

# 豊田 (正)・吉永・合田研究室

実世界・ソーシャルビッグデータ高度処理・可視化基盤



ディペンダブル社会情報プラットフォーム研究センター

インタラクティブデータ解析、地球観測データ工学、適応的言語処理、データプラットフォーム工学

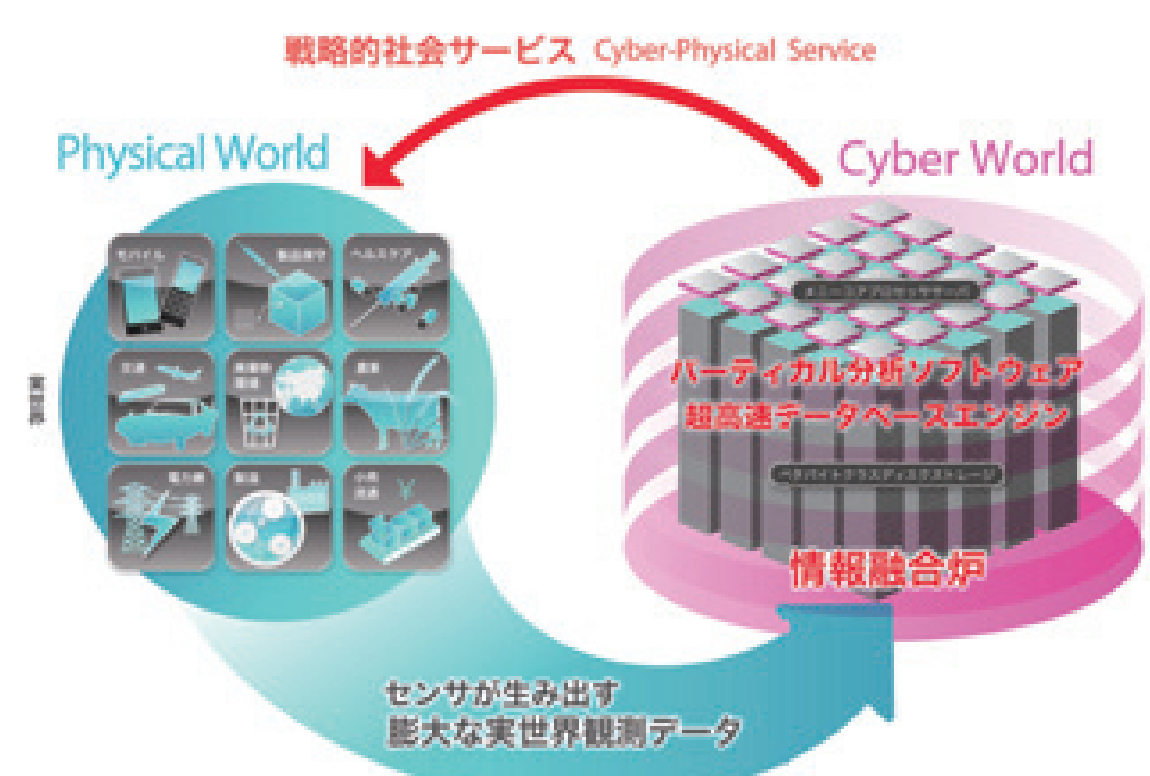
情報理工学系研究科 電子情報学専攻

<https://www.tkl.iis.u-tokyo.ac.jp/>

本研究室では、データベース技術を基盤とし、非順序実行原理なる新方式に基づく超高速最先端データベース、超大規模サイバー空間マイニング、100ペタバイト級超大規模地球環境デジタルライブラリ、言葉から人の心と社会の動きを読み解く自然言語処理技術を始めとし、データ工学における諸課題に対し先進的なシステムソフト、アプリケーション、アルゴリズムの研究を進めている。

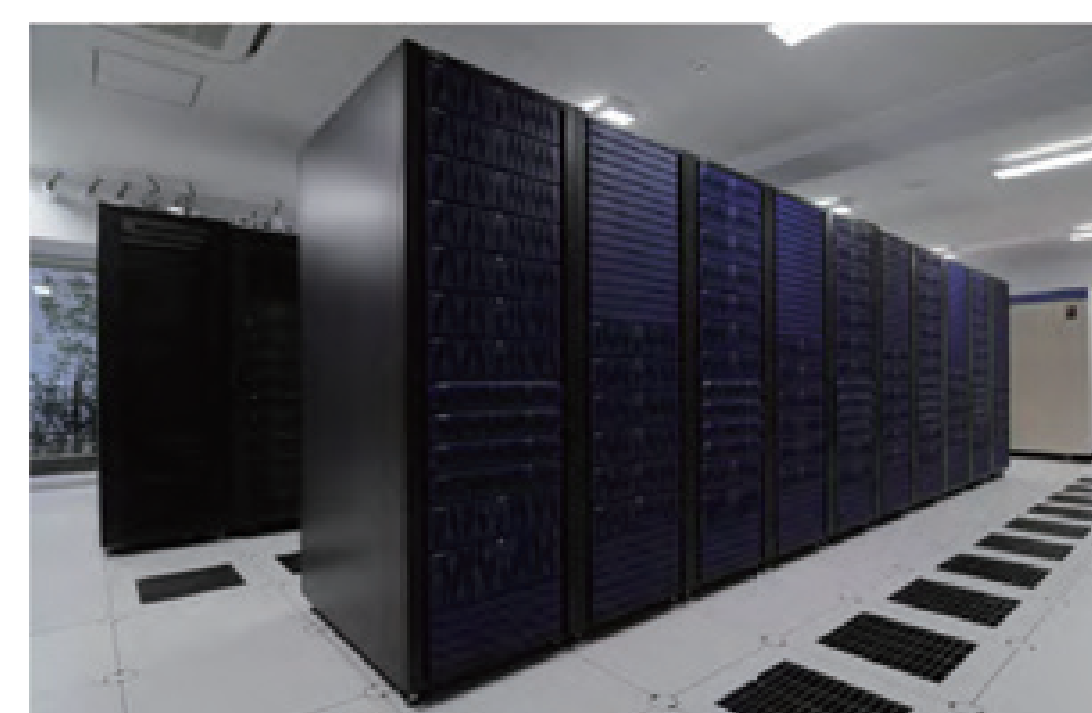
## 超巨大データベース時代に向けた最先端データベースエンジンの開発

「非順序型実行原理」なる新しいソフトウェア実行原理に基づく超高速データベースエンジン（大規模情報を扱うための基盤ソフトウェア）の開発を進めている。ビッグデータ時代のクエリ処理において、現行システム比で飛躍的な高速化を実現しつつある。また、当該データベースエンジンを核として、今までにない先進的なビッグデータ・IoT応用を創出するべく、実証システムの構築に取り組んでいる。加えて、新たな方向性として、データベースエンジンの省エネルギー化を進めている。



## 100ペタバイト級地球環境情報融合システム

各種地球環境情報のデジタルライブラリ化と国際的ポータル構築を進めている。また、大規模データアーカイブ、メタデータ管理等、近年注目されている大規模データ処理のシステム技術に関する研究を進めている。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。



## 超大規模サイバー・フィジカル空間インタラクティブ解析システム

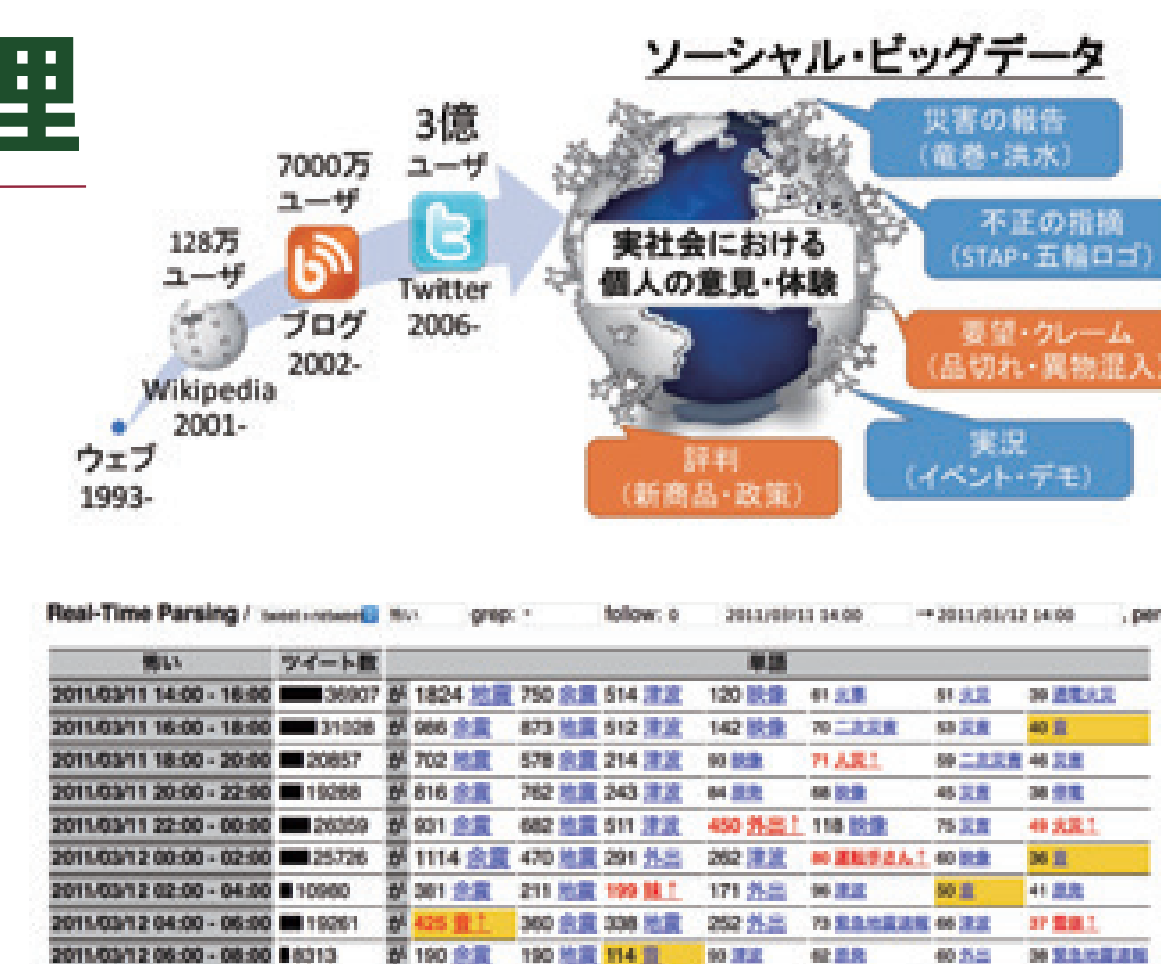
ウェブ・ソーシャルメディア等のサイバー空間と実世界は密接に連動しており、サイバー空間と実世界センサデータの融合解析による社会課題解決を目標とした研究を推進している。1999年から継続的に日本語ウェブページを大規模収集し、数百億URL、数十億ブログ記事、Twitterの数百億つぶやきを含むウェブアーカイブを構築するとともに、ドライブレコーダデータ、交通トラフィックデータ、気象データ等の実世界データの収集・蓄積を行い、その構造、内容、時間変化等を解析するシステムを開発中である。膨大なサイバー空間・実世界データをデータマイニング、機械学習、リンク解析、自然言語処理、画像処理等を用いて解析し、様々な切り口で探索可能な可視化システムを大規模ディスプレイウォール上に実装している。



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

## 言葉から人の心と社会の動きを読み解く自然言語処理

膨大な言葉から人の心と社会の動きを読み解き、人の情報伝達・理解を支援することを目指して、計算機を用いて言葉を速く、正しく「計算」する工学的研究（自然言語処理）に幅広く取り組んでいる。これらの技術の追求が、ときに確率的に、ときに規則正しく振る舞う自然現象である言語の「かたち」に迫る理学的研究（計算言語学）や人間の知能の働きの解明に繋がると考えている。これまで、日々蓄積されるソーシャルビッグデータを実際に解析して社会の動きを捉えるシステムを構築してきた。現在は、大規模言語モデルの応用を開拓すべく、内部機序を分析して、効率を高める研究を行っている。



東日本大震災発生日における恐怖の対象の推移