

CHEMMINING COLLECTION

特許請求の対象となっている分子の検索処理を、手作業で行いますか。それとも、分子や化学物質、生体または病気に関するキーワードの検索を、文書のマイニングで行いますか。内部文書やその情報の中に記載された分子が何であるかわかっていますか。

ChemMiningを使用すると、科学的な知識に基づいて公開文書や内部文書をテキストマイニングでき、構造の詳細について知ることができます。このため、時間の節約が図れ、知的所有権(IP)の状況を調査して重複した研究を回避することができます。

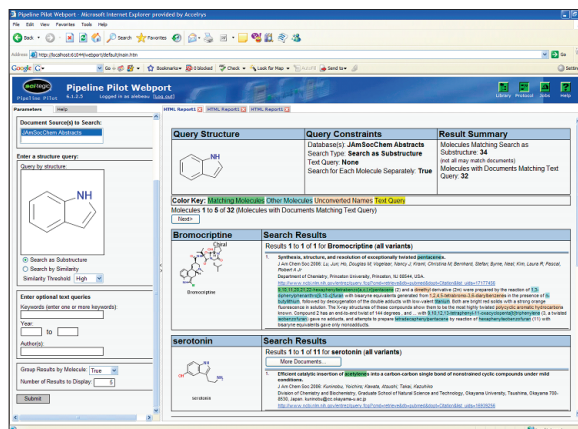
ChemMining Collectionでは、テキスト文書の中に化学名を見つけると、実際の化学構造に変換します。瞬時に複数の文書から検出された構造を表示するレポートを作成したり、大量の文書を処理して検索可能なデータベースを作成します。疾患プロセスや生体分子に関連する用語などの重要な専門用語に対して、化学構造検索とテキスト検索を組み合わせ完全に文脈に依存する、化学的な知識に基づいて検索を行います。

化学物質のテキストマイニングに対する柔軟性の高いワークフローの作成

Pipeline PilotのChemMining Collectionを使用すると、以下のワークフローを作成できます。

- ・ PubMedの抄録や特許を検索して文書内に書かれた化学構造を検索
- ・ 内部文書などのテキストを処理して、個別の化学名や企業IDを含む化学名を検索
- ・ 検出された構造や、その他の関心のある項目(病気や生体分子など)を表示するレポートを作成

- ・ 抄録、特許、および内部文書で検出された構造データベースを作成し、公開済みおよび個人の化学構造を表す包括的な画像を取得
- ・ 文献と化学的性質の分析を組み合わせた柔軟性の高いワークフローを作成



The screenshot shows the Pipeline Pilot Webport interface. On the left, there is a search input field with a chemical structure of a benzimidazole derivative. The main area displays search results for 'Bromocriptine' and 'serotonin'. The 'Bromocriptine' results show a list of documents with their titles and authors. The 'serotonin' results show a list of documents with their titles and authors. The interface includes various filters and options for refining the search results.

ChemMining Collectionを使用すると、構造を含む文書や、他の化学物質やキーワードを強調している文書を検索することができます。

知的所有権 (IP)の調査

PubMedの抄録、US Patents、European Patents など、公開され誰でも利用できる文書の情報源は、大量の重要な化学物質情報を含んでいる非常に価値の高い資産です。しかし、利用できる情報量が大量であることや、自動処理での情報の抽出が複雑で緻密さが要求されるなどの理由で、これらを活用することが困難な場合があります。ChemMining Collection による強力なケミカルテキストマイニング機能を使用することで、知的所有権(IP)を短時間で調査することができます。

重複した研究の回避

ほとんどの組織では、以前に行われた研究で得られた重要情報を含む内部文書が、大量に集積されています。しかし、多くの場合そのデータは効率的に活用されないため、重複した作業となるのが一般的です。別のグループが別々のときに、同じ実験を何度も繰り返すこととなります。ChemMining Collectionを使用すると、以前に行われた研究の構造に関するデータを簡単に利用できるため、重複作業を回避することができます。

よく利用される科学物質の命名法の検索

ChemMining Collectionを使用すると、外部・内部文書を処理し化学物質名などの関連情報を抽出するワークフローをカスタマイズして作成できます。ChemMiningには、文書を処理して、以下のようなさまざまな科学物質名を認識するためのコンポーネントが含まれています。

- ・ IUPAC名
- ・ SMILES文字列
- ・ InChI文字列
- ・ 化学物質の一般名(「アスピリン」や「イブプロフェン」など)
- ・ 内部の個別名と企業 ID

結果を実際の化学物質オブジェクトに変換

ChemMining Collectionでは、サードパーティのアプリケーション統合として、化学物質名を実際の化学物質オブジェクトに変換します。その結果1つまたは複数の文書を処理した後に、検出されたすべての構造を強調した文書を表示したり、実際の化学物質オブジェクトとしてその構造を列挙するレポートを作成できます。

大量文書の処理

ChemMining Collectionは、個人的あるいは少量の文書を単に処理するだけの機能に留まりません。大量の文書を処理して、検索可能な構造データベースを作成することもできます。このため、基礎構造や類似性に基づくクエリーなど、構造によるクエリーを発行したり、クエリーに一致する文書を取得することができます。

化学構造とテキスト検索結果の結合

ChemMining Collectionの重要な機能は、Pipeline Pilot Text Analytics Collectionで行うテキスト検索結果と、化学構造の検索結果を結合することです。疾患プロセスや生体分子に関連する用語など、ドキュメント内にある他の関連項目の検索結果を化学構造の検索結果と結合することで、非常に関連性の高い結果を発見することができます。

要件

- ・ Pipeline Pilot Chemistry Collection
- ・ Pipeline Pilot Text Analytics Collection
- ・ Integration and Reporting Collections (強く推奨)

Pipeline Pilotの詳細については、次の URLを参照して下さい。

<http://accelrys.co.jp/products/pipeline-pilot>