

BATCH REDACTIONS

Sightline Help Documentation



Table of Contents

Table of Contents	2
Batch Redactions	2
Batch Redaction Workflow	3
Saved Search	4
Batch Redactions	5
Choice of Searches for Batch Redactions	5
Analyzing Batch Redactions	6
Pre-Redaction Report	6
Execute Batch Redactions	7
Post Redaction Report	9
Batch Redactions History	10
Review and Confirm the Batch Redactions in DocView	11
Rollback Redactions	12

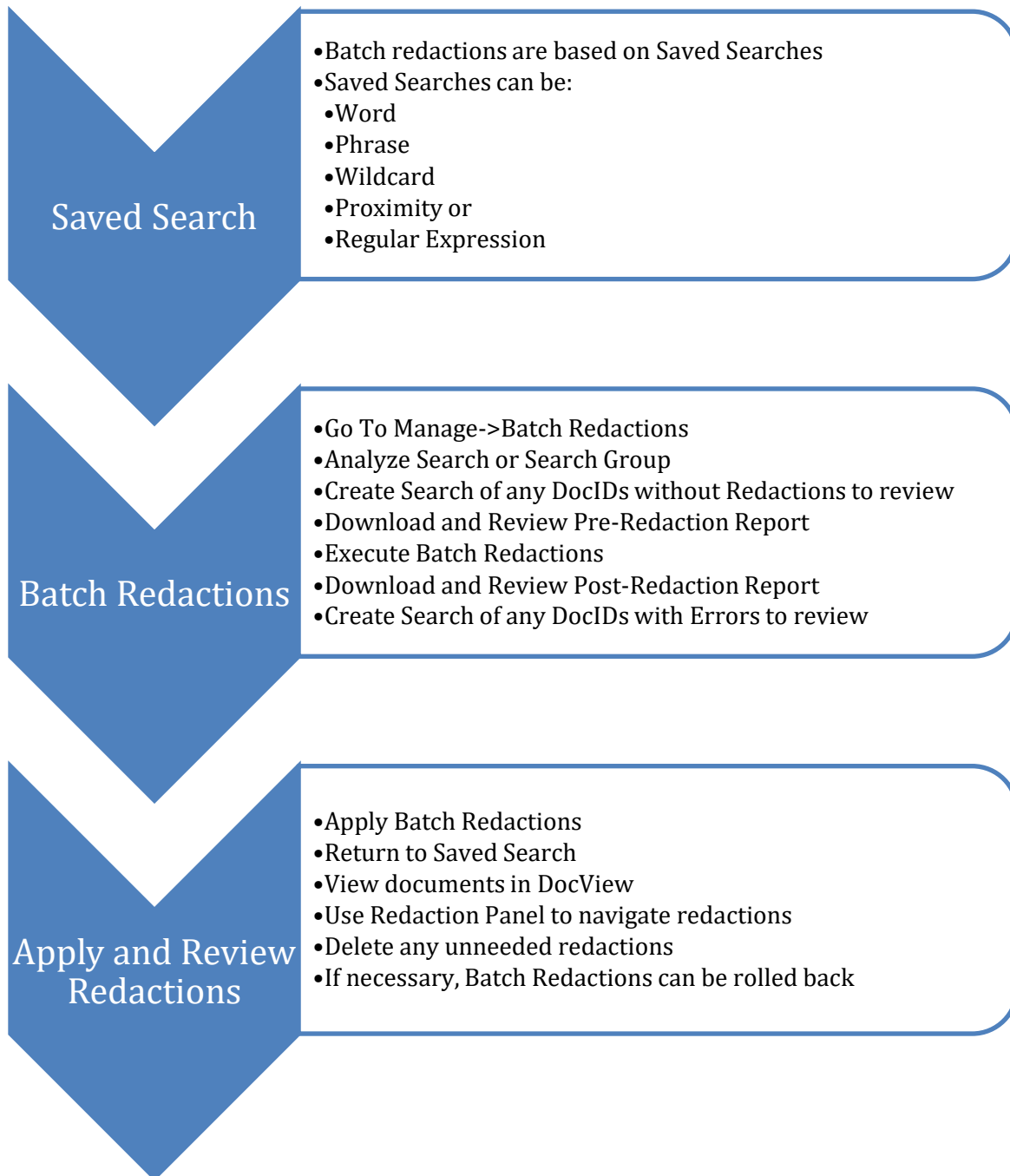
Batch Redactions

Sightline allows Users to use Saved Searches as the basis for text-based, mass redactions. Simply put, a User can execute a Saved Search for any Term, Phrase, or Regular Expression and automatically redact the results of that search across the corpus of data with a few simple clicks. Sightline provides reporting both pre- and post- implementation of the redactions. This allows the User to get a clear, defensible picture of what exactly will be redacted, what will not be and why. Subsequent to the application of the Batch Redactions, the User can perform their own QC of the redactions by returning to the Saved Search, viewing each Batch Redaction in the Document Viewer, and deciding if it should stay or be deleted. If, during the QC, it is determined that the redactions were not what was expected, the User can fully roll back the entire Batch Redaction set with the click of a button.

Please note that Batch Redactions are enabled by default for new projects created after the release of Sightline 7.0. If you need to enable Batch Redactions on an older project, please reach out to your Consilio Representative.

Batch Redaction Workflow

The below chart represents a basic workflow for Batch Redactions. Depending on the needs of the case, this may need to be modified. If you would like to discuss workflow options, please reach out to your Consilio Client Manager or Project Manager.



Saved Search

The Batch Redactions are based on Saved Searches. The User must have already created the Saved Search to apply the redactions. While it is Consilio best practice is to keep Batch Redaction searches simple, there is no limitation as to what kinds of searches can be used. The more complex the search, the more likely it is that the Batch Redaction tool will be unable to apply the redaction.

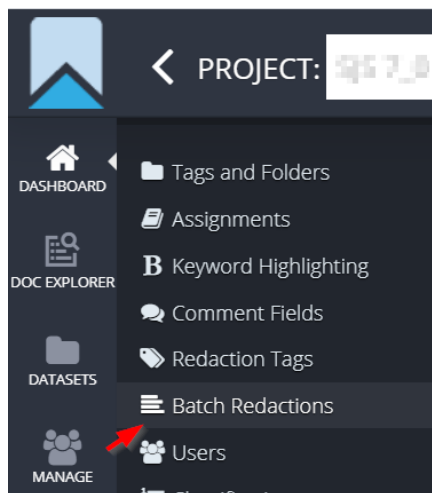
Two common examples of Saved Searches for Batch Redaction are:

- Words and Phrases – Basic word and phrase searches can be used to batch redact a term, phrase, person’s name, business name, etc.
- Regular Expressions – Alpha-numerical combinations and patterns, e.g., phone numbers and Social Security numbers.

Please note the below behavior for certain complex search requests.

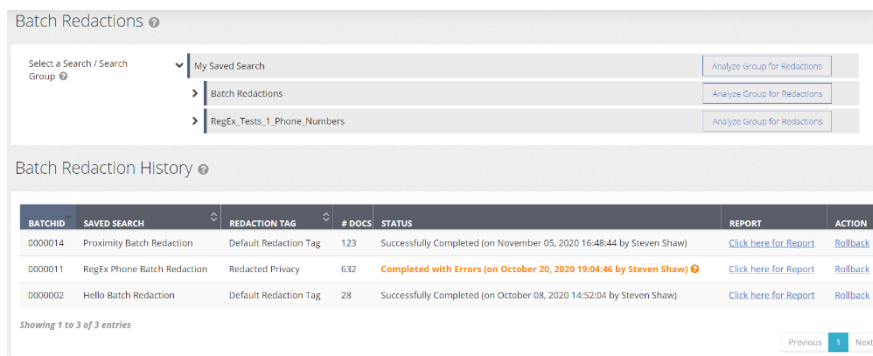
- Wildcard Searches – If there is a search result that hits due to the wildcard expander contains a special character (e.g., the search is company* and the search hit is company.org) the system will make overlapping redactions – one for “company” and one for “company.org”. The term will be fully redacted but will have unexpected redaction text.
- Proximity Searches – If there is a proximity result with overlapping proximity (e.g. “details reservation”~4 will hit twice on the following phrase “Reservation status has changed. Details of the reservation are listed below:” once for “Reservation status has changed. Details” and one for “Details of the reservation”) Sightline will only make the first redaction. To capture both redactions, it is advisable to use two ordered proximity searches (e.g., “details reservation”~4! and “reservation details”~4!)
- Elastic Search, which is Sightline’s search engine, is configured to whitespace most of the special characters e.g., comma, hyphen, period, etc. As a result, a search for “john.smith” not only matches “john.smith” but also matches “john-smith”. However, the Batch Redaction analysis task when analyzing for “john.smith” will only identify “john.smith”. This can result in differences in the document counts for an identified Batch Redaction term across the search results, pre-redaction report and the actual Batch Redaction run and should be reviewed by the User when using this type of search for Batch Redactions.

Batch Redactions



Once the Saved Search has been created, the User is now able to run the Batch Redaction tool. Navigate to Manage -> Batch Redactions from the menu on the left-hand side.

This will open the Batch Redaction Page.



The Batch Redactions page is split into 2 halves; the top half contains the list of Saved Searches available to be analyzed for redactions and the bottom half tracks the Batch Redaction History. Alongside each search the application presents the number of search hits and the time this search was last executed. The User can select either a Search or a Search Group to proceed with the Batch Redaction. When a User selects a Search Group the action will be performed on all the searches under the selected Search Group.

Choice of Searches for Batch Redactions

While the application allows the User to pick any saved search or search group for a Batch Redaction run, it is advisable to pick searches that have meaningful results in the context of a Batch Redaction, i.e., searches with actual document content hits. Batch Redactions will not work with metadata only searches, DocumentContent.Content() searches or pure Work Product searches like (tags:[name:["Responsive"]]). It is also not advisable to use deeply nested proximity query searches. If a User chooses to use a proximity search for Batch Redactions, the resulting applied redactions will be the entire text between the proximity terms, not just the terms specified in the proximity search. For example, if the User uses a proximity search "market transaction"~5, which reads as the term "market" within 5 words of "transaction", then a Batch Redaction involving such a search will redact the terms "market" and "transaction" along with the words between them.

Analyzing Batch Redactions

From the list of Saved Searches, choose the Search or Search Group you wish to analyze for redactions and click the button next to the Group or Search name. To run Batch Redactions, the Search or Search Group must first be analyzed, prior to initiating Batch Redactions. When a Search/Search Group is not yet analyzed the action button against it presents as “Analyze Search for Redactions”/”Analyze Group for Redactions”.

	Analyze Group for Redactions
	Analyze Group for Redactions
Search Hit Count: 259 (October 28, 2020 05:37:38) ⓘ	Analyze Search for Redactions
	View Report & Apply Redactions
Search Hit Count: 163 (October 28, 2020 05:38:16)	View Report & Apply Redactions
Search Hit Count: 470 (October 28, 2020 05:42:22)	View Report & Apply Redactions
	Analyze Group for Redactions

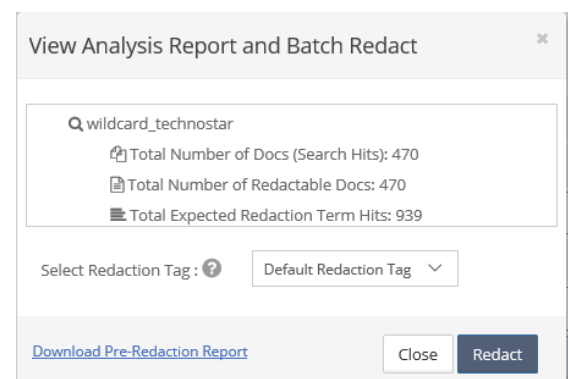
When the User chooses to “Analyze Search for Redactions”, the application analyzes the text of each document of the search results for identifiable redactions based on the search terms in the search query and generates a pre-redaction report comprising the identifiable potential redaction occurrences. While a search is being analyzed, the application presents a progress bar against it. Once analysis is complete, the application sends a notification to the User, and the action button against the search or the search group in the tree changes to “View Report and Apply Redactions”. The User can download the analysis report from the notification.

If any of the searches in a Search Group are yet to be analyzed, the action button against the Search Group of this search will present as “Analyze Group for Redactions”. A User can initiate Batch Redactions for a Search Group only when all the searches in the Search Group have been analyzed.

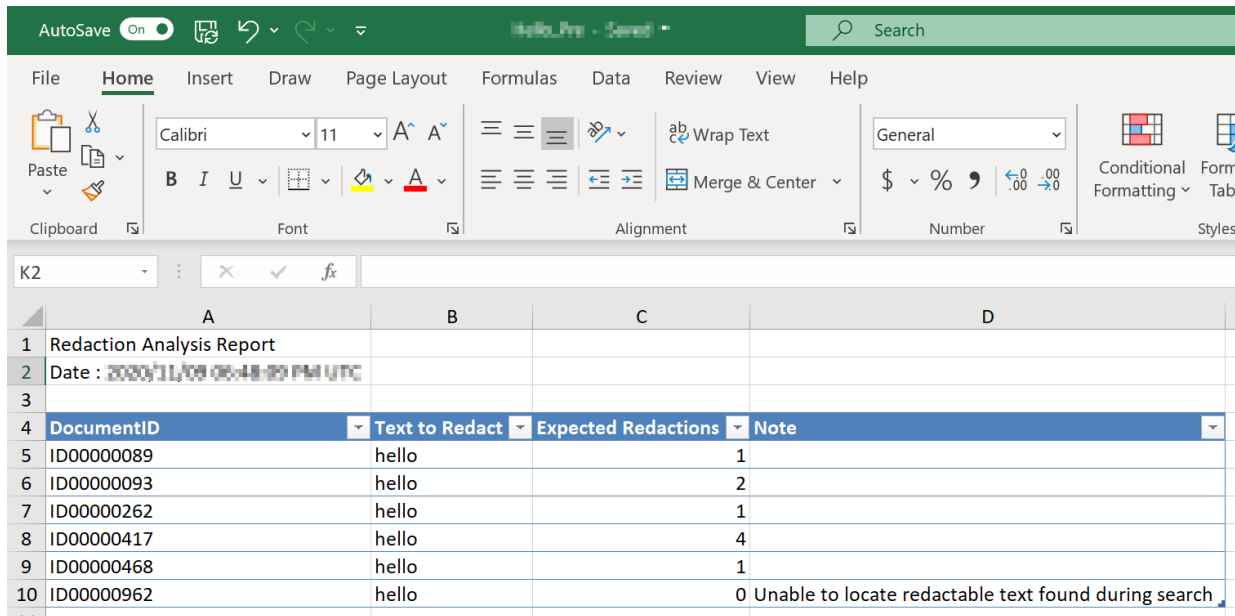
If a User re-executes the saved search that has already been analyzed but not yet subjected to a Batch Redaction, this action button for this search in the Batch Redactions saved search tree will present as “Analyze Search for Redactions” and the User will need to analyze the saved search again before initiating Batch Redactions.

Pre-Redaction Report

The Pre-Redaction report can be downloaded from either the background tasks notification or from the link “Download Pre-Redaction Report” in the “View Report and Apply Redactions” popup. The Pre-redaction report contains: 1) the matching redactable text in the documents of the search results, and (2) the expected count of occurrences of each redaction in each document.



It is strongly recommended to download and review the results of the analysis prior to running Batch Redactions on an analyzed search. This allows the User to confirm the matching redactable terms and provides an opportunity to adjust the search terms, if needed.



DocumentID	Text to Redact	Expected Redactions	Note
ID00000089	hello	1	
ID00000093	hello	2	
ID00000262	hello	1	
ID00000417	hello	4	
ID00000468	hello	1	
ID00000962	hello	0	Unable to locate redactable text found during search

As you can see from the above report snippet, the 6 listed documents contain the search term (in this example, the word, “hello”). Document ID00000093 has 2 instances of the word that are expected to be redacted, while document ID00000417 has 4. Additionally, we can see from the report that document ID00000962 does not have any expected redactions although the document contains a search hit. The note for this document lets the User know that the system is unable to locate redactable text.

It is Consilio’s recommendation and best practice to create a Saved Search of DocIDs where the Expected Redactions are 0. The User can use this search to review the documents that will not have redactions applied and be able to apply any necessary redactions manually.

Execute Batch Redactions

Once a Search has been analyzed for Batch Redactions, a User can initiate a Batch Redaction by clicking on the “View Report and Apply Redactions” button. When a User clicks this button, the application presents a popup that contains:


- “Total Number of Docs” – The number of documents from the saved search identified at the time of executing the search.
- “Total Number of Redactable Docs” – The count of the documents that were identified to contain potential redactions by the “Analysis”.



- “Total Number of Redaction Term Hits” – The count of identified potential Batch Redaction occurrences of all the terms identified by the analysis across the documents analyzed i.e., the total number of redactions that is expected to be applied.

The “View Analysis Report and Batch Redact” popup also presents the option to “Select a Redaction Tag” and lists the available Redaction Tags. The selected Redaction Tag will be associated with each redaction that will be applied during the subsequent Batch Redaction. The User can then click on “Redact” to initiate the Batch Redaction process.

When a Batch Redaction is in-progress, other Users will not be able to initiate a Batch Redaction on this Saved Search, until it is completed. The Batch Redaction process runs as a background task, where it uses the redaction engine to apply redactions on the terms identified during the analysis step. When the Batch Redaction process is initiated, the application presents the current status of the Batch Redaction in the History section.

While the Batch Redaction is in-progress, if the User wants to abort the Batch Redaction, they can do so by clicking on the “Cancel and Rollback” link in Batch Redaction History. When the Batch Redaction background process completes successfully, the User is notified via the bull-horn background tasks notification. The User also gets an email regarding the completion of the Batch Redaction process. In addition, the status of the Batch Redaction run in the Batch Redaction History section presents the completion status of the Batch Redaction run.

Batch Redaction History 

BATCHID	SAVED SEARCH	REDACTION TAG	# DOCS	STATUS	REPORT	ACTION
0000012	term_collateral	Redacted Privacy	161	Completed with Errors (on November 02, 2020 18:15:03 by Guruprasad Guruprasad) 	Click here for Report	Rollback
0000010	wildcard_assurstar	Default Redaction Tag	69	Successfully Completed (on October 28, 2020 10:49:51 by Guruprasad Guruprasad)	Click here for Report	Rollback
0000008	wildcard_assurstar	Default Redaction Tag		Successfully Completed (on October 28, 2020 08:52:38 by Guruprasad Guruprasad)  Rollback Completed (on October 28, 2020 10:37:33 by Guruprasad Guruprasad)	Click here for Report	
0000007	term_rentals	Default Redaction Tag	2	Successfully Completed (on October 28, 2020 08:27:36 by Guruprasad Guruprasad)	Click here for Report	Rollback

All Batch Redaction processes – application, rollback, etc. – are run through a queueing system. It is possible that when running multiple tasks that you will see that one process will need to wait until the previous process completes.

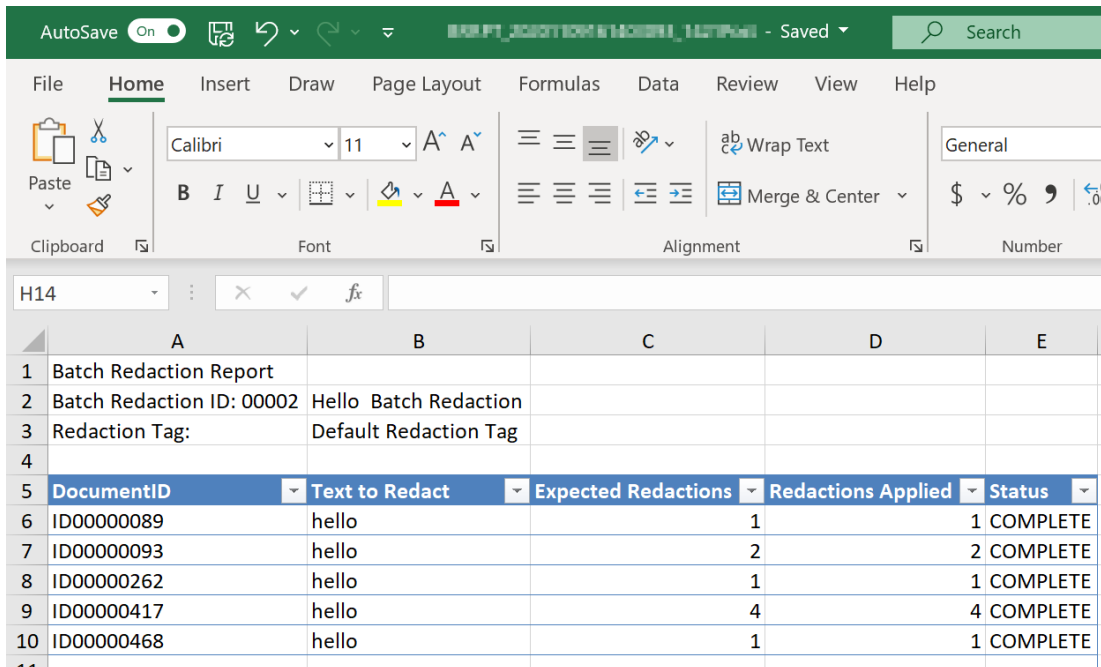
Please note that it is not advisable to apply manual redactions at the same time Batch Redactions are being applied. This may result in lost redactions.

Post Redaction Report

When a Batch Redaction completes, it generates a report which lists each redaction applied during the process and the DocumentID where the redaction has been applied. The report also provides the expected redactions as well as the actual redactions applied, which helps identify any discrepancies in case any expected redactions that were not applied for some reason or if any additional redactions were applied for some reason. The user can download a report for a completed Batch Redaction run (1) by clicking on the link “Click here for Report” in the Report column in the Batch Redactions History section, or (2) from the background tasks notification.

The report header presents the Batch Redaction ID and the Search name on Line 2 and the Redaction tag that was associated with the Batch Redactions applied on line 3. Each row in the report details represents the Batch Redaction that was attempted to be applied during the batch redaction run.

For each batch redaction the following information is presented in the report details:



DocumentID	Text to Redact	Expected Redactions	Redactions Applied	Status
ID00000089	hello	1	1	COMPLETE
ID00000093	hello	2	2	COMPLETE
ID00000262	hello	1	1	COMPLETE
ID00000417	hello	4	4	COMPLETE
ID00000468	hello	1	1	COMPLETE

- **DocumentID:** the DocID of the document
- **Text to Redact:** the text/term that was identified during analysis and attempted to be redacted.
- **Expected Redactions:** The count of redactions for that term in the given document that was expected to be redacted.
- **Redactions Applied:** The actual count of redactions applied for that term in the document. This number may differ from the Expected Redactions if the search hit on a metadata value, as metadata cannot be redacted through Batch Redactions.
- **Status:** Status presented here is based on whether the count of applied redactions matches with the count of expected redactions for that document. The value is “COMPLETE” if the two

counts are equal. The value of is “PARTIAL” if the count of applied redactions is less than the count of expected redactions. The Status can also show, “ADDITIONAL TAGS”, indicating that the count of applied redactions is greater than the count of expected redactions.

It is strongly recommended that the User review this report. This report, and the Pre-Redaction report, will help with the QA and QC of the Batch Redactions.



It is Consilio’s recommendation and best practice to create a Saved Search of any DocIDs in which the Expected and Applied Redactions are not equal. This will allow the User to quickly identify and resolve any redactions that need to be applied or removed.

Batch Redactions History

The “Batch Redaction History” contains a listing of all Batch Redaction that have been executed in the project. Each row in the history corresponds to a Batch Redaction that was invoked in the project and has a unique BatchID associated with it.

Each row here provides the current status of that Batch Redaction along with possible next actions that the User can perform. Consequently, the Batch Redactions History section also provides visibility into the progress of on-going Batch Redactions, with the “Status” and “#Docs” columns presenting the current status of an ongoing Batch Redaction process at that point-in-time. If multiple Batch Redactions were invoked for the same search, each instance of the Batch Redaction run will have its own distinct entry in batch redaction history.

Batch Redaction History

BATCHID	SAVED SEARCH	REDACTION TAG	# DOCS	STATUS	REPORT	ACTION
0000012	term_collateral	Redacted Privacy	161	Completed with Errors (on November 02, 2020 18:15:03 by Guruprasad Guruprasad) 	Click here for Report	Rollback
0000010	wildcard_assurstar	Default Redaction Tag	69	Successfully Completed (on October 28, 2020 10:49:51 by Guruprasad Guruprasad)	Click here for Report	Rollback
0000008	wildcard_assurstar	Default Redaction Tag		Successfully Completed (on October 28, 2020 08:52:38 by Guruprasad Guruprasad)  Rollback Completed (on October 28, 2020 10:37:33 by Guruprasad Guruprasad)	Click here for Report	
0000007	term_rentals	Default Redaction Tag	2	Successfully Completed (on October 28, 2020 08:27:36 by Guruprasad Guruprasad)	Click here for Report	Rollback

Below is information presented for each of the Batch Redactions in the History:

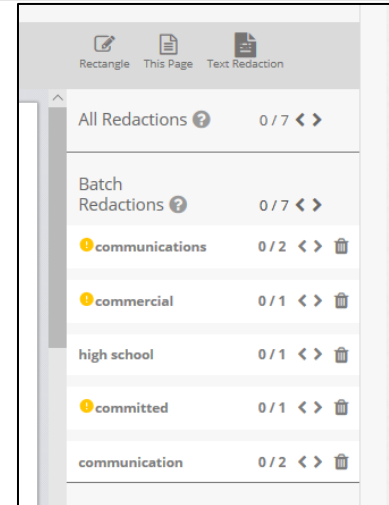
- **BatchID:** ID of the Batch Redaction.
- **Saved Search:** Name of search or searches that have been selected to run the Batch Redaction.

- **Redaction Tag:** This column presents the redaction tag used for the redactions in the batch redaction.
- **# Docs:** This column presents the number of documents for which the redactions that are eligible for redaction based on the Analysis of the Search. For a Batch Redaction in-progress, the #Docs column is updated periodically to present the number of docs that have been Batch Redacted so far. This column presents as x of y, where x is the number of documents for which Batch Redaction is completed at that point in time and y is the total number of documents identified for Batch Redaction during analysis.
- **Status:** The Status field represents the current status of the Batch Redaction at a given point.
 - **In Progress:** The Batch Redaction is still in progress.
 - **Successfully Completed:** The Batch Redaction has completed successfully and presents the User by whom the Batch Redaction was kicked off and when it was kicked off. A completed Batch Redaction means all the expected redactions for this search, as expected by the “Analysis” of the search, were applied successfully.
 - **Failed:** None of the expected redactions were applied as part of the Batch Redaction.
 - **Completed with Errors:** The Batch Redaction has completed but did not apply all the expected redactions.
 - **Rollback Completed:** When a redaction run has been completed successfully or completed with errors and has been subsequently rolled back.
- **Report:** This column presents a hyperlink “Click here for Report”, which the User can use to download the report of the Batch Redaction that was completed successfully or with errors. No link is presented if the batch redaction run is failed.
- **Action:** When a Batch Redaction completes successfully, or partially with some errors, a link to Rollback the Batch Redaction is available. If a user clicks on “Rollback”, the application will undo all the redactions that were applied as part of this particular batch redaction run.

Review and Confirm the Batch Redactions in DocView

A user can select the Saved Search used for the Batch Redactions from the Saved Searches page and navigate to DocView with the search hit documents to review the Batch Redactions applied on these docs.

In DocView, all redactions applied to a document via Batch Redactions are presented in the section labeled “Batch Redactions” in the Redactions Panel. To bring up the Redactions Panel, the User can click on the Redactions icon in DocView. In the Batch Redactions section, each term in the document that is redacted via Batch Redactions is presented along with the count of occurrences in the document. The User can navigate through each occurrence of the redacted term by clicking on the '>' and '<' icons. The count against each term presents in the form of x/y, where y is the total count of occurrences for that term in this document and x is the occurrence number that the user is currently viewing.



For a redactable term, if there are any discrepancies between expected redactions and actual applied redactions, it will be flagged with a Yellow indicator. The User may then manually identify and apply/remove the redactions. While reviewing the Batch Redactions, a user can click on the trash can icon against a term in the batch redactions panel to delete all batch redaction occurrences of the term in the document. A user can also manually edit or delete a specific redaction occurrence.

In addition to the Batch Redactions, the Redaction Panel also presents All Redactions. The All Redactions presents all the redactions on this document, including both manual and batch redactions. The user can navigate through each of these redactions by clicking on the '>' and '<' icons.

Rollback Redactions

If a user wants to undo the redactions applied by a given Batch Redaction for any reason, the user can do so by invoking the Rollback option for a given batch redaction run from the Batch Redactions History. “Rollback” will remove all the redactions applied in the given Batch Redaction run. When a Rollback is in progress for a Batch Redaction, the entry for that row in Batch Redactions History will present the status as “Rollback in Progress”. Once the Rollback is completed, the status for this row will be updated to add “Rollback Completed” along with the date, time and User who initiated the rollback. When a Batch Redaction run has been rolled back, the User can click on the “Report” link to download a Rollback Report. After a Batch Redaction has been rolled back, if the User initiates a Batch Redaction run again using the same Search, this Batch Redaction will be represented by a new entry in the Batch Redaction History section, with a different Batch Redaction ID.

Please note it is not advisable to apply manual redactions while a Batch Redaction Rollback is in process. This may result in lost redactions.