Introductions

Who are you?
Analytics Best Practices
Relativity Analytics Features

**Structured**
- Email threading
- Textual Near Duplicate ID
- Language identification

**Conceptual**
- Concept searching
- Categorization
- Clustering
- Keyword expansion
## Use Case Features

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrowing the review set</td>
<td>Email threading</td>
</tr>
<tr>
<td></td>
<td>Foreign language identification</td>
</tr>
</tbody>
</table>
# Use Case Features

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrowing the review set</td>
<td>Near duplicate identification</td>
</tr>
<tr>
<td>Quality control</td>
<td>Cluster visualization</td>
</tr>
</tbody>
</table>
# Use Case Features

## Use Case
- Narrowing the review set
- Quality Control
- Investigation

## Feature
- Keyword expansion
<table>
<thead>
<tr>
<th>Use Case</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrowing the review set</td>
<td>Clustering</td>
</tr>
<tr>
<td>Investigation</td>
<td>Categorization</td>
</tr>
<tr>
<td>Quality control</td>
<td></td>
</tr>
<tr>
<td>Organizing large sets of data</td>
<td></td>
</tr>
</tbody>
</table>
Common Objection - “Analytics is best only for the largest cases.”

- Messaging with analytics in e-discovery has focused on large case wins.
- There are a number of uses for a majority of cases. For example:
  - **Batching** – Reviewing conceptually related documents increases review speed
  - **Production prep** – Analytics can help to avoid mistakenly producing privileged docs
  - **Keyword sampling** – Address keyword issues to help identify other potentially relevant documents
  - **Threading** – Only review inclusions and reduce the volume of email
Email Threading
Email Threading

What is it?
• Identifies and arranges emails that were part of a single thread or conversation.

What is it used for?
• Allows you to:
  – Easily see the order of each email in a thread.
  – See which emails are inclusive (i.e. have unique content).
  – Identify email duplicate spares (i.e. emails with the same content).

How will it help me?
• Sort and organize emails by thread for more intuitive review.
• Saves time if only reviewing the non-duplicative inclusive emails.
  ◊
Best Practices and Considerations

- Profile Setup
- Conversation ID
- Completeness of data
- Attachment ID
- English Language header information
- Bates Numbers
- Views that display Inclusive only
- Production specifications
- Recipients not considered
- QC using email threads
Near Duplicate
Can you spot the difference?
Textual Near Duplicate Identification

What is it?
• Identifies documents with highly similar text and places them into relational groups.

What is it used for?
• Allows you to:
  – Use near dupe groups in searching or filtering.
  – Conflict check coding decisions amongst near dupes prior to production.

How will it help me?
• Saves time by identifying very similar documents prior to the start of review. You can also use the near dupe groups for review and QC.
Best Practices and Considerations

• Run instead of or in place of email threading
• Use with Compare function
• Not meant to eliminate items but as prioritization and grouping
• Use for QC, comparison of datasets
• Include Numbers?
Language Identification
Language Identification

What is it?
• Determines a document’s primary language and up to 2 secondary languages.

What is it used for?
• Allows you to see how many languages are present in your collection, and the percentages of each language by document.

How will it help me?
• Easily filters documents by language and batch out files to native speakers for review.
• Determines if translation is needed. ◊
Best Practices and Considerations

- Footer information
- Header Information
- Segment dataset for desired reviewer
Conceptual Analytics
Best Practices and Considerations Index

- Minimum text
- Maximum text
- Repeated Content
What is Conceptual Analytics?

Relativity Analytics is a mathematical approach to indexing documents.

Terminology is understood based on its usage in your documents.
  - No outside word lists
    • Dictionaries, thesauri, etc.
  - Language-agnostic
  - Term co-occurrence, not term location ◊
Value of Concept Search

• Avoids term mismatch issues
  – Pop vs. soda
  – Football vs. soccer

• Avoids intentionally confusing use of language
  – Code words

• Finds documents even if exact language differs
  – Misspellings
  – Synonyms
Keyword Expansion
Keyword Expansion

What is it?
• Uses the concept space to allow users to submit terms and returns conceptually related words

What is it used for?
• Investigating the language of the workspace using known keywords

How will it help me?
• Allows you to find code words
• Assists in expanding the keyword list
• Familiarize yourself with the language of the case.
Best Practices and Considerations

- Concept or term submission
- Copy to dtSearch
Clustering
Cluster Browser
Heat Maps Show You Where Your Data Lives

FIND YOUR COUNTY
Choose a state…

KEY
Unemployment Rate
- More than 13%
- 10-12.9%
- 7-9.9%
- 0-6.9%
What is it?
• Use the power of the conceptual index to identify groups of conceptually related documents.

What is it used for?
• This can be used as a tool for investigation, analysis, review, or QC.

How will it help me?
• Investigate a large unknown dataset
• Cull out non-relevant documents quickly
• Speed up a linear review by batching conceptually related documents together
Best Practices and Considerations

- Cluster sub group
- Cluster all documents
- Batch by cluster
- QC with Clusters
Real World Challenges
Challenge #1

You have a discovery deadline quickly approaching. You were on target for your deadline until you were just dropped with 100 GB of data to review. How will you get through this data in time for your deadline?
You have a discovery deadline quickly approaching. You were on target for your deadline until you were just dropped with 100 GB of data to review. How will you get through this data in time for your deadline?

1. Use your existing coded documents as examples in a categorization set.
2. Create a search to find documents categorized as Responsive with a rank higher than 80.
3. Batch these out to second level review.
4. Create a search to find documents categorized as Not Responsive and rank greater than 80, or where they are not categorized.
5. Batch these out to first level review.

The key is prioritization.
Challenge #2

Your case team received 5 paragraphs from a subject matter expert depicting potential conversations among three people that corporate counsel believes to be important. How will you find these types of conversations between these three custodians?
Challenge #2 – Solution

Your attorney received 5 paragraphs from a subject matter expert depicting potential conversations among three people that corporate counsel believes to be important. How will you find these types of conversations between these three custodians?

1. Create a search to find these three custodians’ documents.
2. Create a categorization set.
3. Add each paragraph as an example, using Text Excerpts.
4. Run categorization against the documents in Search #1.
Challenge #3

You need to QC your production to make sure no privileged documents go out the door. How will you speed up this process to be as efficient as possible?
Challenge #3 – Solution

You need to QC your production to make sure no privileged documents go out the door. How will you speed up this process to be as efficient as possible?

1. Run Textual Near Duplicate Identification against all documents.
2. Run a search using metadata fields to find privileged documents.
3. Include Textual near duplicates on the search and filter down to find inconsistencies.
Challenge #4

You’ve already coded your own documents, and you just received a production from the opposing counsel. You’ve been data dumped! How will you find the relevant documents that you need?
Challenge #4 – Solution (Option A)

You’ve already coded your own documents, and you just received a production from the opposing counsel. You’ve been data dumped! How will you find the relevant documents that you need?

Option A:
1. Cluster the received production data.
2. Evaluate the clusters to see if there is any junk that can be quickly eliminated from review.
You’ve already coded your own documents, and you just received a production from the opposing counsel. You’ve been data dumped! How will you find the relevant documents that you need?

Option B:
1. Use your existing documents as examples in a new categorization set.
   a. Designation, Issues, Privilege, Hot Docs, etc.
   b. Use the production as the “Documents to be Categorized” search
2. Cluster the received production documents.
3. Use Pivot and compare each cluster with the categorized issues.
One More Thing…
Mobile Beta Program

- New iPad app separate from Relativity Binders
- Mobile-friendly design for saved searches, views, document lists, and coding layouts as they exist in Relativity
- All coding on mobile is synchronized to and viewable in Relativity
Join the Beta!

- Experience the new app before it becomes publicly available.
- Explore new features added every month.
- Guide the development of the product – a direct-line to Product Management.
- Email MobileBeta@kcura.com to participate.
In the Future

- Administer Relativity from your smartphone
- Case strategy and construction
- Reporting dashboards
- Support for additional devices and operating systems
Relativity Project Management Specialist Certification

75-question test
Take it online or in-person
Sign up by August 16 to save 50%
One More “One More Thing”…
RELATIVITY FEST 2016
OCTOBER 09-11 | CHICAGO, IL

Get 3 Days of e-Discovery & Information Governance Education

relativityfest.com