

教員名	豊田正史准教授	研究場所	生研	研究分野	ウェブ工学
-----	---------	------	----	------	-------

## 社会現象を読み解くための大規模ウェブ分析

ウェブは 1 兆以上ともいわれる膨大な文書がリンクで結合された複雑なネットワーク構造を持ち、ニュースサイト、個人ブログ、マイクロブログ等により文書の生成、更新、削除が行われ日々変化し続けています。近年のソーシャルメディア普及により、この変化は現実世界の事象を反映する傾向を益々強めています。

本研究室では、膨大なウェブ関連情報を収集解析し、文書の内容、リンク構造、さらにそれらの動的な変化から有用な情報を抽出し、可視化技術等を駆使して社会現象を読み取る手法を研究しています。更に近年では、交通情報データ等多様な実世界データとウェブ情報との関連性を探る研究も進めています。検索エンジンにとどまらない、新しいウェブ解析技術、ウェブサービスおよびアプリケーションの構築に興味のある方の参加を期待します。

### ■主な研究内容

**ウェブ・ソーシャルメディアの空間構造分析：**ウェブ上では、互いに関連するページ同士がハイパーリンクにより密に結合され、ソーシャルメディアではユーザ同士が友人関係で結合されています。こうしたネットワーク構造を分析することで、同じテーマに関するウェブサイトのコミュニティや、友人間を情報がどう拡散するかなどを調べることが可能です。図 1 は、大震災時に Twitter のユーザがどのように避難場所情報を拡散し共有したかを可視化したもので、情報拡散の中心人物や、拡散のパターンをモニタリング可能にしています。

**スパム分析：**検索エンジンにおいて自サイトの順位を不当に上げようとしたり、ソーシャルメディア上で意図しないページへユーザを誘導したりするスパム行為が蔓延していますが、これを分析することでスパムをフィルタするだけでなく、スパム挙動を把握したり、新しいスパムを発見したりすることが可能になります(図 2)。

**多メディア分析：**ウェブの時系列的な変化からは、新しい話題の発生や、話題の盛り上がりおよび収束の状況を知ることができます。これをテレビ等の複数のメディアにおける話題の盛り上がりと比較する技術を開発しています。例えばブログ上に頻繁に現れる画像をテレビ映像のデータベースと高速に照合しほぼ同一の画像がどのようにメディア間を伝搬したかを分析することが可能です(図 3)。

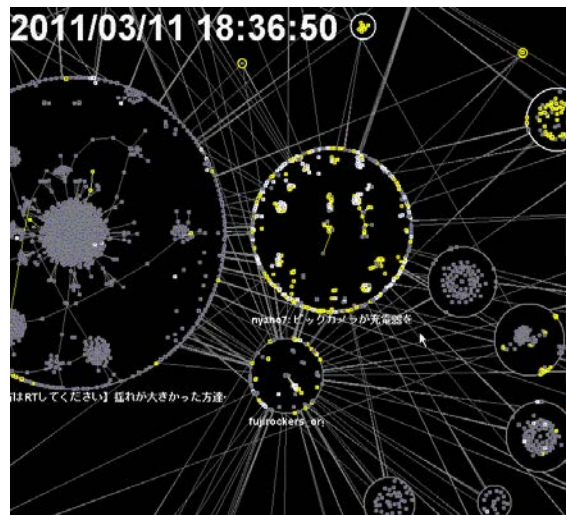


図 1: 大震災時の避難場所情報の拡散可視化

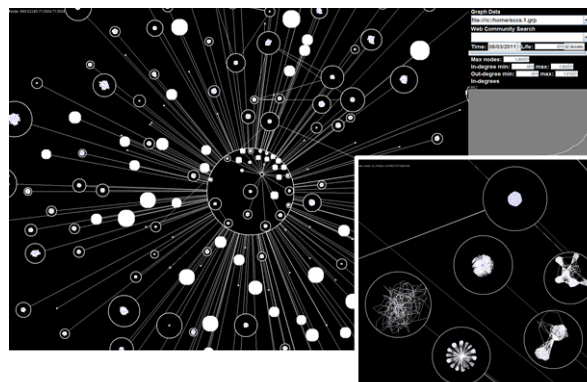


図 2: 検索エンジンスパムの構造分析

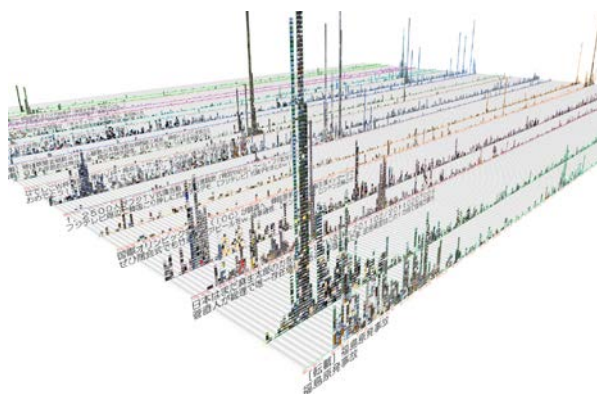


図 3: 多メディア分析

E-mail: [toyoda@iis.u-tokyo.ac.jp](mailto:toyoda@iis.u-tokyo.ac.jp)  
 HP: <http://www.tkl.iis.u-tokyo.ac.jp/~toyoda/>