RightFind™ XML for Mining

Copyright Clearance Center

Frequently Asked Questions

1. What does 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 CCC created the XML for Mining solution which provides a single source of XML full-text content for text mining researchers.

CCC's XML for Mining solution enables researchers to quickly identify and download full-text article collections from multiple publishers through a single source. Content metadata is normalized to make it easy to integrate XML content into your existing text mining workflow and tools. Users can access a web-based user interface or a robust RESTful API for workflow integrations and custom application development.

2. How does the solution work?

CCC collects XML content from multiple publishers and normalizes the metadata to the JATS standard format for mining. XML for Mining enables researchers to search the full-text of all the content to identify an article corpus for mining. Results are checked against company subscriptions and the full-text content of the subscribed articles are available for download along with the abstracts and metadata from those materials to which the company is not subscribed.

CCC uses an automated means to obtain your company's subscription information from the publisher websites. The solution automatically downloads subscription holdings (A-Z title lists) from publisher websites on your behalf to ensure copyright compliance.

3. 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 our system. The software uses that information to automatically obtain your company's subscription information from the publisher's websites on your behalf. The solution automatically downloads and updates subscription holdings from the publisher website and constructs an A-Z title list to ensure copyright compliance.

4. Does the solution allow full-text downloads from the collection of available articles?

Your company's holdings information determines whether a particular article is subscribed or unsubscribed within each project. Downloads of unsubscribed articles contain only metadata, including abstracts, and do not contain the full text of the article. Downloads of subscribed articles contain the full text. Each organization is limited to a maximum number of total unique full-text downloads based on the usage plan they choose.

5. How does this solution work with text mining software like Linguamatics I2E or IBM Watson?

XML for Mining is specifically designed to allow users to access and obtain machine readable content formatted in XML for loading into text mining systems such as Linguamatics I2E or IBM Watson. CCC uses the JATS format for its XML files, enabling mining tools to easily ingest the content from our system. For more information on JATS, visit http://dtd.nlm.nih.gov/. Each downloaded article contains normalized metadata and the structured XML full text from the publisher.

6. How can I use the downloaded content?

You can make the following uses of the content obtained from CCC's XML for Mining solution:

- Store the XML files, 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 your 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 substantially comprised of your text mining output intended for use by persons other than yourself;
- Use downloaded content to create a library or collection to substantially replace the need to subscribe to or otherwise obtain a particular item of content;
- Reformat the downloaded content into PDF, EPS, DOC or other formats intended substantially for reading or for uses other than text and data mining.

7. Can I purchase XML full-text of articles I am not subscribed to?

Currently you can only download metadata and abstracts for not-subscribed-to articles. Users will be able to purchase a license to download full-text XML of articles they are not subscribed to in a future release of the solution.

8. If I find an article in the XML for Mining solution 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 solution.