results. This is very similar to how * (wildcard) or ? (wildcard for any single character) can...
character) work. dtSearch reserves the % sign as an operator. Thus, even though adding % to the alphabet and removing it from the Ignore list enables % to be a searchable character, % can't be successfully used as a search term. However, you can employ the following Regular Expression to bring back apple%:

"##apple\W"

The word apple% here is brought back because the regular expression metacharacter \W matches for any non-alphanumeric character, which includes the % sign.

80.2.3.8 Searching for &

Perform the following steps to search for the & character:

1. Make & a searchable character in your dtSearch index.

2. Delete & from the [Spaces] string, so dtSearch doesn't interpret it as a space (word break) in the index.

3. Perform a full build of the index.

You can now search for the & character in your workspace.

- Searching Quick Reference

81 Using tokens to customize stamps in a production

Occasionally, you might want your Bates numbers to include a confidentiality footer, or other information in the same stamp location. Using tokens — codes that represent metadata for records in the database — you can arrange multiple fields together in a free text box.

81.1 Recipe overview

This recipe describes how to use tokens for stamping on a production in a free text field instead of pulling down a particular field in the database.
81.2 Requirements

This recipe is applicable to all versions of Relativity.

81.3 Directions

Using production sets you can stamp documents with various pieces of information. Typically, there is one field selected from the drop-down for each stamp location (e.g., the right footer or left header). However, if you want to produce with more than one piece of metadata information, you can to use tokens to customize your stamps.

You can use tokens for any data field and you can use them in combination with static text and/or other tokens. The syntax to use a token is `{!Field:ArtifactID!}. You can find the ArtifactID for a field by adding the ArtifactID field to your Fields view.

An example of combining static text and a token would be: Custodian: `{!Field:1036657!}. In this example, Custodian is the static text and acts as a label for the token, the custodian field.

An example of combining multiple tokens would be: `{!Field:1036657!} {!Field:1037086!} {!Field:1037082!}. In this example, you are branding a string of three fields together on an image. You can use hard returns or spaces between the tokens. You can also add separators, such as carriage returns, hyphens, or commas, between the fields.

Production sets

82 Using the field tree to view folders

Occasionally, system admins need to use the folder browser view for security and still want to see the visual arrangement of folders as the document was originally stored. Using the Field Tree, we
can mimic the folder browser.

### 82.1 Recipe overview

This recipe shows you how to create a multi-choice field and open it to the Field Tree. Folder choices may be created when users perform their initial data load, or later on as an import overlay. Users may move a document from one folder choice to another (or display the document in multiple folder choices) by creating a document layout and editing as desired.

### 82.2 Requirements

- All versions of Relativity
- Folder paths less than 200 characters

### 82.3 Directions

1. Create a new field named **Folders** and open it to the Field Tree.
2. Use a load file with slashes between folder names.
3. Import into the multiple choice field. This loads into the field with semi-colons separating items.

![Field Tree with folders](image1)

The appearance of folders and folder choices on the field tree look very similar with the exception of the icon.

![Field Tree comparison](image2)
The image on the left is Folder Browser. The image on the right displays folder choices in the Field Tree.

To change the location of a document within a folder choice, or to display the document in more than one folder choice, create a layout on the document level, as displayed below.

The layout below only has the folders field available. Use the checkboxes just like any other multiple choice selection, adding or removing documents from choices to simulate movement between folders.

- Fields
83 Using the Update Duplicate Status Script

If global deduplication has not been applied to documents prior to loading them into a workspace, you can use the Update Duplicate Status script, which is included in the Processing Duplication Workflow solution, to identify duplicate documents.

This recipe provides additional guidance for using the Update Duplicate Status script. The script tags documents as either master, unique or duplicate. Using these tags, you can filter out duplicates, and pass only master and unique documents, together with their family members, on to the review team.

83.1 Requirements

- Relativity 9.1 or higher
- The Processing Duplication Workflow solution has been deployed in the workspace.

83.2 Directions

1. Publish or load documents into a workspace.
2. Set up a saved search that returns parent level emails and parent level loose documents. The duplicate status field will be set on the parent level documents, and then family members will be included with the parents, as described below.
   1. If using Relativity Processing, parent documents can be identified by including a search condition where the Level field equals 1.
3. Create a single choice Duplicate Status field with 3 choices: Unique, Master, and Duplicate.
4. Run the Update Duplicate Status script:
   1. Use the saved search described above.
   2. Use the duplicate status field created above.
   3. For the Relational Identifier, select a file duplicate hash field. If using Relativity Processing to load data, the Processing duplicate hash field can be used.
   4. Leave Duplicate Sort Order blank, unless you have set up a Custodian field to maintain a custom priority sort order. When this field is blank, the system sorts on document artifact ID, so the first document loaded in the workspace becomes the Master document when duplicates are identified.
5. After the script runs, the duplicate status will be populated for top level, parent documents.

6. At this point, you have a couple of options for including family members:
   1. Option 1 - Create a search that returns all documents where the duplicate status is set to Master or Unique, and include family.
   2. Options 2 - Copy the parent document duplicate status value to child documents and then create a search that returns all documents where the duplicate status is set to Master or Unique. You would not need to include family in this search.
      1. One way to copy the duplicate status values is to deploy the ‘Propagate Coding Post-import’ solution and run the script, Propagate Coding Post-Import. This script will copy the duplicate status value from the parent to each child document so that all documents in a family will have the same duplicate status value. For more information, see Propagating to one field on demand.
      2. An alternative way to copy the duplicate status values to is to construct searches to return each status, include family, and then use a mass edit to update the child documents with the status value of the parent.
      3. After copying the parent duplicate status value to children, create a search that returns all non-duplicates. Family does not need to be included in this search, since all family members will now have a duplicate status value.

7. Make the non-duplicate documents identified in the above steps available to reviewers.

83.2.1 Publishing or loading additional documents

After publishing or loading additional documents into the workspace, follow the steps described above to identify Master, Unique and Duplicate documents and to make non-duplicate documents available to reviewers.

When you are publishing or loading additional documents, keep the following in mind:
Run the Update Duplicate Status script against all parent level documents in the workspace for which you want to identify duplicates - both the newly published/loaded documents and the existing documents. Running the script clears the duplicate status field and then re-populates it.

The duplicate status for an existing document could change from Unique to Master if a duplicate of an existing, unique document is subsequently published/loaded into the workspace. This change in status will not impact the workflow described above, as all non-duplicate records (both Master and Unique) are being identified for review.

84 Validating RegEx filters in Structured Analytics

This recipe explains how you can access the Analytics server UI in order to validate Regular Expressions filters.

84.1 Directions

1. Run a full analysis of your Structured Analytics set BEFORE applying any Regular Expressions filters so you can see how the extracted text appears in the Analytics pipeline.
2. Navigate to the URL for the Analytics server user interface (UI), or log in to your Analytics server. If you do not have the credentials for your Analytics server, contact your IT department for assistance.
3. Open the Analytics server UI.
   - If you're on the Analytics server, the URL will be: https://localhost:8443/nexus/r1/
   - If you're accessing the UI directly from an external URL, the URL pattern will be: https://AnalyticsServerName:8443/nexus/r1/
4. Upon logging in to the UI, a list of available Analytics item stores is displayed.

5. On this item store list, find your Structured Analytics set by searching for the ArtifactID of that Structured Analytics set. The ArtifactID can be found by navigating to the Structured Analytics set in your Relativity workspace and looking at the URL.
6. Click Ctrl+F to search for your Structured Analytics set ArtifactID on the Analytics server UI, and then click the appropriate link.

7. You are now viewing the Analytics server's staging area for your Structured Analytics set. From here, you can look at the text in any document.

Before proceeding, note the URL for this page.

This is the breakdown of the URL:

https://AnalyticServer:8443/nexus/r1/staging/CaseArtifactID_StructuredAnaSetArtifactID

- CaseArtifactID = 1020998
- StructuredAnaSetArtifactID = 1056609
8. Navigate to a list of all documents in your Structured Analytics set by clicking item data in "Manage staging area item data".

```
CAAT 3.16.1.GA - Staging Area: 1020998_1056609
GET https://psi-rit-9-0:8443/nexus/r1/ staging / 1020998_1056609
```

View staging area configuration.
View staging area resources.
Manage staging area tasks.
Manage staging area item data.
Manage staging area item ingestion.
Run searches on this staging area.
Move this staging area to another device.
Rename this staging area.
Run diffs on items in this staging area.

Use a DELETE to delete this staging area:
Delete this staging area

```
{
  "connectorINGesting": false,
  "dateCreated": 1437585750395,
  "dateLastDataIngested": 1437585774145,
  "numItems": 657,
  "numProblems": 0,
  "stagingAreaid": "1020998_1056609"
}
```

9. You now have a list of links to all documents in the Structured Analytics set. Click any document ID to view how text for that document appears in the Analytics pipeline.

```
Found Item IDs
[
  "RK_CDC_01_012275",
  "RK_CDC_01_012279",
  "RK_DNC_01_012277"
]
```

You can also navigate to a document it by pasting its ID to the end of the staging area's URL, after the word "item" as follows. Add a forward slash to separate "item" from the document ID:

- https://AnalyticsServer:8443/nexus/r1/staging/1015598_1248495/item/EN000017
10. Once you have selected a document, click **text** toward the bottom of the Item screen.

![Image showing a document with text highlighted](image1)

11. You can now see how the extracted text of a document appears in the Structured Analytics back end. You must craft your RegEx filter to match how the text appears here.

![Image showing a web page with text](image2)

12. Run a Full Analysis again with your RegEx filter applied. Repeat steps 2–10. At step 10, you should see the text you matched with the RegEx disappear. If not, this means your RegEx has failed to match and you need to make adjustments to the expression.

Searching with regular expressions

### 85 Verifying Analytics filters

*Published November 6, 2012*

Using Relativity Analytics filters, you can optimize indexes and refine your results. The following process shows you how to confirm filters have been successfully applied.
85.1 Requirements

Relativity Analytics

85.2 Directions

1. Identify a document where filters are needed.
2. Create a test saved search that includes that document. Only return extracted text.
3. Create a new Analytics profile, applying the desired filters.
4. Create a new index using the saved search as the training and searchable set.

   Note: Analytics indexes require a minimum amount of text to build, so an error message will indicate if there is insufficient text and more documents need to be added.

5. Populate the index, but do not perform a build.
6. Navigate to the CA server.
7. Open the CAAT index directory. It should look like this:
   C:\ContentAnalyst\indexes\1015199_9(this is just an example)\Items\Features.dat

   Note: We recommend that you copy the file to another location to look at it.

8. Open the file. This shows you what the documents look like when they are sent to the concept search index. There will be random characters around each word, but you can see what is processed from this.
9. Compare those documents to the originals in Relativity to see what was filtered out.

   ■ Saved search
   ■ Analytics
Proprietary Rights

This documentation ("Documentation") and the software to which it relates ("Software") belongs to kCura LLC and/or kCura’s third party software vendors. kCura grants written license agreements which contain restrictions. All parties accessing the Documentation or Software must: respect proprietary rights of kCura and third parties; comply with your organization’s license agreement, including but not limited to license restrictions on use, copying, modifications, reverse engineering, and derivative products; and refrain from any misuse or misappropriation of this Documentation or Software in whole or in part. The Software and Documentation is protected by the Copyright Act of 1976, as amended, and the Software code is protected by the Illinois Trade Secrets Act. Violations can involve substantial civil liabilities, exemplary damages, and criminal penalties, including fines and possible imprisonment.

© 2017. kCura LLC. All rights reserved. Relativity® and kCura® are registered trademarks of kCura LLC.