

既存のWeb的資産を仮想的にグリッド化

データベース・グリッドの研究開発

インターネットに代表される情報化の進展に伴い、知的基盤としての情報は増大の一途である。例えば、我々つくばWANデータ共有委員会の調査では、つくば地域の研究機関だけで数百に上る研究情報のWeb公開データベースが存在する。そのため、地理的・組織的に分散した多様な情報を、高度に統合・連携させることで新たな知見を得たいという要求が高まりつつある。

一方、最近のグリッド技術、特にWebサービス技術と結びついたグリッドの研究開発によって、分散したデータベースの統合や連携の実現が容易になりつつある。また、グリッド上での分散計算と結びつけることで、単なる分散検索に留まらない高度なデータマイニングなどのブレイクスルーも期待することができる。

我々は、主に科学技術関連のデータベース処理を応用対象として、データベース・グリッドというグリッド上での異種・分散データベースの統合・連携についての研究を行っている。今回、その成果の一つとしてOGSA-WebDBというシステムを開発した。これは、Webで公開されている既存のデータベースには手を入れることなく、仮想的にグリッド上のデータベースとして利用可能なシステムである。それぞれの

Web上のデータベースは、仮想的には関係データベースとしてSQL言語でアクセスされ、単独では不可能なデータベース同士の結合や比較を行うことができる。

この開発により、既にインターネット上に存在する膨大な知的資産を、グリッド上で仮想的に統合・連携すると共に、グリッド内のデータベースや計算処理とも容易に結びつけることが可能になりつつある。

本システムはOGSA (Open Grid Service Architecture) と呼ばれるグリッドの基本構成に基づいたツール等で構築され、高い互換性と拡張性を有している。また、すべての機能がグリッド上のデータ提供サービスとして実装されているので、新しいデータベースの登録などの情報管理もインターネットを介して行うことができる。開発したソフトウェアは昨年11月のSC2003会議などで展示・発表し、Web上の文献データベースや医薬データベースの統合検索に優れた性能を実証した。引き続き異種データの変換・統合を含めた高度な統合手法や、膨大な数のデータベースが存在する環境での情報サービスやデータベース発見等の手法の研究等を、主につくば地域の研究関連データベースを対象として進めていきたい。

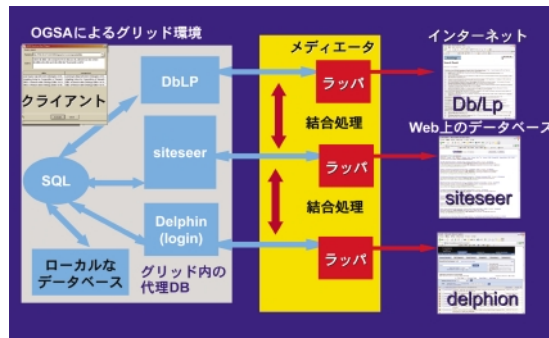


図1 OGSA-WebDBのシステム構成図

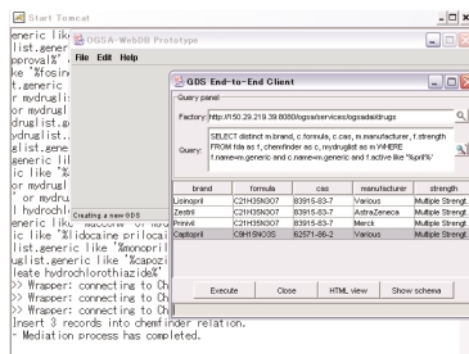


図2 Web上の医薬DB (fda, chemfinder) と、手元のDB (mydruglist) の統合実行例



こじま いさお
小島 功
kojima@ni.aist.go.jp
グリッド研究センター

関連情報

- 共同研究者: サイド・ミルザ・パレビ (産総研特別研究員) .
- <http://www.gtrc.aist.go.jp/dbgrid/>
- I. Kojima, S. Mirza: The Future of Grid Data Environment Workshop, GGF10 (2004).
- <http://www.aist.go.jp/infobase/tsukubawan/> (仮) つくば知的資源サイバーモール.