

# RightFind<sup>®</sup> XML for Mining

## Frequently Asked Questions

### Software FAQs

#### **What does RightFind XML for Mining do?**

Researchers struggle to gain access to full-text articles for mining. When they do get the full-text they must contend with multiple formats and inconsistent license terms — all of which inhibit text mining efforts. To address these issues, Copyright Clearance Center (CCC) created RightFind XML for Mining, which provides a single source of full-text XML content for text mining researchers.

CCC's RightFind XML for Mining solution enables you to quickly identify and download a collection (or corpus) of full-text articles from multiple publishers through a single source. Content and metadata is normalized to make it easy to integrate XML-formatted content into your existing text mining workflow and tools. You can access RightFind XML for Mining through a web-based user interface or a RESTful API for workflow integrations and custom application development.

#### **How does the service work?**

CCC collects content in XML format from multiple publishers and normalizes the content and metadata to the JATS standard format for mining. For more information on JATS, visit <https://jats.nlm.nih.gov/>. RightFind XML for Mining enables you to research the full-text of all the content in the solution to identify an article corpus for mining. Results are checked against company subscriptions and Open Access and the full-text content of the subscribed articles is available for download along with the abstracts and metadata from those materials to which your company is not subscribed.



### **How does CCC know whether my company subscribes to a particular journal?**

The librarian or information manager at your company stores credentials for the various publisher platforms in RightFind XML for Mining. The product uses that information to automatically obtain your company's subscription information from the publisher's websites on your behalf. RightFind XML for Mining automatically downloads and updates subscription holdings from the publisher's website and constructs an A-Z title list to ensure copyright compliance.

### **Does RightFind XML for Mining allow full-text downloads from the collection of available articles?**

Your company's holdings information determines whether you are subscribed to a particular article within each project. Downloads of not subscribed articles contain only metadata, including abstracts, and do not contain the full text of the article. Downloads of subscribed articles will contain the full text.

## **Licensing FAQs**

### **What is a unique download?**

A unique download refers to an individual full-text article contained in a download archive (compressed in .zip or tar.gz format) that has not been included in any previously-created download archive.

Download archives are created when a user selects "Create Download" from the Project Results page.

### **If I download an item twice, is it counted?**

No. An article previously downloaded by anyone in your organization does not trigger another charge for your organization. The XML for Mining system "remembers" you have previously downloaded the article and includes it in your new download archive so you can mine against it without an additional charge.

### **I received an email that my download archive is expiring. What should I do?**

If you have not downloaded the content, follow the link in the email notification to download the archive file. Download archives are available for 90 days after the creation date. After that time, we remove the archive from our system and you must create a new archive.



### **Can clients purchase full-text XML-formatted articles to which they are not subscribed?**

Yes. Clients can either download metadata and abstracts at no cost for unsubscribed articles, or purchase the full-text XML article.

### **If a client purchases unsubscribed articles for one project, must they purchase them again for use on other projects?**

No. After a user has purchased an article, that article is available in full-text XML to all other users and for all other projects at no additional charge.

### **How can I use downloaded content that is subscribed, unsubscribed, or purchased?**

You can make the following uses of the XML content obtained through RightFind XML for Mining:

- Store the XML files internally for the term of the RightFind XML for Mining agreement, and create and store up to two additional copies solely as reasonably necessary to text mine the content.
- Perform text and data mining across the content using your own tools or third party software.
- Integrate the results of their text mining into an internal database for internal business purposes.

You may not:

- Externally distribute or make available any copies of the content;
- Create, license, sell, distribute, or otherwise make available a database or other product comprised, in large part, of their text mining output intended for use by persons outside your organization;
- Use downloaded content to create a library or collection to substitute for subscriptions or purchases of the content that your organization ordinarily relies on;
- Reformat the downloaded content into PDF, EPS, DOC or other formats intended for reading, or for uses other than text and data mining.

Note: The above terms do not apply to Open Access (OA) articles



obtained through the RightFind XML for Mining solution, which in each case are subject to the specific terms of the relevant OA license.

#### **How can I use content marked as Open Access?**

Articles identified as "Open Access" are subject to the terms of the specific Open Access (OA) license under which the rightsholder has made the article available. The RightFind XML for Mining license terms and conditions do not apply to such OA articles. However, downloads of OA articles will count against the number of downloads available to your organization under the RightFind XML for Mining agreement.

Open Access articles included in the RightFind XML for Mining service are made available under a Creative Commons CC-BY or similar license that allows for commercial use of the content.

#### **Are downloads of Open Access content counted as unique downloads? Why?**

Yes, downloads of Open Access articles are counted as unique downloads. There is a cost involved in storing, delivering, normalizing, and indexing these articles. Because of these costs we charge the organization. The same is true for all full-text article types. Users have the option to filter out Open Access content if they choose to reserve downloads for non-OA full-text content.

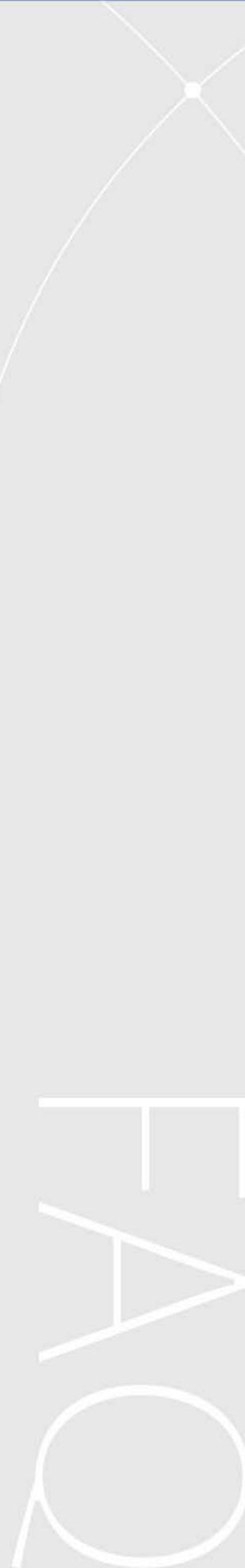
#### **If I find an article in RightFind XML for Mining that I want to read, can I purchase a PDF?**

Assuming you do not have a subscription, you can purchase readable versions of articles from the publisher site or from within CCC's RightFind® content workflow solutions. RightFind XML for Mining clients who also use RightFind Enterprise can be configured to fulfill human-readable versions of your project results within the Enterprise platform via an OpenURL icon. You can access content from existing subscriptions, Open Access, or library copies, or acquire articles through document delivery, all while contributing valuable usage metrics to your organization's Enterprise instance.

#### **Integration FAQs**

##### **How does RightFind XML for Mining work with text mining software like Linguamatics I2E?**

XML for Mining is specifically designed to allow you to access and obtain



machine-readable content formatted in XML for loading into text mining software such as Linguamatics I2E. CCC uses the JATS format for its XML files, enabling text mining software to easily ingest the content from RightFind XML for Mining. In cases when we have an integration partnership with a text mining software product (for example, Linguamatics I2E), users can easily export their corpus of articles directly into the text mining tool.

### **What is SciBite DOCstore?**

SciBite DOCstore is a powerful semantic search application that enhances the discovery and retrieval of scientific information. Users can interrogate scientific literature at scale and uncover new relationships previously hidden in unstructured text using a simple user interface or RESTful API. The integration of RightFind XML for Mining with DOCstore enables:

- Automatic indexing of semantically enriched full-text content into DOCstore from CCC's database of millions of articles from more than 50 publishers.
- Storage and management of article content for all literature-based semantic enrichment projects, with minimal administrative effort.
- Semantic search of full-text content with entity- and class-based sentence and document queries.

### **How does RightFind XML for Mining work with SciBite DOCstore?**

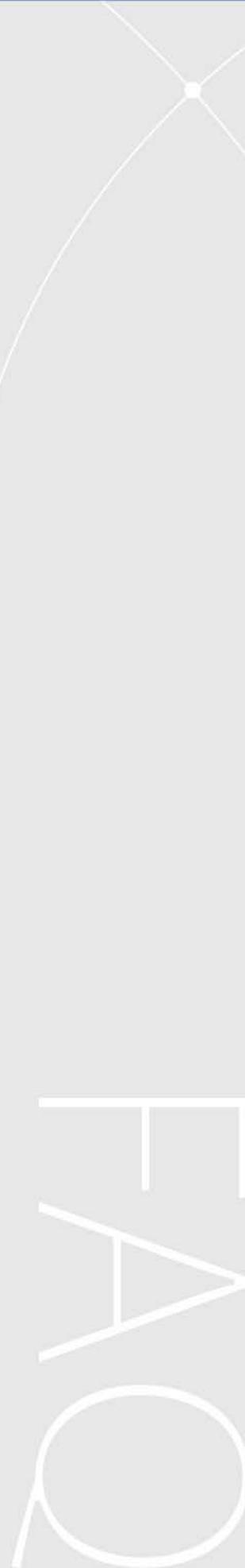
Users create a corpus of normalized, full-text XML articles in RightFind XML for Mining. With a single click, users can semantically enrich the corpus to be queried within SciBite DOCstore, using a flexible user interface or RESTful API.

### **What Is semantic enrichment?**

Semantic enrichment helps users find new ideas and concepts by tagging unstructured text with information about its meaning. This process of tagging makes content more discoverable and increases relevancy in search results.

### **What are the benefits of pairing RightFind XML for Mining work with text mining software like SciBite DOCstore?**

The integrated solution provides a broad corpus of normalized full-text content and a fully hosted semantic workflow with simple yet robust interfaces – reducing administrative overhead and enabling quicker time to value.



- **Discover vital data and assertions.** Through a broad corpus of normalized full-text content, unearth connections that can be found only in full text, and synthesize knowledge using powerful semantic search.
- **Save time and money.** Spend more time on analysis and discovery and less time wrangling data sources and implementing technology. The fully hosted RightFind XML for Mining with DOCstore solution reduces time and costs associated with article licensing, content management, index creation and administration.
- **Simplify copyright compliance.** All content in RightFind XML for Mining is pre-authorized for commercial text mining. This means peace of mind that text mining projects comply with copyright and minimize infringement risks.

**If I create an I2E or DOCstore index, are the indexed articles counted as a download?**

Yes. Articles indexed in I2E or DOCstore are first put into a download archive, at which time we count the unique downloads and your organization is charged for these.

