

情報化学システム構築ツール

Daylight-MOE・KDEインテグレーションの紹介

本年4月より販売を開始したDaylight CIS社製 Daylightツールキットは、創薬研究支援のための強力なシステム構築ツール群です。お客様のニーズに合わせて柔軟で高速な化合物情報処理システムを構築できます。今回は弊社で取り扱っているCCG社製 MOEとInforSense社製 KDEを組み合わせたDaylightデータベースの使用例を紹介します。

◆ Daylightツールキット

Daylight製品はSMILESやフィンガープリント、THOR/Merlinというデータベースを扱うためのツールキットから構成されます(図1)。SMILESはDaylight社で考案された表記法なので、ツールキットによりその全機能を使用することができます。

これらのツールキットを自由に組み合わせることで独自の化合物情報システムを構築できます。従来のパッケージ型のデータベースシステムでは、細かなカスタマイズなどが実施しにくく、硬直なブラックボックスになりがちです。しかし、Daylightツールキットの場合は目的や使い勝手に合わせてシステムを柔軟に調整することが可能です。

しかも、Daylightツールキット最大の特長であるデータ検索スピードは、独自にカスタマイズを行っても損なわれることはありません。例えば、500万件程度の化合物エントリに対する類似構造検索も、1秒足らずで完了します(使用PC CPU: Intel Pentium 4 2.5GHz, メモリ: 1GB)。

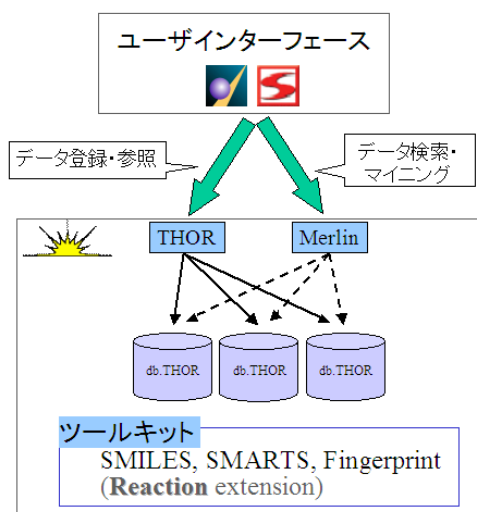


図 1 Daylight使用イメージ

Daylightツールキットはパッケージソフトのような簡便さよりも柔軟性を徹底的に重視します。このため、システム構築やメンテナンスにかかる時間と労力が懸念されるかもしれません。しかし、Daylightツールキットを大量のデータを扱う化合物データ処理エンジンとして活用しながら、CCG社製 MOEやInforSense社製 KDEなどのサードパーティ製品をインターフェースとして利用することでそれぞれの長所を活かした研究環境を構築できます。

◆ Daylight-MOEインテグレーション

CCG社製 MOEとDaylightツールキットを組み合わせることにより、MOEのGUIによる良好な操作性とDaylightの高性能なデータ検索を実現できます。操作性の優れたMOEパネルにより効率的に大量データを絞り込み、探していた化合物に素早くたどり着きます。

Merlinパネルでの検索が非常に高速なため、検索条件を変更すると、検索結果もリアルタイムで変化します(図2)。数値を入力したり、スライダーで値

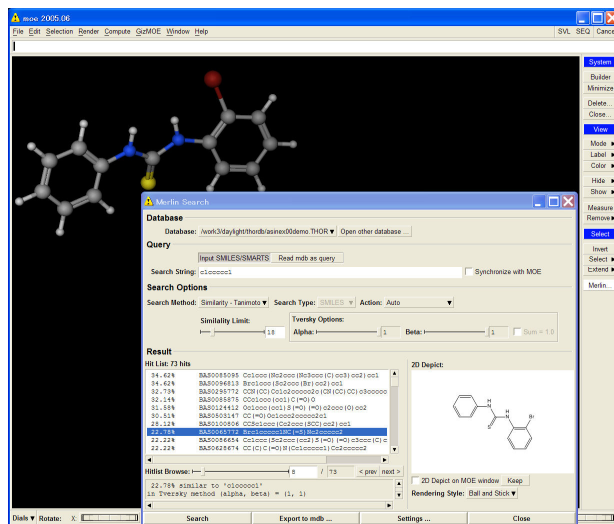


図 2 MOE 2005.06用 Merlinパネル

を変更すると、その動きに合わせて結果が反映されます。

検索の仕方も一つ一つのクエリー分子を入力するだけでなく、MDBファイル(MOEのデータベース形式)から一括して入力することも可能です。また、新バージョンのMOE 2005.06で搭載された分子二次元描画機能も積極的に利用しています。検索された化合物をSMILES形式や化合物名だけでなく分子構造として表示できるので、研究者の直感にあった使い方が可能になります。

Merlinパネルは通常のMOEパネル同様にSVL言語により記述されていますのでカスタマイズも容易です。機能を拡張したり使い勝手を簡単に向上できるのも、柔軟なDaylightツールキットとMOEならではの長特です。

◆ Daylight-KDEインテグレーション

InforSense社製 KDEとDaylightツールキットを組み合わせることで強力な化合物データマイニング環境を構築することができます。

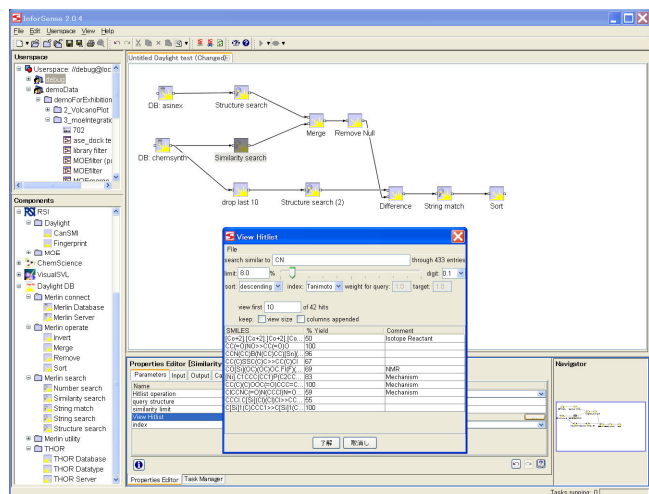


図3 InforSense KDEのMerlinワークフロー

KDEでは1枚のパネルではなく、ワークフローによって検索条件を表現します(図3)。このため、Daylight-MOEでは難しい複雑な検索条件の作成も簡単に実現できます。例えば、以下のような処理が可能になります。

- ・分子構造以外の(数値・文字列)検索
- ・複数のデータベースを利用する検索
- ・AND, OR, XORなどの複合検索

条件が複雑になっても、処理の流れが視覚的に確認できるので、内容の把握が容易です。むしろ、ワークフローを使い込んで処理が複雑になればなるほど、その重要性が認識できます。

また、試行錯誤の結果得られた検索条件・結果はもちろん重要ですが、その作業を行った目的や経緯などの方が重要な場合も多々あります。そのような場合でも、KDEは結果だけでなく試行錯誤の過程ごと保存できるので過去の知見を有効に活用できます。マイニングの過程が共有できる形で残るので、他の研究者が行った作業を参考にしたい時や、時間が経って忘れがちな処理内容を思い出したい時にも重宝します。これらはKDEとの組み合わせによりもっとも軽減される作業です。

KDEではDiscovery PortalというWebポータルモジュールも用意されています。このため、Daylightツールキットを使用するワークフローをWeb上に公開するのも容易です。複雑になってしまったワークフローでも、随時変更を加えたいパラメータのみを残して、独自のDaylightポータルを作成できます(図4)。

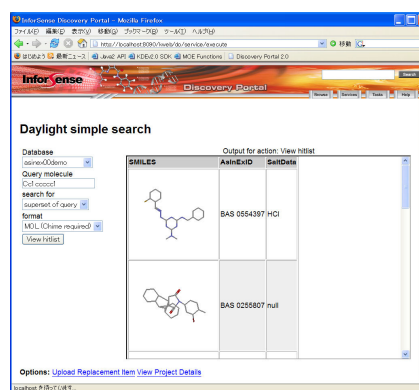


図4 Webに公開したDaylight用ポータル

◆ Daylightデータベース開発

Daylight-KDEのノードで可能な処理はMerlinデータベースへの検索が中心ですが、これに加えてTHORデータベース管理ノードも利用できます。このノードはデータベース作成の手間を軽減するツールとしても、あるいはデータベース構成管理環境としても活用することもできます。データベースを構築する過程や実行結果をワークフローとして保管できますので、構成管理の労力や時間も大幅に削減できます。