ワークフロー解析基盤技術の構築

○福井一彦, 田代俊行, 矢葺幸光, 浅井潔

●研究の目的

ライフサイエンス研究分野では、その細分化や専門化に伴い、様々なデータベース、解析ツール、予測ソフトウェアが開発され膨大な数に達しており、これらを有効かつ効率よく利用するには様々な課題を克服する必要があります。生命情報工学研究センター（CBRC）では、統合データベースプロジェクトに参画し、これら分散するソフトウェア解析やデータベース検索等を、一つ一つユーザが実行するのではなく、一連の処理の流れを定義することにより、効率的に短時間に実行するためのワークフローに関する技術開発を進め、統合化を目指した実用的な情報基盤開発、利用環境の整備（RDF化）及び解析結果の統合化された可視化に取り組んでいます。（図1）

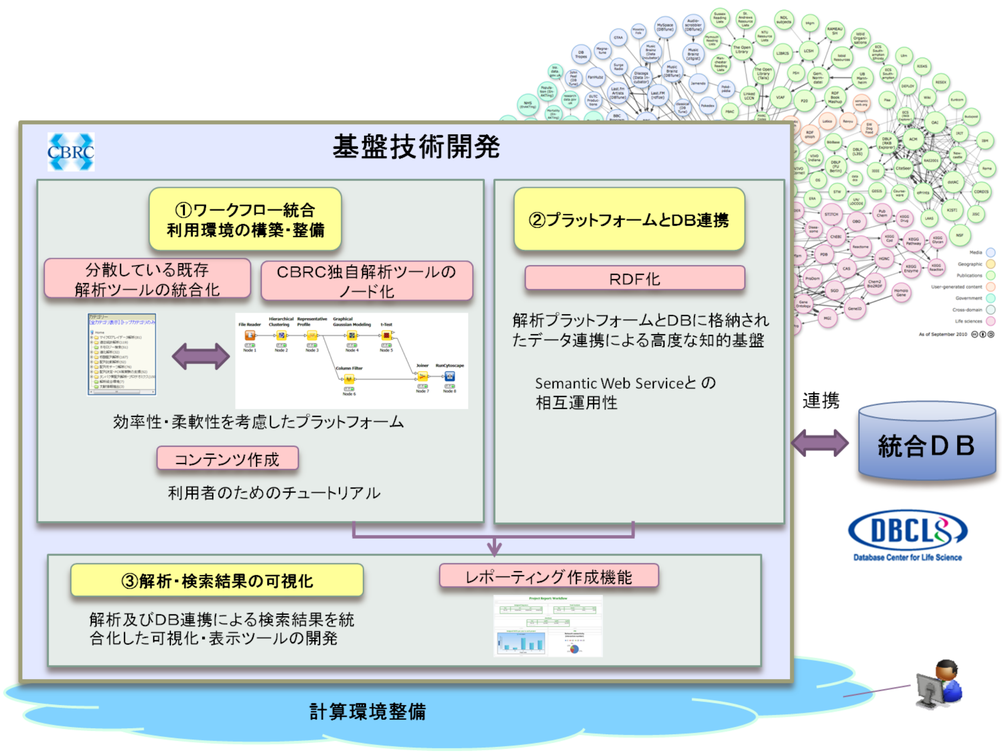

図1　基盤技術開発プログラムにおける開発概要

●研究の成果

プラットフォームとしてドイツのコンスタンツ大学で開発された無償配布版であるKNIMEを利用しています。KNIMEはeclipseベースのワークフロー型フラットフォームであり、一つの処理をノードという形式で表します（図2）。ユーザーはこれらのノードを結合し組み合わせることで、プラットフォーム上でワークフローを構築し、データの読み込み、計算、解析、可視化を行うことができます。またセマンティクWEB技術としてSADIフレームワークを用い、ダイナミックな解析プラットフォームの開発に取り組んでいます。

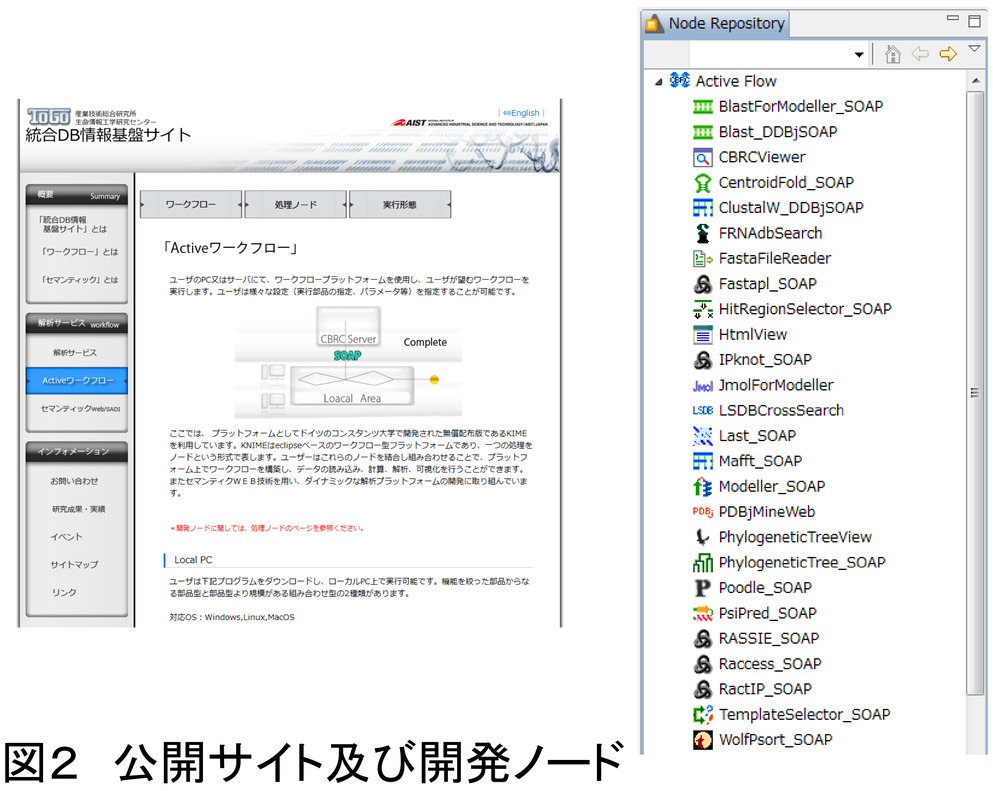

図2　公開サイト及び開発ノード

●将来の展望

●ライフサイエンス研究分野における専門化・細分化されたデータベースや解析ツールの迅速な変化や新規情報技術へ柔軟に対応できるワークフロー解析として期待できます。

●セマンティクWEB技術の利用によりデータベースと解析ツールを組み合わせたシームレスな大規模大量データ解析を可能とする技術開発へと繋がります。

●参考

ワークフローは統合DB情報基盤サイトから利用・ダウンロード可能です。

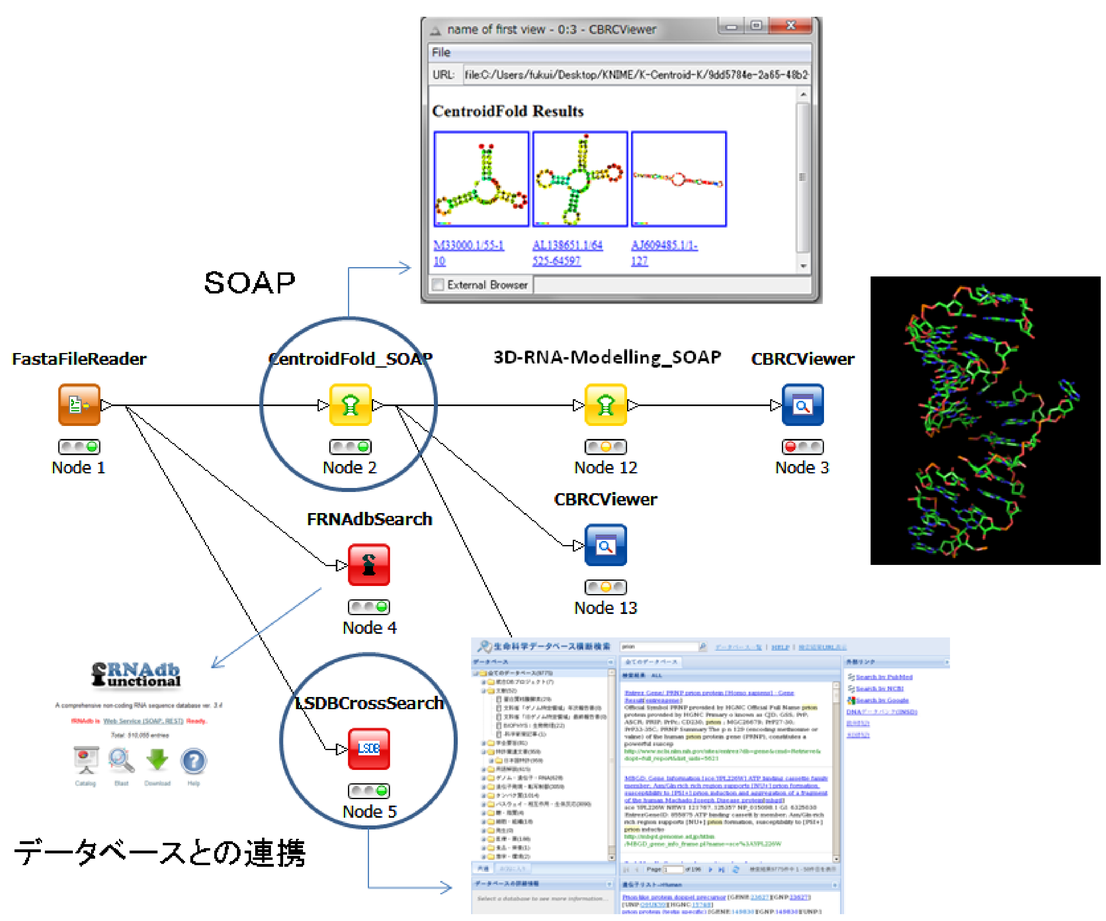

図3.　ワークフロー解析

統合DB情報基盤サイト　http://togo.cbrc.jp

Licensed under a Creative Commons
表示2.1日本 license (c)2011
福井一彦 (CBRC)