

TEXT ANALYTICS COMPONENT COLLECTION

Text Analytics Collectionは、文献検索とテキストマイニングの機能をPipeline Pilot に加えるものです。デフォルトで様々な文献データベースや検索エンジンのコンポーネントを実装しています。単独で使用しても非常に強力なテキストマイニングのツールですが、ChemistryやBioinformaticsの解析フローと統合するときに、もっとも大きな効果を発揮します。インタラクティブなキーワード検索、大量の文献情報に対するテキストマイニングなど、どちらも対応可能です。

TEXT ANALYTICS COLLECTIONを使用すると、次のことが実現できます。

- ・ PubMed、パテント、NIHグラント、Google、あるいはローカルファイルの検索が可能です。
- ・ 語句、分野、類義語、ワイルドカードでの一致検索を単一のクエリー言語で実行できます。
- ・ 文書とオンライン文献の間にあるキーコンセプトあるいは相関性を見出すことができます。
- ・ オンライン文献からのアノテーションや関連付けを、既存のChemistryやBioinformaticsのプロトコルに対して加えることができます。
- ・ 新しいトレンド/相関性の発見、パテントの解析などを行い、また見出した傾向のトラッキングも可能です。

オンライン文書の検索

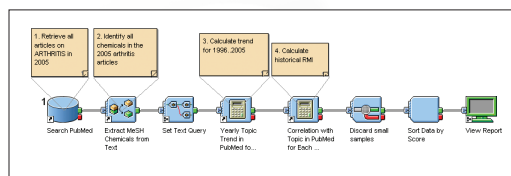
Text Analytics Collection は、オンライン文書リソースから情報を抽出する強力な武器となり、PubMed、米国パテント、Wikipedia、Yahoo (その他検索エンジンもユーザー追加定義が可能)などへアクセスできます。これらデータベースへの検索はインタラクティブなクエリーでも、あるいは沢山の文書を一括ダウンロード/解析するルーチン処理のように自動マイニングすることも可能です。

ローカル文書の検索

オンライン文書の検索と同様に、ローカル文書の検索も全く同様に実行可能です。Word、HTML、PDF、あるいはテキストファイルといったローカル文書をインデックスして検索することもできます。検索対象の文書タイプはユーザーが追加定義することもできます。また、オンライン検索の結果をローカル保存して、文書の高速取り出しやポスト処理に活用することも可能であり、ローカルデータベースのフォルダーの中身を自動的にモニターして追加・編集を更新し、常に最新状況を保持することができます。

解析結果へのアノテーション

Text Analytics Collectionを使用すると、結果データをクエリーとして文献データベースを検索して有用な関連情報を追加することが、非常に簡単にしかも全自動で行うことが可能です。例えば、配列解析コレクションにより遺伝子クラスター解析を行い、PubMedにてトップ引用されている



各遺伝子のサマリー情報をアノテーションする一連の処理を自動で簡単に実現できます。これにより、さらに深いレポート考察が可能となります。

新しいトレンドの抽出

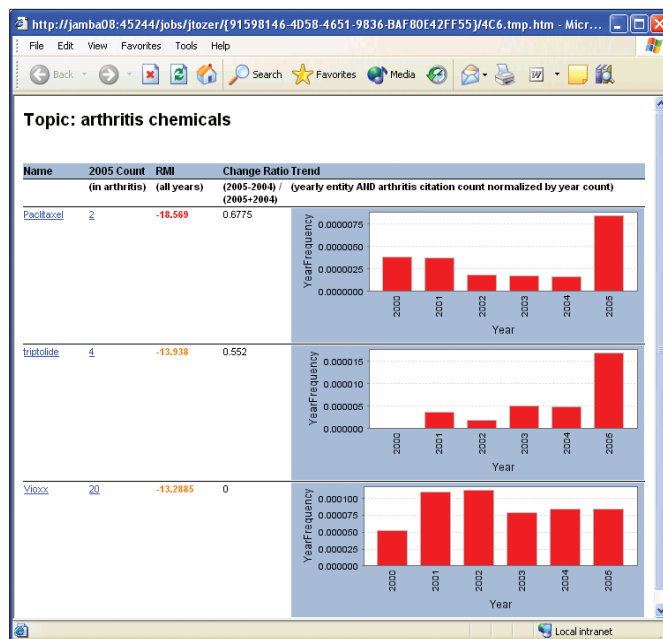
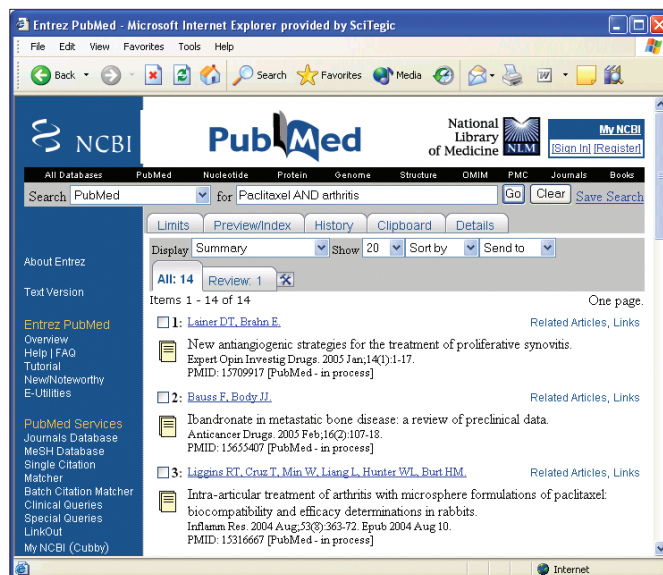
Text Analytics Collection を使用すると、調査したいトピックスについての文献動向のモニタリングが可能となり、新しいコンセプトの萌芽をアラートすることさえ可能です。具体的には、関心のあるトピックスの新しい文献を検索し、その中のコンセプトを表現しているキーワードを検出し、これらキーワードとトピックスの関連性解析を定点観測することで可能となります。これにより幅広いトピックスに関して新たなトレンドの立ち上がりモニターでき、それが周知の事実となる前にキャッチすることが可能となります。

パテントデータベースのマイニング

Text Analytics Collectionは、調査対象分野における研究トレンドの解析ツールにもなります。パテントデータベースをマイニングすることで、何件のパテントがあるのか、分野は何か、それはどの企業なのか、などを反映するトレンドを見出すことができます。たとえば、「fuel cells」の分野のパテントマイニングするプロトコルを作ることで、この分野が非常に早く立ち上がりつつある分野であることを知ることができます。また自動車企業がこの分野を支配していることや、ホンダとGMがイノベーションをリードしていることなどがわかります。

PIPELINE PILOTの概要

Pipeline Pilotは、さまざまな場所に保存されているデータから科学的価値を引き出し、科学的ワークフローを自動化して、より広範な科学コミュニティでのコラボレーションを促進することにより、研究開発組織の技術革新を支援する、拡張性に富んだ大規模サイエンティフィック・インフォマティクス・プラットフォームです。Pipeline Pilotのコンポーネントコレクションはプラットフォームの科学的な構成要素であり、科学的なカテゴリや機能でグループ化されています。コンポーネントをグラフィカルに組み合わせることで、データの取得、フィルタリング、分析、レポート作成のワークフローを作成できます。



Text Analytics Collectionにより文献やローカルファイルの中から着目する重要な文書を見出しキーとなるコンセプトを抽出し、トレンドや相関性から新たな知見を得ることが可能となります。

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>