

## 時空間データ統合システム

秋山 祐樹<sup>1</sup>, 柴崎 亮介<sup>2</sup>

<sup>1</sup> 東京大学 新領域創成科学研究科 社会文化環境学専攻, <sup>2</sup> 東京大学 空間情報科学研究センター  
連絡先: <aki@iis.u-tokyo.ac.jp>

- (1) **動機:** 近年インターネットの普及により新鮮な空間情報の利用者が急増している。また平成 19 年 5 月に地理空間情報活用推進基本法が制定され、デジタルの地理空間情報を活用するための法的な枠組みも構築された。これに伴い日本では様々な自治体や企業が多種多様な空間データを整備公開する状況になりつつある。しかしながら多くのデータが未だ各々独自のフォーマットでデータを作成しているため、データ相互の統合や時系列化は困難な状態である。各種空間情報を自在に統合出来る環境は今後不可欠な技術となる。
- (2) **アプローチ:** ソースデータの種類によってそれらが保有する空間情報の詳細さは異なる。そこで本システムは空間情報の名称情報(施設名称, 店舗名称等)と位置情報(住所, 経緯度, 建物情報)のみを使って異なる種類のデータ同士の空間統合および、異なる時点間の時系列化を可能にした。個店単位のデータ統合や、二時点間の時系列変化情報(存続, 入替, 新設, 消滅)の取得が可能となる。
- (3) **意義:** 一般的な GIS ソフトウェアの空間結合は空間データの二次元空間情報しか参照出来ない。住所表記の不完全さによるアドレスマッチング精度低下は直ちに空間統合精度の低下に繋がる。本システムは住所表記や建物情報(入居する階や部屋番号)も考慮に入れた三次元空間統合を実現するこ

とでこの課題を解決した。また統合後のデータの名称情報の同一性判定機能を搭載することで、名称が一致しないデータの発見や、二時点間の時系列変化情報の取得も可能になった。また一般的な GIS ソフトウェアでは入出力データの容量が大きい。そこで本システムでは入出力データを CSV とすることで大容量のデータも扱えるようにした。

### (4) 特徴:

- ・ 入居する建物情報(階, 部屋の自動認識)や住所表記を考慮出来る三次元空間統合技術。
- ・ 名称表記の揺れを考慮した同一性判定機能。
- ・ 名称同一性判定の障害となる雑音語(地名, 駅名等)の辞書開発とそれを用いたクリーニング。
- ・ 局所的に出現する頻出語「ローカル頻出語」の発見とクリーニング。
- ・ 入力データ CS とし、テキスト処理に徹する構造にすることで、メモリ使用量を抑え大容量データの処理を実現。

### (5) その他:

- ・ システムの一般公開(Web)を検討中。
- ・ 本システムを用いて時系列化した各種時空間データのダウンロードサービス(Web)を検討中。
- ・ 本研究は、国土交通省および財団法人国土計画協会「国土政策関係研究支援事業」の支援を受けている。

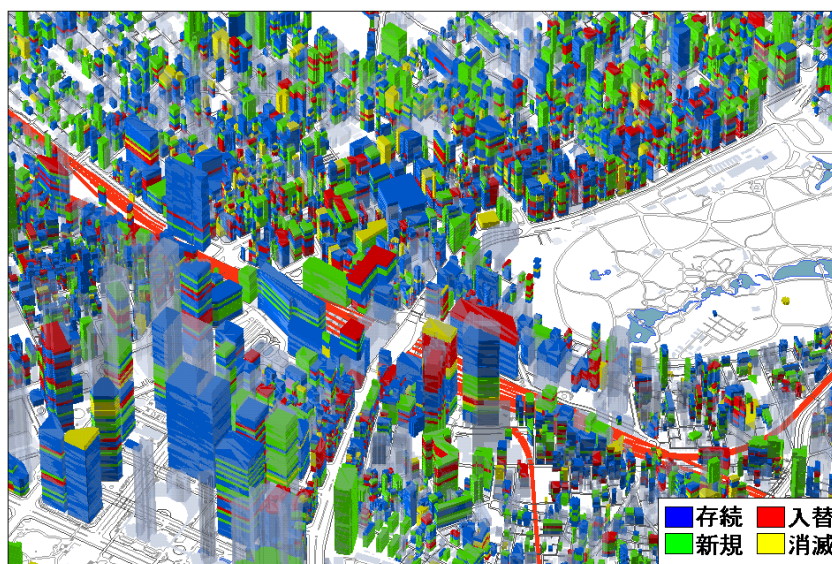


図 1: 時空間データの例(新宿駅周辺の時系列変化)

2000 年と 2005 年の電子電話帳(タウンページデータベース)を時系列化したデータを作成し、それを電子住宅地図(Zmap TONW II)と統合しビジュアライズした。ソースデータが部屋番号情報を持つ場合は、部屋単位のデータを作成出来る。