

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スする手法を開発した⁶⁾。

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^{7,8)}。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い、操作が容易な環境を作り出すことができる⁹⁾。

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。

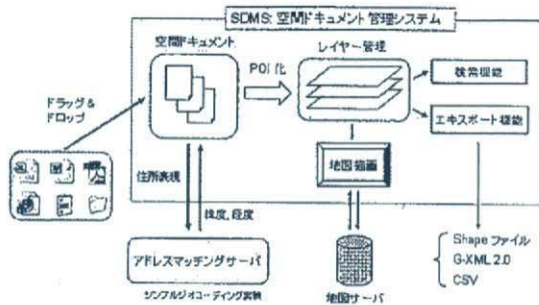


図4. SDMSのシステム構成図

4.1. SDMS のデータ処理の主な流れ

- (1st) 一般ドキュメントあるいはそれらが入ったフォルダをユーザがSDMSにドラッグ&ドロップする。
- (2nd) 一般ドキュメントをプレインテキストに変換する。
- (3rd) プレインテキストを自然言語解析して名詞句を抽出する。
- (4th) 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。
- (5th) プレインテキスト中で、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。」
- (6th) <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。
- (7th) 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。
- (8th) 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。

- (9th) 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に、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きている。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集合を混乱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: プリンタ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。

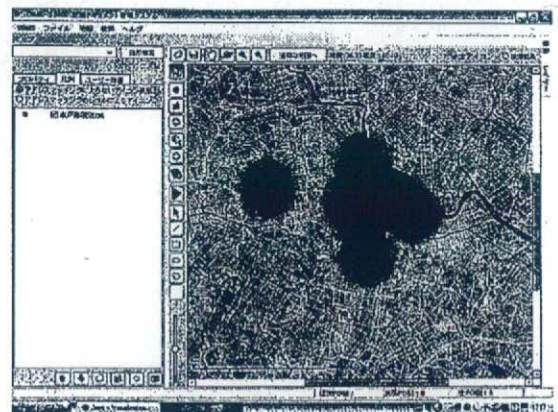


図5. POIの多重度に応じた濃淡表示機能の例

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ーシステム (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、道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ルが行われる。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、うまくジオコーディ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 ヘッドラッグ&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¹⁰⁾。

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¹¹⁾ 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¹²⁾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い。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かと考えている。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¹³⁾。

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.