

# 喜連川, 豊田, 根本, 生駒, 合田, 吉永, 伊藤研究室

## [実世界・ソーシャルビッグデータ高度処理 ・学習・可視化基盤]

生産技術研究所 ソシオグローバル情報工学研究センター

データ工学・地球観測データ工学・ウェブ工学・適応的言語処理

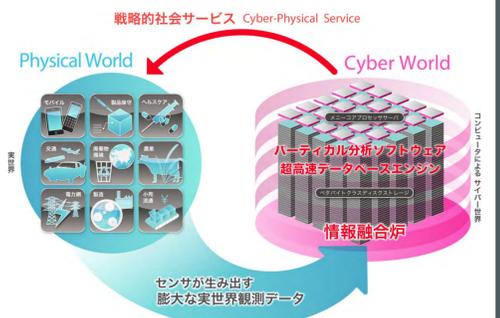
電子情報学専攻

<http://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、25ペタバイト級超大規模地球環境デジタルライブラリ、ソーシャル・ビッグデータ利活用のための言語処理基盤を始めとし、データ工学における諸課題に対し、『誰も見たことが無い、ドキッとする』システムソフト、先進アプリ、ハード、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。



### 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

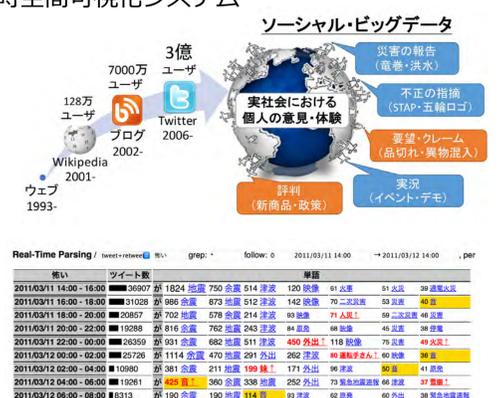
ウェブ・ソーシャルメディア等のサイバー空間と実世界は密接に連動しており、サイバー空間と実世界センサデータの融合解析による社会課題解決を目標とした研究を推進している。1999年から継続的に日本語ウェブページを大規模収集し、数百億URL、数十億ブログ記事、Twitterの数百億つぶやきを含むウェブアーカイブを構築するとともに、ドライブレコーダデータ、交通トラフィックデータ、気象データ等の実世界データの収集・蓄積を行い、その構造、内容、時間変化等を解析するシステムを開発中である。膨大なサイバー空間・実世界データをデータマイニング、機械学習、リンク解析、自然言語処理、画像処理等を用いて解析し、様々な切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

### ソーシャル・ビッグデータ利活用のための言語処理基盤

マイクロブログなどを通して人々が発信する莫大で多様なソーシャル・ビッグデータを人間の代わりに読み解き、社会に役立つ知見を得るため、実際的な自然言語処理技術の研究を進めている。超高速・高精度で適応的な基礎言語解析技術の開発、実世界テキストから得られた知識を矛盾なく集積するための方法論の提唱を行い、言語情報を基に非言語情報を整理して構造化するフレームワークの構築に取り組んでいる。一方で、超大規模データを活用し、機械翻訳や対話システムなど自然言語を介した知的アプリケーションの高度化も進めている。



東日本大震災発生日における恐怖の対象の推移