

研究成果の刊行に関する一覧表

雑誌

発表者氏名	論文タイトル	発表誌名	巻号	ページ	出版年
片岡裕介、浅見泰司、 浅利靖、郡山一明	需要密度に対する供給 効果を最大化するAED の最適配置地点	地理情報システム学会 GIS理論と応用	14-2	73-81	2006
白石陽、有川正俊、 相良毅、浅見泰司	空間ドキュメント管理 システムの設計と実装	DEWS	WEB	B7-10	2007

需要密度に対する供給効果を最大化するAEDの最適配置地点

片岡 裕介・浅見 泰司・浅利 靖・郡山 一明

Optimal allocation of AED's by maximization of supply effect on demand density

Yusuke KATAOKA, Yasushi ASAMI, Yasushi ASARI and Kazuaki KOHRIYAMA

GIS—理論と応用

Theory and Applications of GIS

Vol.14 No.2/2006

抜刷

40

GIS/A 地理情報システム学会

需要密度に対する供給効果を最大化する AED の最適配置地点

片岡裕介・浅見素司・浅利 靖・郡山一明

Optimal allocation of AED's by maximization of supply effect on demand density

Yusuke KATAOKA, Yasushi ASAMI, Yasushi ASARI and Kazuaki KOHRIYAMA

Abstract: A probability density function of demand for automated external defibrillators (AED's), which is based on the distribution of demand points by kernel density estimation method, is estimated from the data on occurrences of cardiopulmonary arrests. The probability that someone can be saved by AED's is formulated using probability of survival to hospital discharge. "Supply effect" is expressed as demand density multiplied by this probability of saving, and the locations of AED's are optimized by maximizing supply effect in the entire region. The results show that optimal locations tend to cover the center of the city and that the supply effect is approximately proportional to the number of AED's.

Keywords: AED (automated external defibrillator), 密度分布 (density distribution), 最大被覆 (maximal covering), 空間パターン (spatial pattern), 生存退院率 (probability of survival to hospital discharge)

1. はじめに

AED (自動体外式除細動器) は、突然的な心停止状態への救急措置に有効とされる医療機器である。取り扱いが簡便で、既に一般市民による使用が認められている状況のなか、全国の自治体には設置に向けた早急な対処が求められている。現状の AED の設置場所としてその多くは駅、空港やイベント会場などの大規模な集客施設、もしくは医療機関が主となる傾向がみられる。人命に直接的に関わる事態である AED の設置に際しては、地域の社会的そして経済的な事情も加味されなければならない。そこで殊に強調されるべきは、より多くの装置を用意する努力もさることながら、装置の配置地点に関する慎重な検討であろう。

AED 設置に関する最近の研究をみると、Malcom et al. (2004) は、地域の人口密度と心停止が発生した場所との関連についての調査を行った。その結果として、人口密度が高い場所であるほど、病院外で心停止が発生する割合が減少する一方で、自宅外で発生する割合については増加する傾向があったことが報じられている。また、Crocco et al. (2004) では、集客施設で必要とされる、AED の設置数に関する数理モデルの提案がなされた。AED の効果が注目されつつある現在の社会的状況を鑑みても、その設置地点に関する検討が今後求められていくであろうし、この要求に応える最も直接的で有効な方法は、理論的に得られた設置地点が示されることであろう。ならびに GIS の分野としても、危機情報の管理および解析等の面から、より一層の貢献が望まれるものと考えられる。

そこで本稿では、AED の最適配置地点に関する考察をおこなうにあたり、主に以下の点を目的とし

片岡：〒277-8568 千葉県柏市柏の葉 5-1-5
東京大学空間情報科学研究センター
Center for Spatial Information Science, University of Tokyo
E-mail ukataoka@csis.u-tokyo.ac.jp

ている。まず、心停止発生地点（需要点）の空間パターンにもとづいた地域全体の装置の需要量を表す密度分布に対して、装置との位置関係で決定される救命確率をもって各配置地点（供給点）周辺で重み付けした「供給効果」を最大化する配置問題を提案する。さらに、実際の心停止発生地点をデータとして用いた場合の最適配置地点を示すとともに、地域の救命確率および「供給効果」の地理的分布についても表すことで、本稿で提案される方法の有効性を実証する。

2. 需要量の空間パターンと供給効果の定式化

本節では、まず地域における需要量の空間パターンを把握する。さらに、装置の設置に伴って装置周辺に生じると考えられる救命確率を定義し、需要量に救命確率を加味した「供給効果」を定式化する。

2.1. 需要点にもとづく密度分布の推定

通常、需要量を把握する際には、対象となるものの特性を考慮し、それに応じて適切な方法が用いられる。ここでは、実際に装置が必要とされた地点、つまり過去に心停止が発生した地点をAEDの「需要点（各点で均質）」とし、需要点の分布パターンにより地域の需要量が決定されるものと考えた。とはいえ、心停止が発生した地点のみに需要があるとするのは、同一地点で再発する可能性を考えても妥当性を欠く。よって、過去に数多く発生した地域を潜在的に需要が高い地域であると見なすこととする。従って各需要点周辺の需要量は、需要点で最大であり、周囲に拡がるにつれて減衰する確率密度分布で表現することができよう。ただし、ごく短期間のうちでは、心停止の発生が偶然であることを免れない。そのため需要点数については、需要を代表するのに十分な数である必要がある¹⁾。

需要量の地域全体の確率密度分布についてはカーネル法を用いて推定する (Silverman, 1986)。カーネル法とは、点分布を形成する各点の中心に、カーネルと呼ばれる密度分布の山を置き、それら全てのカーネルの合計をもって全体の確率密度関数を求める密度推定法である (図1参照)。心停止発生地点は、

自宅、医療・福祉施設等といった一般に日常生活の多くを過ごす場所での発生、および外出先での発生からなると考えられる。つまり、地域の人口分布のみならず他の要因をも加味した、AEDの潜在的な需要量を表す密度分布がカーネル法で推定される。

なお、本稿ではカーネル関数のパラメータとなるバンド幅選択に際して、Sheather and Jones (1991)の方程式解によるプラグイン法を採用する²⁾。

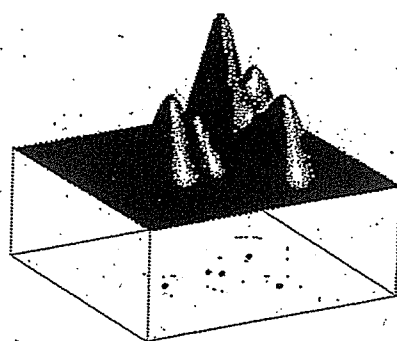


図1 点分布と密度分布

カーネル法によると、平面上の点 (x, y) における確率密度関数 $\hat{f}(x, y)$ は、点分布の各点 i の座標 (x_i, y_i) を用いて以下のように表せる (Bowman and Azzellini, 1997)。

$$\hat{f}(x, y) = \frac{1}{n} \sum_i K\left(\frac{x-x_i}{h_x}\right) K\left(\frac{y-y_i}{h_y}\right) \quad (1)$$

$(K(\cdot))$: カーネル関数, n : 点の総数,

h_x, h_y : 各座標値に関するバンド幅)

また式(1)で、カーネル関数として正規分布を用いて、各需要点の座標が $(x_{1,i}, y_{1,i})$ とすると、点分布にもとづく地域全体の確率密度関数は、以下の式(2)のようにできる。

$$\hat{f}(x, y) = \frac{1}{2\pi h_x h_y n_1} \sum_{i=1}^{n_1} \exp\left(-\frac{(x-x_{1,i})^2}{2h_x^2} - \frac{(y-y_{1,i})^2}{2h_y^2}\right) \quad (2)$$

(n_1 : 需要点数)

式(2)は、需要点から推定された確率密度関数であると同時に、需要量の地理的分布を表す空間パターンとして捉えられる。

2.2. 供給効果の定式化

次に、設置される装置周辺の需要量に対して、装置による効果の程度を表す「供給効果」について考える。何よりも装置が有効に使用されるためには、心停止発生時点から限られた時間内に処置が施される必要がある。その際、各施設立地点から一定距離内にある領域に含まれる需要量を最大化する配置問題として知られる。最大被覆問題 (Maximal Covering Location Problem) があげられる。最大被覆問題は、Church and ReVelle (1974) により提案された後、救急施設などを対象とした離散平面における配置問題として発展するなかで、様々な解法がこれまでに提案されてきた (Brotcorne et al., 2003)。ただ先述のとおり、地域の需要量は点分布であるよりも、各地点の潜在的な需要を表す確率密度分布で表現される方が適切であろう。さらに AED の場合では、心停止に始まり処置に至るまでの時間に、救命確率が大きく依存するとされるため (American Heart Association, 2000)、被覆される領域内においても場所によっては効果の程度は著しく異なると考えられる。同様に Alsalloum et al. (2006) は、救急車の配置モデルにおいて、通常用いられる一様に被覆される領域を、定めた制限時間内に被覆される確率に置き換えた方法を提案しているが、AED の場合においても類似した状況にあると考えてよい。

そこで、装置から距離 d の地点における救命確率は、経過時間に依存する救命確率を、距離的な問題に置き換えた以下の関数型で近似できるとする³⁾。

$$g(d) = \exp(-ad^2) \quad (3)$$

(a : パラメータ)

本稿では、心停止が発生した現場に居合わせた「救助者」が装置を取りに行くことを前提としている。救助者の置かれた状況を見ると、設置場所がわかっている装置については、装置に向かって取りに行くことが可能である。ところが、当然ながら救助者は装置の全設置点を把握しているわけではない。事前知識も不確実であることに加え、このように一刻を争う事態では行動が直感に大きく左右され、状況として通常の施設選択の場合とは異なるともいえ

る。つまり、救助者は必ずしも現場から最も近くにある装置を常に把握しているわけではなく、ある地点において装置が心停止にもたらす効果は、心停止発生地点から最も近くにある装置との関係性のみに影響するとは限らない⁴⁾。すると装置の「認知度」の拡がりや、装置からの距離に伴う救命確率の減衰に代替されているとしたときに、近くに設置されている装置が多いほど、設置に伴う効果は上昇するものとも考えられる。以上より、ある地点における AED 設置によって与えられる影響の程度は、各装置の救命確率の和として説明できるものと仮定する。

よって式(3)を用いると、地点 (x, y) での救命確率の和は、各装置の設置地点の座標 $(x_{2,j}, y_{2,j})$ を用いて以下のようにできる。

$$g(x, y) = \sum_{j=1}^{n_2} \exp\left(-a\left((x-x_{2,j})^2 + (y-y_{2,j})^2\right)\right) \quad (4)$$

(n_2 : 装置の設置数)

以上をふまえ、装置の効果が期待される装置周辺において、救命確率を需要量に加味することで得られる、需要と装置の効果に依存する量を本稿では「供給効果」と定義する。なお、理解を容易にするために一次元で考えると、需要量と「供給効果」の関係は以下の図2のようになる。

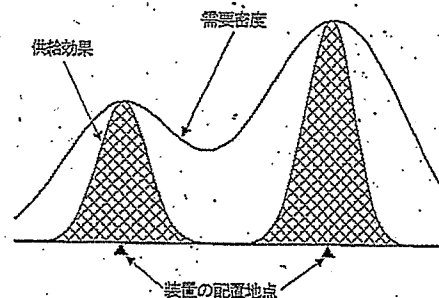


図2 需要密度に対する供給効果

これより、ある地点における需要に対して救命可能な量となる「供給効果」は、その地点の需要量に救命確率の和を乗じたもので表されることから、「供給効果」の地域全体の総和 (図2の網掛け部分) は、以下の式(5)のように定式化される。

$$\begin{aligned}
L_1(x, y) &= \iint \hat{f}(x, y) \cdot g(x, y) dx dy \\
&= \iint \left\{ \frac{1}{2\pi h_x h_y n_1} \sum_{i=1}^{n_1} \exp\left(-\frac{(x-x_{1,i})^2}{2h_x^2} - \frac{(y-y_{1,i})^2}{2h_y^2}\right) \right. \\
&\quad \left. \times \sum_{j=1}^{n_2} \exp\left(-a((x-x_{2,j})^2 + (y-y_{2,j})^2)\right) \right\} dx dy \\
&= \frac{1}{4n_1 \sqrt{1+2ah_x^2} \sqrt{1+2ah_y^2}} \\
&\quad \times \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \exp\left(-\frac{a(x_{1,i}-x_{2,j})^2}{1+2ah_x^2} - \frac{a(y_{1,i}-y_{2,j})^2}{1+2ah_y^2}\right) \\
&\quad \times \operatorname{erf}\left(\frac{x-x_{1,i}+2ah_x^2(x-x_{2,j})}{h_x \sqrt{2+4ah_x^2}}\right) \\
&\quad \times \operatorname{erf}\left(\frac{y-y_{1,i}+2ah_y^2(y-y_{2,j})}{h_y \sqrt{2+4ah_y^2}}\right)
\end{aligned} \tag{5}$$

式(5)において、各項における積分定数については省略した。また、 $\operatorname{erf}(\cdot)$ は誤差関数をあらわし、以下のように定義される。

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \tag{6}$$

3. 供給効果を最大化する配置問題

本節では、2節で定式化された「供給効果」を最大化する、AEDの最適配置問題を提案する。

まず、式(5)で積分区間を無限領域としてよいことから、以下の式(7)を得る。

$$\begin{aligned}
L_2(x_{2,1}, x_{2,2}, \dots, x_{2,n_2}, y_{2,1}, y_{2,2}, \dots, y_{2,n_2}) \\
= \frac{1}{n_1 \sqrt{1+2ah_x^2} \sqrt{1+2ah_y^2}} \\
\times \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \exp\left(-\frac{a(x_{1,i}-x_{2,j})^2}{1+2ah_x^2} - \frac{a(y_{1,i}-y_{2,j})^2}{1+2ah_y^2}\right)
\end{aligned} \tag{7}$$

式(7)で定式化された「供給効果」を、このように設置場所の候補地が全くない状況で、単純に最大化しようとするならば、需要量の多い領域を幾度も被覆してしまう可能性がある。それは、同一地点に装置を複数台設置するという結果を得ることを意味する。ところが、ある装置のごく近辺で、複数の心停止患者がほぼ同時に発生する状況は想定しづらく、いわば「容量の制限がない」と条件付けられるAEDの問題に限っては、同一地点に解が重複して

得られる状態は好ましくない。そのため、ここでは装置間の距離についての制約を設けることにより、救命確率の和が1を超える地点が存在しないものとする⁵⁾。つまり、装置の効果が期待できる各装置の設置地点からの限界の距離を d_{\min} とおき、ある装置から d_{\min} の距離内には他の装置が設置されることがないものとする。よって、各装置間の距離を $d_{s,t}$ で表し、制約条件として式(9)とおくことで、最適化されるとき各配置地点の座標値は、式(7)が最大化されるとき解集合として与えられる。つまり、本稿で定式化された配置問題は、以下の多変数最大化問題に帰着する。

$$\text{Maximize } L_2(x_{2,1}, x_{2,2}, \dots, x_{2,n_2}, y_{2,1}, y_{2,2}, \dots, y_{2,n_2}) \tag{8}$$

subject to

$$d_{s,t} = \sqrt{(x_{2,s} - x_{2,t})^2 + (y_{2,s} - y_{2,t})^2} \geq d_{\min} \quad \forall s, t \tag{9}$$

4. 適用事例

前節で提案されたAEDの配置問題の適用事例を扱う。需要点として弘前市の心停止発生地点のデータを用いた場合の、AEDの最適配置地点を示す。

4.1. 心停止発生地点とその密度分布

需要点となる心停止発生地点としては、弘前地区消防事務組合が過去3年間に扱った内因性心肺停止で、原因疾患が心疾患と推定された411例を分析に用いた。

始めに、当データ内の属性として含まれている住所表記から位置情報を取得した⁶⁾。各需要点の位置と、式(2)より得られる各地点の確率密度を表す等値線(間隔は 1.0×10^{-9})を図3に示す。なお、カーネル関数のバンド幅選択は、2.1.で述べたように、Sheather and Jones (1991)によるプラグイン法を採用した。

図3の需要点にもとづく密度分布について見てみると、弘前駅や弘前城の南側に隣接する市役所から少し離れた地域、および土手町と呼ばれるメインストリート付近を中心として、心停止発生密度が非常に高いとされる地域が拡がっていることが確認された。また、市街地部周縁である地域帯で密度分布が

大きく変化しており、概して人口分布の変化に対応しているものと理解される。

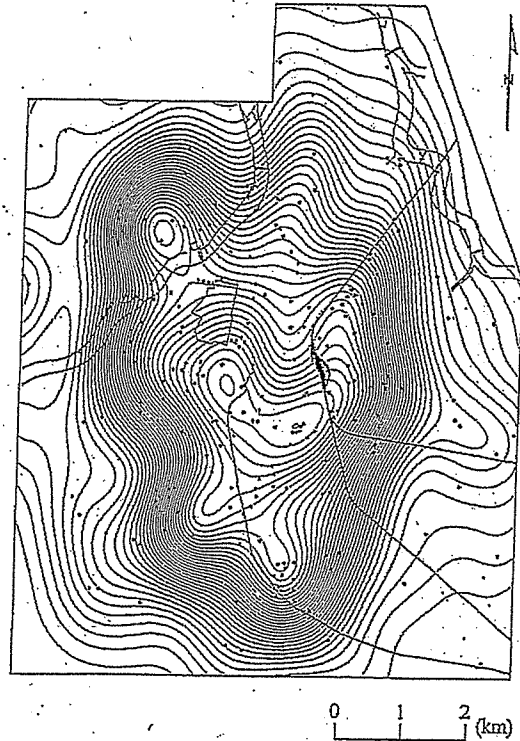


図3 心停止発生地点とその密度分布
(図中の表記については以下の通り。●: 需要点, および需要量密度を示す等値線。背景地図として, 河川, 鉄道, 弘前城を表記。)

図3 心停止発生地点とその密度分布

4.2. 処置までの時間にもとづく救命確率式

本稿では、式(3)で定式化された装置からの距離にもとづく救命確率式を求めるために、処置までの時間に依存する「生存退院率」を用いることとする。「生存退院率」とは処置による蘇生の成功率を意味し、倒れてから除細動の開始までに要する時間が1分経過するごとに約10%低下するとされることから⁷⁾、AEDのような問題においては国内外を問わず重要な知見とされている。これにもとづき、10分経過した時点で救命確率がほぼ0になるように式(3)のパラメータを求めた ($a=4.32 \times 10^{-5}$)。なお、実際のパラメータの決定においては、本稿で設定された仮定にもとづき、分速80mで装置に向かっ

て直進するとし、往復にかかる時間を考え5分経過した400mの時点で救命確率が 1.0×10^{-3} となるとした⁸⁾。

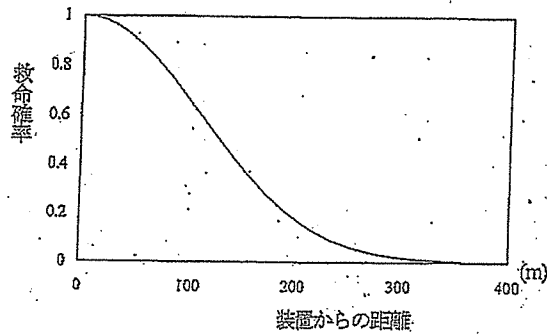


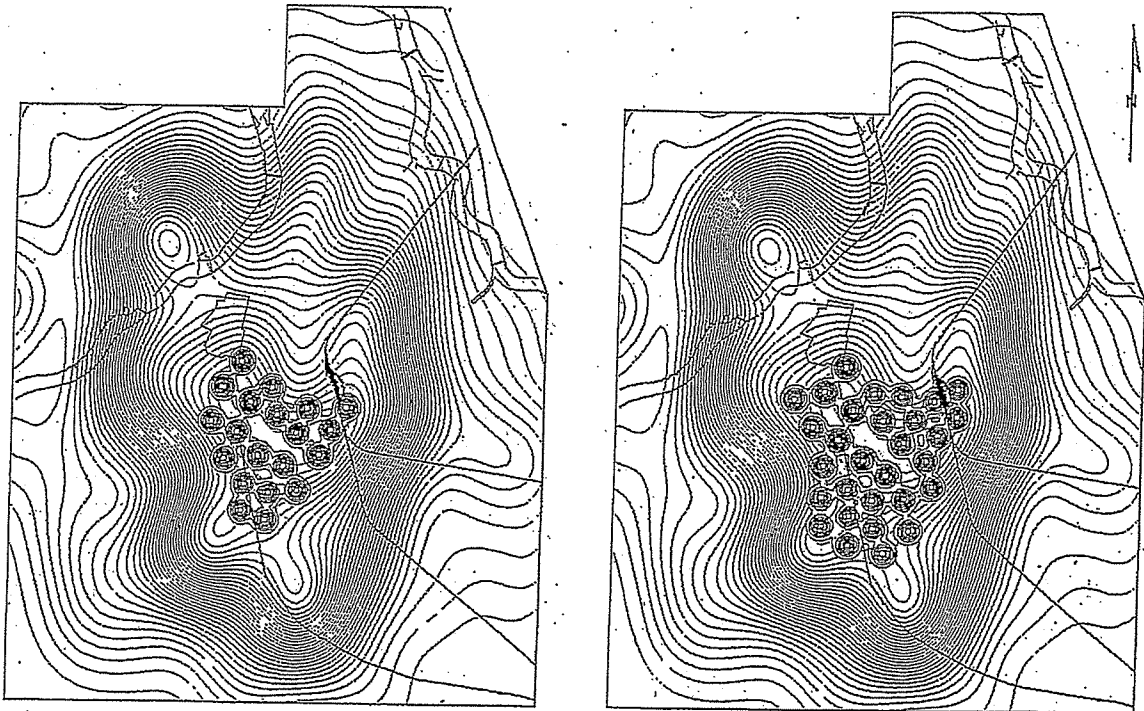
図4 各装置からの救命確率

4.3. AEDの最適配置地点

次に式(8)を用いて、AEDの設置数を1点から30点までの間で所与とし最適地を求める。前項での考察をふまえて、式(9)において $d_{min}=400m$ とし、式(8)を多変数の制約条件付き最大化問題として数値計算をおこなった結果⁹⁾、分析の対象とした設置数の全ての場合で各最適地が対象地域内に得られた。そのなかで、設置数が「20点」、「30点」のときの各最適地、および任意の地点における「供給効果」の密度分布を示す等値線(間隔は 5.0×10^{-3})を図5に、また、装置による救命確率の密度分布を示す等値線(間隔は 2.0×10^{-3})を図6に示す。参考のために、図3の需要点の密度分布についても各図に併せて示した。

得られた最適配置地点の分布の特徴として、4.1.で装置の需要密度が高いとされる地域において、設置数が増加するに伴い、最適地が密度分布の「山」の頂点付近からその周辺に向かって、満遍なくカバーするように得られるという傾向が表れている。

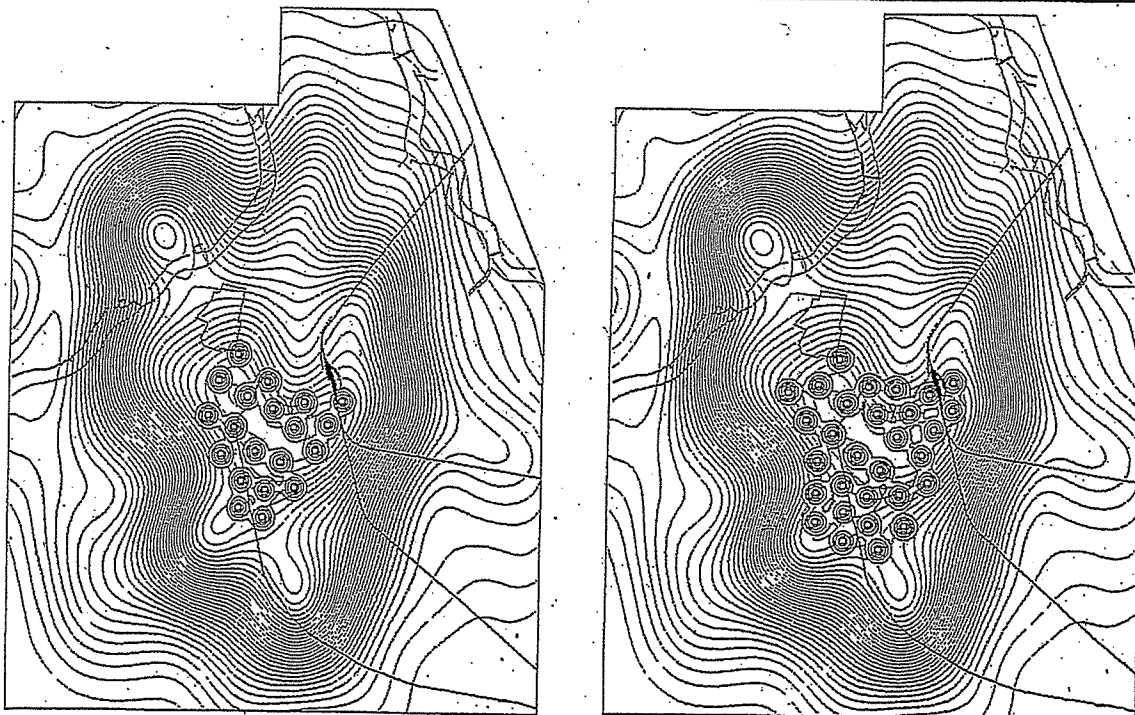
また、図6の救命確率の密度分布は、需要分布に依らない各地点での装置の効果を示しており、各等値線が救命確率で表された装置までの一定の到達距離の分布を意味している。従って、一刻を争う状況下では、視覚化された判断基準として特に有効な手がかりとなることも期待できるであろう。



20点配置

30点配置

図5 AEDの最適配置地点と供給効果の密度分布



20点配置

30点配置

図6 AEDの最適配置地点と救命確率の密度分布

0 1 2 (km)

(図5,6中の表記については以下の通り、□: AEDの最適配置地点, および各々の密度(濃線), 需要量密度(淡線)を示す等値線. 背景地図として, 河川, 鉄道, 弘前城を表記.)

なお、図5と図6における密度分布が非常によく似た様相を呈している理由は、最適地が得られる範囲内で需要量に大きな変化がないことによる。

4.4. 設置数に伴う効果

図7に、各設置数と式(8)の目的関数値との関係を示す。例えば、「供給効果」を表す目的関数値が0.01となるときでは、最適地に設置された装置によって地域全体の需要量の1%がカバーされることを示している。

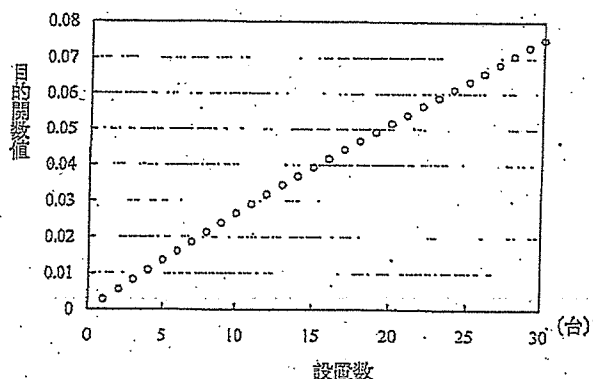


図7 設置数と目的関数値

図7を見る限りでは、設置数の増加に従って「供給効果」はほぼ線型に増加することから、少なくとも設置数が30点までにおいては、需要がほぼ同等に高い地域に装置が割り当てられていると考えられる。「供給効果」の変化の割合は、結局のところ需要量の分布状況に大きく依存するものであるが、一般的には、設置数が少ない初期の段階では急激に増加し、その後緩やかに増加を続けることが予測できる。言い換えれば、設置数が増加するにつれ、需要密度の高い地域から需要密度の低い地域へと、装置で新たにカバーする領域が移行して行くといえる。

さらにここで得た結果は、より多くのAEDを地域全体として有することで、地域医療に貢献できることを意味するとともに、供給効果の上昇の程度が鈍らないこの範囲の設置数においては、AEDの配置の検討が特に重要とされるべきであることを示唆している。

従来の限られた施設以外にも設置が検討される状

況にあるなか、適切な配置地点についての議論は一層重要性を増すことが予想される。現実には、すべての最適地にAEDを配置することは非常に困難ではあるが、問題の重要性からも現状の配置状況の適切性は常に問われるものである。そのためには、何よりも理論的な最適配置地点、およびその効果については厳密に示されている必要がある。

5. おわりに

本稿では、AEDの配置地点に対する検討が必要であるとの認識にたち、実際の心停止発生地点からなる需要点より推定された確率密度分布に、生存退院率にもとづく救命確率を加味したAEDの最適配置問題を提案した。なかでも、地域全体の需要量に対する「供給効果」を定式化するとともに、配置地点の座標値に関する多変数最適化問題としてその解を数値的に得た。AEDの最適配置問題では、装置がカバーできる範囲が限られているとともに、装置までの距離が非常に重大な要因である。本稿では、対象地域内で密度分布が比較的高い場所が広範囲に広がっていたために、装置の最適配置地点は特定の地域に集中して得られる傾向が顕著であった。しかし、心停止発生地点に影響を及ぼすとされる、高齢者の人口分布や、人的そして物的な集積度に大きな偏りがある地域では、地域全体としての需要量の地理的な変化が激しくなる。その際には、求まる最適地点もある程度分散するといった、今回とは異なる結果を得ることも予想される。要は、AEDの最適配置地点は直感的に導き出すことが困難であるだけに、十分な検討が望まれる問題であると考えてよいであろう。

また、AEDの場合に、設置に伴う用地取得や周辺環境への配慮などの問題は、通常の施設立地の場合ほど重大ではないと考えられる。装置に関する維持管理などの問題はあっても、明らかに立地不可能であるような条件を除けば、この点で大きく性質が異なることが理解される。つまり、現段階において候補地として装置の設置地点を事前に限定するのではなく、連続的な平面上に得ようとする本稿の試みは、現状に対しても有意義で実効性の高い問題と

位置付けて良いと言える。

最後に、今後の課題について述べる。AEDの設置が希望される施設として、「駅」、「学校」、「公共施設」などの、現段階で積極的に設置が進んでいる施設を挙げる声が一般には大きいものの、「コンビニエンスストア」あるいは「交番」への設置を希望する意見も少なくない¹⁰⁾。「コンビニエンスストア」や「交番」は、地域住民に比較的認知されている施設であり、例えば深夜のような時間を問わない緊急時への対応の面でも期待できる。また、市街地のみならず居住地域も含めて幅広く分布していることから、有効なAEDの配置地点として、需要量との関係性を含めた検討をおこなうことも重要であると考え、これらをふまえて今後は、既存施設へのAEDの設置を視野に入れた配置問題についても考察していきたい。

謝辞

本研究をおこなうにあたり、国立保健医療科学院技術評価部丹後俊郎部長、東京大学空間情報科学研究センター丸山祐造助教授および東京大学大学院工学系研究科都市工学専攻貞広幸雄助教授より貴重なコメントを頂いた。また、弘前地区消防事務組合には救急車による搬送データの提供をはじめ、多大なご協力を頂いた。

本研究は、厚生労働科学研究費補助金（健康科学総合研究事業）「地理及び社会状況を加味した地域分析方法の開発に関する研究」（代表：浅見泰司）の支援を受けた。ならびに、東京大学空間情報科学研究センターの研究用空間データ利用を伴う共同研究（研究番号75）として、以下のデータを利用した。（株）ゼンリン提供：ZmapTownII 青森県弘前市 記して謝意を表す。

補注

1) 本稿の適用事例で扱っている心停止発生数に関しては、季節などの時期による発生数の変動の可能性や、設置するAEDの数との関係を吟味し、過去3年分の411例を本稿で分析を行うのに十分なサンプル数とした。実際の密度分布を

見る限りでも、市街地のなかで人口が多い居住地域ほど密度が高くなる傾向を示すことなどから、需要量をよく代表しているものと判断した。

- 2) Sheather and Jones のプラグイン法では、式(1)の f の推定値が、理論的な考察によるバンド幅からではなく、データから推定されるという点に特徴がある。Simonoff (1996) では、一般に最良なバンド幅選択法が発見されていないことを踏まえつつも、プラグイン法が「最先端」の方法であるとともに、多方面で有効な選択法であると紹介されている。
- 3) 処置に至るまでの時間の経過と後生の生存退院率との関係に加えて、道路状況などの地理的な条件を考慮した結果、距離の増大に伴う効果の減衰を示す関数式として適当であると判断した。
- 4) 例えば、本稿で想定されるAEDの場合とは異なり、救急車の出動が前提となるような事例では、発生地点から最も近くにある立地点が選択されるのが妥当であろう。
- 5) 計算上、被覆される領域外でも1を超える地点は存在するが、本稿でも述べるように影響を無視できるほど微少であることから、装置から一定距離以上の地点での効果は無視できることが前提とされている。
- 6) 位置情報の取得に際しては、東京大学空間情報科学研究センターが提供するアドレスマッチングサービスを用いた。
- 7) American Heart Association (2000) によれば、心停止発生から1分経過するにつれ、生存退院率が約7%~10%で減少するとされている。これに基づいて本稿では、10分経過した時点で生存退院率がほぼ0になるとした。
- 8) AEDの効果が期待できる範囲がごく限られていることに加え、対象地域では道路密度の粗密が分析に大きな影響を与えないと考えられることから、本稿では直線距離を採用した。
- 9) 最適化計算では、MATLABのOptimization Toolboxを使用した。

- 10) AED (自動体外式除細動器) 普及を推進する会
による「AED の設置場所に関するアンケート」
より < [http://www.aedjapan.com/vote/qnaire2.
cgi](http://www.aedjapan.com/vote/qnaire2.cgi) >.

参考文献

- American Heart Association (2000) Guidelines 2000 for cardiopulmonary resuscitation and emergency cardiovascular care: International consensus on science, *Circulation*.
- Alsalloum, O.I. and Rand, G.K. (2006) Extensions to emergency vehicle location models. *Computers & Operations Research*, 33, 2725-2743.
- Bowman, A. and Azzelini, A. (1997) *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press.
- Brotcorne, L., Gilbert Laporte, G. and Semet, F. (2003) Ambulance location and relocation models. *European Journal of Operational Research*, 147, 451-463.
- Church R.L. and ReVelle C.S. (1974) The maximal covering location problem. *Papers of the Regional Science Association*, 32, 101-18.
- Crocco, T.J., Sayre, M.R., Liu, T., Davis, S.M., Cannon, C. and Potluri, J. (2004) Mathematical determination of external defibrillators needed at mass gatherings. *Prehospital Emergency Care*, 8, 292-297.
- Malcom, G.E., Thompson, T.M. and Coule, P.L. (2004) The location and incidence of out-of hospital cardiac Arrest in Georgia. *Prehospital Emergency Care*, 8, 10-14.
- Sheather, S.J. and Jones, M.C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society series B*, 53, 688-690.
- Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Simonoff, J.S. (1996) *Smoothing Methods in Statistics*. New York: Springer-Verlag.

空間ドキュメント管理システムの設計と実装

白石 陽 有川 正俊 相良 毅 浅見 泰司

2007
『DEWS』B.7-10

空間ドキュメント管理システムの設計と実装

白石 陽[†] 有川 正俊[†] 相良 毅[‡] 浅見 泰司[†]

[†] 東京大学 空間情報科学研究センター 〒277-8568 千葉県柏市柏の葉 5-1-5

[‡] 東京大学 生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

E-mail: [†] {siraisi, arikawa, asami}@ccsis.u-tokyo.ac.jp, [‡] sagara@tkl.iis.u-tokyo.ac.jp

あらまし 本稿では、住所情報を含むデジタルドキュメントを、直感的なインタフェースで扱うことのできるシステムとして、空間ドキュメント管理システム (SDMS: Spatial Document Management System) を提案する。SDMS は、ドラッグ&ドロップという簡単な操作で、即座に、ドキュメント中の住所情報を、点として地図上に表示することができる。SDMS では、非定型の一般的なデジタルドキュメントから、住所情報を抽出し、アドレスマッチング技術を利用して POI (Point of Interest) を生成し、地図上に表示する。SDMS を用いることにより、多くの一般的なユーザが、簡単に、デジタルドキュメントを空間的に整理し、地図化することが可能となる。

キーワード 空間ドキュメント, 位置情報, GIS, ジオコーディング, ジオパース, Point of Interest (POI)

1. はじめに

近年の位置情報関連技術の進歩は目覚ましい。位置測位技術の発達、GPS 搭載携帯電話の普及、位置情報サービスの高度化、地図 API の公開、Web コンテンツへの緯度経度情報の埋め込みなど、位置情報を扱う様々な枠組みが整ってきている [1,2,3,4]。緯度経度という絶対的な位置情報は、こうしたサービスを支える基盤となる情報であるが、多くの人にとっては、ただの数字の列にしか見えないため、必ずしも人間にやさしい表現とは言えない。それに対して、住所や地名などの位置情報は、そのままでは地図に表示することはできない間接的な表現であるが、人間の理解しやすいものである。日常的に利用される自然言語による表現は、コミュニケーションにおいても効果的であり、案内文書、報告書、店舗一覧、施設一覧、光化学スモッグ発生情報など多くのデジタルドキュメント上で利用されている。今後、緯度経度のような“直接位置参照情報”が目に見えない形で埋め込まれたドキュメントが増えていくことは容易に想像できるが、依然として、住所や地名などの“間接位置参照情報”を含む様々なドキュメントが、多くの場面で利用されていくことは間違いない。本研究では、こうした住所や地名などの位置情報を含むデジタルドキュメントを「空間ドキュメント」と呼ぶ。インターネット上の Web ページ、メールの添付ファイル、自分のコンピュータの中のファイルなど様々な空間ドキュメントが存在する。そのファイル形式も、HTML, Word, Excel, TEXT など多様である。

本稿では、一般ユーザが日常的に利用している、このような多種多様な空間ドキュメントを、容易に管理、検索、地図化できるツールとして、空間ドキュメント管理システム (Spatial Document Management System,

以下 SDMS と呼ぶ) を提案する。SDMS は、ドラッグ & ドロップという単純な操作で、住所や地名などを自動的に抽出し、緯度経度を算出して、Point of Interest (POI) として地図上に表示することができる。SDMS を用いることによって、インターネット上やパソコン内に存在する様々なドキュメントを空間的に整理し、把握することができる。また、地域行政、保健医療、公衆衛生、危機管理などの現場において、こうした場所情報を含むドキュメントは、貴重な情報源であるにも関わらず、地図化する簡易な手段がなかったために、死蔵されている場合も多く、空間ドキュメントを扱うことのできる SDMS に対する潜在的なニーズは高いと考えられる。

以下、本稿では、2 章で関連研究を挙げ、提案システムの位置付けについて述べた後、3 章で提案システムの説明を行う。4 章では、動作例を説明した後、考察と今後の課題を述べる。最後に、5 章でまとめを行う。

2. 関連研究

2.1. 地理情報システムとの比較

住所、地名、郵便番号などの間接位置参照情報を緯度経度などの直接位置参照情報に変換することをジオコーディング (geo-coding) と言う。住所を緯度経度に変換する処理は、特に、アドレスマッチングと呼ばれる。それに対して、ドキュメント中から住所や地名などの地理情報を抽出することをジオパース (geo-parse) 処理と呼ぶ。

空間情報を管理できるシステムとして地理情報システム (GIS: Geographic Information System) [5] がある。GIS は、様々な空間情報の管理、解析、統合、可視化を行うことのできる便利なツールである。ジオコード

ィングは、GIS 分野でも古くから重要な課題であり、現在では、商用のサービス、あるいは、フリーの API を介してジオコーディングサービスを利用できる。既存の GIS では、入力データとして表として整理された、すなわち構造化された空間データを必要とする。多くの研究者は、GIS を利用する前に、緯度経度を含む表形式のデータを作成する必要がある。元データが住所である場合には、CSV などの表形式で住所情報を整理した後、ジオコーディングサービス（例えば、CSV アドレスマッチングサービス[6]）を利用することにより、緯度経度の付与された構造化された空間データを準備することができる。このように GIS では、GIS を利用する前の手順が、住所情報の収集・抽出も含めて、非常に手間のかかる仕事となっている。

空間ドキュメントのファイル形式は、HTML, Word, Excel, TEXT など多様であり、半構造や未構造の空間データである。既存の GIS では、構造化された空間データを必要とするため、こうした非定型（半構造あるいは未構造）のドキュメントを、そのままの形で直接扱うことはできない。多数のドキュメントを対象とする場合、ドキュメント中に多数の住所が含まれている場合には、大量の住所情報を手動で抽出し、構造化データとして整理するのは現実的ではない。

本稿で提案するシステムでは、空間ドキュメント中から住所を抽出して、緯度経度に変換し、地図上に表示できる必要があるが、空間ドキュメントに対するこうしたジオコーディングとジオパースの処理を、ユーザの手を返さず、自動化できると望ましい。また、GIS は、高度な機能を持つため、専門家にとっては強力なツールとして利用できるが、その反面、一般のユーザが操作を習得するためには訓練が必要であり、日常的に使いやすいものとは言えない。

空間ドキュメントを管理するためのシステムの要件としては、(a)種々の形式のファイルを扱えること、(b)非定型のドキュメントから自動的に住所を抽出して緯度経度を算出できること、(c)算出した緯度経度情報をドキュメント中に埋め込むこと、それに加えて、(d)それらの処理を簡単な操作で実行できることが必要である。

2.2. 住所情報の抽出とドキュメントの構造化

本研究では、住所情報を抽出するために「住所地名テーブル」を作成し、ドキュメント中の文字列とテーブル内の地名語とを比較することでジオパース処理を行っている。WWW 文書から住所情報を抽出し、文書を構造化したり、地理情報を付与 (geo tagging) したりすることで、位置指向の検索やコンテンツの整理が

可能となる[7,8,9]。Amitay らは、地名辞典 (gazetteer) を利用して、Web コンテンツを地名語 (place name) と関連付ける手法を提案している[7]。横路らは、WWW 文書のための位置指向検索手法を提案しており、文書中から住所情報を抽出し、文書の構造化を行っている[8]。その際、住所名を辞書に加えて形態素解析を行い、各形態素を住所辞書と比較することで、住所抽出を行っている。住所辞書には、都道府県名、市区町村名、町字などを単独もしくは上位の住所階層からの省略のない形で登録している。そして、住所を正しく抽出するために、「都道府県名から省略なく正確に記述されている」、「都道府県や市区など住所を明確に示す接尾辞がついている」などのルールを適用することで、住所抽出の精度を向上させようとしている。形態素解析のみを利用する住所抽出では、その精度が辞書に登録されている住所情報に依存し、登録された文字列が長いと検索漏れが生じ、文字列が短いと住所認識の精度が下がる可能性がある。長屋らは、国土交通省の提供する「街区レベル位置参照情報」[10]に基づいて、住所抽出のためのキーワードを生成し、キーワード群の状態遷移表を用いて文字列マッチングを行っている[9]。ただし、キーワードとしては、「丁目」以降が削除された住所文字列を選択し、「丁目」以降の文字列は構文解析によって処理している。一般的なドキュメントに含まれる住所情報では、上位の住所階層が省略されることが多いため、そういった住所情報を漏れなく抽出できることも重要である。地域行政、地方自治体、地域ポータルといった対象を考えると、上位階層の住所表記は省略される可能性が高い。本研究では、文字数の比較的短い住所文字列をテーブルに登録することで住所抽出の再現率を向上させ、アドレスマッチングの変換結果の精度に基づいて最終的な住所判定を行うことで住所抽出の精度を向上させようとしている。

さらに、最近では、GeoRSS[3]や Microformats [4]など、ドキュメント中にタグを用いて属性情報を埋め込む枠組みについての議論が行われている。住所や地名などを属性としてドキュメント中に埋め込み、ドキュメント中のオブジェクトと関連付けることができれば、その位置情報を活用して様々な応用が期待できる。本研究においても、ジオパース処理とジオコーディングの結果である位置情報をタグの属性としてドキュメント中に埋め込むことで、ドキュメントの管理や POI の検索を行う。位置情報タグを用いてドキュメントを半構造化する手法については、先行研究[11]においても提案し、プロトタイプシステムの実装を行っているが、本稿では、よりユーザビリティの高いシステムの実現を目指す。

3. 空間ドキュメント管理システム(SDMS)

3.1. 人間中心型システムとしての設計

本研究は、既存の GIS のような専門家が利用するツールではなく、一般のユーザが日常的に利用できるツールの開発を目指している。基本的な設計方針は、次の通りである。

- ファイルのドラッグ&ドロップという単純かつ直感的な操作によって、ドキュメントの入力を簡便化する
- 入力された非定型のドキュメントに対して、自動的に、住所を抽出し緯度経度に変換する仕組みを開発する
- 抽出した緯度経度情報を、タグを利用してドキュメント中に埋め込むことで、空間ドキュメントの管理や検索に役立てる

空間ドキュメントに対するジオコーディングを自動的に行うためには、ジオパース機能が不可欠であり、ジオコーディング機能と連携させる仕組みが必要である。本稿では、ジオコーディング機能そのものについては、既存のサービスを利用するものとして、ジオコーディングのインターフェースとなる部分に焦点を当てる。当然のことながら、提案システムのジオコーディングの性質や性能は、利用するジオコーディングサービスに依存するが、人間中心型のシステムを開発するためには、ユーザとインタラクションを行う部分と、ジオコーディングプロセスとの仲介を行う部分の設計も重要である。

3.2. システム構成

図 1 に、SDMS のシステム構成を示す。

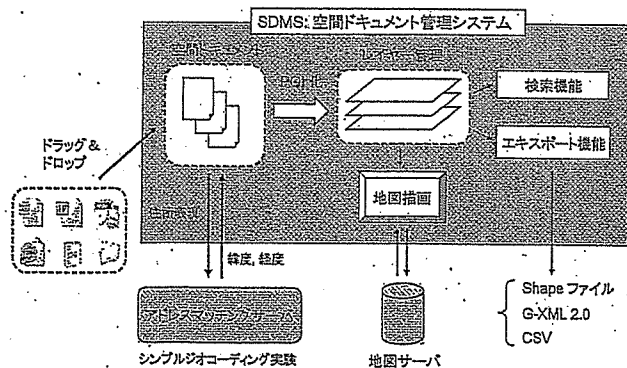


図 1: システム構成

SDMS は、ドラッグ&ドロップされたドキュメントから住所を抽出し POI に生成する部分 (POI 生成部)、抽出した POI をレイヤーとして管理する部分 (レイヤー管理部)、POI レイヤーと背景地図を描画する部分 (地図描画部) などから構成される。POI 生成部は、

ネットワークを介してリモートのアドレスマッチングサーバと通信することにより緯度経度変換を行う。地図描画部は、リモートの地図データサーバから要求した縮尺の地図画像データを取得し、背景地図として表示する。その他、レイヤー管理された POI に対する検索機能とエクスポート機能を持つ。

3.3. 処理の流れ

次に、SDMS における処理の概略 (図 2) を説明する。

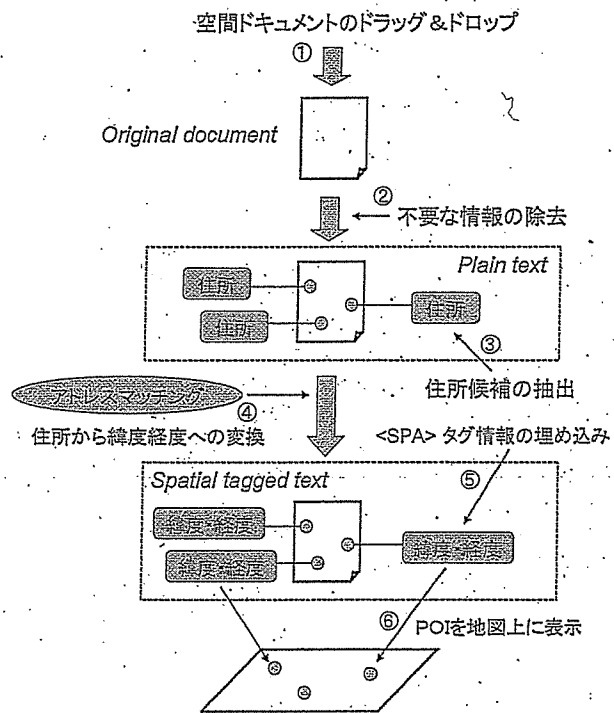


図 2: 処理の流れ

- ① ファイルアイコンをドラッグ&ドロップすることにより、ドキュメント (original document) を SDMS に入力する。この際、複数のドキュメントを含むフォルダアイコンをドラッグ&ドロップすることにより、複数のドキュメントをまとめて入力することも可能である。
- ② SDMS は、入力されたドキュメントから、不要な情報を除去して、プレインテキスト (plain text) を生成する。例えば、HTML ドキュメントを対象としている場合は、そこに含まれる HTML タグを取り除く。
- ③ プレインテキストを先頭から走査し、住所候補を見つけるごとに、その住所候補文字列をアドレスマッチングサーバに送信する。この部分の処理の詳細については、3.3.4 にて説明する。
- ④ アドレスマッチングサーバは、送信された文字列

に対してアドレスマッチングを行い、緯度経度情報を生成し、SDMSに返信する。

- ⑤ アドレスマッチングが成功した場合には、プレインテキスト中の該当する部分に、タグを埋め込み、緯度経度とともにサーバより返信された情報(3.3.1参照)をタグの属性として埋め込む。サーバに送信した文字列が住所でない場合や変換結果の精度が十分でない場合には、タグの埋め込みを行わない。SDMSは、このタグ付けされたテキスト中の各タグの属性情報をPOIとして管理する。複数のドキュメントがドラッグ&ドロップされている場合には、各ドキュメントが一つのレイヤーとして管理され、各ドキュメント中のPOIの集合が別々のレイヤーとして管理される。SDMSは、この空間タグ付けテキスト (spatial tagged text) の集合をPOIのデータベースとみなして、POIに対する種々の操作を行う。
- ⑥ SDMSは、空間タグ付けテキストを参照し、生成されたPOIを点分布として地図上に表示する。背景地図は、ネットワークを介してリモートの地図サーバから取得される。フォルダアイコンがドラッグ&ドロップされた場合には、各ドキュメント中のPOIがすべて地図上に表示されるが、ドキュメントごとに異なる色のアイコンとして表示される。

3.3.1. シンプルジオコーディング実験

SDMSは、アドレスマッチングサーバとして、東京大学空間情報科学研究センターで運用している「CSIS シンプルジオコーディング実験[12]」を利用している。シンプルジオコーディング実験は、日本語で記述された住所・地名・駅名・公共施設名を緯度経度に変換し、結果をXML形式で返す実験的なサービスとして運用されており、SDMSでは、住所候補を緯度経度に変換するために利用する。国土交通省より無償で公開されている「街区レベル位置参照情報[10]」をアドレスマッチングのためのテーブルとして利用しているため、変換精度は街区レベルまでであるが、無料で利用できるサービスとなっている。さらに、表記の揺れや省略に強いといった特徴も持っているため、SDMSも同様の利点を持つ。より詳細な号レベルの精度を持つ有償のサービスもあるが、非常に高価であり、対象地域が広がるほどコスト負担も大きくなる。多くのユーザーによる利用を考えると、街区レベル位置参照情報に基づくマッチングは、現時点での望ましい選択肢と言える。

シンプルジオコーディング実験は、RST (Representational State Transfer) サービスとして実装されており、住所候補文字列を含むクエリを送信する

と、アドレスマッチングの結果をXML形式で返信する。例えば、「千葉県柏市柏の葉5-1-5」という文字列を送信すると、次のXMLが返される。

```
<?xml version="1.0" encoding="EUC-JP" ?>
<results>
  <query>千葉県柏市柏の葉5-1-5</query>
  <geodetic>wgs1984</geodetic>
  <iConf>5</iConf>
  <converted>千葉県柏市柏の葉5-1</converted>
  <candidate>
    <address>千葉県/柏市/柏の葉/五丁目/1番</address>
    <longitude>139.935455</longitude>
    <latitude>35.901711</latitude>
    <iLvl>7</iLvl>
  </candidate>
</results>
```

図3: シンプルジオコーディングによる住所変換結果

<query> は送信文字列を表し、<converted>は送信文字列の中で変換に成功した文字列を表している。<iConf>は変換結果の信頼度を表し、iConfが1や2の場合はエラーである。<candidate>には、変換候補が格納される。「中央区」や「富士見」などの地名は二本中に数多く存在するため、住所文字列の長さが短い場合には、複数の変換候補が返される。<address>は正規化された住所情報、<longitude>は経度、<latitude>は緯度、<iLvl>は変換された住所の階層レベルを表している。

なお、「柏の葉5-1-5」という文字列を与えた場合にも、同様の緯度経度が返される。都道府県名や市区町村名が省略された場合でも、変換候補が一意に決められる場合には精度の高い変換が可能である。

3.3.2. POI (Point of Interest) 情報の保持

本システムでは、一般的なドキュメントに対して、タグを用いて位置情報を埋め込み、ドキュメントの管理に活用する。具体的には、シンプルジオコーディング実験より返信されたアドレスマッチングの結果(緯度経度および関連情報)に基づいて、次のようなタグ(<SPA>タグ)が埋め込まれ、空間タグ付けテキストが作成される。

```
「柏キャンパス〒277-8568<SPA id="1" address="千葉県 柏市 柏の葉 五丁目 1番" level="7" lat="35.901711" lon="139.935455" label="千葉県柏市" content="〒277-8568 千葉県柏市柏の葉5-" message
```



```
=" url="" alias="" lastmodified = "20070106171427"
dictionaryid = "-1" converted = "千葉県柏市柏の葉
5-1-">千葉県柏市</SPA>柏の葉 5-1-5 東京大学空間情
報科学研究センター]
```

<SPA>タグの属性のうち、address は正規化された住所、level はマッチングレベル、lat は緯度、lon は経度、converted は変換された住所表現を表し、いずれも、シンプルジオコーディングによるマッチング結果に基づいて設定される。content は、変換に成功した住所に文章中の前後の文字列を追加したものが格納され、POI ごとの検索を行う際に利用される。

SDMS は、空間タグ付けテキスト中の <SPA> タグの属性の一つの POI 情報として管理し、表示や検索の対象とする。

3.3.3. 住所地名テーブルの作成

本システムでは、非定型のドキュメントから住所を抽出するために、一定の長さの地名語（住所の部分文字列）を蓄積したテーブルファイル（「住所地名テーブル」）を用意する。例えば、「柏市柏の葉」、「目黒区駒場」といった文字列をファイルに格納し、ドキュメント中から取り出した文字列と比較することで、取り出した文字列が住所候補かどうか絞り込むことができる。

アドレスマッチングサービスによって文字列を変換した結果十分な精度が得られなければ、その文字列は住所ではないと考えられるので、ドキュメントの先頭から、一文字ずつシフトしながら、ある一定の長さの文字列を順番にアドレスマッチングサービスに送信する方法でも、ドキュメント中の住所を特定し緯度経度を算出していくことはできる。しかし、この方法では、大量の住所でない文字列をサーバに送信することになり、無駄が多く、処理時間も長くなると考えられる。

本システムでは、国土交通省の「街区レベル位置参照情報」[10]に基づいて、「hook.txt」と呼ばれるファイルを作成し、最長5文字の約20万語の地名語を登録し、住所候補の絞り込みに利用する。この住所候補テーブルを利用することで、アドレスマッチングサーバに送信する文字列の数を大幅に制限することができる。この際、単純に街区レベル位置参照情報（住所と緯度経度の対応表）の各レコードに含まれる住所文字列の連続する5文字を取り出すわけではなく、次に述べる方針にしたがって、地名語を作成し、テーブルに登録する。

日本における住所表記は、都道府県、市区町村、町丁字、丁目というように、住所階層レベルの異なる複数の文字列から構成されているが、hook.txt には、「千葉県柏市」、「柏市柏の葉」など、2階層を結合した文

字列を住所候補として登録する。その結合された文字列が5文字より長い場合には5文字までの部分文字列を登録する。逆に、市区町村や町丁字の名前が短い場合も存在するため、「柏の葉-5」や「柏1」といった5文字より短い長さの文字列も登録されている。5文字より長い住所文字列を登録することで、さらに絞り込むことも可能ではあるが、現実的ではない。また、「丁目」の省略に対応するために、「駒場4丁目」は「駒場4」という形式で登録する。

住所地名テーブルを用いたマッチングの目的は、住所らしい文字列を漏れなく抽出することであり、この段階で住所であることを確定する必要はない。本研究のアプローチでは、アドレスマッチングサーバ側の緯度経度変換の結果から住所であるかどうか判断するため、同様の処理をクライアントであるSDMS側で実施するのは冗長であり、効率的でない。

3.3.4. ジオパース処理のアルゴリズム

ここで、プレインテキスト中の i 番目から j 番目までの文字列を $S(i,j)$ と表記する。3.3の手順③では、プレインテキストを先頭から走査しながら、次の手順を繰り返す。なお、 i の初期値は、1であり、 $S(1,5)$ はプレインテキストの先頭5文字を表す。図4に、このジオパース処理のアルゴリズムの概略を示す。

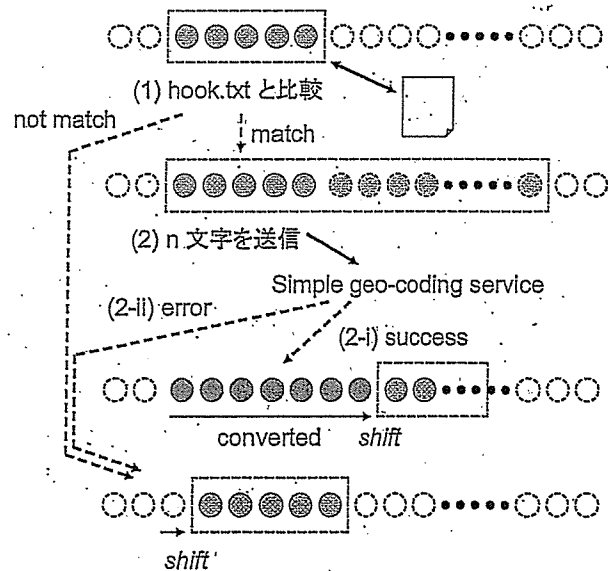


図4: ジオパース処理のアルゴリズムの概略

- (1). プレインテキストから、5文字の文字列 $S(i, i+4)$ を取り出して、hook.txt 中の文字列と比較する。hook.txt には、5文字以内の異なる長さ L_m ($m \leq 5$) の文字列が登録されているため、各 L_m に対して、文字列 $S(i, i+m-1)$ との比較を行う。 $S(i, i+m-1)$ は、 $S(i, i+4)$ の最初の文字から m 番目の文字までを含む。

(2). hook.txt 中に、 $S(i, i+m-1)$ とマッチするものがあれば、住所候補として、 $S(i, i+4)$ を先頭に含む長さ n の文字列 $S(i, i+n-1)$ をサーバに送信する。現在の実装では、 $n=30$ としている。

(i). サーバからマッチング結果が返され、信頼性が高い。すなわち、iConf が 3 以上の場合には、住所変換が成功したのものとして、<SPA> タグに属性情報を設定し、変換に成功した文字数分 k (3.3.1 の <converted> の文字数に相当) だけシフトして ($i=i+k$)、(1) の処理へ進む。

(ii). iConf が 0 もしくは 1 の場合には、1 文字シフトして ($i=i+1$) して、(1) の処理へ進む。

(3). hook.txt にマッチするものがない場合には、1 文字シフトして ($i=i+1$)、(1) の処理へ進む。

3.4. 空間ドキュメント管理システムの特徴

その他の機能も含めて整理すると、SDMS は、次の特徴を持つ。

- (a) 単純な操作による空間ドキュメントの入力
ファイルやフォルダのアイコンのドラッグ&ドロップだけでなく、URL 指定による Web ページの入力もサポートしている。
- (b) ロバスタなアドレスマッチングの提供
シンプルジオコーディングサービスを利用することにより、表記の揺れや省略に強いアドレスマッチングを提供する。街区レベルではあるが、日本全国をカバーしている。
- (c) 空間ドキュメントのレイヤー管理
空間ドキュメントに含まれる POI の集合は一つのレイヤーとして管理される。複数のドキュメントが入力されている場合には、各ドキュメントに含まれる POI が別々のレイヤーとして管理され、地図上に表示される。
- (d) 日本全国の背景地図の提供
背景地図は、ネットワークを介してリモートの地図サーバから取得する。国土地理院発行の数値地図 25000 をベースとした複数の縮尺の地図画像を利用できる。
- (e) エクスポート機能のサポート
GIS や統計ソフトなどの他の解析ツールとの連携を考慮して、地図上の POI を別形式としてエクスポートすることが可能である。出力形式としては、Shape ファイル、G-XML、CSV 形式をサポートしている。
- (f) テキスト検索や空間検索のサポート
ドキュメント単位、POI 単位でのキーワード検索をサポートしている。

4. 実装及び考察

4.1. 実装

SDMS は、Java 言語を用いて実装しており、Windows OS 上で動作する。アドレスマッチングおよび背景地図管理の機能は、リモートサーバに配置しているため、ネットワークを利用できる環境が必要となるが、それらの機能を外部に持つため、本体プログラムのサイズは比較的小さく、可搬性は高い。また、前述した通り、街区レベル位置参照情報を利用しているため、アドレスマッチング機能を無料で利用できる。アドレスマッチングの精度が重要になる場合もあるが、街区レベルのマッチングでも十分な状況が多いと考えられる。例えば、SDMS を利用して、概略的に、空間的な分布と傾向を把握し、その後で、時間をかけて、高精度のアドレスマッチングと高機能の GIS を使って、より詳細な分析を行うといった役割の分担も考え得る。多くのユーザに SDMS を利用してもらうためにも、無料での配布・利用という点は重要である。

4.2. 動作例

ユーザは、コンピュータ内のファイルやフォルダのアイコンを SDMS にドラッグ&ドロップするだけで、そこに含まれる住所情報を、空間的な分布として見ることができる。以下、具体例について説明する。

図 5 は、ファイルのドラッグ&ドロップの結果である。入力されたドキュメントは、あらかじめパソコン中に保存された東京都消防庁の消防署一覧のページ (http://www.metro.tokyo.jp/ANNAI/TOCHO/SOSHIKI/s_houbo_c.htm) であり、ドキュメント中に含まれている各消防署の住所が POI に変換され、地図上に点分布として表示されていることがわかる。

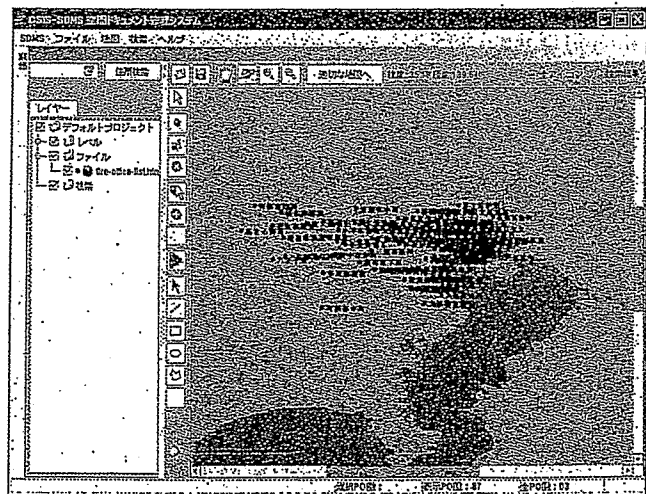


図 5: ファイルアイコン (消防署一覧の HTML ファイル) をドラッグ&ドロップした時の実行結果

図6は、複数のファイルを含むフォルダのドラッグ&ドロップの結果である。このフォルダには、光化学スモッグの発生に関する情報（東京都福祉保健局、光化学スモッグによる被害と思われる発生状況について、http://www.fukushihoken.metro.tokyo.jp/kanho/smog/higa_iindex.html）が日付ごとに保存されており、各ドキュメントには発生場所に関する住所情報が含まれている。

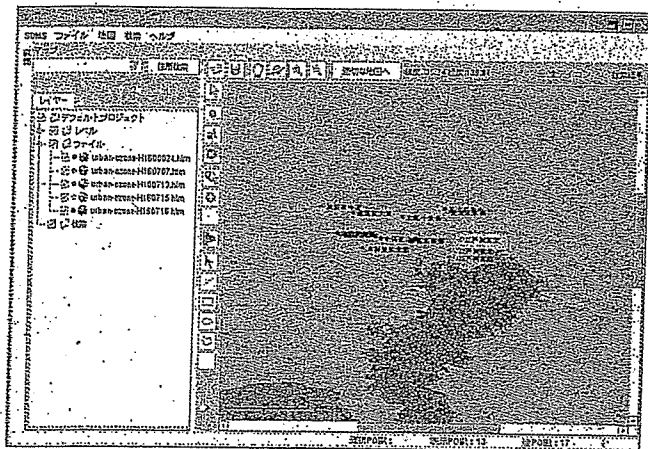


図.6: 複数のファイルを含むフォルダのアイコンをドラッグ&ドロップとした時の実行結果。

図6の左側のレイヤー管理部において、各ドキュメントが別々のレイヤーとして管理され、それぞれのドキュメント中のPOIが色の違うアイコンとして地図上に表示されていることがわかる。地図上の各アイコンは元ドキュメントと関連付けられており、マウス操作で、そのPOIを含むドキュメントを簡単に開くことができる。

4.3. 考察

3.3.1.で述べた通り、ドキュメント中に含まれる住所文字列の長さが短い場合には、複数の住所候補がサーバから返信される。SDMSでは、複数の住所候補から一つの住所に絞り込むために、次のいずれかの基準を用いる。一つ目の基準では、変換されている緯度経度と地図の中心座標とを比較して、地図の中心点に近いものを選択する。もう一つの基準では、「東京都」や「柏市」といったキーワードを指定し、そのキーワードを含む住所を選択する。例えば、「駒場4丁目」という住所文字列に対して、「茨城県取手市駒場4丁目」と「東京都目黒区駒場4丁目」が候補として返されるが、あらかじめキーワードとして「東京都」を指定しておくことで、後者のみを選択し表示することが可能である。地域単位での分析や自治体など地域単位での利用を考えた場合には、都道府県名や市区町村名が省略される場合が多いため、複数の住所候補が返される可能性が

高くなるが、このような場合、キーワード指定によるフィルタリングが有効であると考えられる。

本研究では、住所候補を抽出するために、hook.txtを用いるアプローチを取っているが、ごくまれに、住所でない文字列が住所として認識される場合がある。例えば、「昭和45年」「平成6年」といった文字列を含む場合、「昭和4」や「平成6」といった地名が、hook.txtに登録されているので、住所として認識され、地図上に表示されてしまう。ただし、これらの地名をhook.txtから削除してしまうと、本来住所であるものも住所として抽出されなくなってしまうので、省略された住所表現に対応するためには、hook.txtにこうした文字列を登録しておくことは必要である。この場合でも、前述したキーワード指定をすることで、ある程度の絞り込みは可能である。

現在、SDMSは、国立保健医療科学院の健康危機管理支援情報システム（H-CRISIS）[13]でダウンロード可能となっている。H-CRISISは、保健所などの地方の公衆衛生機関が危機管理情報を閲覧し、情報を共有するためのポータルサイトである。公衆衛生や健康危機管理といった場面では、場所情報が重要であるにも関わらず、多くのドキュメントが死蔵されてしまっている現状があり、本システムの潜在的なニーズは非常に高い。健康危機に関する空間ドキュメントをSDMSによって地図化することができれば、過去の情報を活性化できるだけでなく、災害時のリアルタイムでの意思決定支援にも役立つと考えられる。専門性を必要とする高価なGISソフトウェアではなく、日頃から利用しているドキュメント管理ツールの方が非常時や緊急時にも効果的に活用されることが考えられる。

4.4. 今後の課題

今後、ツールの配布に向けて準備を進めていく予定であるが、一般ユーザからのフィードバックを考慮して、インタフェースを改善し、より使いやすいツールとして開発を進めたいと考えている。また、現状では、地図上に表示されるのは点だけであるが、線や面などの幾何オブジェクトへの変換や地図上での重ね合わせ表示ができると空間的な関係性を把握するのに有用であると考えられる。

SDMSは、ドキュメント中から住所情報を抽出して緯度経度に変換することができるが、現時点では、郵便番号、電話番号、地名などの他の間接位置参照情報はサポートしていない。特に、地名に関しては、応用分野に特有の地名用語辞書を作成し、その辞書を選択・共有することで、SDMSを専門的な用途で活用できると考えられる。変換したPOIの緯度経度情報は、地図上のドラッグ操作により容易に修正することがで

きるため、SDMSを利用して、修正されたPOIの位置情報を登録しながら、ユーザ辞書を整備していくことも考えられる。

SDMSは住所情報の抽出と表示は可能であるが、その住所が対応している実世界オブジェクトの情報をドキュメント中から探し出すことは現時点ではできない。例えば、消防署一覧の例で考えても、消防署の点分布は表示されるが、各点がどの消防署に対応しているかは自明ではない。全自動で対応するのは難しいとしても、ユーザの負荷を軽減するような対策は考案したいと考えている。

火災や交通事故などの災害事故情報、光化学スモッグ発生情報などのセンサ情報、インフルエンザ発生情報などの健康危機管理情報などを公開しているサイトでは、掲載されている情報が定期的に変化するため、そこに含まれている住所情報も変化する。空間的な分布が、時間的にどのように変化していくかを把握するためには、指定されたURLに定期的アクセスして、Webページ中のドキュメントから住所を抽出して、生成したPOIをスナップショットとして保存し、アニメーションとして再生できると望ましい。現在、その時系列管理機能を実装中である。

5. まとめ

本稿では、住所情報を含むデジタルドキュメント（「空間ドキュメント」）を、直感的なインタフェースで扱うことのできるシステムとして、空間ドキュメント管理システム（SDMS: Spatial Document Management System）を提案し、実装したシステムについて説明を行った。SDMSは、ドラッグ&ドロップという簡単な操作で、ドキュメント中の場所情報を点として地図上に表示することができる。空間情報を扱うことのできるシステムとして、地理情報システムがあるが、定型書式を入力とするため、空間ドキュメントを直接扱うことは難しい。SDMSでは、非定型の一般的なデジタルドキュメントから、場所情報を抽出し、アドレスマッチング技術を利用してPOI(Point of Interest)を生成し、地図上に表示することができる。SDMSは、Java言語を用いて実装されており、HTML、Excel、TEXTなど多くのデジタルドキュメントを扱うことができる。SDMSは、日常的に利用できる人間中心型のツールとして開発を進めている。SDMSによって、多くの一般的なユーザが、簡単に、デジタルドキュメントを、空間的に整理し、地図化する環境が整備されていくと考えている。

謝辞

本研究は、厚生労働省科学研究費補助金（健康科学総合研究事業）「地理及び社会状況を加味した地域分析手法の開発に関する研究（代表：浅見泰司）」の支援を受けている。

文 献

- [1] Google, Google Maps API Version 2 Documentation, <http://www.google.com/apis/maps/documentation/>
- [2] Yahoo!JAPAN, Yahoo!地図情報, <http://map.yahoo.co.jp>
- [3] GeoRSS, <http://www.georss.org/>
- [4] Microformats-geo, <http://microformats.org/wiki/geo>
- [5] P.A. Langley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind, Geographic Information Systems and Science, John Willy&Sons, 2001.
- [6] 東京大学空間情報科学研究センター, CSV アドレスマッチングサービス, <http://spat.csis.u-tokyo.ac.jp/cgi-bin/geocode.cgi>
- [7] E. Amitay, N. Har'el, R. Sivan, and A. Soffer, "Web-a-where: geotagging web content", SIGIR 2004, pp.273-280, 2004.
- [8] 横路誠司, 高橋克巳, 三浦伸幸, 島健一, "位置指向の情報の収集, 構造化および検索手法", 情報処理学会論文誌, Vol.41, No.7, pp.1987-1998, 2000.
- [9] 長屋 務, 森本泰貴, 藤本典幸, 出原 博, 萩原兼一, "Google Maps API を応用したロボット型施設検索システムの試作", データ工学ワークショップ 2006, 5B-i6, 2006
- [10] 国土交通省国土計画局国土情報整備室, 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/>
- [11] T. Sagara, M. Arikawa, and M. Sakauchi, "Spatial Document Management System Using Spatial Data Fusion", IIWAS2001, pp.399-409, 2001.
- [12] 東京大学空間情報科学研究センター, CSIS シンプルジオコーディング実験 (街区レベル位置参照情報), <http://pc035.fkl.iis.u-tokyo.ac.jp/~sagara/geocode/modules/simple-geocode/>
- [13] 国立保健医療科学院, 健康危機管理支援情報システム H-CRISIS, <http://h-crisis.niph.go.jp/hcrisis/index.jsp>