RightFind XML for Mining Quick Start Guide

Copyright Clearance Center

RightsDirect

Contents

CCC Customer Service

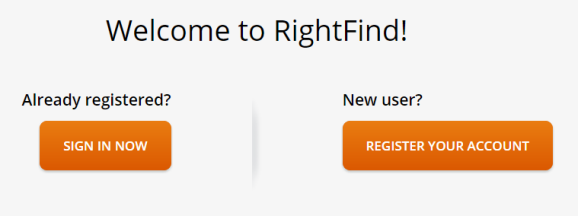
You can contact CCC's award winning customer service via phone or email.

- xmlformining@copyright.com
- 978-646-2777

Access RightFind XML for Mining

To access XML for Mining, visit https://rightfind.copyright.com.

You'll see the 'Welcome to RightFind' page. To sign in, click the 'Sign in now' box.

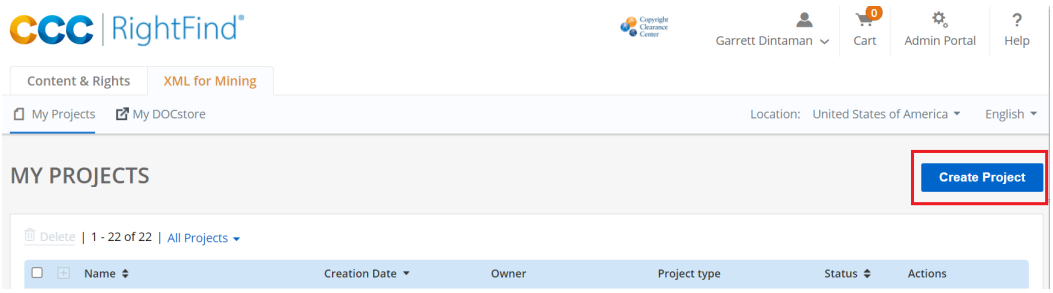


After logging into to RightFind, click the XML for Mining tab as shown below.

Create a Project

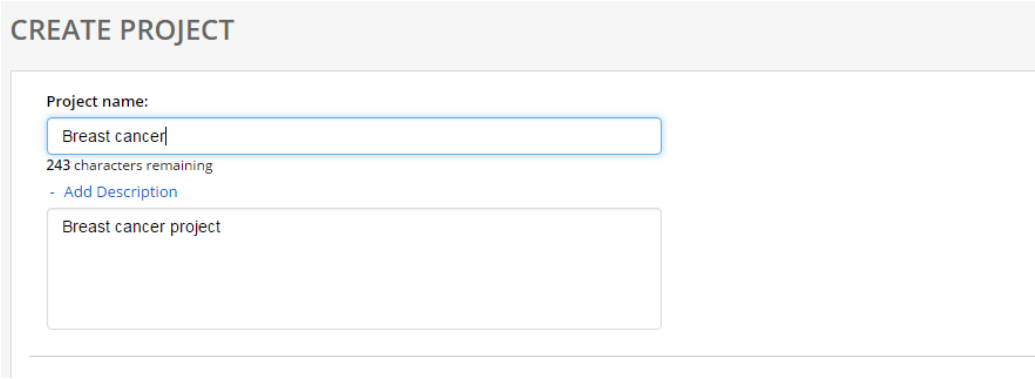
Creating an XML for Mining *project* is how you'll generate a subject-specific corpus of full-text XML content to mine against in your preferred text mining software.

To create a new project, click 'Create Project' from the 'My Projects' page.

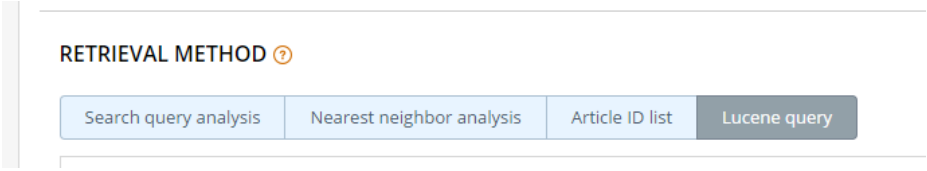


On the 'Create Project' page

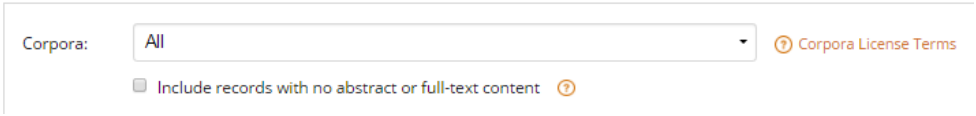
1. Define a project name (required). The project name is a short descriptor that relates the documents you are gathering in XML for Mining to the research you intend to do in your text mining tool
2. Enter a project description (optional). Include additional details about the project that may help to differentiate similar projects from one another.



3. Select the method for identifying the articles in your project's full-text XML corpus. Choose from the following four search retrieval methods:
   - *Search query analysis* – Create a traditional Boolean keyword search
   - *Nearest neighbor analysis* – Upload document(s) of interest to retrieve similar articles
   - *Article ID list* – Load a list of PMIDs or DOIs you have previously identified as articles of interest (e.g., using an external discovery tool) to retrieve the full-text XML of those articles
   - *Lucene query* – Create a complex and flexible query that refers to any indexed field, using unary or binary logical operators in Lucene syntax
     - You can find more information about Lucene syntax and index fields here: https://rightfind.copyright.com/rs-ui-web/build/dist/assets/lucene-query-project-tip.pdf

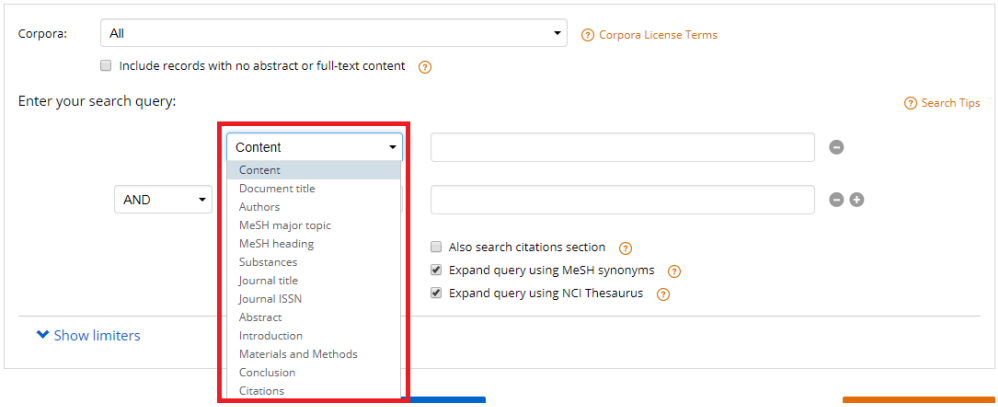


4. Select the corpora for your project. Select from the main corpus of full-text articles or the Medline abstracts corpus. Also select whether to include results that are metadata records only, with no abstract or full-text content; these are excluded by default.



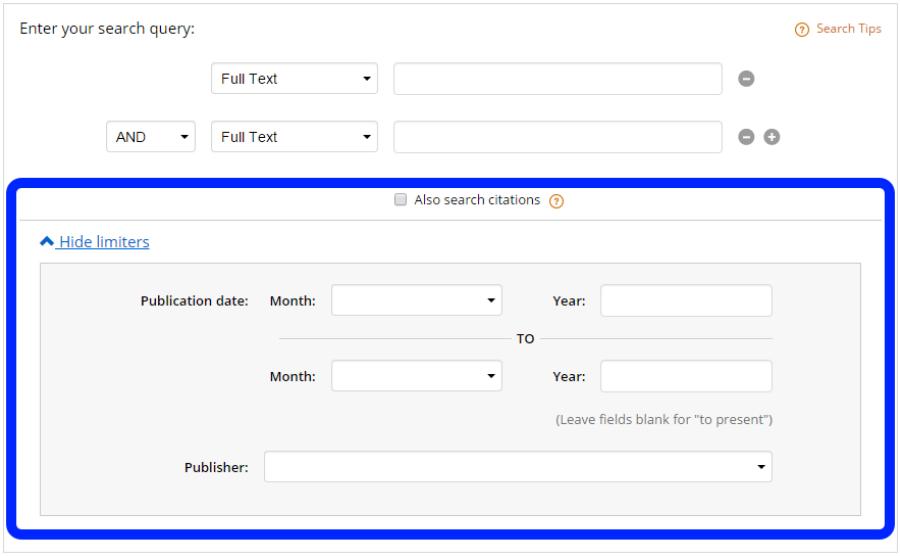
To create your project using 'Nearest neighbor analysis', 'Article ID list,' or 'Lucene query' methods, please see the Quick Reference section.

When using the 'Search query analysis' method, you can search within specific sections of the full-text article:

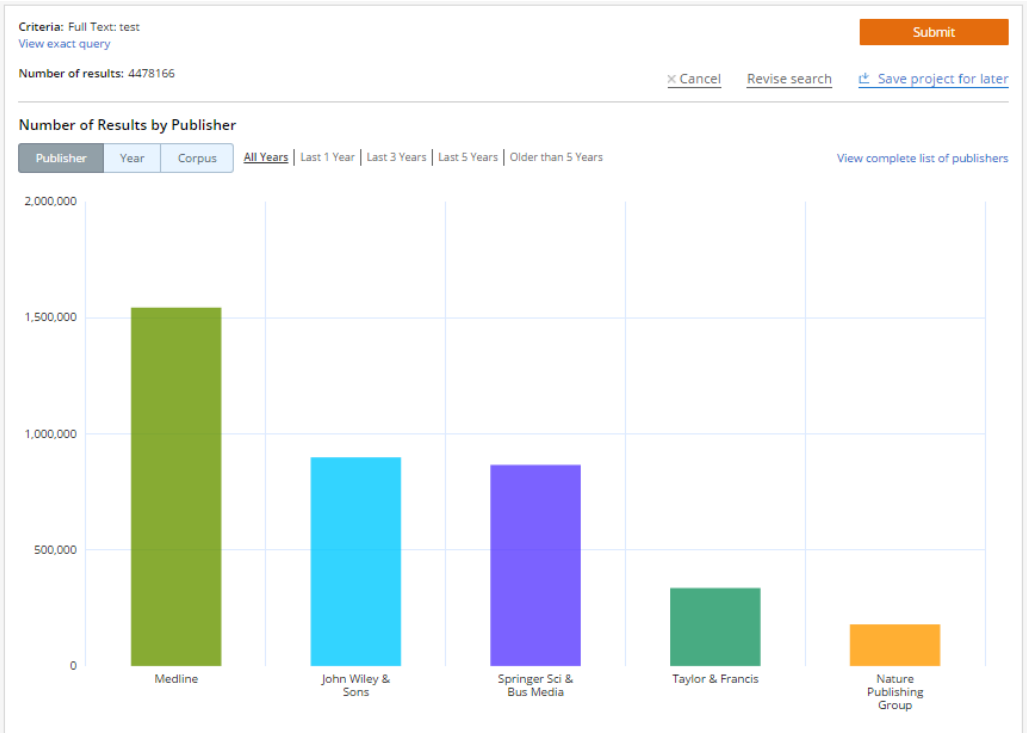


Be sure to specify whether XML for Mining should expand your query by applying synonyms from the NCI Thesaurus and MeSH synonym lists (e.g., searching for *mammary carcinoma* in addition to *breast cancer*). If you de-select both of the synonym options, only a stemmer will be applied to your query to ensure different forms of the word are captured as hits (e.g., *run, runs, running*).

… and optionally restrict your results by publication date or publisher:

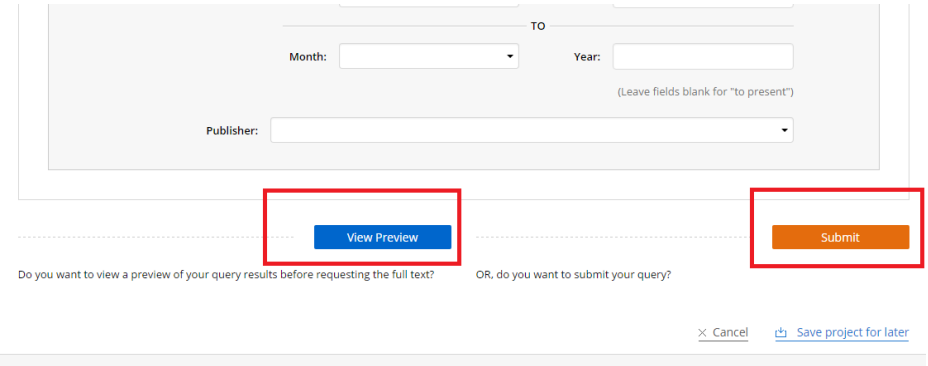


5. After providing your Boolean query, click the 'View Preview' button to see high-level statistics on the results you can retrieve for your query and get an idea of the size of the corpus you are building. You can use the interactive bar graphs to learn how your results vary by publisher, year, corpus, and journal.



NOTE: You can go back and forth between the 'Create Project' and 'Project Preview' pages to further refine and modify your query, as needed.

6. After you have refined your initial search, click the 'Submit' button to begin the retrieval process. During retrieval, XML for Mining obtains the articles that satisfy your request and conducts a lookup in your subscription holdings to create an initial project corpus for your review. No downloads are charged to your account at this time. You will have a chance to further refine your search before building your final corpus for use in your text mining tool.



NOTE: Checking the articles you have submitted against your subscriptions will take some time to process. You will receive an email when your project is ready to review, and you'll have an opportunity to filter the results further.

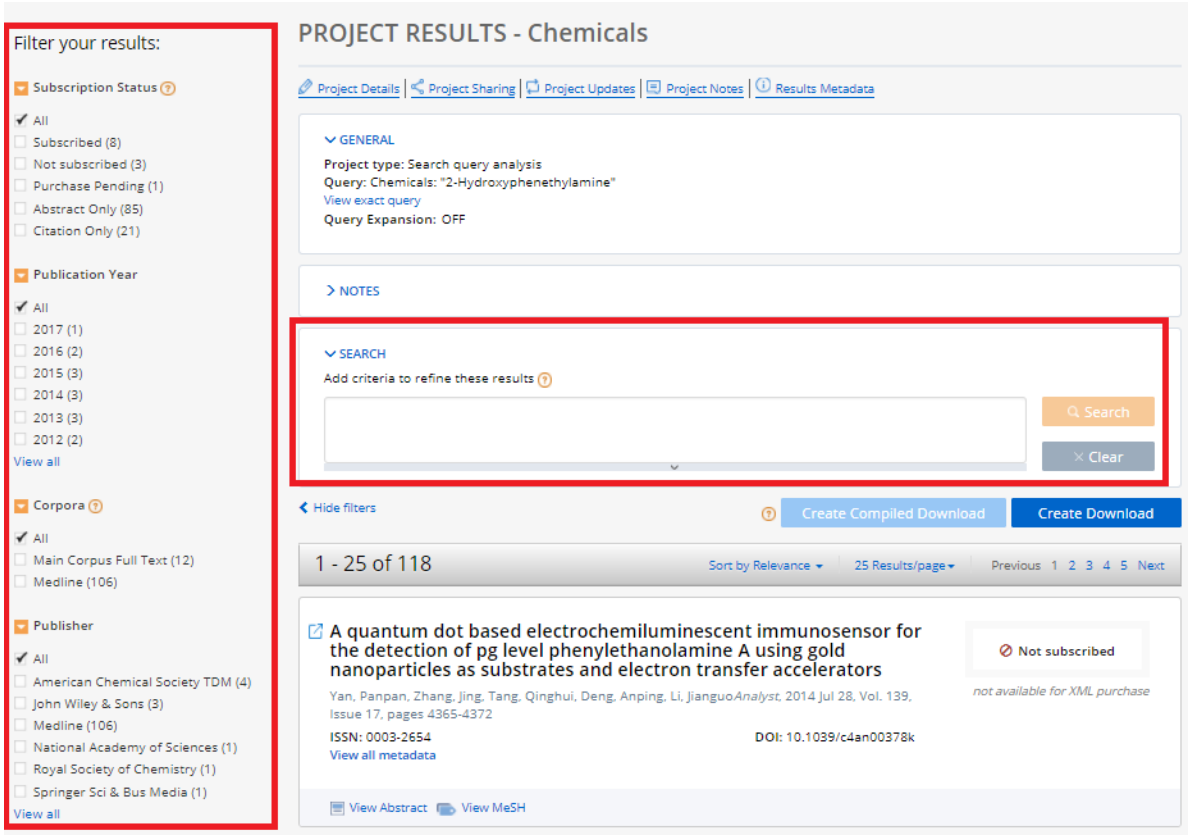
After you have submitted project, you can only modify the project query by cancelling the project processing. You can cancel processing on the 'My Projects' page.

Define a Corpus

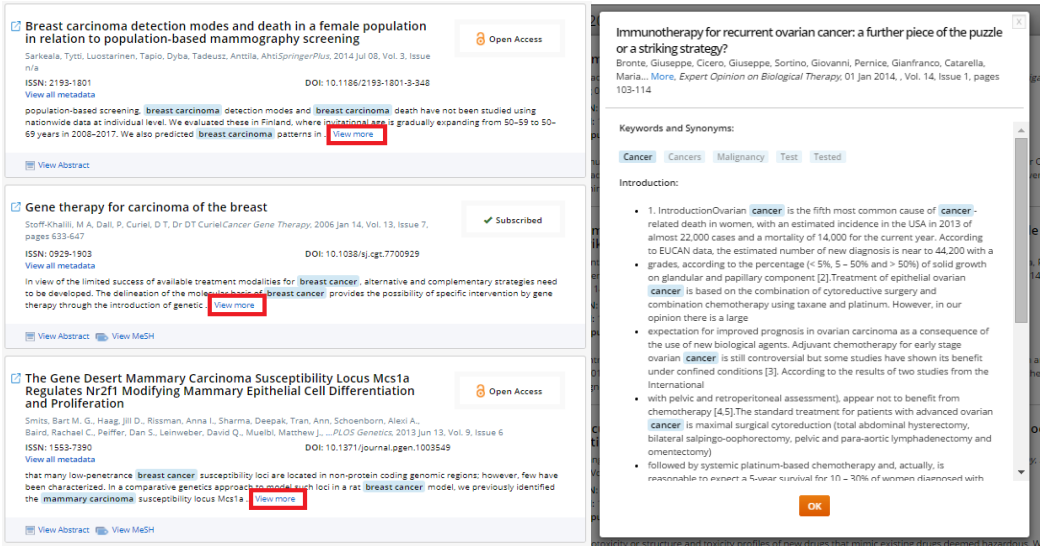
You will receive an email when your project results are ready to review. From the 'My Projects' page, click on the name of your project to go to the 'Project Results' page. Here you can further examine and filter your results to define a final corpus for your project and download or export the collection into a text mining tool.

On the 'Project Results' page you can use the 'Subscription status,' 'Publication years,' 'Corpora,' and 'Publisher' filters in the left column to further narrow your results.

You can also enter secondary search criteria using Lucene query syntax to further refine your search results. For more information on Lucene syntax, click here.

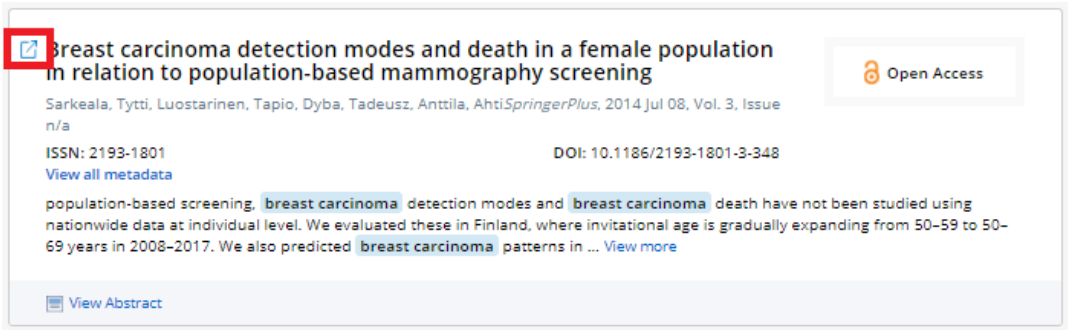


You can determine the relevance of particular results by examining the article title and associated metadata (including MeSH tags, when available). Click the 'View more' link for a given article to see examples of keywords, in context, from different article sections.



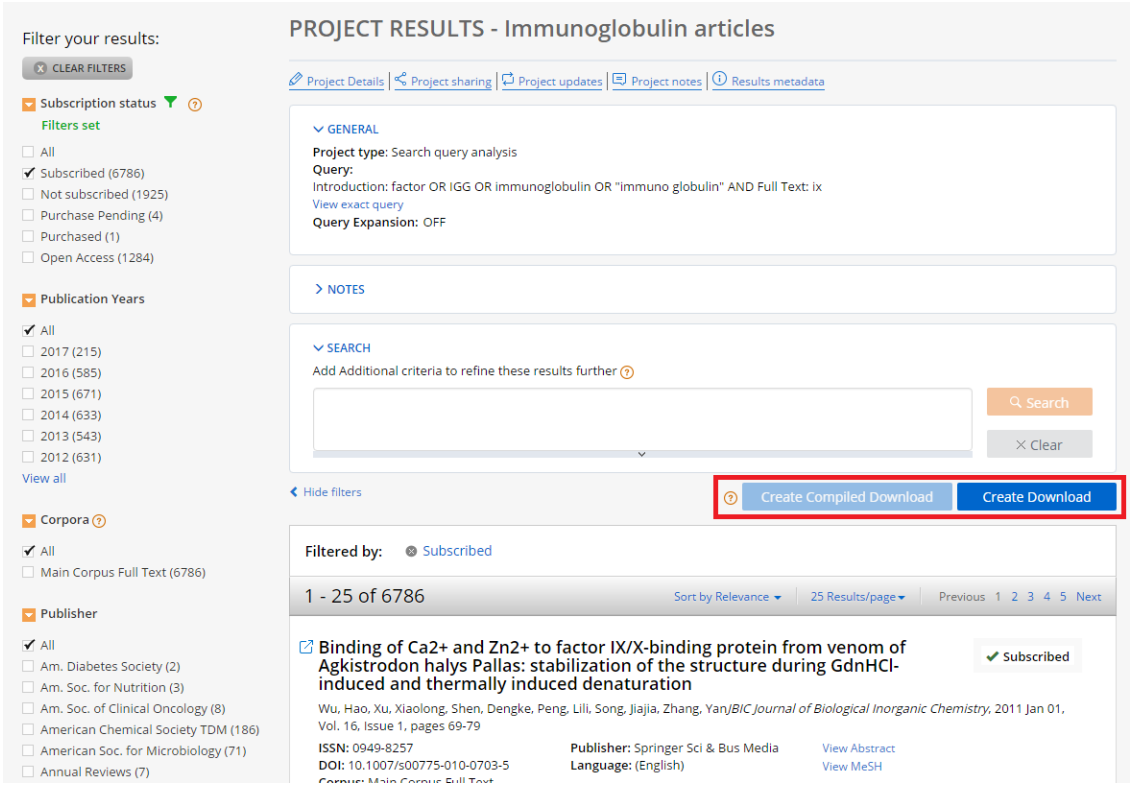
Within the ‘view more’ window, click on each keyword to see highlighted examples of the word in context. Note that all keywords and synonyms used in the final query are available here for you to examine.

If your organization has been configured with the RightFind Enterprise OpenURL integration, you can also path from XML for Mining into your RightFind Enterprise instance, to fulfill a human readable (PDF) version of the article from subscriptions, open access, library copies, or document delivery. Click the ‘linkout’ icon next to the article title.



Export a Corpus from RightFind XML for Mining

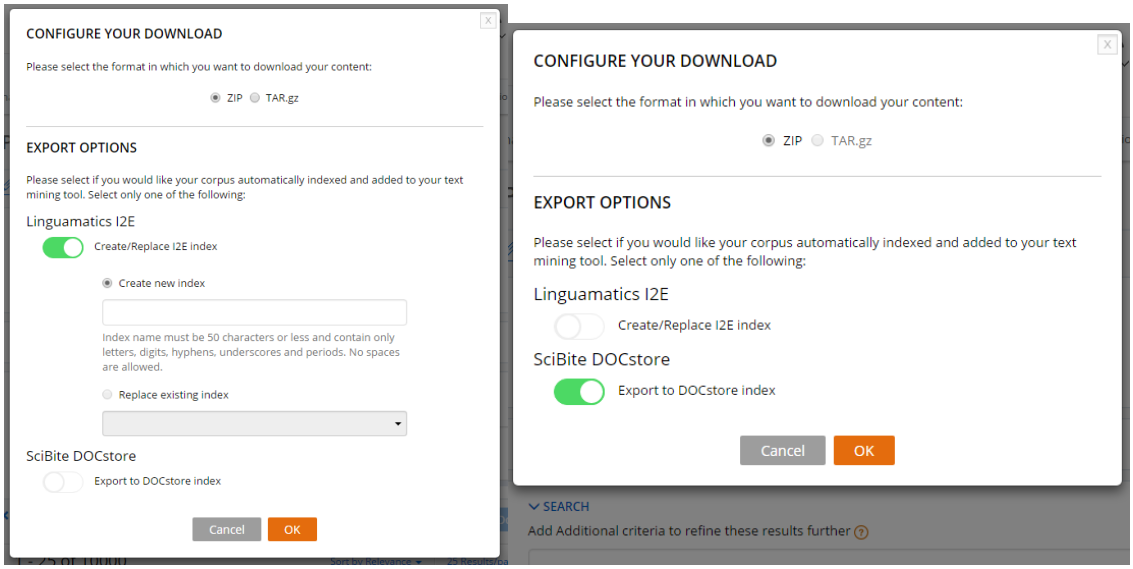
After filtering your project results to a desirable corpus, click ‘Create Download’ to begin downloading the corpus for import into your text mining software. You can also export your corpus directly into your text mining software.



XML for Mining will ask you to confirm your download. If you have any unsubscribed articles in your project results for which full-text XML is available, you will be asked whether you'd like to purchase those articles.

After confirming your download count, select the format of your download.

If your instance of XML form Mining has been configured to integrate with SciBite DOCstore or Linguamatics I2E software, you can adjust your export settings to directly index your corpus into these applications.



You will receive an email when your download and/or index is ready.

The confirmation email contains a link enabling you to immediately download your XML articles. You can also click the 'Download' icon for any project in a 'Ready for download' state.

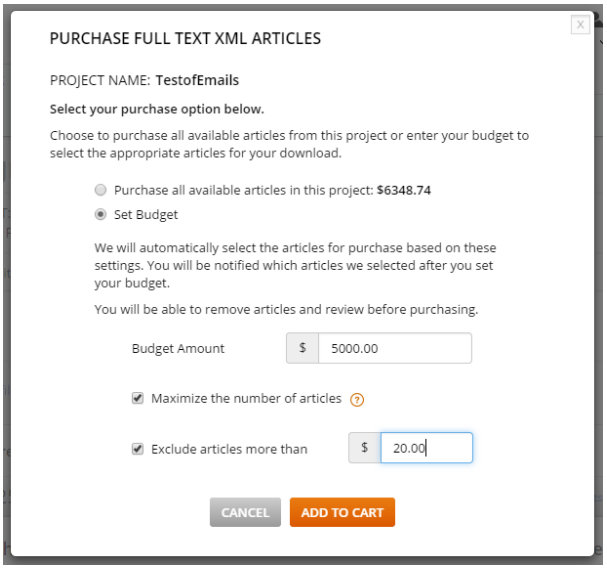
When downloaded, each article will be a single XML file.

Purchase Unsubscribed Articles

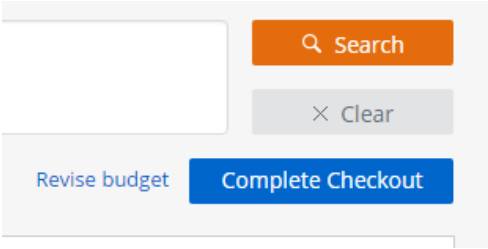
When you attempt to download or export project results that include articles to which your company doesn't subscribe, XML for Mining enables you to purchase the full-text XML articles for an additional fee. If you opt not to purchase the articles, you can proceed to download the article title and abstract at no charge; abstract-only downloads are also not counted against your annual download cap.

You may purchase all unsubscribed articles (total purchase price shown), or you can set a budget cap on the purchase. If you choose to set a cap, you will be able to set preferences for applying a budget. You can choose to maximize the number of articles or exclude articles above a certain price.

If you set a budget and make no other selections, XML for Mining will add articles to your cart until it hits your budget amount, in relevancy order.

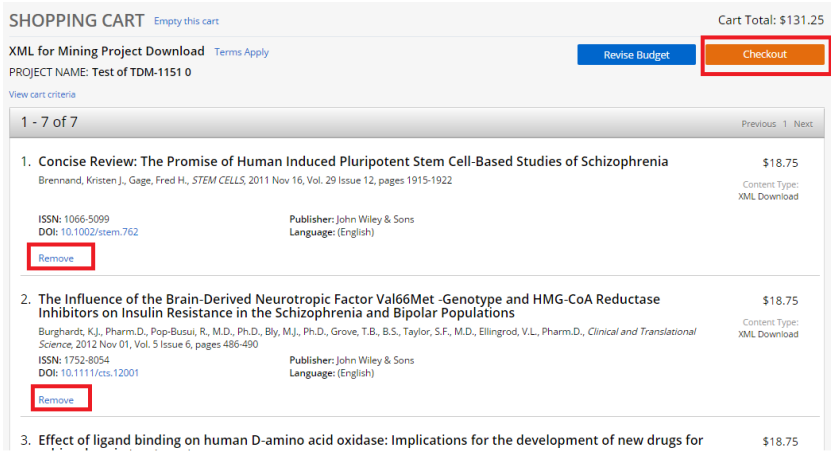


Click 'Add to Cart' to add XML articles to your shopping cart that match your preferences. Your project will be in "Adding to cart" status and you will receive email notification when your cart is ready for review. To checkout, proceed to the 'Project Results' page and click 'Complete Checkout.' You can also choose to re-load you cart by revising your budget.



Before you checkout, you will have the opportunity to review the items in your cart and remove them individually.

Click 'Checkout' to complete the article purchase process. Depending on how your organization is configured, you will be asked to confirm your order (in the case of invoice configuration) or you will need to enter a credit card before you confirm. Organizations can also have their accounts configured to require administrative approval prior to checkout.



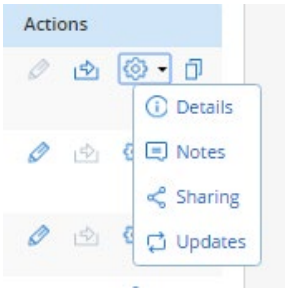
After confirming your order, the items in your cart will be processed. You will receive email notification when this is done. Proceed to the 'Project Results' to obtain your download.

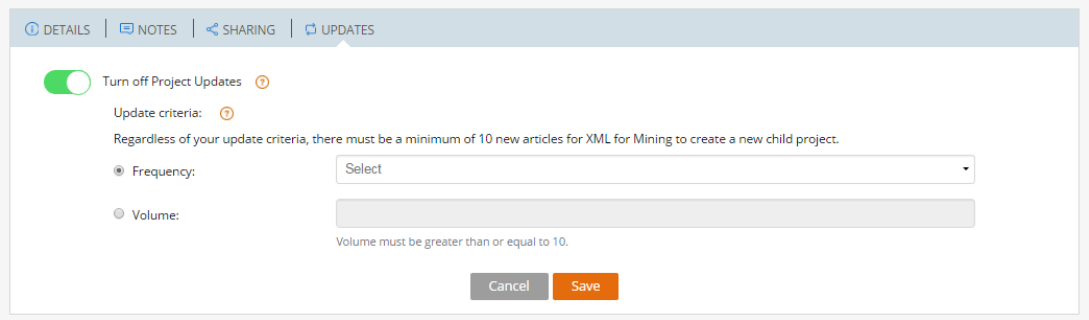
Configure Project Updates

After a project is completed, you have the opportunity to configure settings for project updates; this allows you to automatically receive notifications and create projects when new articles are loaded to XML for Mining that match your original search criteria.

Note: Project updates can be configured only for Search Query Analysis or Lucene Query projects. If you create a project using the Nearest Neighbor or Article ID methods, you cannot receive automatic updates.

After your project has completed, click on the 'Details & Settings' gear icon and navigate to 'Updates':





Here you can define the parameters for your project updates:

- Turn on project updates:
  - Frequency: New projects are created on a frequency of your choice (weekly, monthly, quarterly), as long as new articles satisfying your original search criteria are present.
  - Volume: New projects are created when the number of new articles matching your original search criteria meets your volume threshold.

Note: No matter which project update criterion you select, there must be at least 10 new articles matching your original search criteria in order for a new project to be created.

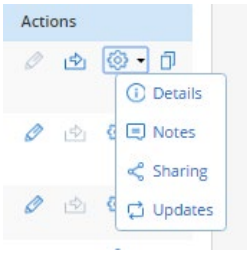
After you have configured your project update settings, click 'OK'.

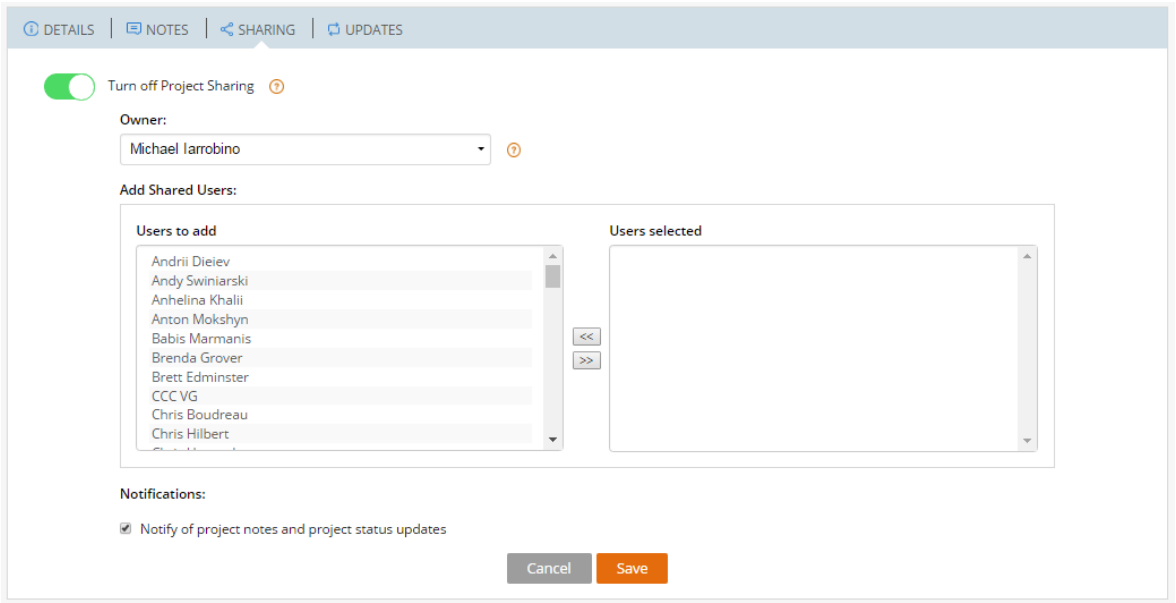
When your update criteria are satisfied, a new project (the "child" project) will be created automatically and you will receive an email notification. You can manually download results from your new child projects or compile all parent and child project results into one download export.

Share Projects

When a project is saved, you have the opportunity to configure settings for project sharing and can also add project notes; this allows you to work collaboratively with other named users of XML for Mining.

After your project is saved, click on the project settings icon and navigate to the 'Project Updates' tab:





Here you can define the parameters for your project sharing:

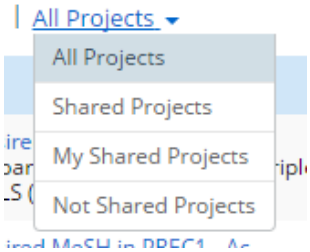
- Turn on project sharing:
  - Owner: The owner of the project is you by default. Only project owners can create downloads/exports from a project or purchase not subscribed articles. When you set another user to be the owner of the project, you automatically become a shared user.
  - Shared users: Specify users you would like to share the project with; these users will receive an email notification that you have shared the project with them. Shared users can modify project queries and filters but cannot create downloads/exports or purchase not subscribed articles.

- Notifications: Select whether you would like to receive emails about new notes on their project, or changes to project status. Regardless of this setting, you will receive notifications when sharing configuration settings have changed for the project.

Projects will show in your My Projects page with new values under the 'Owner' column, as follows:

- Not shared. Project sharing is turned off for this project.
- Your username. Project sharing is on for this project and you are its owner.
- Another username. Project sharing is on for this project and the referenced user is its owner; you are a shared user.

You can filter the list shown in My Projects using the drop-down menu at the top of the list:



Quick Reference

Additional Methods to Create a Project Corpus

Nearest Neighbor Analysis

Nearest neighbor analysis enables you to upload document(s) of interest to retrieve similar articles. To do this:

1. Select 'Nearest neighbor analysis' on the 'Create Project' page.
2. Enter the following parameters:
    a. *Similarity factor:* Set the similarity factor 'close' to ensure your retrieved corpus us more similar to your uploaded document(s). Set it 'distant' to decrease the similarity.
    b. *Maximum number of documents:* Select a number of documents to be returned, in increments of 1,000. NOTE: You may receive fewer documents than specified if your similarity factor is set very 'close.'
3. Upload documents. You can upload one or more TXT, WORD, PDF, or XML files.
4. When you are satisfied with your file upload and your project criteria, click the 'Submit' button to retrieving your article results.

Search Query Analysis

Search query analysis enables you to specify a traditional Boolean query. For more information on creating a project using search query analysis, see the examples in the 'Create a Project' section.

Article ID List

The article ID list retrieval method allows you to provide DOIs or PMIDs obtained from an external search service. To do this:

1. Select 'Article ID list' from the 'Create Project' page.
2. Select the way in which you'd like to provide your list:
    a. *Upload list.* Click the 'Choose File' button to select a .txt file from your local computer with each DOI or PMID on a separate line. There is no limit to the number of IDs that can be matched using this method.
    b. *Enter list.* Enter a new DOI or PMID on each line of the text input box.
3. When you are satisfied with your article ID entry or file upload, click the 'Submit' button to begin retrieving your article results.

Lucene Query

The Lucene query project type allows you to create projects based on complex but flexible queries, using

any indexed field and combining clauses using unary or binary logical operators in Lucene syntax. You can find more information about Lucene syntax and index fields here: https://rightfind.copyright.com/rs-ui-web/build/dist/assets/lucene-query-project-tip.pdf

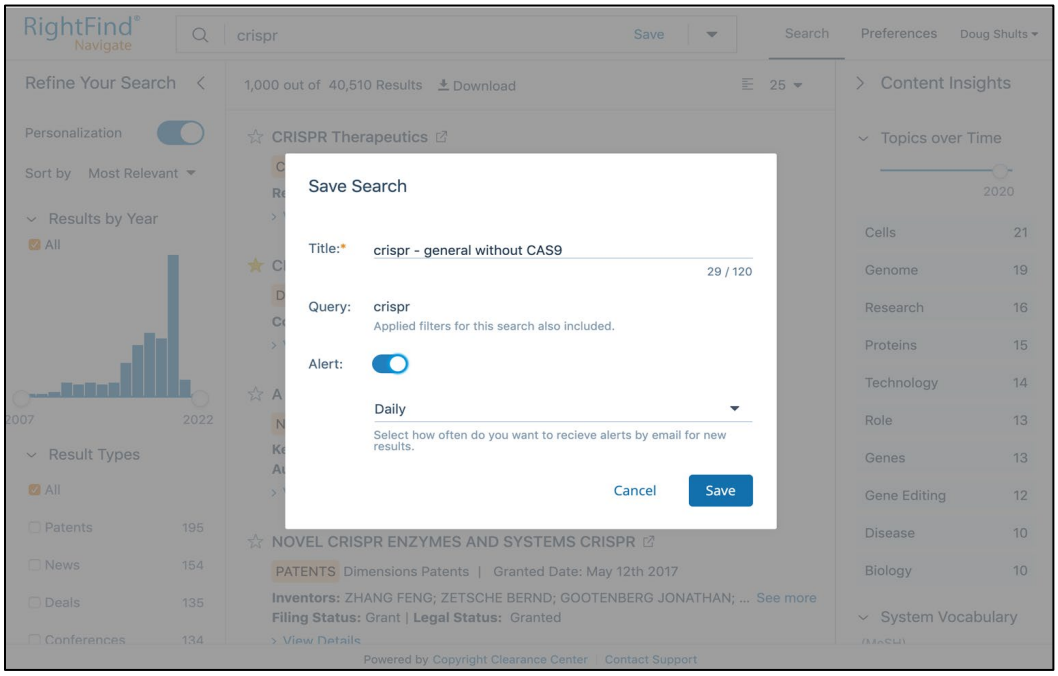
Article Subscription Status

Article Subscription Status	Description
Abstract Only	XML for Mining does not have access to the full-text article. You will receive the abstract and metadata.
Citation Only	XML for Mining does not have access to an abstract or the full-text article for this record. You will receive the article metadata.
Subscribed	Article is in a journal that is within your organization's holdings. When downloading, you will receive the full-text XML.
Not Subscribed	Article is in a journal that your organization does not subscribe to. When downloading, you will receive the abstract and metadata, or can purchase the full-text XML separately.
Open Access	Article is open access (CC-BY). When downloading, you will receive the full-text XML.
Purchase Pending	Article is being loaded to or is in a shopping cart that hasn't yet checked out. When downloading, you will receive the abstract and metadata, or can complete purchase of the full-text XML.
Purchased	Article has been purchased previously by you or someone in your organization. When downloading, you will receive the full-text XML.

Project Status

Project Status	Description
Open	Project name, description, and query are saved but not submitted
Submitted	Project name, description, and query are saved and request for full text has been submitted
In Queue	Project name, description, and query are saved and request for full text is in the queue to be processed

Project Status	Description
Execution	The system is performing search activities or clustering algorithms, depending on the project type.
Processing	Project name, description, and query are saved and request for full text is in process
Completed	Project name, description, and query are saved and request for full text has been completed
Ready to Download	A download is available for this project.
Adding to Cart	XML downloads are being added to a shopping cart for purchase.
Cart Ready	Your shopping cart is ready to be checked out.
Purchase Pending	You have a purchase that is pending; no additional purchases can be made.



In addition to individually scheduling email alerts for your saved search queries, you can also schedule a digest which organizes and collects results across all your saved searches and sends an update in a single digest email alert.