

## ウェブを対象としたロボット型検索による 指定地理座標周辺の住所関連情報検索手法の提案

森本 泰貴<sup>†</sup> 藤本 典幸<sup>†</sup> 萩原 兼一<sup>†</sup>

<sup>†</sup> 大阪大学大学院情報科学研究科コンピュータサイエンス専攻  
〒560-8531 大阪府豊中市待兼山町1-3  
E-mail: †{hiroki-m,fujimoto,hagihara}@ist.osaka-u.ac.jp

あらまし 近年、GPS機能の普及により、携帯端末のユーザは自分の現在地を地理座標として取得することが可能となっている。そのため、指定された地理座標を原点とし、その周辺の地理情報を検索するシステムが有用であると考える。そこで我々は、以前に開発したロボット型住所関連情報検索システムを応用し、住所文字列をキーワードとしたロボット型検索により、指定地理座標周辺の住所関連情報を検索する手法を提案する。提案手法は指定地理座標周辺の住所を取得し、取得した住所文字列を以前に開発したシステムの入力として住所情報を検索する。実験の結果、住所をキーワードとしたロボット型検索という直接的な手法に少し改良を加えることで、指定地理座標周辺の住所関連情報検索に有効な手法となりうるという結論を得た。

キーワード WWW, 地理情報検索, 周辺情報検索, GPS, Google Earth

## A Method for Web Retrieval of Address-Related Information around a Given Geographical Point

Hiroki MORIMOTO<sup>†</sup>, Noriyuki FUJIMOTO<sup>†</sup>, and Kenichi HAGIHARA<sup>†</sup>

<sup>†</sup> Graduate School of Information Science and Technology, Osaka University  
1-3 Machikaneyama-machi, Toyonaka-shi, Osaka, 560-8531 Japan  
E-mail: †{hiroki-m,fujimoto,hagihara}@ist.osaka-u.ac.jp

**Abstract** Current and next generation mobile phones are equipped with a GPS unit, enabling users to know their current geographical location. Given this ability, a system to retrieve Web information based on location is useful. In order to implement such a system, we propose a method that performs robot-type keyword retrievals using address strings as keywords. We present experimental results to evaluate the effectiveness of this method. The results demonstrate that our method with future improvement will be effective for implementing web retrieval for a given geographical scope.

**Key words** WWW, Geographical search, Scope focused search, GPS, Google Earth

### 1. はじめに

2007年4月より携帯電話へのGPS機能搭載が義務化されるなど、GPS機能の普及が進んでいる。これにより、GPS機能を搭載した、インターネットに接続可能な携帯端末が一般的となり、携帯端末のユーザは自身の現在地を地理座標として取得することが可能となっている。このような環境において、ユーザの現在位置を入力とし、ウェブ上からその周辺の地理情報を検索するシステムが有効であると考えられる。そこで我々は、地理座標を入力とし、その周辺に位置する住所とその関連情報を検索して地図を用いて表示する、範囲指定式地理情報検索シ

ステムを開発したいと考えた。このシステムは経度  $x$  と緯度  $y$  からなる指定地理座標  $(x, y)$  と周辺住所を検索する範囲の距離  $r$  を入力とし、 $(x, y)$  を中心とした半径  $r$  の円形の範囲内に位置するすべての住所を指定地理座標の周辺住所として取得する。そして周辺住所に含まれる各住所について、住所とその関連情報を検索してユーザに提示する。このとき検索する情報は、飲食店や駅などといった特定の種類の情報ではなく、取得した各周辺住所に関するあらゆる住所関連情報を検索する。検索が終了すると、範囲指定式地理情報検索システムは検索した住所関連情報を、Google Earth [2] などを用いて視覚的な形でユーザに提示する。

本論文では、範囲指定式地理情報検索システム実現の方法として、住所文字列をクエリとしたウェブ検索による情報検索を提案する。提案手法では、まず全国の全ての住所とその地理座標の対応を記したデータベースを用意し、これを用いて指定地理座標の周辺住所に対応した住所文字列リストを取得する。そして、取得した各住所文字列について、住所文字列をクエリとしたウェブ検索を行い、住所関連情報を取得する。住所関連情報の検索については、我々が以前に開発したロボット型住所関連情報検索システム GMPSearch [1] の手法を用いる。この手法は、ウェブ上をクロールしてクエリとして与えられた文字列に適合するウェブページを収集し、収集したウェブページから住所表記の認識および認識した住所表記の説明記述の抽出を行い、住所と説明記述の組を住所関連情報として出力する。

提案手法には2つの問題があると考えられる。まず第1に、GMPSearchの手法は一般的なクエリ（神社、温泉など）に対しては実用的な時間で実用的な数の住所関連情報を検索できることを確認しているが、住所文字列をクエリとした検索と同様に実用的な数の情報を得られるとは限らない。特に、地番を住所文字列に含める場合、地番の表記方法には「1-2-3」や「一丁目二番三号」など多彩なパターンが考えられるため、住所文字列によるウェブ検索がどの程度機能するかは疑問である。もう1つの問題は住所関連情報の検索にかかる時間である。指定された地理座標の周辺住所は数多く存在すると考えられる。実際のところ、入力 $r$ を1kmとした場合には周辺住所は地番までの細かさでみると数百件となる。よって、その全てについてウェブ検索を行うには大きな時間を要し、効率的な検索が行えない可能性がある。

そこで、本研究では実際に住所文字列をクエリとしたウェブ検索による範囲指定式地理情報検索システムを試作し、試作システムを用いて住所関連情報検索を行い、その性能を評価する実験を行った。この実験により、住所文字列をクエリとしたウェブ検索によって実用的な数の情報が得られるかを確認するとともに、地番を省略した住所文字列による検索によって地番のみが異なる複数の住所に関して同時に情報を検索し、周辺住所に含まれる住所に関する情報をどの程度正確に取得できるかを検証した。本論文ではこうした実験の結果をもとに、住所文字列をクエリとしたウェブ検索による範囲指定式地理情報検索システムの実現において工夫すべき点について考察する。

以降、2節で試作システムの構成とアルゴリズムについて述べ、3節で評価実験について述べる。その後、4節で関連研究を紹介し、5節でまとめと今後の課題を述べる。

## 2. 試作システム

### 2.1 試作システムの構成

試作システムは、指定地理座標の周辺住所の取得と、住所文字列をクエリとした住所関連情報検索をそれぞれ別のシステムで処理している。周辺住所の取得には新規に作成したシステムを用い、住所関連情報検索にはGMPSearchを用いる。なお、GMPSearchへのクエリ入力は、取得した周辺住所を参照して手動で行う。

試作システムにおける指定地理座標の周辺住所リスト作成システムは、経度 $x$ と緯度 $y$ の組からなる地理座標 $(x, y)$ および検索範囲の半径 $r$ を入力とする。入力を受けたシステムは、入力を元に指定地理座標の周辺住所を取得し、指定地理座標からの距離順にソートしたリストとして出力する。なお、周辺住所は地番までの細かさ（〇〇市××町X丁目Y番Z号ならば〇〇市××町X丁目Y番まで）で取得している。

周辺住所リスト作成システムを実行した後、取得されたリスト中の周辺住所のそれぞれについて、その住所文字列をクエリとしてGMPSearchを実行し、各住所文字列を検索する。周辺住所を検索するには、周辺住所リストに含まれる全ての住所について検索を実行する必要があるが、1つの地理座標に関する周辺住所は多数存在すると考えられ、その全てについて検索を実行するには多くの時間を要する。よって、範囲指定式地理情報検索システムの実現にあたってはより効率的な検索方法を考案する必要があると考えられる。本論文では簡単な実装で範囲指定式地理情報検索システムを試作し、効率的な検索方法を探るために実験を行った。

### 2.2 試作システムのアルゴリズム

#### 2.2.1 指定地理座標周辺住所リスト作成システムのアルゴリズム

周辺住所の取得には、国土地理院が無料で提供しているデータファイル[3]を用いて作成したデータベースを用いる。提供されているデータファイルでは、地番までの細かさでの全国の住所について、1つの住所に代表点として1つの地理座標が与えられている。よって、指定地理座標 $(x, y)$ と代表点との距離が $r$ 以内となる住所が周辺住所に該当するといえる。作成したデータベースは、地番までの細かさでの全国の住所をインデックスとして、各住所の代表点の経度および緯度をレコードとしている。総レコード数は1400万強、ファイルサイズはインデックスファイルとデータファイルを合わせて1.6GB強である。

座標間距離の計算については、地理座標の場合、厳密に計算すると地球の丸みなどを考慮せねばならず、計算が煩雑になる。しかし、本研究の目的は周辺住所の検索であり、比較的短い距離が計算できれば十分である。よって、試作システムでは地理座標を平面座標に近似して計算を行っている。平面として考えた場合、1km離れた二点間の地理座標は、緯度が等しい場合は経度の差が $\frac{1}{80}$ 度となり、経度が等しい場合は緯度の差が $\frac{1}{120}$ 度となることが知られている。これより、2つの地理座標（東経 $x_1$ 度、北緯 $y_1$ 度）と（東経 $x_2$ 度、北緯 $y_2$ 度）との座標間距離は

$$\sqrt{(80(x_1 - x_2))^2 + (120(y_1 - y_2))^2} \text{ km}$$

で求められる。

#### 2.2.2 ロボット型住所関連情報検索システム GMPSearch のアルゴリズム

住所関連情報の検索にはGMPSearchを用いている。以下ではGMPSearchが用いるアルゴリズムを処理の順序に従って簡単に説明する。GMPSearchのアルゴリズムの詳細については文献[1]を参照されたい。

### ● 制限時間付きトピック主導型クローリング

我々が以前に開発したアルゴリズム [4] を用いてウェブページのクローリングを行う。このアルゴリズムは与えられた収集時間の間、与えられたクエリに適合するウェブページの収集を行うものである。

まず、Google Web APIs [5] あるいは Yahoo! Search APIs [6] を用いて種ページを取得する。そして種ページに加えて、種ページからリンクされているウェブページをクロールする。ただし、アンカーテキスト以外の領域にキーワードを含まないページはクエリに適合しない傾向があるため、そのようなページのリンクはたどらない。

### ● 住所表記の認識

収集したウェブページ中に出現する住所表記を認識する。住所表記の認識は、全国の全ての住所をキーワードとする文字列マッチングとして実現している。

国土地理院が無料で提供しているデータファイル [3] から抽出した全国の住所表記から、丁目以降を削除した住所表記をキーワードとして文字列マッチングを行う。丁目以降の文字列の認識は多彩な表記が考えられるため、構文解析によってそれらを全て認識している。丁目以降を削除した住所表記は約 12 万件弱が存在するため、多数のキーワードを線形時間で同時に検索する文字列マッチングアルゴリズムである AC 法 [7] を用いている。

### ● 住所の説明記述の抽出

ウェブページ中に出現する住所に関する説明記述（施設の住所であれば施設名やその解説など）を抽出する。システムが最終的に出力する住所関連情報は、認識した住所と抽出した説明記述の組のリストである。

住所とその説明記述は、大抵の場合はウェブページ中で固まって記述されると考えられるので、HTML 構文木において住所と説明記述は 1 つの部分木をなすはずである。そこで、ウェブページを HTML 構文木とした場合に、住所表記を 1 つだけ含む部分木の極大集合を求め、求めた部分木に含まれる TEXT ノードの内容を、その部分木が含む住所の説明記述として抽出する。

### ● 住所の経度・緯度への変換

Google Earth [2] などを用いて地図上に住所関連情報を提示するため、住所の経度・緯度への変換（geocoding）を行う。指定地理座標周辺住所リスト作成システム（3.2.1 節）で用いているのと同じデータベースによって、geocoding を実現している。

## 3. 実験

### 3.1 実験内容

この実験の目的は、住所表記をクエリとしたロボット型住所関連情報検索の有用性を評価することである。

有用性の評価基準として、検索の結果得られる情報数を用いる。GMPSearch による住所関連情報検索の経験から、「温泉」や「駅」、あるいは「そば」や「うどん」といった一般的なクエリを用いた検索では、ウェブページ収集時間を 60 秒とした場合に 100 件の情報が得られれば十分に実用的な数の情報が得

られるといえる。

しかし、周辺住所リスト作成プログラムによって取得する周辺住所は地番までの細かさであるため、1 つの地理座標に関して多数の周辺住所が存在し、その全てについて検索を行うと検索の試行回数が多くなり、大きな時間を要する。また、取得する周辺住所において地番の記述方法は半角文字である（例：1-3）が、実際には地番の表記方法には様々なパターンが考えられる（例：1-3、一丁目三番、1 丁目 3 番など）ため、取得した周辺住所をそのままクエリとして検索した場合、表記方法の違いから検索できない情報が存在すると考えられる。さらに、実際の住所表記においては都道府県名を省略することがしばしばあるため、都道府県名を省略した住所文字列による検索を行い、都道府県名を省略しない場合と比較する必要がある。

実験では、実際に周辺住所リスト作成プログラムによって周辺住所リストを獲得し、リストから住所を選出して GMPSearch のクエリとして検索を行い、検索結果を解析した。まず、地番を省略した住所文字列、および地番と都道府県名を省略した住所文字列による検索を行い、検索結果の違いを評価する。このとき、クエリに地番が含まれないため、地番以前の文字列が同じである複数の住所に関して同時に情報を取得できる。その後、地番を省略した住所文字列による検索の結果得られた情報の各住所について、地番を含めた住所文字列による検索を行い、地番を省略した場合との検索結果の違いを評価する。このとき、地番を省略した文字列による検索では、地番の表記方法に関わらず情報を収集できるが、地番を含めた住所文字列による検索では地番の表記方法がクエリと異なるような情報は検索できない。よって、単純にそれぞれの検索で獲得した情報数を比較するのみならず、地番を省略した場合の検索結果のうちウェブページ中の住所表記において地番が半角文字で記述されていた情報の数と、地番を含めた場合の検索結果の情報数を比較して評価する。

### 3.2 実験環境

実験に用いた環境は以下の通りである。

CPU Intel Pentium 4 - 2.8 GHz

メモリ 1 GB

OS Microsoft Windows Server 2003

ネットワーク環境 Gigabit Ethernet

### 3.3 実験結果

ここでは指定座標を東経 127.687161 度、北緯 26.107876 度の点、検索範囲の半径を 1km とした場合の実験結果を記載する。なお、指定座標は番地までの細かさの住所では沖縄県那覇市樋川 1-15 に相当する。この場合、周辺住所リスト作成システムによって、780 件の周辺住所が得られた。

得られた周辺住所のうち、

- 最も指定座標に近い住所：沖縄県那覇市樋川 1-15
- 最も指定座標から遠い住所：沖縄県那覇市牧志 3-20
- 地番より上の単位でみた場合に最も遠い住所：沖縄県那覇市宇国場 779

の 3 つの住所に注目して、それぞれの住所から地番を省略した住所文字列、および地番と県名を省略した住所文字列で住所関

表 1 地番を省略した住所文字列による住所関連情報の検索結果  
 Table 1 Address-related information retrieved by an address string without a lot number

住所文字列	収集時間 (秒)	検索情報数		適合情報数	
		住所数	情報数	住所数	情報数
沖縄県那覇市桶川	60	35	151	34	150
	180	39	260	37	257
	300	39	393	37	390
沖縄県那覇市字国場	60	44	144	0	0
	180	48	251	0	0
	300	52	281	0	0
沖縄県那覇市牧志	60	32	225	11	131
	180	36	428	※	※
	300	43	711	※	※

表 2 県名および地番を省略した住所文字列による住所関連情報の検索結果

Table 2 Address-related information retrieved by an address string without both a prefecture name and a lot number

住所文字列	収集時間 (秒)	検索情報数		適合情報数	
		住所数	情報数	住所数	情報数
那覇市桶川	60	31	185	31	185
	180	41	436	41	436
	300	41	539	41	539
那覇市字国場	60	56	159	0	0
	180	64	310	0	0
	300	68	353	0	0
那覇市牧志	60	41	453	13	278
	180	47	804	13	479
	300	51	1051	15	565

表 3 地番を含む住所文字列による住所関連情報の検索結果  
 Table 3 Address-related information retrieved by an address string followed by a lot number

住所文字列	検索情報数			
	地番	地番を省略した文字列		地番を含む文字列
全て		地番表記が半角のみ		
沖縄県那覇市桶川	1-1	31	14	25
	1-2	1	0	0
	1-4	2	0	2
	1-5	18	15	29
	1-6	4	1	1
	1-7	2	0	0
	1-9	8	0	0
	1-12	2	0	0
	1-13	16	2	2
	1-14	42	12	58
	1-15	106	15	20
	1-16	20	0	4
	1-17	1	1	0
	1-18	4	0	7
	1-19	1	0	1
	1-22	3	3	3
	1-26	3	0	0
	1-27	12	1	9
	1-28	8	3	15
	1-29	2	0	0
	1-30	1	0	0
	1-32	3	1	1
	1-33	7	0	9
	1-34	3	0	0
	1-36	10	0	1
	2-1	4	1	6
	2-2	6	3	5
	2-3	8	5	10
	2-6	20	5	6
	2-7	1	0	1
	2-8	19	15	20
	2-9	2	1	1
	2-10	3	2	2
	2-11	3	0	0
	2-12	6	0	5
	2-14	3	3	5
2-16	5	2	7	

連情報を検索した。その結果を表1および表2に示す<sup>(注1)</sup>。表1は地番のみを省略した住所文字列、表2は地番と県名を省略した住所文字列での検索結果をそれぞれ表す。なお、ウェブページ収集時間は60秒、180秒、300秒の3通りで検索しており、表1および表2にはそれぞれの検索について、クローリングしたウェブページから検出した住所関連情報の総数を記載している。住所数は住所関連情報を1件以上検出した住所の総数であり、情報数は1つの住所に関する複数の情報も数えた住所関連情報の総数である。また、検索情報数は検出した全ての住所数ならびに情報数、適合情報数は検出した情報から周辺住所に含まれる住所に関する情報のみを選出した場合の住所数と情報数である。

実験の結果、指定座標に近い「沖縄県那覇市桶川」に関しては、60秒の収集時間で150の適合情報を得られた。文献[1]において行った実験と比較すると、これは十分に有用な情報数といえる。また、不適合な情報が非常に少ないが、これは沖縄県那覇市桶川の全ての地番が周辺住所に含まれているためである。検索範囲を円形とするため、指定座標に近い住所に関してはこのような事例が多くなると考えられる。なお、ここでの不適合な情報は、「沖縄県那覇市桶川」という文字列を関連情報に含む全く別の住所に関する情報であった。このような情報の例としては、別の市の市役所の住所に「那覇市桶川の子供とわが市の子供の交流会を開いた」という一文を含む説明記述をあわせた住所関連情報が挙げられる。

また、「沖縄県那覇市牧志」に関しても、60秒の収集時間で131の適合情報を得られ、情報数は十分である。しかし、周辺住所に含まれない住所に関する情報も多く取得している。これは、同じ沖縄県那覇市牧志に属する住所であっても地番によって周辺住所に含まれる住所と含まれない住所が存在するためである。

さらに、「沖縄県那覇市宇国場」に関しては検索情報数自体は少ないが、適合情報が一切得られなかった。これは、沖縄県那覇市宇国場には1000を超える地番が存在し、周辺住所に含まれるのはそのうち4つだけであったためである。極端な事例ではあるが、沖縄県那覇市牧志に関する検索結果とあわせて考えれば、地番を省略した住所文字列による検索の問題点が認識できる。すなわち、周辺住所に含まれる住所のうち、指定地理座標からの距離が遠い住所に関しては、地番を省略した住所文字列による検索の適合率が低くなる。これは住所を地番より上の細かさでみた場合、その住所の一部のみが周辺住所の範囲に含まれることが原因である。周辺住所の範囲を指定地理座標を中心とした円形としているため、指定地理座標から遠い住所ほど、地番より上の細かさでみた場合の住所における周辺住所に含まれない部分、即ち地番までの細かさでみた場合に周辺住所に含まれない住所の割合が大きくなりやすい。

また、県名を省略した住所文字列による検索は、県名を省略

しない住所文字列による検索と比較して一般に情報数が多くなった。これはウェブ上に存在する住所表記において県名を省略した表記が存在するためだと考えられる。

表3は、「沖縄県那覇市桶川」で検索した結果得られた情報のうち適合である各住所について、地番を含めた住所文字列で検索した結果を示したものである。なお、ウェブページ収集時間はいずれも300秒としている。また、検索情報数の「地番を省略した文字列」は「沖縄県那覇市桶川」で検索した結果得られた情報数を、「地番を含む文字列」は地番を含む住所文字列で検索した結果得られた情報数を、それぞれ住所毎に表したものである。さらに「地番を省略した文字列」では、全ての情報数に加えて、ウェブページ中の住所表記において地番が半角文字で記述されていたものの情報のみの数も記載している。

表3より、地番を含む住所文字列で検索したとき、地番が半角文字で記述されているウェブページの情報、すなわち住所文字列に適合するウェブページの情報という観点でみた場合においては、地番を省略した住所文字列で検索したときと比較して情報取得数が一般に多くなることがわかる。しかし同時に、ウェブ全体に存在する住所表記において、地番を半角文字で記述している住所表記の割合は決して高くないことも推測できる。

### 3.4 考 察

実験結果から得られた知見をまとめると以下ようになる。

- 住所文字列をクエリとした検索でも十分な数の住所関連情報を得られる
- 県名を省略した住所文字列による検索のほうがより多くの情報を得られる
- 地番を省略した住所文字列での検索は指定座標から遠い周辺住所に対しては信頼性が低い
- ウェブ上の住所表記には地番を半角文字以外で記述しているものが比較的多い

以上の知見より、住所文字列をクエリとしたウェブ検索による指定地理座標周辺の住所関連情報検索において有効な手法を考察する。

最も適合率が高い方法としては、周辺住所リストに含まれる各住所について、それぞれ地番まで含めた住所文字列による検索を行うことである。ただし、ウェブ上における住所表記は地番の記述方法に様々なパターンが考えられるため、考えられるあらゆるパターンに対応しなければ検索可能な情報が少なくなる。よって、1つの住所に関する検索において、地番の表記方法を変更した複数のクエリが必要になる。このため、ただでさえ多いクエリがさらに多くなり、ウェブページ収集の効率を考えると現実的ではない。

よって、ウェブページ収集のクエリには地番を省略した住所文字列を使うべきだと考えられる。前節の実験では周辺住所は780件であったが、地番を省略した住所文字列でみると19通りしかない。また、地番を省略しているため、地番の表記方法の多彩さを考慮する必要が無い。このため、ウェブページ収集の効率が改善される。この方法の問題点は、地番より前の文字列は周辺住所に含まれるものの、地番までの細かさでみると周辺住所に含まれない、という住所に関する情報まで取得してし

(注1)：表1内の※欄については、試作システムのエラーにより取得した情報の詳細が見えず、測定できなかった。原因は収集したウェブページの中にURLに2バイト文字を含むページがあったためと思われる。

まうことである。そこで、地番を省略した住所文字列で住所関連情報を検索した後、周辺住所に含まれない住所に関する情報を削除する方法が考えられる。GMPSearch では住所関連情報を出力する際に、ウェブ上の表記方法に関わらず地番を半角文字で表す形で住所文字列を正規化している。よって、得られた住所関連情報の住所文字列と周辺住所に含まれる住所文字列の完全文字列マッチングによって、周辺住所に含まれるか否かを判別できる。

以上より、提案システムの実現手法として、

- 都道府県名および地番を省略した住所文字列をクエリとして住所関連情報を検索

- 得られた情報から周辺住所に含まれない住所に関する情報を除去

という手順を地番を省略した住所文字列のそれぞれについて行う、という手法があげられる。ただし、各住所文字列に関して逐次的に検索を行うと時間がかかるため、ウェブページの収集処理を並列化して各住所文字列に適合するウェブページを同時に収集するなど、高速化のための工夫が必要となるだろう。また、周辺住所に含まれる住所数が少ない住所文字列に関しては、地番を含めた住所文字列で個別に検索する、あるいはあえて無視するという方法によって、全体の処理の効率化をはかることも考えられる。

#### 4. 関連研究

指定地理座標の周辺情報を検索するサービスとして、Yahoo!JAPAN が提供するウェブサービスである Yahoo!地図情報 ローカルサーチ [8] がある。このサービスでは、キーワードを入力として住所、郵便番号、施設名のいずれかがキーワードに適合する施設を検索する機能と、地理座標と距離を入力として指定した座標から指定距離内にある施設を検索する機能を持つ。このうち後者の機能が、指定地理座標周辺情報検索を実現している。

[8] が出力する情報は、検索した住所あるいは施設の名称および地理座標である。検索対象が施設である場合は、市区町村レベルでの施設の住所もあわせて出力する。また、Yahoo!JAPAN は商店などの施設で特典を利用できる Yahoo!クーポンを発行しているが、[8] では指定地理座標周辺情報検索によってこの Yahoo!クーポンの対象施設を検索することも可能である。

我々が提案するシステムは、データベース化されていないウェブ上の情報をウェブクローリングによって取得するという点で [8] と異なる。

#### 5. まとめと今後の課題

ロボット型ウェブ検索によって指定地理座標周辺の住所関連情報を検索する範囲指定式地理情報検索システムの作成を目的として研究を行った。方針として、指定地理座標の周辺住所リストを取得し、各住所について住所文字列をクエリとしたウェブ検索によって住所関連情報を検索する手法を提案し、提案手法の有用性を検証するために実験を行った。

実験の結果、地番を省いた住所文字列をクエリとしたウェブ

検索により、範囲指定式でない住所関連情報検索の場合と同程度の情報数が得られることがわかった。しかし、地番を省いた住所文字列による検索では、周辺住所に含まれない地番の住所に関する情報まで取得してしまい、検索精度が低下する可能性がある。また、地番を含む住所文字列による検索は、周辺住所全てに対して検索を行うと住所数が多く非効率であるほか、ウェブ上での住所表記における地番の記述方法が多様であるために検索できない情報が多数存在すると考えられる。以上より、地番を省いた住所文字列によってウェブ上から住所関連情報を検索し、その後で周辺住所に含まれない住所に関する情報を除去する、という手法が現実的であるといえる。

今後は、本研究で提案した手法を用いて、ウェブを対象とした範囲指定式地理情報検索システムを開発する予定である。必要ならばさらに改良を加え、実用的なシステムとして完成させたいと考えている。

#### 文 献

- [1] 森本泰貴, 藤本典幸, 長屋務, 出原博, 萩原兼一, “Web を対象としたロボット型住所関連情報検索システムの開発”, 情報処理学会論文誌, Vol.J90-D, No.2, pp.245-256, 2006
- [2] Google Earth, <http://earth.google.co.jp/>
- [3] 国土地理院, 街区レベル位置参照情報ダウンロードサービス, <http://nlftp.mlit.go.jp/isj/>
- [4] 藤本典幸, 萩原兼一, “ウェブマルチメディア検索のためのパーソナルシステム”, 夏のデータベースワークショップ (DBWS2005), 電子情報通信学会技術研究報告, DE2005117, pp.61-66, 2005
- [5] Google APIs, <http://www.google.com/apis/>
- [6] Yahoo! デベロッパーネットワーク Yahoo! 検索, <http://developer.yahoo.co.jp/search/>
- [7] D. Gusfield, “Algorithms on Strings, Trees, and Sequences : Computer Science and Computational Biology”, Cambridge University Press, 1997
- [8] Yahoo! デベロッパーネットワーク Yahoo! 地図情報 ローカルサーチ, <http://developer.yahoo.co.jp/map/localsearch/V1/localsearch.html>