

## 歩道ネットワークと地名辞書を基本とした日常的な場所表現を 対象としたジオコーディング手法

野秋 浩三<sup>†</sup> 相良 毅<sup>‡</sup> 有川 正俊<sup>†</sup>

<sup>†</sup> 東京大学空間情報科学研究センター 〒153-8904 東京都目黒区駒場四丁目 6 番 1 号

<sup>‡</sup> 東京大学生産技術研究所 戦略情報融合国際研究センター 〒153-8505 東京都目黒区駒場四丁目 6 番 1 号

E-mail: <sup>†</sup> {noaki, arikawa}@csis.u-tokyo.ac.jp, <sup>‡</sup> sagara@iis.u-tokyo.ac.jp

**あらまし** 我々は、住所情報を様々な文書から抽出し、経緯度の情報をそれらの文書に付加する研究を行ってきた。このような間接位置参照情報を直接位置参照情報へ変換する処理をジオコーディングという。本稿では、対象の範囲を広げ、道案内文のような日常的な場所表現のジオコーディングの枠組みを提案する。具体的には、基盤情報としての歩道ネットワーク・データベースに、一般ユーザでも収集可能な POI(Point Of Interest)データベースを統合して利用できる枠組みを整理し、また経路記述の形式言語 FRS(Formal Route Statement)を提案し、歩道ネットワーク・データベースに基づいた FRS のジオコーディングの手法を議論する。

**キーワード** Web とインターネット、空間 DB、データの可視化、GIS

## Geocoding method for daily local expressions using sidewalk databases and geographic name dictionaries

Kouzou NOAKI<sup>†</sup> Takeshi SAGARA<sup>‡</sup> and Masatoshi ARIKAWA<sup>†</sup>

<sup>†</sup> Center for Spatial Information Science, at the University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8904 Japan

<sup>‡</sup> Institute of Industrial Science, at the University of Tokyo 4-6-1 Komaba Meguro-ku, Tokyo 153-8505, Japan

E-mail: <sup>†</sup> {noaki, arikawa}@csis.u-tokyo.ac.jp, <sup>‡</sup> sagara@iis.u-tokyo.ac.jp

**Abstract** A lot of data have been delivered through the Internet including HTML documents and Internet mails. Many of the data on the Internet include geo-reference descriptions. We have studied a method of converting such descriptions like addresses into their corresponding coordinates, that is, tuples of longitude and latitude. The process of converting descriptions into coordinates is called geocoding. In this paper, we focus on daily local expressions as a new type of descriptions to geocode. We first introduce a framework to embed users' collecting POI (Point of Interest) data in sidewalk network databases, then propose FRS(Formal Route Statement) as a query language for the sidewalk network database.

**Keyword** Web and Internet, Spatial DB, Data Visualization, GIS

### 1. はじめに

インターネットの普及により、地球上のどこにいてもアクセス可能な巨大情報空間が実現した。インターネット上には HTML 文書や Eメールなど、数十億というデジタル文書が流通しており、その多くは住所や地名といった現実世界の場所を参照する記述を含んでいる。これらの記述をジオコーディング (geocoding)、つまり現実世界の位置と対応させれば、実空間と連携した情報活用が可能になる。現在、インターネットの利用環境は PC から携帯電話や PDA などのモバイル端末まで多様に広がっている。屋外で移動しながら利用するケ

ースも既に珍しくなく、様々な情報を統合した歩行者用ナビゲーションサービスも一部実現されている。このような背景を踏まえ、都市生活において位置情報をキーとした情報アクセスへのニーズはますます高まっていると言える。

飲食店情報を提供するポータルサイトでは、店舗の情報として店名やメニュー、お店の雰囲気をはじめ、住所、電話番号、郵便番号などの形式化された場所記述までがフォーマットに従って整備されているが、道案内のような文章では、待ち合わせ場所や店舗が面している通りの名前など、住所のように定形化された書き方ではなく、

くだけた記述がなされる場合が多い。住所、電話番号、郵便番号などの形式化された記述はアドレスマッチングなどのジオコーディング技術によって、安定して一意の空間と対応づけることができる。一方で、「渋谷駅から道玄坂を上り道玄坂上交番を右へ約200m」のように、ランドマークや通りとの空間関係によって表現され、可視化に際して点だけでなく線で表現すべき道案内文のような記述も存在する。

本稿では、計算機上で自然言語による場所記述をジオコーディングするための基礎として、文書の中でも特に道案内文のような日常語による経路記述のジオコーディングを実現する手法を議論する。

## 2. 歩道ネットワーク・データベース

本節では、道案内文のような記述の特徴を捉え、ジオコーディングするために必要なデータについて説明する。

### 2.1. 経路記述の構成

道案内のような文章では「丸井方面へ進み、井の頭通りを左折し直進、スペイン坂を上がってすぐ右側」や「JR渋谷駅東口より、宮益坂をのぼって約5分、左手」のように、ランドマーク(「丸井」、「JR渋谷駅東口」)や通り(「井の頭通り」、「スペイン坂」)などの名前で場所を特定し、空間的な関係(「方面へ進み」、「を左折し直進」、「を上がってすぐ右側」)によって目的地までの道筋を表現している。

このような人の通っていく道筋を記述した文を総称して経路記述と呼び、本稿では特に都市空間における経路記述を議論の対象とする。表1に渋谷をモデルとして、ランドマークとなる主な地物についてまとめた。これらのランドマークは経路の始点、経由点、終点となり得るが、文脈によっては「丸井方面」のように単に方向を示すために用いられる場合もある。また、表2に経路を決定する主な記述と具体例をまとめた。このような記述は複数の選択可能な経路がある場合に移動経路を決定する重要な情報となる。

表1: ランドマークとなる主な地物と渋谷での具体例

地物の種類	具体例
店舗	丸井, QFRONT, 渋谷 HMV
交番	道玄坂上交番, 渋谷駅前交番
交差点	渋谷スクランブル交差点
駅出入口	JR 渋谷駅(中央改札口, 東口, 南口, 新南口, ハチ公口, 宮益口, 玉川口), 地下鉄銀座線渋谷駅, 地下鉄半蔵門線渋谷駅, 京王井の頭線渋谷駅
通り	明治通り, 明治通り, 公園通り, 井の頭通り, 文化村通り, 青山通り, 宮益坂, 金王坂, スペイン坂, センター街, オルガン坂

表2: 経路を決定する主な語の例

経路を決定する語の種類	具体例
進行方向(直進, 右折, 左折)	進み, 上がって, 直進, 右, 右側, 右手, 右折, 左, 左側, 左手, 左折
移動距離	200m, 5分

### 2.2. 歩道ネットワーク・データベースのスキーマ

以上のような特徴から、自然言語による経路記述を実空間と対応させるためには、ランドマークの名称やその経緯度、位相関係の情報を持ったネットワーク・データベースが必要となる。このデータベースは、特に人が移動するうえで出発地や目的地、目印となる地物で構成されるため、本稿では歩道ネットワーク・データベースと呼ぶ。

歩道ネットワーク・データベースのスキーマを図1に示す。特に「交差点」や「通り」は、歩道ネットワーク上では複数のノードとリンクによって構成される。例えば、十字路の交差点を表現する場合は少なくとも4つのノードが必要となる。また、特定の呼称がつけられた通りも複数のノードとリンクの集合によって構成されるものとする。このようなノードとリンクには属性情報の1つであるin属性にそのグループを代表する地物名を入力して管理する。なお、歩道ネットワーク・データベースのすべてのノードは必ず1つ以上のリンクでつながれているものとする。

SidewalkDB:	(Nodes, Links)
Nodes:	nodeの集合
Links:	linkの集合
node:	(id, name, coordinate (longitude, latitude), in, class, incoming_link, outgoing_link, poi)
link:	(id, start_node(id), end_node(id), direction, distance)
node.id:	node識別子
node.name:	名称
node.coordinate:	座標(経度, 緯度)
node.in:	グループ名称
node.class:	地物の型
node.incoming_link:	入リンク(link識別子)
node.outgoing_link:	出リンク(link識別子)
link.id:	link識別子
link.start_node:	始点nodeの識別子
link.end_node:	終点nodeの識別子
link.direction:	方向(度)
link.distance:	距離(m)

図1: 歩道ネットワーク・データベースのスキーマ

### 3. 自然言語による経路記述の問い合わせ

経路記述は、図2で示すように、始点、経由点、終点の各種ノードとそれらをつなぐリンクおよび各属性情報によって記述されるため、可視化に際して空間関係を判断しながら複数の地物を線で結ぶ必要がある。また、目的地は文書中にすでに含まれている場合が多く、省略されることが多いため検索を行う必要がある。そこで、歩道ネットワーク・データベースに対して、問い合わせを行うための幾何的記述として、FRS(Formal Route Statement)を提案する。

「JR 渋谷駅東口より、宮益坂をのぼって約5分、左手」

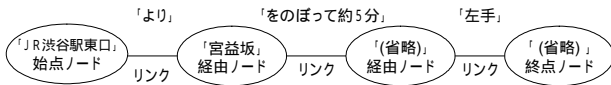


図2: 経路記述の構成イメージ

#### 3.1. 経路の幾何的記述

自然言語による経路記述を計算機で処理するためには、幾何的な記述を定義し変換する処理が必要となる。ここではすべての経路記述はノードとリンクの集合によって表現されると仮定し、経路の幾何的記述である FRS を提案する。図3に FRS の文法を示す。

前節の表2に挙げた進行方向を決定する語はリンクの属性情報 (link.connect)へ変換するために表3のような簡単なテーブルを用意する。また、移動距離が時間で記述されている場合はメートルに換算するものとする。以下に、「JR 渋谷駅東口より、宮益坂をのぼって約5分、左手」を FRS に変換した例を示す。

例1. 入力ツールとして、フォームを利用して FRS を入力した場合

```
node_desc(0) = "JR 渋谷駅東口"
link_desc(0) = "より,"
node_desc(1) = "宮益坂"
link_desc(1) = "をのぼって約5分,"
node_desc(2) = ""
link_desc(2) = "左手"
FRS = node_desc(0):link_desc(0):node_desc(1)
      :link_desc(1):node_desc(2):link_desc(2)
```

FRS ::=	node_desc(0):link_desc(i):node_desc(i)* [i={1,...,n}];
node_desc(i) ::=	node(i).node_attribute_list
node_attribute_list ::=	none   node_attribute_value (& node_attribute_value)*
node_attribute_value ::=	node_attribute = value
node_attribute ::=	id   name   coordinate   in   class   status
value ::=	numerical_value   sting_value   url   status_values   connect_values
status_values	start   end   via
connect_values ::=	straight   right   left

link_desc(i) ::=	link(i).link_attribute_list
link_attribute_list ::=	none   link_attribute_value (& link_attribute_value)*
link_attribute_value ::=	link_attribute = value
link_attribute ::=	id   start_node(id)   end_node(id)   direction   connect   distance

図3: FRS(Formal Route Statement)の文法

表3: 進行方向を決定する語の変換

変換する語	link.connect の値
進み, 上がって, 直進	straight
右, 右側, 右手, 右折	right
左, 左側, 左手, 左折	left

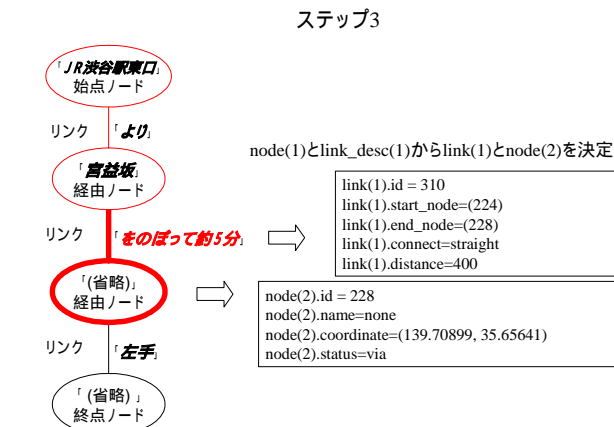
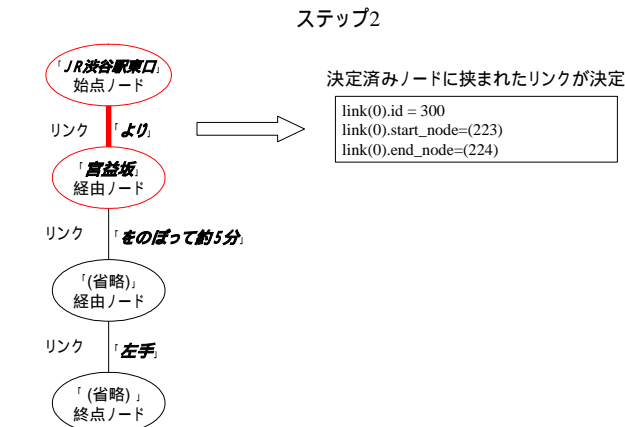
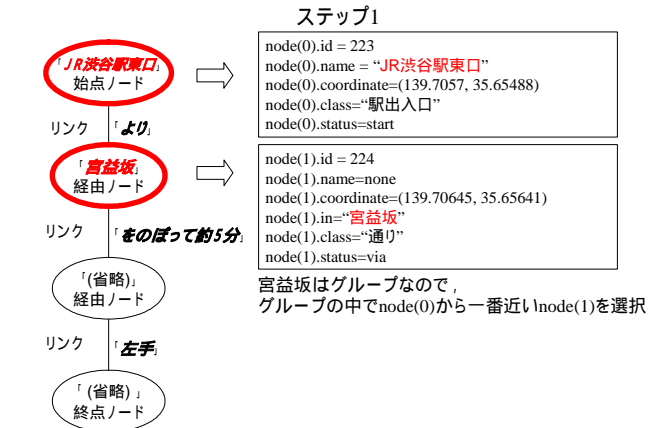
#### 3.2. 問い合わせ処理の流れの概略

将来的には様々な文書中の文章から経路記述を抽出し、前節で提案した FRS へ変換する一連のアルゴリズムが必要となる。本稿では経路記述を切り出して入力するまでは人手によって行う例を仮定して、前節の例で示した FRS の問い合わせ処理の概略を図4と共に以下に説明する。

- 入力した並びから 始点と終点のノードの status 属性はそれぞれ start, end となり、その間の経由ノードはすべて via に決定する。 node\_desc(0) = "JR 渋谷駅東口", node\_desc(1) = "宮益坂" の文字列を歩道データベースの name 属性, in 属性に対してパターンマッチングを行い node(0), node(1) を特定する。 in 属性にマッチングする "宮益坂" は複数ノードによって構成されるため "JR 渋谷駅東口" から最も近いノードを特定する。
- (1) で特定された node(0) から node(1) までの最短経路から、その間の link(0) を特定する。
- 隣接ノードの名前が省略されていたり、名前を持たない場合、link\_desc(i) の文字列を表3のような空間関係辞書に問い合わせ、変換した結果の link.connect や link.distance からリンク、ノードを特定する。 link\_desc(1) = "をのぼって約5分" は link(1).connect=straight, link(1).distance=400 に変換し、特定済みの前ノードからその条件を満たす link(1) と node(2) を特定する。
- link\_desc(2) = "左手" を link(2).connect=left に変換し、特定済みの前ノードからその条件を満たす link(2) と node(3) を特定する。

以上の処理によって、経路記述から幾何的記述に変換することができる。各ノードは経緯度情報を持っているので、単にグラフを書くのではなく実空間の位置に対応した経路を示すことができる。よってこの処理は経路記述のジオコーディングを可能にする。

上記の例では、FRS で記述された問い合わせに対する解は1つであったが、実応用の経路記述のジオコーディングの結果は、解が複数個になり、またそれぞれの解が精度(accuracy)、またはランキングを持つ。例えば、「5分」は「300m~500m」に換算され、時間的な点が距離的な区間へ展開され、ほぼ400mが最も確からしい距離と考えられる。このような例に関しても、上記のアルゴリズムを拡張して実現できる。



#### ステップ4

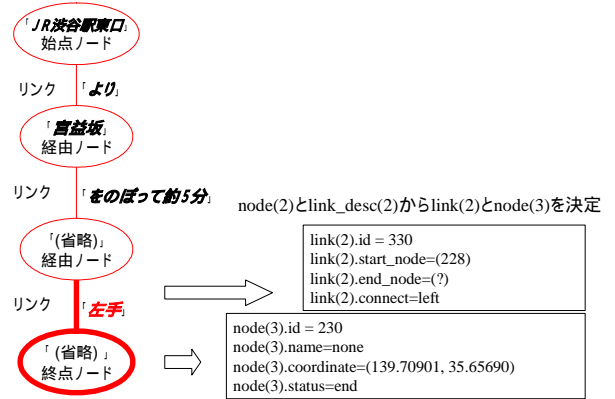


図 4: 処理の流れの一例

#### 4. まとめ

本稿では、歩道ネットワーク・データベースを用いた経路記述のジオコーディングの基本的な枠組みを提案した。現在、この枠組みの有効性を検証するために、Web 文書を対象に商用の歩道ネットワーク・データベースとユーザ定義の POI データベースを用いて実験を行っている。

今後、様々な文書中に表れる経路記述を分析し、都市空間での経路記述という用途に沿った空間関係の記述を整備していく予定である。

#### 文 献

- [1] 相良毅 “非構造・半構造空間情報の高度利用に関する研究”，博士論文，2003。
- [2] 平松薫，“地理的関連性を用いた Web 検索技術の研究”，博士論文，2002。
- [3] Annette Herskovits，“空間認知と言語理解”，堂下修司，西田豊明，山田篤(共訳)，オーム社，東京，1991。
- [4] 長尾真，佐藤理史，黒橋禎夫，角田達彦，“自然言語処理”，長尾真(編)，岩波書店，東京，1996。