

健康危機管理のための空間ドキュメント管理システム

浅見泰司, 有川正俊, 白石陽, 相良毅

2008

保健医療科学

(国立保健医療科学院)

## 特集：地域診断・症候サーベイランスに向けた空間疫学の新展開

### 健康危機管理のための空間ドキュメント管理システム

浅見泰司<sup>1)</sup>, 有川正俊<sup>1)</sup>, 白石陽<sup>1)</sup>, 相良毅<sup>2)</sup>

<sup>1)</sup> 東京大学 空間情報科学研究センター, <sup>2)</sup> 東京大学 生産技術研究所

### Spatial Document Management System (SDMS) for Health Risk Management

Yasushi ASAMI<sup>1)</sup>, Masatoshi ARIKAWA<sup>1)</sup>, Yoh SHIRAIISHI<sup>1)</sup>, Takeshi SAGARA<sup>2)</sup>

<sup>1)</sup> Center for Spatial Information Science, The University of Tokyo,

<sup>2)</sup> Institute of Industrial Science, The University of Tokyo

#### 抄録

健康危機が発生した場合には、どこで何が起きているかを的確に把握し、効率的な対応方法を検討することが重要である。ところが、このような場所に関する情報は、これまであまり有効には利用されてこなかった。その原因としては、場所情報自体のIT化の遅れがあげられる。場所情報を扱うシステムとしては、地理情報システム (Geographic Information Systems, GIS) と総称されるアプリケーションがあるが、一般に高度な数理解析を行うために用いられるものであり、操作にあたっては専門的な知識や技能が必要とされることから、誰でもが扱えるようなシステムにはなっていない。また、場所情報の表現には多くのバリエーションがあり、GISで利用できるフォーマットに加工するためにも経験や知識を要するため、健康危機発生などの緊急時に必要な場所情報がGISに容易に利用できるような書式で届けられることを期待することはできない。場所情報は多くの場合、人間が分かるアナログ形式で、つまり住所や地名や経路記述などの自然言語で記録されている。そのため、データ形式を変換したり、高度な操作技術が必要になるなどの理由で、これらの貴重な場所情報は有効に利用されずじまいになってしまうことがほとんどである。

この問題点を克服するために、従来のGISとは異なる新しい枠組みとして、人間中心情報デザインの観点から、空間ドキュメント管理システム (Spatial Document Management System, SDMS) を基本設計し、ソフトウェアツールとして実現した。そして、その有効性を実利用を通して検証してきた。SDMSは、一般ユーザがパソコン上で日常的に利用しているデジタルドキュメント (電子メール、テキスト、ウェブページ、ワード・ドキュメント、エクセル・ドキュメント、PDFドキュメントなど) をアプリケーションウィンドウにドラッグ&ドロップするだけで自動的に住所や地名を抽出し、経緯度を算出して、地図上で空間分布を提示できるソフトウェア環境を提供する。一般に、従来のGISが専門家向けの枠組みであったのに対し、SDMSは専門知識無しに誰でもがデジタル空間情報を日常的に取り扱える新しい空間情報利用環境であり、本論文ではSDMSの提案の背景と意義、その基本原理を紹介し、今後の展望について述べる。

キーワード： 非定型データ、ジオコーディング、ジオパーシング、空間ドキュメント、人間中心情報デザイン

#### Abstract

Information concerning "where" is essential to effectively tackle the outbreak of health hazards. However, only limited use of such information has been practiced till now. Slow development of information technology related to spatial information may be accused to this fact. Geographic information system (GIS) is the software specially designed to deal with spatial information for advanced spatial analyses, which unfortunately shuts novice users away due to heavy needs of skills for its operation. Moreover, so much variation of expression for spatial information is available, and its transformation to GIS-ready format needs professional experience and knowledge. Therefore, in the event of health hazard emergency, well formatted data is rarely obtainable which is ready for GIS use. In many cases, the collected data is described in natural language, such as

〒277-8568 千葉県柏市柏の葉5-1-5

5-1-5 Kashiwanoha Kashiwa, Chiba, 277-8568 Japan.

addresses, area names and route descriptions. The necessity of data conversion and special skills to use GIS often discourages people effectively utilizing spatial information.

To remedy this situation, we developed new software called spatial document management system (SDMS), which is designed as a human centered tool for information processing in a framework different from traditional GIS. The SDMS has been validated through its actual usage. SDMS attains automatic process of extraction of addresses and area names, calculation of latitude and longitude, and visualization of their spatial distribution on a map by only drag and drop operation of digital documents (email, text, web page, MS-word document, MS-excel document, PDF document, etc.) that are handled on a PC on a daily basis. Compared to GIS which is designed for experts, SDMS is designed for ordinary people without expert knowledge. This article is devoted to exposition of background and its significance of development of SDMS, its basic principle of operation, and prospect of its future development.

**Keywords:** unstructured data, geocoding, geoparsing, spatial document, human-centered information design

## 1. はじめに

臨界事故に見られる原子力ハザード対策, パンデミック被害やテロなどに起因するバイオ・ハザード, 事故や犯罪による化学薬品ハザードなど, 現代社会において健康危機リスクは多様になってきており, かつ, 素早い対応をしないとその被害は大きくなりかねない. このため, 健康危機イベントが発生した際に, それへの素早い対応体制整備は厚生労働行政上, 喫緊に行わなければならない重要な課題となっている.

事実, このような事態に対応するために, 危機発生時に使える大規模システム構築の開発などが行われてきた. しかし, 緊急時にしか使われないシステムは, 往々にして, ユーザがその操作法を未習熟であるケースが多く, 十分に活用できない事が多い. そのため, 平常時にもよく使われ, 緊急時にも利用できるようなシステムが, 実は危機管理上, 有効であると言える.

緊急時には, ただでさえ情報伝達が難しい状況であることが多く, 情報発信者には受け手側の便宜をはかって書式を整えるなどといった余裕はないと考えた方がよい. このため, 定型ではなく, 様々な非定型ドキュメントを有効に活用できるシステムが望まれていると言えよう.

本論で紹介する空間ドキュメント管理システム (SDMS: Spatial Document Management System)<sup>1,2,3,4)</sup> は, 上記の目的を具現化するために開発されてきたものである. 空間ドキュメント管理システムは, テキストドキュメントに分解しうる非定型ドキュメント (たとえば, 電子メール, テキスト, ウェブページ, ワード・ドキュメント, エクセル・ドキュメント, PDF ドキュメントなど) をドラッグ&ドロップという簡単な操作だけで, 地図化し, かつ GIS (地理情報システム) などに取り込めるようなファイルに変換するシステムである. 自由な書式でありながら, 汎用性に優れたデータ書式で空間情報を流通できるシステムが確立できるため, 健康危機情報の収集から空間データ化するまでの時間や手間を大きく省くことができる. 健康危機の初動時に被害場所分布を速やかに一覧したい際に有効である

が, 実際には, 地図化の作業は通常業務でも多々必要とされることから, 平常時でも頻繁に使われるはずの汎用的システムとなっている.

すでに, SDMS の試行版を国立保健医療科学院の健康危機管理支援ライブラリーシステム (H-CRISIS) で公開し, 保健所などで利用可能な状況になっている. ただ, 現在の SDMS では地図の操作性が低い, 線や面など高次元のイベントには対応できない, 集中している地点を簡易に表示する機能がないなどの課題がある. そこで, 現在の SDMS をさらに改良するべく開発を行っている. これにより, 空間データ化された情報をそのまま一次解析することができ, 健康被害症候の特異性を早期に検出し, またその原因や健康被害の空間的拡散の特色を適切に推測する支援システムを構築でき, リアルタイムでの知的地域診断システムの構築およびその行政的利用が可能となる. さらに, ユーザに対しては各種時空間情報の GIS への統合に伴い現状把握, 情報の相互関係, 問題点把握等に関してより理解しやすい形での表示を可能とする等により, 健康で安全な社会の構築の礎となる. このシステムの応用により, 迅速な健康被害に対する警告発信, 原因究明の初動期支援, 平常時における健康関連情報の発信などに寄与するものと思われる.

## 2. SDMS の概要と基本原理

### 2.1. SDMS 開発の背景と意義

パソコンを用いた日常的な業務で一般的に扱うドキュメント, たとえば電子メール, テキスト, ウェブページ, ワード・ドキュメント, エクセル・ドキュメント, PDF ドキュメントなどには, 住所や地名などの多くの場所情報が記載されている. これらの場所情報は人間にとって可読性・伝達性が高く, 頻繁に利用されている. 一方, 従来の GIS では, 経緯度のような座標値といった, 機械に扱いやすい数値による場所情報を主対象としてきた. その結果, 多くの場所情報を含む一般ドキュメントは, 従来の GIS では扱うことができない.

座標のように機械に向く場所情報記述を直接位置参照情

報と呼び、一方、住所や地名のように、人間にやさしい場所情報記述を間接位置参照情報と呼ぶ。住所を経緯度に変換するように、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呼ぶ。われわれの研究では、ジオコーディングにより、一般ドキュメントをパソコン上で簡単に地図と連携させて扱える枠組み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・実現した。専門家のツールである GIS と違い、医療従事者を含む一般ユーザが SDMS を用いることにより、日常的に利用しているドキュメントを空間的に管理できるようになる。

たとえば、重要な会議のお知らせを電子メールで受け取った場合を考えてみる。会場は今まで行ったことがない場所で、メールには会場の住所と名称しか書かれていないとする。このような場合、PC を使いこなしているユーザであれば、メール文中の住所の部分をクリップボードにコピーして、インターネットの地図検索サイトや PC にインストールしてある地図ソフトウェアなどにその住所をペーストし、会場周辺の地図を表示させることができる。SDMS は、住所を抜き出して該当する座標に変換し、地図を表示するという一連の処理を自動で行うため、ユーザはただメールを SDMS のウィンドウにドラッグ&ドロップするだけでよい。

## 2.2. SDMS のコンセプト

パソコンで扱われているデータは、人間が作成し、人間が読むドキュメントが中心である。そのようなドキュメントの作成と管理を行う環境がパソコンと言ってもよいだろう。一方、現在パソコンで地図を扱う上で一般的な地図ソフトウェアや専門家が利用する GIS は、地図を中心にデータを管理しようとするため、ユーザ自身がドキュメントに含まれる場所情報を抽出して専用のデータを作成し、その後のデータの追加や編集なども全て地図の上で行わなけれ

ばならない。ほとんどのパソコンユーザにとって、地図は見たいときに見られればよいものであって、わざわざ手間をかけて専用のデータを作成し、維持管理するほどの必要は感じないだろう。

本論文で提案する SDMS は、一般ユーザが日常的に作成・利用するドキュメントを中心に考え、上述のように簡単に自然に地図と連携させることにより、ワープロ・電子メール・ウェブブラウザを使う感覚で、日常的に空間情報を利用・作成する環境を広く提供する。これが Spatial Document Management System, 空間ドキュメント管理システムという名前の所以である。

たとえば、Windows のエクスプローラなど、ドキュメントを整理するためのファイルマネージャと呼ばれるソフトウェアでは、五十音順や時間順による表示機能は不可欠であり、これに基づいてユーザはドキュメント管理を行っている。同様に、ドキュメントを「空間順」で表示・管理するソフトウェアが SDMS の主概念であり、SDMS を用いれば文章中の地理的な内容やドキュメントを作成した場所を表現する位置印 (location stamp, 時刻印 time stamp に対する概念) に基づいて空間順に表示できる。もともと、空間情報は特別なものではなく、われわれ人間は活動の記録を残したり、活動の予定を立てる場合には、必ず地球上の場所に関する情報を使ってコミュニケーションを行っており、むしろ場所情報記述が使えないと日常活動に支障をきたす。この状況は、IT が出現する以前から変わりはないが、古典的な IT では、人間同士がコミュニケーションで使う空間情報記述を取り扱う能力がなかった。SDMS はそれを可能として、人間が日常的に取り扱っているさまざまなドキュメントを地図上で利用できるようにする環境を実現した点で意義がある (図 1)。

また、広義のドキュメントである画像やムービーなどのマルチメディアコンテンツも空間順で管理する重要性は認められている。たとえば、写真ファイルに関しては、GPS

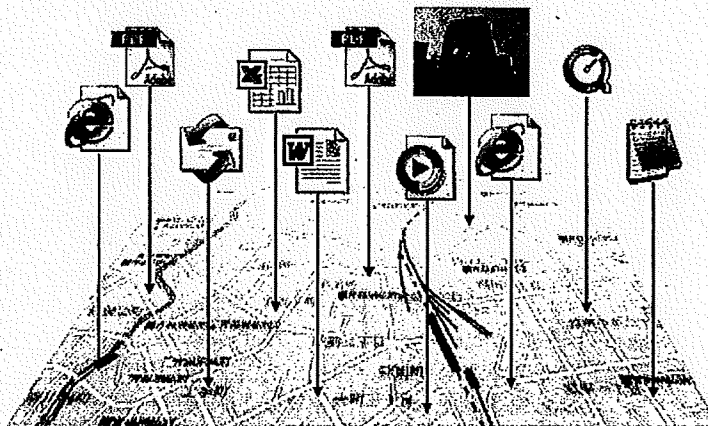


図 1. 空間ドキュメント管理システム (SDMS) のイメージ

日々利用している一般ドキュメント (MS ワード、テキスト、メール、ウェブ、MS エクセル、PDF など) を簡単に地図上の位置に結び付ける利用環境を実現する。ファイルをアルファベット順、時間順で見ることがよく、空間順で見ることができる利用環境を実現する。将来は、この機能はデスクトップへ統合されると考えられる。また、EXIF や MPEG7 などの国際標準により、画像やムービーへの位置記述は一般化してきている。

から取得した位置情報を EXIF 形式でメタデータとして埋め込む仕組みはほとんど全ての GPS 付携帯電話で一般的になっている。動画に関しても、ロケの位置情報をメタデータとして記述する国際標準 (MPEG7 など) もすでにある。

### 2.3. SDMS の使用例

SDMS の利用方法は極めて簡単で、一般ドキュメントをデスクトップ上で、SDMS の地図ウィンドウへドラッグ&ドロップするだけである。ここでは2つの使用例の表示結果を示す。

1つめの例は、東京大気公害裁判原告の住所リストが記載されたエクセルのドキュメントを SDMS で表示した結果である (図2)。この図から本郷通りや早稲田通り、首都高速4号線と5号線の沿線に居住している原告が多いことが分かるので、まずこれらの道路から調査を始めればよいと考えられる。このように健康危機が発生した場合、その要因が地理的なものかそれ以外のものかを迅速に判断する必要がある。

1852年ロンドンで流行したコレラの発生源を突き止めた John Snow は、患者の居住地を地図上にプロットしてその中心にある水道ポンプに原因があることを示した<sup>5)</sup>。このエピソードは疫学調査による地理的な解析が健康危機に有効であることを示した最初の例として有名であるが、多数の患者の居住地を一つずつ地図にプロットする作業には、地図が用意されているとしても長い時間がかかる。しかし SDMS を利用すれば、図2の例では600人以上の住所を地図にするのにほんの数秒しかかからない。

今日の疫学では、地図化した後に各種の数値解析を行い、より詳しい推定を行う必要がある。しかし、まず地図にしてみるというこれまで時間のかかった作業が何の事前準備もなく一瞬で終わるため、手軽におおよその見当をつけることができ、無駄な作業や解析を行って時間を浪費することを防ぐことができる。また、SDMS には変換した点データを GIS にエクスポートする機能もあるので、より詳しい解析は GIS で継続して行うこともできる。

2つめの例は、ウェブ上に公開されている伊丹市交通事故情報の例である。交通事故情報は、1週間ごとにまとめて表形式として、ウェブページに公開されており、発生日時や事故内容とともに発生場所の住所が記載されている。この場合、このウェブページをいったんデスクトップ上に保存して、SDMS にドラッグ&ドロップすればよい。ウェブページを保存するのが難しければ、上部にあるメニューバーから「ファイル」を選び、そのプルダウンメニューの中から「URL 指定」を選択し、ウェブブラウザのようにウェブページの URL を直接入力して、新しいドキュメントとして登録することもできる。これは場所情報を簡単にネットワーク上で共有する方法として利用することもできるが、専用の地図サーバなどは一切不要で、既存のウェブサーバにファイルをアップロードするだけでよい。

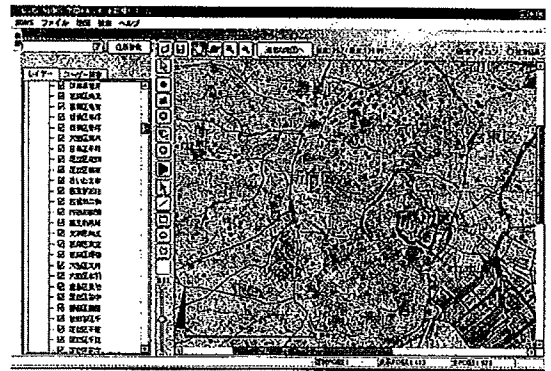


図2. 空間ドキュメント管理システムの例  
東京大気汚染公害裁判原告の住所リストから作成した空間分布図。東京大気汚染公害裁判に関する詳しい説明は以下を参照していただきたい。http://taiki-tokyo.web.infoseek.co.jp/  
本裁判は、2007年8月8日に和解が成立している。

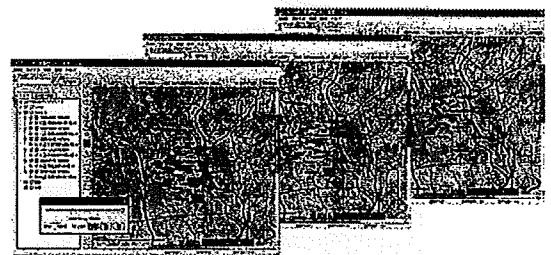


図3. SDMS の時計ブックマーク管理の例  
伊丹市交通事故情報のウェブページに掲載してあった一週間ごとの交通事故発生一覧 (6時系列) から空間分布図を作成し、アニメーションで重量表示させた例。  
(データソース) http://www.itami.fm/seikatu/jiko/index.html

1つのドキュメントは、1つのフォルダとして、左側の空間ブックマークに登録される。図3の例では、それぞれのブックマークを、時計ブックマークにして時系列情報として保存した結果をアニメーションとして再生する様子をイメージ化したものである。

## 3. SDMS の技術

### 3.1. 一般ドキュメントから POI への処理手続き

ドキュメント中心の SDMS では、GIS のような数値による場所情報データを持っていない。そのため、一般ドキュメントから POI (Point of Interest) つまり地図上の点地理オブジェクトの自動生成が中心的な情報処理となる。この処理は、主に2つの部分処理、(a) ジオパーシング (geoparsing) と (b) ジオコーディング (geocoding) から構成されている。以下、それぞれについて説明する。

#### 3.1.1 ジオパーシング

ジオパーシングは、一般ドキュメントから住所や地名の抽出を行う処理である。たとえば電子メールの文章から会議場の住所の部分だけを見つけ出して、切り出す処理と考えると分かりやすい。技術的には、自然言語で書かれた文章から場所に関する名詞の部分適切に抽出する処理であり、仮名漢字変換などに用いられている自然言語処理技術の一つである。しかし住所や地名には「田中」や「中村」

など人名と紛らわしいものも多く、場所に関する名詞であることを機械で保証するためには地名辞書を用意する必要がある。また、構文解析により、このような名詞句の抽出処理を高速化するのも難しいことは知られている。われわれは、独自のアルゴリズムを用いて、高速に適切にジオパースする手法を開発した<sup>6)</sup>。

### 3.1.2 ジオコーディング

ジオコーディングは、住所や地名などの間接位置参照情報を、経緯度 (x,y) などの直接位置参照情報へ変換する処理であり、住所に限った場合にはアドレスマッチング (address matching) と呼ばれる。ジオコーディングの基本的な処理は、住所や地名などの間接位置参照情報と経緯度が対になった表 (データベース) を使って、自然言語の中の場所記述を間接位置参照情報とマッチングして、対応する経緯度を出力するという単純なものである。

しかし、住所や地名にはさまざまな表現がある。たとえば「竜ヶ崎」は「龍ヶ崎」と表記されることがあり、「一丁目二番三号」は「1-2-3」と省略されることがある。さらに、平成の大合併により多くの自治体が統廃合され、それに伴い「浦和市仲町」が「さいたま市浦和仲町」になり、その後「さいたま市浦和区仲町」になるといった時間経過による住所表記の違いも起きている。われわれは、このような省略や表現の揺れに強く、昭和45年以降の住所の変化に対応し、かつ高速な処理が可能なアルゴリズムを開発し、東京大学空間情報科学研究センター (CSIS) のウェブサービスとして実現した。SDMS ではこのウェブサービスを利用しているため、ドキュメントに含まれている住所に表記の揺れや省略があったり、古いドキュメントで地名が昔のままであっても、そのまま地図上にプロットすることができる。

なお、このジオコーディングのウェブサービスは、CSIS シンプルジオコーディング実験という名称で、SDMS 以外のアプリケーションからでも利用できるように一般公開している。CSIS シンプルジオコーディング実験では、日本全国を街区レベル (\*) のマッチングまでカバーし、大量のジオコーディングのリクエストを高速に処理できる<sup>7,8)</sup>。このウェブサービスでは、場所記述をサーバに渡すと、マッチングされた住所や地名、マッチングの度合いの値、正規化された記述などを XML 形式で返す。

(\*: 街区レベルまでのジオコーディングしか実現できない理由は、アドレスマッチングのデータとして、国土交通省国土計画局 国土情報整備室が無料で一般公開している街区レベル位置参照情報を利用しているからである。もし有料の民間のアドレスマッチングデータを利用すると、より詳細な号レベルのマッチングも可能である。東京大学空間情報科学研究センター内では、株式会社ゼンリンの住宅地図データを利用した、号レベルのアドレスマッチングを利用した空間情報科学の研究も行われている。)

同様のサービスは Google でも提供しており、号レベルまでのマッチングが無料で利用できるが、一日に利用できる回数に制限があることや、自治体合併により変則的な名称変更が行われた一部の旧住所が正しく変換できない (田

無小学校の旧住所、東京都田無市本町4-5-21など) といった長所短所がある。

### 3.2. 使いやすさとユーザ負担軽減の追求

SDMS はパソコンを日常業務で利用している一般ユーザを対象としているため、できる限り平易な操作を追求している。たとえば、一般ドキュメントを POI 化する処理は、ドキュメントをウィンドウにドラッグ&ドロップするだけという単純な操作で実行することができる。このような単純で直感的な操作は、SDMS および場所情報を簡単に利用できる環境を実現する大きな要因と考えている。このドラッグ&ドロップという操作は、SDMS 固有のものではなく、デスクトップ型のパーソナルコンピュータのユーザインタフェースでは標準的な対話プロトコルであり、一般ユーザにとって違和感無く使え、操作が容易な環境を作り出すことができる<sup>9)</sup>。

SDMS では、インターネット接続が利用必須条件であり、インターネットを介して、CSIS の日本全国のジオコーディングサービスと、日本全国の背景地図サービスを利用している点も、一般ユーザが場所情報を簡単に利用する環境を実現するのに大きく貢献している。商用の GIS を利用する場合、まず背景地図のデータを揃えて、それからその上にプロットするデータをドキュメントから抽出して作成する必要があるが、それらの空間データを揃えるのはデータの選定やフォーマットの指定など初心者にとって戸惑うことの多い大変な作業であり、この段階で挫折してしまってデジタル場所情報の利用を断念するケースが多い。SDMS の場合には読み込んだドキュメントの点分布に応じて適切な背景地図を自動選択するので、苦勞することなく地図を作成することができる。

また、ジオコーディングサービスと背景地図サービスをウェブサービス化することにより、クライアントで大きな地図データを持つ必要もなく、クライアントのソフトウェアパッケージを小さくでき、インストールが簡単になっている点も特徴である。さらに住所や背景地図は時間とともに変化するが、メンテナンスはサービス側で行われるので、ユーザは地図データの再購入やアップデートなどを行う必要がない。

このように、SDMS は、住所や地名が入ったドキュメントさえあれば、すぐにデジタル場所情報を取り扱うことができる。つまり、従来の GIS に比べて、ユーザは環境整備に時間と費用を費やすことなく、本題の作業に直接的に簡単に入っていくことができる。

商用の地図ウェブサービスは、印刷物に載せる場合など著作権上複雑な手続きや使用料が必要となる。ところが SDMS の背景地図は、無料で公開されているものなので、地図の出典さえ明示すれば、印刷物でも、インターネットでもほぼ自由に公開することができるという点も実用性が高いと言える。ただし、SDMS の地図は、現時点では商用の地図に比べると、精度や見映えの点では劣るという欠

点がある。

#### 4. データ処理の主な流れおよび補助機能の紹介

SDMS のデータ処理の概要は、図4のとおりである。以下、基本的なデータ処理の手続きを示す。

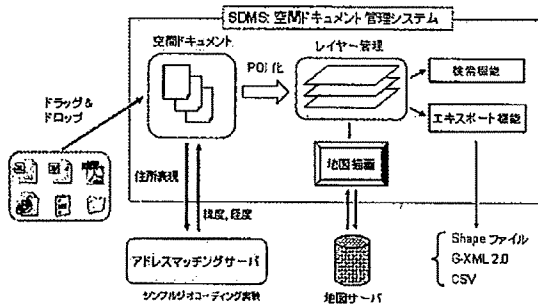


図4. SDMSのシステム構成図

##### 4.1. SDMS のデータ処理の主な流れ

- (1 st) 一般ドキュメントあるいはそれらがいったフォルダをユーザがSDMSにドラッグ&ドロップする。
- (2 nd) 一般ドキュメントをプレーンテキストに変換する。
- (3 rd) プレーンテキストを自然言語解析して名詞句を抽出する。
- (4 th) 抽出した名詞句が地名であるかどうかを、アドレスマッチング・サーバに問い合わせる。このアドレスマッチングの特徴は、住所のゆれや省略に頑強であり、また処理速度が高速である点にある。地名である場合は、正規化された地名と座標値とマッチングレベルが返される。マッチングレベルとは、県・市・区・丁目・番地など、どのレベルまでマッチしたかを示す値である。
- (5 th) プレーンテキスト中で、地名と判定された部分に、<spa>タグを挿入し、地名として識別された部分の情報を残す。この場合、<spa>タグの属性として、address, level, lat (itude), lon (gitude), urlなども同時に設定される。urlは、もとのドキュメントの格納場所を指す。  
(例)「発生場所は <spa id="5" address="東京都北区十番台一丁目" level="5" lat="35.756924" lon="139.723083" label="北区十番台" url="http://juban/"> 北区十番台1丁目 </spa> 付近だ。」
- (6 th) <spa>タグの属性値の部分が、POI (Point of Interest) として、空間データベースに登録される。
- (7 th) 日本地図上に抽出されたPOI集合が表示される。また、POI集合を表示するのに最も適切な範囲と縮尺が自動設定され、地図が表示される。
- (8 th) ドラッグ&ドロップの単位である、ドキュメントかフォルダごとに、フォルダ（またはレイヤ）が設定され、POI集合がフォルダごとにブックマークと

して管理される。このフォルダごと、およびフォルダ中のPOIごとに表示・非表示・削除・アイコン画像設定の制御が可能である。

- (9 th) POIに対して、キーワード検索も可能であり、キーワード検索結果は、新しいレイヤのPOI集合として表示される。たとえば、「インフルエンザ」や「食中毒」のキーワードが含まれるドキュメントのPOIだけを生成することも可能である。ドキュメントから生成されたPOI（点地理オブジェクト）は、経緯度のような位置情報だけではなく、地名としてマッチングされた部分テキストの周辺、たとえば、前後30文字のテキストも、周辺テキストデータとしてPOIの1つの属性データとして持たせている。（この30文字という周辺テキストデータの長さは、環境設定のダイアログを使って文字数の変更を行うことができる。デフォルト値は30文字である。）位置データだけでは検索できない、建物名や分類などの関連する情報がこの周辺テキストデータに含まれていることが多く、これを用いてPOIを検索・絞り込みができる機能は便利である。
- (10th) エクスポート機能：POIをShape, CSV, G-XML 2.0の形式で出力可能である。これにより、商用のほとんどのGISとの連携が可能となる。高度な空間解析や視覚化は、商用のGISで行うという役割分担を想定している。

##### 4.2. SDMS の補助機能

以下に、4.1のデータ処理の主な流れで出てこなかった、重要な補助機能の説明を行う。

- (a) 地図範囲・縮尺自動選択機能：選択したPOI集合を表示するのに適切な背景地図の自動選択を実現するために、縮尺・移動操作の自動機能を実現している。また、ブックマークや地図ウィンドウ上で単数あるいは複数のPOIを選択し、「適切な地図へ」のボタンを押すことにより、POI集合を表示する適切な縮尺の地図が表示される。
- (b) 地図データキャッシュ機能：一度ダウンロードした地図データはローカルに保存し、その後の地図表示において、すでにローカルに地図データがある場合は、地図サーバからダウンロードせず、ローカルの地図データを利用し、地図表示の高速化を実現している。この機能により、ネットワークに接続されていなくとも、ローカルに地図データが存在すれば地図表示が可能となる点は都合が良い。また、実験結果として、キャッシュに地図データがある場合は、表示スピードは約10倍程度速くなることが確認できた。
- (c) POIとドキュメントの表示機能：POIアイコンを選択し、マウスの右ボタンから「詳細表示」を選択す

- ることにより, POI の中身つまり構成属性値 (経緯度, ラベル, 住所, ソースドキュメントの URL, アドレスマッチングレベルなど) を表示させたり, あるいは変更することが可能である。また, POI アイコンをダブルクリックすることにより, POI が含まれるもとのドキュメントを表示することができる。
- (d) POI の多重度管理・表示機能: ドキュメントを POI に変換する際, 住所が同じ場合など, 同一の POI が複数個生成される場合があり, それらの多重度の情報もすべて 1 つの POI の情報として管理している。地図上では, 1 つの POI アイコンにカーソルを重ねると, その POI の多重度がポップアップで表示されるようになっている。現在, 多重度や点密度をより自然に直感的に可視化する機能を開発している (図 5)。
- (e) ブックマーク管理機能: ドキュメントを単位にしたレイヤ管理をブックマークの枠組みで実現している。最小単位は POI であり, それらをフォルダの概念で階層的に管理している。フォルダや POI のアイコンで, 表示・非表示の設定や凡例のアイコンや色の設定も可能である。
- (f) アニメーション機能: ブックマークの特殊型として, 時計ブックマークがあり, 時計ブックマークには時刻印を記述することができ, その時刻印順に, 時計ブックマークに対応する地図の集合をアニメーションとして逐次表示できる。たとえば, 時系列で区分された POI 集合をアニメーションで表示することにより, イベントの拡散や減衰などの現象を視覚的に効果的に表現できる。
- (g) プロジェクト管理機能: プロジェクト管理機能を SDMS に追加したことにより, ユーザは目的ごとにプロジェクトを使い分けることによって, ドキュメント集合と POI 集合を混雑することなく利用できる分かりやすい環境を実現できる。ブックマークは, その作成時点で即座にファイル (不揮発性メモリ) に記憶される。つまり, 明示的に保存操作を行わなくても保存される。SDMS のプロジェクトをブックマークと同様の枠組みにし, データ操作した結果は自動的に保存されるようにすることにより, 作業内容を間違えて消去することが無い環境を実現した。
- (h) ユーザ空間辞書: ユーザが独自に地名と座標を登録できる機能。たとえば, 現在の SDMS では, 街区レベルのジオコーディングしか用意していないが, SDMS により生成された POI を手動で移動して, 地図上の正しい位置へ POI アイコンを移動した場合, この情報をユーザ空間辞書に登録可能である。同じ住所記述を含む別のドキュメントを新たにドラッグ & ドロップした場合, ユーザ空間辞書に登録された新しいアドレスマッチング情報が適用され, 高精度な経緯度を持つ POI が生成される。また, 空間辞書の中に無い住所以外の場所の名前 (たとえば建物名など) をユーザ空間辞書に登録することにより, 高精度な経緯度を持つ POI 化の処理を実現できる。
- (i) グラフィックス機能: 地図上に, 追加的に, 文字列, 点図形, 線図形, 面図形, 画像などを生成・配置でき, 地図上にユーザが任意に作図を行うことができる。
- (j) 印刷機能: プリント機種非依存にするために, SDMS 画面の出力は画像ファイルの出力として実現している。その画像ファイルをプリンタで出力するという方針を採っている。
- (k) 環境設定: 操作のデフォルト値を環境設定パネルで変更可能である。代表的な環境変数を以下で紹介する。
- ① 複数マッチングの優先順位付けの設定: 住所や地名でマッチングした結果が複数個あった場合の選択方法として, 初期設定では, 現在表示している地図の中心の位置から最も近いものに変換されるように設定されている。これ以外に, ある都道府県や市区町村に限定して地図表示をしたい場合など, その都道府県や市区町村といった住所の接頭語を指定してアドレスマッチングの範囲を限定することが可能である。
  - ② ドキュメントとアプリケーションのバイディング設定: ドキュメントの種類ごとに, ダブルクリックした場合に起動するアプリケーションを設定できる。
  - ③ 保存場所のデフォルト値の変更: プロジェクトやアイコン画像の保存場所を明示的に設定できる。
  - ④ プロキシサーバー設定: プロキシウェブアクセスする場合に, プロキシサーバーを利用している部署のために, プロキシサーバーの設定も可能にしている。

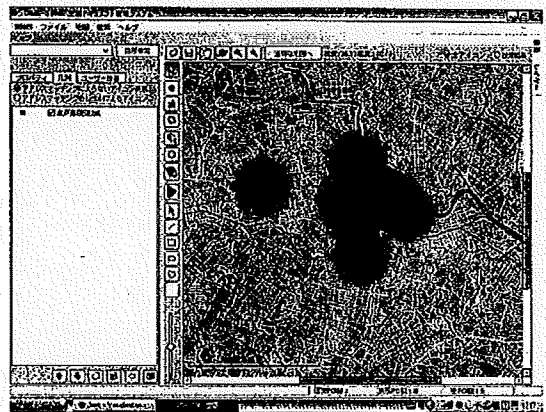


図 5. POI の多重度に応じた濃淡表示機能の例



## 5. ソフトウェア配布に関して

### 5.1. SDMS の動作条件

- SDMS には現在主な2つのバージョンがある。
- SDMS Ver. 2.5 (2006年7月時点の最新版)
- SDMS Ver. 3.1 (2008年3月時点の最新版)
- 動作する基本 OS: Windows XP (Windows Vista や Windows 2000でも動作することが確認できている。)
- インターネットへのブロードバンド接続を必要とする。接続スピードが遅いと、ジオコーディングと地図表示に時間がかかり、使い勝手が悪い環境になる。
- ウェブを閲覧できる環境を必要とする。(ただし、ウェブを見るために、プロキシサーバの設定が必要な場合は、SDMS の環境設定パネルでその設定を行う。)
- メモリ: SDMS Ver. 2.5の場合、512MBでも動作することは確認できたが、SDMS Ver. 3.1では、1GB以上のメモリを推奨する。メモリは大きい方が処理速度が向上し、安定して動作する。

### 5.2. 公開状況とインストール方法

SDMS は、2008年7月時点では、国立保健医療科学院の健康危機管理支援ライブラリシステム (H-CRISIS) にアカウントを持っているユーザだけに、SDMS Ver. 2.5 を公開している。2008年9月には、H-CRISIS に加えて、東京大学空間情報科学研究センターの SDMS ホームページ (<http://sdms.csis.u-tokyo.ac.jp/>) を通して、最新バージョンである SDMS Ver. 3.1 を公開する予定である。

ここでは、H-CRISIS のホームページを通して、SDMS をダウンロードする手順を紹介する。まず、H-CRISIS にログインする。次に、H-CRISIS のトップページから、工具箱の地理情報システムのページに移動すると、インストーラー一式 (SDMS Ver.2.5.1) とマニュアルがダウンロードできる。インストーラー一式から、インストールファイルをダブルクリックすることにより、ほぼ自動的にインストーラーが行われる。SDMS は、Java Application であり、Java がインストールされている必要がある。しかし、SDMS をインストールするときに、Java も同時にインストールするオプションも選べるようになっているので、実際には、Java をほとんど意識しなくても簡単にインストールできる。

### 5.3. サポートと制約

- (a) SDMS は、現在のところ、バイナリを無料で公開していく方針である。ただし、再配布は禁止する。
- (b) ソフトウェアはボランティアで作成された無料の実験的な公開ソフトウェアであり、きめ細かい十分なサポートを行うことはできない。また、SDMS のユーザが、これを使って損害を被っても、SDMS の提供者は何も保証はできない。
- (c) 2008年9月の公開に向けて、SDMS のホームページ

を準備する予定であり、ここで、FAQ や掲示板で問い合わせを受け付けるサービスを予定しているため、ボランティアによる小さなサポート体制は作って行く予定にしている。

- (d) SDMS は、いつも完全に動作するとは限らず、間違った答えも出すことがある。たとえば、ウェブドキュメントなど、タグが多重化している場合、うまくジオパーシングできない場合がある。これを回避するためには、前処理として、手動でドキュメントをまずテキストとして保存した後に、そのテキストファイルを SDMS へドラッグ&ドロップすると良い結果が得られる。
- (e) アドレスマッチングに関しても、思い通りの結果が出ないこともある。たとえば、「昭和」、「平成」、「明治」は年号として書いたとしても、同名の地名があるため、それらの年号を地名として認識して、それぞれ POI として生成することもある。これを回避するためには、SDMS の環境設定で、県名などの住所の接頭語を入力し、アドレスマッチング範囲を事前に限定する使い方が勧められる。
- (f) アドレスマッチングの精度: 街区レベルまでをカバーしている。SDMS が使っている CSIS アドレスマッチングサービスは、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に基づくものである。
- (g) 地図データは、国土交通省 国土地理院の数値地図 25000 (空間データ基盤) を使っている。この地図データ以上の精度やこの地図データに含まれない地理情報は提供することができない。
- (h) SDMS の特徴は、簡便さにある。より高度な情報処理や視覚化を実行したい場合は、エクスポート機能を利用して、データ移動を行い、商用の GIS などを利用すると良い。

## 6. 終わりに

現在、インターネットの世界で普及しているサーチエンジンは、検索のために特別にデータベースを作成したものではなく、人間が人間に読んでもらうために作成した一般ドキュメントを自然言語処理技術と情報検索技術により、大量の一般ドキュメントを機械的にデータベース化して、キーワードなどの簡単な検索方法で利用可能にし、また情報の適切さ (relevance) を自動計算して、検索結果を適切な順に提示する枠組みである<sup>10)</sup>。

場所情報を扱う現在の主流の情報システムである GIS を考えてみると、人間中心のシステム<sup>11)</sup> とは言い難い。われわれは、自然言語の場所記述を含むデジタルドキュメントを対象として、ドラッグ&ドロップという簡単で自然な操作だけで、デジタル地図を作成・操作できる人間中心の場所情報管理システムを検討し、実際に試験システム SDMS を構築して、保健医療現場での場所情報利活用の

促進を実現する現実的な枠組みの体系化に関して研究を行ってきた。つまり、われわれのアプローチは、GISを人間中心の観点から再設計する試みと見なすことができ、従来の空間データ中心システムから、空間コミュニケーション中心システムへと変革させるものと位置付けている。

SDMSを現在試験的に公開しており、当初は多くのユーザが同時にSDMSを利用した場合、サーバの負担が重くなるなどの障害が起きるといった懸念があったが、実際にはそのような障害は起きずに、多くのユーザが同時に利用してもうまく動作できることも確認できている。SDMSのソフトウェアとしての品質は、商用のソフトウェアと比較するとまだ劣るが、その基本機能と基本枠組みはユーザから比較的高い評価を得ているので、今後も品質を上げるための努力を続ける。また、最先端の空間解析や視覚化のアルゴリズム<sup>12)</sup>をSDMSに取り込み、従来のGISに置き換わるように、ソフトウェアツールの完成度を向上させた。最終的には、保健医療従事者が日常的に気軽に利用できる空間情報コミュニケーションツールへと発展させるための改良の努力を行っていききたい。

将来展望としては、このSDMSを音声認識やメールリーダーと組み合わせることにより、音声や電子メールから、関連する地図を動的に表示するような高度な利用環境につながると考えている。また、携帯電話などを中心とするユビキタス環境でも、人々の会話で自動的に地図を提供するなどの気の利いたサービスを実現することも可能と考えられる<sup>13)</sup>。

## 謝辞

本研究は、厚生労働科学研究費補助金（地域健康危機管理研究事業）「地域の社会情報及び地理情報を加味した健康危機情報の分析と支援システムに関する調査研究」（代表：浅見泰司）（H19-健危-一般-009）の支援を受けている。アドレスマッチング処理の一部では、国土交通省国土計画局 国土情報整備室が提供している「街区レベル位置参照情報」を利用させていただいている。背景地図は、国土交通省 国土地理院が提供している「数値地図25000（空間データ基盤）」を利用させていただいている。ソフトウェア配布に関しては、国立保健医療科学院 健康危機管理支援ライブラリーシステム（H-CRISIS）に支援いただいている。

## 参考文献

- 1) 浅見泰司, 有川正俊, 白石陽, 片岡裕介, 相良毅. 空間ドキュメント管理システムの設計と開発に関する研究. 東京大学空間情報科学研究センター第8回シンポジウム (CSISDAYS 2005), 全国共同利用研究発表大会; 2005. p.43.
- 2) 有川正俊, 白石陽, 相良毅, 浅見泰司. 人間-人間コミュニケーション世界で流通するドキュメントの空間化利用環境 - SDMS (Spatial Document Management System) -. 東京大学空間情報科学研究センター第9回年次シンポジウム (CSISDAYS 2006), 全国共同利用研究発表大会; 2006. p.74.
- 3) Sagara T, Arikawa M, Sakauchi M. Spatial document management system using spatial data fusion. In: Third Int'l Conf. on Information Integration and Web-based Applications and Services; 2001; Linz, Austria. p.399-409.
- 4) Shiraishi Y, Arikawa M, Sagara T, Asami Y. Spatial document management system for ubiquitous mapping. Int'l Journal of Urban Sciences 2006; 10(2): 106-14.
- 5) 中谷友樹. 保健医療と地理情報科学. 中谷友樹, 谷村晋, 二瓶直子, 堀越洋一, 編. 保健医療のためのGIS. 東京: 古今書院; 2004. p.1-31.
- 6) 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. 電子情報通信学会 第18回データ工学ワークショップ; 2007; web掲載.
- 7) 相良毅, 有川正俊, 坂内正夫. ジオレファレンス情報を用いた空間情報抽出システム. 情報処理学会論文誌「データベース」2000; 41(SIG6 (TOD7)): 69-80.
- 8) 相良毅, 有川正俊, 坂内正夫. 分散位置参照サービス. 情報処理学会論文誌 2001; 42(12): 2928-40.
- 9) Morville P. Ambient findability. O'Reilly Media Inc.; 2005.
- 10) Langville AN, Meyer CD. Google's pagerank and beyond: the science of search engine rankings. Princeton University Press; 2006.
- 11) Endsley MR, Bolte B, Jones DG. Designing for situation awareness: an approach to user-centered design. CRC; 2003.
- 12) 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』2006; 14(2):1-9.
- 13) Arikawa M, Noaki K. Geocoding natural route descriptions using sidewalk network databases. In: IEEE Computer Society, Int'l Workshop on Challenges in Web Information Retrieval and Integration 2005. p.136-144.

空間ドキュメント管理システム：SDMS

片岡裕介

2010

測量

(社団法人日本測量協会)

最近の情報収集では、インターネット上をはじめとして、電子化されたファイルやデータを扱う機会が非常に多い。効率良く進めるためにも、電子ファイルの情報の中から、自分にとって重要な要素を手早く拾い集めたいところである。

ここで、ご紹介するSDMS (Spatial Document Management System) は、一言で表現すると、電子ファイル内に含まれる住所や地名を、地図上で自動的にマッピングしてくれるソフトである。つまり、ドキュメント内で住所や地名を含む文字列を検索し、それを緯度経度といった位置情報に変換して、地図上に表示するのである。位置情報への変換は、ネットワークを介して行われ、変換ツールとして東京大学空間情報科学研究センター (以下、CSIS) が提供する、「アドレスマッチングサービス」(<http://newspat.csis.u-tokyo.ac.jp/geocode/>) と同じものを使用している。

SDMSの主な特徴として、①ファイルやフォルダをSDMSのウィンドウに「ドラッグ&ドロップ」するのみ、という非常にシンプルな操作 (図-1) ②ワード、エクセル、電子メール、ウェブページなどの形式に対応、が挙げられる。その他に、点分布密度の視覚化や、RSSへの対応をはじめとした、ユーザビリティの向上がこれまでに図られてきた (図-2)。SDMSはCSISの健康危機情報プロジェクト (代表：浅見泰司教授) で開発され、これまで実験的配付となっていたが (図-3)、本年3月に3.5版として、<http://sdms.csis.u-tokyo.ac.jp/> から一般公開される予定である (本稿執筆時)。

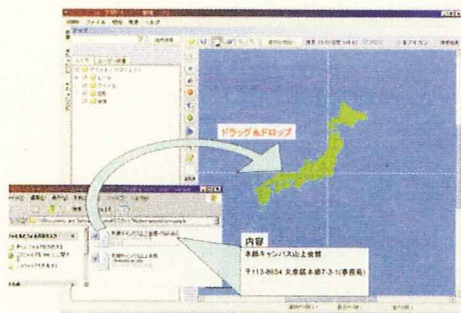


図-1 SDMSの基本操作 (自動マッピング)

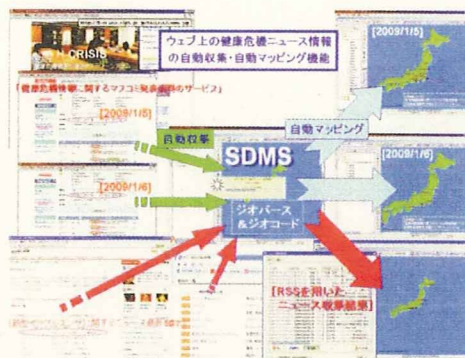


図-2 ウェブ上のニュース情報の自動収集機能



図-3 SDMSウェブサイト (<http://sdms.csis.u-tokyo.ac.jp/>)

SDMSの特徴である「マッピング機能」を実際に使ってみると、予想もしなかった発見をすることがあるのではないかと。どの場所にあるのか、どのように分布しているのか、といった問題について、地図上で距離や方角を頼りに、視覚的・直感的に捉えることができるためである (身近な例としては、Google マップ)。空間情報のリアルタイム管理の面からも、今後は様々な場面での活用が期待される。

(東京大学 片岡裕介)

