

超大規模データを利活用する最先端技術群：新原理による超高性能データベースエンジン、サイバー空間・実世界データ融合活用基盤、地球環境情報融合システム、クラウドサービス連携技術

☎ <http://www.tkl.iis.u-tokyo.ac.jp/>

■ 研究の概要

インターネット時代の主役はネットワーク上のコンテンツにある。ウェブには約1兆以上ものページが存在し、更に驚異的な速度で増加し続けており、情報の宝庫である一方で、効率的な管理・解析の技術に対する要求はますます高まっている。また、高度なセンサ技術の進展により、膨大な数のセンサが「しゃべる」時代にはいり、従来とは比べものにならない細密な観測が可能となってきた。すなわち、超可観測の世界が構築されると想定され、これら膨大なデータを格納する高性能なデータベースエンジンと多様な分析ソフトウェアから構成されるシステムが不可欠となる。本研究室では、データベース技術を基盤とし、最先端データベース、超大規模ウェブマイニング、地球環境デジタルライブラリを始めとするデータ工学の課題に対し、システムソフトウェア、先進アプリケーション、ハードウェア、アルゴリズムに亘る、あらゆる観点から研究を進めている。

■ 主な研究テーマ

超巨大データベース時代に向けた最先端データベースエンジンの開発とサイバーフィジカルシステム (CPS) の研究：

実行順序を実行時に動的に決定する「非順序型データベース実行原理」に基づくマルチコアプロセッサ並びに大規模ディスクアレイを想定した超高性能データベースエンジンを開発する (Fig.1)。同時に当該データベースエンジンを核とし、巨大データ活用により可能となる次世代戦略的サービス (サイバーフィジカルサービス) の実証システムを構築し、当該エンジンの有効性を明らかにする。

超大規模ウェブアーカイブ：

喜連川研では約14年以上にわたり継続的に日本語ウェブページを大規模収集し、およそ300億URL、20億ブログ記事、Twitterの200億つぶやき等を含むウェブアーカイブを構築しており、その構造、内容、時間変化を解析するシステムを開発中である。これら膨大なウェブ情報を、リンク解析、自然言語処理、画像処理等を用いて解析し、様々な切り口で探索可能な可視化システムを構築し大規模ディスプレイウォール上に実装している (Fig.2)。また、豊田研究室との連携のもと、話題の伝播、言語の変化、スパム活動などといった多様な社会活動の分析実現に取り組んできた。Web情報は、画像・映像等への多メディア化が急速に進むと同時に、放送映像等の実世界情報と相互に及ぼし合う影響も拡大し続けており、多様な解析を可能とする多メディアWeb情報解析基盤の構築と社会分析ソフトウェアの研究も進めている。さらに、交通情報等の実世界データとウェブデータ等の異種データを融合した実社会分析システムの研究開発を進めている (文科省・実社会ビッグデータ利活用のためのデータ統合・解析技術の研究開発)。

20ペタバイト超地球環境情報融合システム：

地球観測データを利用して、地球環境の理解を深め、予測能力を高め、危機管理や資源管理等における健全な政策決定に資する情報を提供することが国際的に求められている。このようなニーズに応えるため、地球環境研究者と共同で、20ペタバイト超ストレージ層からアプリケーション層まで構成される地球観測データ統合・情報融合システムの構築を行っている。様々な地球観測データを統合・解析することにより、科学的・社会的

的に有用な情報に変換できるシステムの実現を目指している。(Fig.3)。

東京大学地球観測データ統合連携研究 (<http://www.editoria.u-tokyo.ac.jp/>)

クラウド時代の大規模計算技術の構築：

大規模な計算機群から、必要な計算資源を必要に応じて即座に利用することができる「クラウドコンピューティング」を利用し、負荷変動や障害に頑健なアプリケーションを構築するための研究を進めている。特に、大規模探索などこれまでは適用が難しかった問題においても、オンデマンドな分散計算が容易に行えるような基盤技術の確立を目指している。

■ 研究環境

計算機環境はとても充実しており、ハイエンドサーバ・大規模クラスタによる演算処理能力や、巨大ディスプレイ、ペタバイトのアーカイブ装置などを利用した先進的な研究が可能である。研究室の構成は教授、特任准教授に加え、ポスドクなど学位保持者が多く、やる気のある学生への支援を惜しまない。海外での発表機会も多数ある。

✉ [kitsure@tkl.iis.u-tokyo.ac.jp](mailto:kitsure@tkl.iis.u-tokyo.ac.jp)

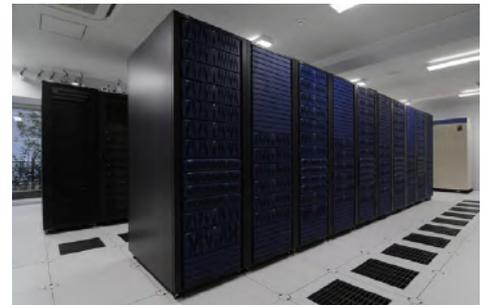


Fig.3 地球環境20ペタバイト超ストレージ



Fig.1 超高性能データベースエンジン実験システム

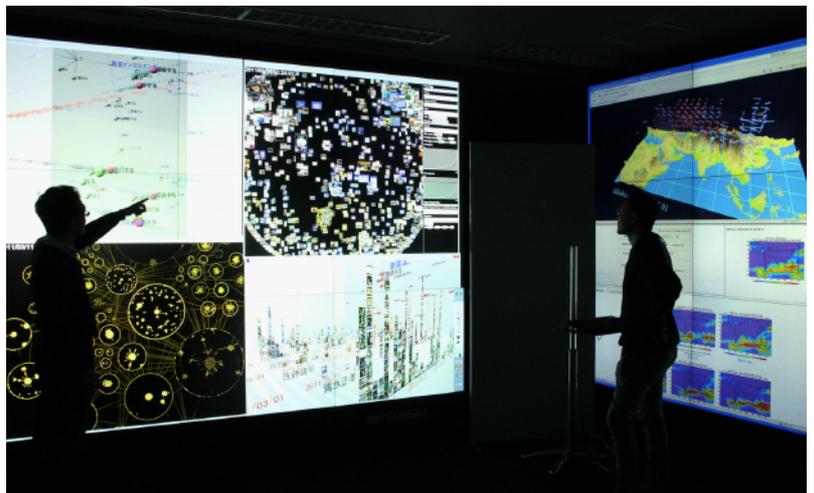


Fig.2 ディスプレイウォール上の大規模時空間可視化システム