

Chemical intelligence  
in text mining

Find, manipulate and  
analyze chemicals in  
documents

Survey the IP  
landscape or mine  
internal documents  
for chemical entities

## ChemMining Collection

Do you manually process patents to find which molecules have been claimed? Or, would you like to mine documents to find molecules of interest, as well as other molecules, chemical entities and biological or disease keywords? Do you know what molecules exist in internal documents and the context of those documents?

You can save time, survey the IP landscape and avoid duplicating research efforts by using ChemMining to perform chemically-intelligent text mining of public and internal documents, thereby learning more about your structures of interest.

The ChemMining Collection lets you find chemical names in text documents and convert them to live chemical structures ready for further manipulation. Explore a few documents at once and create a report showing the structures found, or process a large set of documents and create a searchable database of structures found. Combine your chemical structure searches with textual searches for key terminology, such as terms related to disease processes or biological molecules, to perform true context-specific, chemically-intelligent searching.

### Create flexible workflows for chemical text mining

With the ChemMining Collection for Pipeline Pilot, you can create workflows that:

- Search PubMed abstracts and patents to find chemical structures present in the documents
- Process internal documents and other text to find chemical names, including custom chemical names and corporate IDs

- Create reports showing the structures found, as well as other terms of interest, such as diseases and biological molecules

- Create databases of structures found in abstracts, patents, and internal documents, attaining a comprehensive picture of published and private chemical structures

- Create flexible workflows that combine literature and chemistry analyses

The screenshot displays the Pipeline Pilot Webport interface. On the left, there are search parameters including 'Document Source(s) to Search', 'Enter a structure query' (with a chemical structure of a benzimidazole derivative), and 'Enter optional text queries'. The main area shows search results for 'Bromocriptine' and 'serotonin'. The 'Bromocriptine' section shows 'Results 1 to 1 of 1 for Bromocriptine (all variants)' with a detailed abstract snippet. The 'serotonin' section shows 'Results 1 to 1 of 11 for serotonin (all variants)' with a detailed abstract snippet. The interface includes navigation buttons like 'Back', 'Forward', and 'Search', and a 'Local internet' indicator at the bottom right.

With the ChemMining Collection, you can search for documents containing a structure of interest, highlighting other chemicals and keywords as well.

### Explore the IP landscape

Published, open access document sources, such as PubMed abstracts, US Patents, and European Patents represent hugely valuable assets containing a large amount of critical chemical information. However, unlocking the value of these resources can be difficult due to the amount of information available, and the complexities and subtleties of extracting the information in an automated process. The powerful chemical text mining capabilities offered by the ChemMining Collection will help you quickly explore the IP landscape.

### Avoid duplication of research efforts

Most organizations have large repositories of internal documents containing critical information from previously conducted research. However, that data is often not effectively used, typically leading to duplication of effort, as the same experiments are conducted over and over again by different groups, at different times. The ChemMining collection can help organizations avoid duplication of efforts by making it easier for researchers to access data about work that has already been conducted on structures of interest.

### Search using popular chemical naming conventions

With the ChemMining Collection, you can build customized workflows to process external and internal documents and extract chemical names and other relevant information. ChemMining contains components for processing documents and recognizing a variety of chemical names, including:

- IUPAC names
- SMILES strings
- InChI strings
- Common chemical names (such as “aspirin” and “ibuprofen”)
- Internal custom names and corporate IDs

### Convert results into live chemistry objects

The ChemMining Collection integrates with third-party applications for converting chemical names to live chemistry objects. As a result, after processing one or more documents, you can create reports that show the document(s) highlighted with all the structures found, as well as enumerate the structures as live chemistry objects.

### Process large numbers of documents

With the ChemMining Collection, you’re not limited to simply processing individual or small sets of documents. You can also process large numbers of documents to create a searchable structure database. This allows you to query by structure and retrieve documents matching a query, such as a query based on substructure or similarity.

### Combine chemical structure and textual searches

A critical feature of the ChemMining Collection is its ability to combine chemical structure searches with text searches powered by the Pipeline Pilot Text Analytics Collection. By combining your chemical structure searches with searches for other relevant terms within documents, such terms related to disease processes and biological molecules, you can find highly relevant results.

### Prerequisites

- Pipeline Pilot Chemistry Collection
- Pipeline Pilot Text Analytics Collection
- Integration and Reporting Collections (strongly recommended)