

## 論文の内容の要旨

論文題目 非構造・半構造空間情報の高度利用に関する研究

氏名 相良 毅

### 1. 本論文の背景と概要

カーナビシステムや PC 用地図ソフト, 携帯電話・PHS での位置情報サービスなど, 個人レベルでの空間情報利用が進んでいる。これらのアプリケーションは, 病院やレストラン情報などの空間コンテンツデータの検索や閲覧を主目的としている。しかし, 代表的な空間情報システムである地理情報システム(GIS)は, 空間解析や高精度な空間情報管理を目的としており, 空間コンテンツデータを管理するにはオーバースペックとなっている。その結果, データ作成にコストがかかりすぎることや, 個人ユーザには操作が難しいことなどが利用上の障害となっている。

本研究ではこの問題の一解決策として, 個人情報や Web 上・タウン情報誌などに豊富に存在する共有コンテンツ情報を空間情報として利用することを考え, そのための要素技術として次の 2 つの手法を研究し, 実利用可能なプログラムを開発した。

- 1) 住所参照手法: 地名や住所という形で記述された場所の情報を座標値に変換する手法
- 2) 空間情報抽出手法: 自然言語で記述された場所記述を含む文(ワープロ文書や Web ページ)から住所らしい部分を抽出し, マークアップする手法

これらの手法を利用して, これまで十分に活用されていなかった非構造・半構造化空間情報を容易に検索・閲覧することが可能な新しい空間情報システム, 「空間文書管理システム」を開発し, その上に各種アプリケーションを構築して有効性を示した。

## 2. 日本の住所体系に適した住所参照手法

住所で記述された場所の情報をコンピュータで空間情報として扱うには、経緯度のよう  
に数値化された座標値に変換しなければならない。この手法は住所参照手法と呼ばれ、欧  
米諸国では地理情報システムの基礎的なツールとして広く利用されている。しかし、日本  
では以下に示す2つの理由により、実用的な手法が確立していなかった。そこで本研究で  
は、日本の住所体系や表記上の特徴を整理し、日本の住所体系に適したアルゴリズムを開  
発、実装を行った。このアルゴリズムは住所以外の一般的な地名にも適用可能である。

### 1) 日本の住所体系の複雑性

日本の住所体系は、「不動産登記法」と「住居表示に関する法律」による体系が混在して  
いる。また、町村は「郡」に含まれるが市は含まれない、政令による指定都市は区を置く  
ことができるなどの規則も体系を複雑化している。以下にいくつかの住所記述の例を示す。

- 宮城県仙台市青葉区国分町三丁目7-1 (地番, 指定都市部)
- 東京都多摩市関戸六丁目12番地の1 (地番, 市部)
- 東京都西多摩郡檜原村467-1 (地番, 町村部)
- 大阪府大阪市西区千代崎三丁目中2番1号 (住居表示, 街区方式)
- 山形県東根市板垣北通り25号 (住居表示, 道路方式)
- 京都府京都市下京区綾小路通烏丸西入 (京都市通名)

### 2) 日本語の特徴に起因する住所記述の曖昧性

日本語では単語間に空白やカンマなどの区切り文字を使わないため、住所を正しく単語  
に区切る必要がある。また、都道府県名や市町村名の省略などに対応するため、2種類の  
木構造(地名階層木と地名インデックス)を用いた検索手法を用いた(Fig.1, 2)。また、アラビ  
ア数字と漢数字の混在や、「1-2-3」のような省略記法にも対応する手法を示した。

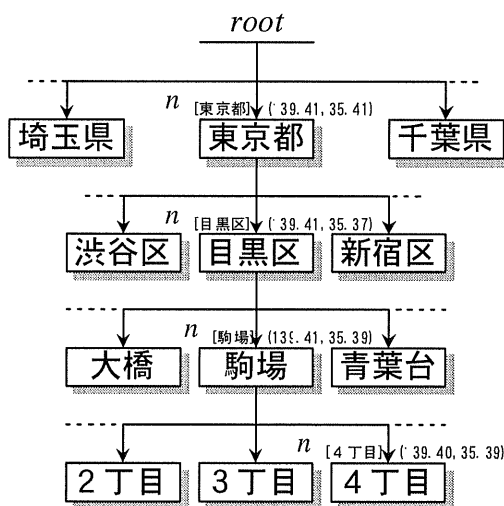


Fig.1 地名階層木

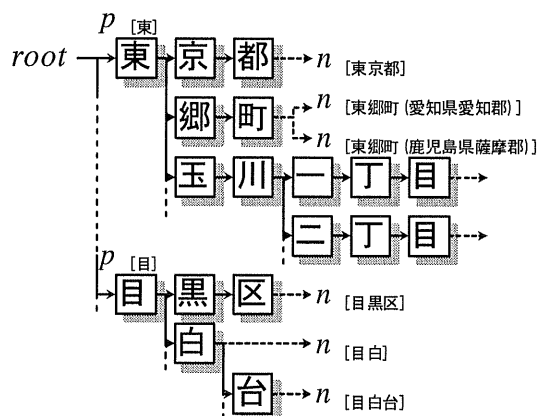


Fig.2 地名インデックス

### 3. ネットワーク分散システム化と自治体の分置廃合への対応

住所参照手法は、地名と座標値の対応表を集中的に整備・維持するコストが高いため、大都市部など一部の地域や、丁目レベルまでの照合に限定されることが多かった。そこで地名照合システムをネットワーク分散サーバ化し、各サーバが地理的に住所の管理を受け持つシステムとした(Fig.3)。これにより整備コストの負担を分散すると同時に、検索負荷も分散されるため、詳細な住所も効率的に検索可能になった。

また、日本では自治体の分置廃合により頻繁に住所が変わってしまうが、その度に対応表を作成しなおすのではコストがかかりすぎる。そこで地名階層木を拡張し、新旧の住所をつなぐリンクを作成することで(Fig.4)、新旧の住所が混在するデータも変換できる住所参照システムを実装した。

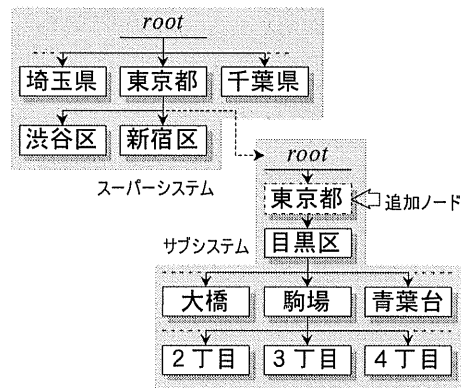


Fig.3 ネットワーク分散システム

#### 分置廃合の例

- 名称変更, 市制移行  
西多摩郡羽村町→羽村市
- 合併  
保谷市+田無市→西東京市
- 政令指定都市化による区政施行  
千葉市稲毛→千葉市稲毛区稲毛
- 合併に伴う下位地名の名称変更  
(保谷市) 本町→(西東京市) 保谷町, (田無市) 本町→(西東京市) 田無町

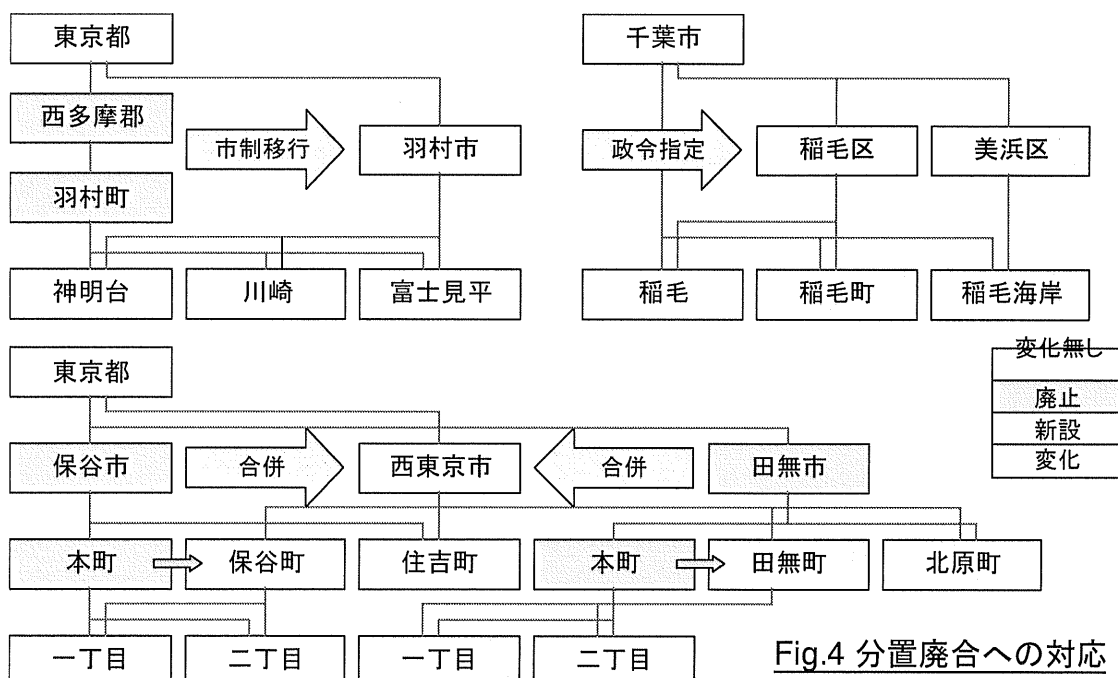


Fig.4 分置廃合への対応

## 5. 空間情報の抽出と半構造化手法

自然言語で記述されたテキスト文書（ワープロ文書，メール，Web ページ・・・）に含まれる地名を自動的に抽出できれば，住所参照手法を利用して，これらの文書を絶対的な位置にリンクすることができる．空間検索・空間演算を適用すれば，場所に依存した人ナビ用の空間コンテンツ情報として利用する，などの高度利用が可能になる．

地名を文章中から抽出するには形態素解析手法が利用できる．理想的には全ての地名が固有名詞として分類されるはずだが，実際には大部分が複合名詞や人名と区別できない．そこで，まず品詞レベルで地名らしい部分文字列を取り出し，住所参照システムに問い合わせることで地名であることを確認し，抽出するシステム「芭蕉」を構築した．

## 6. 空間文書管理システム

位置参照手法と半構造化手法を利用することで，一般文書を場所にリンクすることができる．この手法を，Web や実世界の各種情報源から収集した文書を空間データに変換するために利用すれば，地理情報システムで利用可能な点集合を自動作成するという目的に利用することも可能である．しかし，ナビゲーションのように人間に情報を提供することを目的とした場合，収集した文書そのものが閲覧できた方が（人間にとって）より分かりやすい．そこで，文書をそのままの形で管理・検索・閲覧できる新しい空間情報システム，「空間文書管理システム」を考案した(Fig.5)．

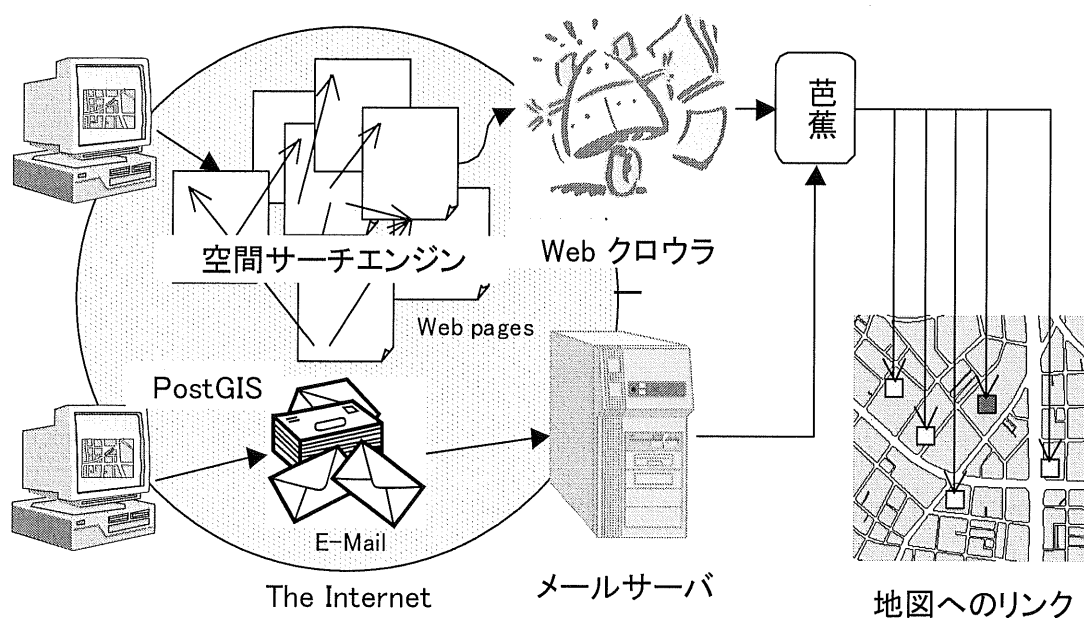


Fig.5 インターネット指向 空間文書管理システム