

地図への文書自動配置機能を有する

空間的文書検索システムの開発

-位置情報サービスのコンテンツ基盤技術-

1. 背景

本プロジェクトを実施する背景には、現在、次の状況が生じてきたためである。

(状況-1) 現在、広域コンピュータ・ネットワークを介してアクセス可能な空間情報を含む情報源は増大している。それらの情報源に含まれる空間情報を介した新しい情報獲得技術の確立が重要となっている。特に、インターネットを介してアクセス可能な情報源の約8割には、本プロジェクトが対象とする地名を含む空間情報が含まれていると言われており、コンピュータ・システムを利用した情報提供サービスが情報源に含まれる空間情報を扱うことは本質的課題として位置づけられる(参考文献1)。

(状況-2) GPS に代表される位置認識ハードウェア技術や携帯電話などの無線ネットワーク技術の発展・普及によって、カーナビや人ナビに代表される位置情報サービス(LBS:Location-Based Services)が生み出され、その価値は日増しに増大している(参考文献2)。しかしながら、位置情報サービスが提供可能な情報が特定の情報源に含まれるものに限定されており、そのことが位置情報サービスの適用範囲を抑えている状況を生み出している。それゆえ、位置情報サービスのためのコンテンツ技術を広域ネットワーク環境に分散配置された情報源を対象とするものとして実現することが本質的課題として位置づけられる。

参考文献1：有川正俊，位置情報サービスとサイバースペースの融合，日本バーチャルリアリティ学会誌第7巻3号，pp.120-125 (2002)

参考文献2：<http://www.zdnet.co.jp/mobile/news/0105/29/goo.html>

2. 目的

本プロジェクトでは、広域ネットワーク環境に分散配置された情報源を、その情報源に含まれる位置情報が指す地図上の位置に自動配置し、地図を介して検索するための技術を開発する。これにより、背景に述べた状況が抱える課題を解決する。特に、本プロジェクトでは文書データからなる情報源を対象とし、文書データをそれに含まれる地名が指す地図上の位置に自動配置し(以後、この機能を「地図への文書自動配置機能」と定義する。)、検索するための技術の開発を行う。

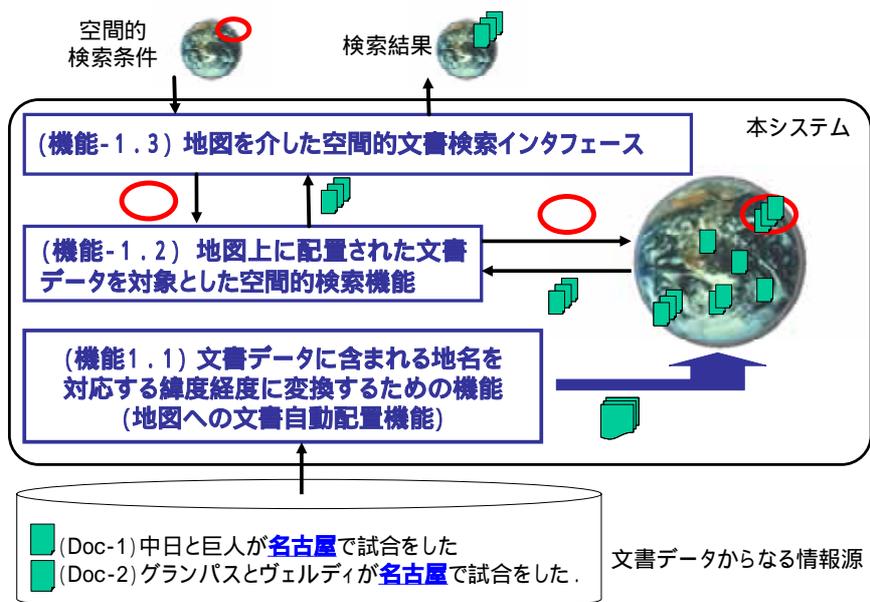


図1: 地図への文書自動配置機能を有する空間的文書検索システムの構成

3. 開発の内容

ここでは、本プロジェクトにおいて開発を行った「地図への文書自動配置機能を有する空間的文書データ検索システム」について述べる。

図1は、地図への文書自動配置機能を有する空間的文書データ検索システムの構成図を表す。本システムの新規性は(機能-1.1)の実現方式にある。具体的には、文書データに含まれる地名を、地図データに登録されている住所名と直接マッチングするのではなく、地図に含まれるランドマークに特徴を表す単語群をメタデータとして割り当て、そのメタデータと文書データの相関量を計算することにより、文書データと高い相関のあるランドマークを自動選択する機能を実現する。これにより、一般に、文書データに現れる地名には、図1の「名古屋」のように、地図データに登録されている完全な住所名として記述されていないものが多く存在するが、(機能-1.1)は、そのような地名から対応する緯度経度を自動的に決定することが可能になる。例えば、図1の「名古屋」について、(機能 1.1)は、Doc-1の「名古屋」を中日の本拠地であるナゴヤドームの緯度経度に、Doc-2の「名古屋」をグランパスの本拠地である瑞穂競技場の緯度経度に変換する。その実行手順については、(参考文献3)、(参考文献4)に詳細に記述したので、そちらを参照されたい。

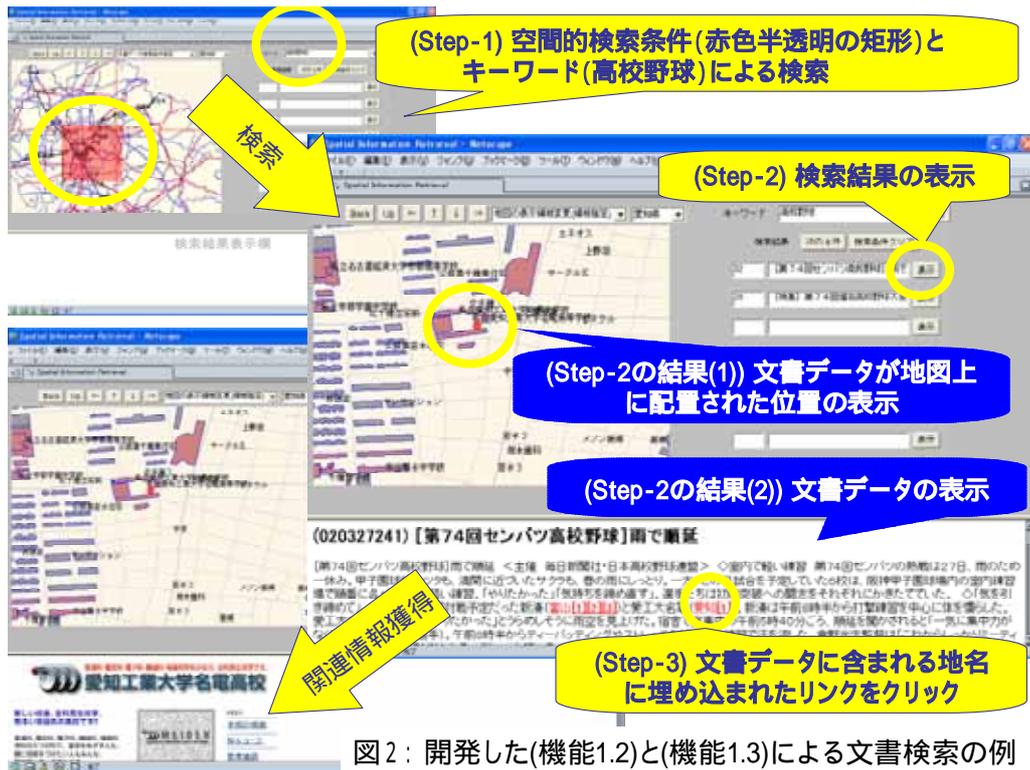


図2：開発した(機能1.2)と(機能1.3)による文書検索の例

図2は、(機能-1.2)と(機能1.3)による文書データ検索の実施例を表す。その実施例における検索手順は次のとおりである。

(Step-1) 空間的検索条件(名古屋市中心部を含む赤色半透明矩形)とキーワード(“高校野球”)を与え、実行ボタンを押すことにより検索を行う。ここで、空間的検索条件に対応する赤色半透明の矩形は、検索者がマウスを用いてその対角2点をクリックすることによって作成したものである。検索される文書データは、与えられたキーワードをその文書データ内に含み、空間的検索条件として指定された矩形内に配置されているものである。その結果は、検索結果リスト表示部に表示される。図2では、2件の文書データが検索されている。

(Step-2) 最上位にランクされた文書データ(内容はセンバツ高校野球に関するもの)を閲覧するために、その文書データの表示ボタンを押す。その結果、その文書データが配置された位置が空間的検索条件指定部に描画されている地図の赤色マーカにより示され(図2:Step-2の結果(1))、さらに、検索された文書データ表示部にその内容が表示される(図2:Step-2の結果(2))。

(Step-3) 文書データに含まれる地名をクリックすることにより、検索者は関連する情報獲得を行う。本システムでは、開発のポイントでも述べたように、検索された文書データ表示部に表示される文書データに、その地名が出現する位置に関連する Web ページへのリンクを埋め込むので、この例のように、その地名をクリックすることにより、検索者は関連する情報獲得を行うことが可能になる。

開発した地図への文書自動配置機能の配置精度に関しては、1000件の文書データ(そ

れらに含まれる地名総数は3952)を対象として評価を行った。その結果、配置候補を3件までに拡大すると、7割強になることを確認した。これより、開発した地図への文書自動配置機能の実現方式に関して妥当であることを明らかにした。

本ソフトウェアは、Java と C 言語によって開発を行い、Web サーバ(具体的には Tomcat)の機能を拡張することによって、実装を行った。また、空間的検索機能については、PostgreSQL のユーザ定義関数として実装を行った。その他に、前文検索エンジン Namazu と ChaSen を利用した。これらのソフトウェアが動作する環境において、本ソフトウェアは実行可能である。

(参考文献3)細川宜秀: 地図への文書自動配置機能を有する空間的文書検索システム. 情報処理学会 夏のプログラミング・シンポジウム「データの達人」報告集, pp.1-8 (2003)
(参考文献4) Hosokawa, Y. and Naohisa, T.: "A context-dependent geo-coding method for document databases," to appear in Information Modeling and Knowledge Bases (IOS Press), Vol. 16 (2004).

4. 従来技術(または機能)との相違

本システムの特徴は、先の例のように、文書データに含まれる地名が完全な住所名かランドマーク名として記述されていなくても、文書データの文脈を認識して、その地名が指すランドマークを特定できる点にある。現在、住所名やランドマーク名から地図上の位置を検索する WWW サイトがあるが、この特徴は、それらのシステムとの最も大きな違いである。このシステムの特徴的な利用方法の1つとして、お店の口コミデータを、そのお店の周辺にいる人に即座に配信することが考えられる。このとき重要なことは、口コミデータを書く人が、店名や代表メニューなどのそのお店を表す特徴と、そのお店がある市名などの大まかな地域名を知っていれば、そのお店の住所や緯度経度を知らなくても、その口コミデータを地図上に配置できることである。これは、そのお店の正確な位置を気にせずに記述された口コミデータを、そのお店の周辺にいる人に配信することができることを意味する。さらに、口コミデータを必要とする人にとっては、位置に基づいて必要なお店の口コミデータのみを選別できるようになる。

5. 期待される成果

本開発で実現する文書検索システムは、主に次の2つの社会的貢献に寄与するものと考えられる。

(貢献-1) 文書データからの知識情報獲得の精度向上

文書データは、知識情報を保持・配布するための主要なメディアであるので、文書データからの知識獲得技術の確立が重要課題として位置づけられる。現在、単語ベースの文書検索システムは、社会に浸透し、その価値が認められている。本開発で実現するシステムは、単語ベースの文書検索システムの問題とする検索結果中のノイズの多さを空間的関連性評価によって削減するものである。したがって、本システムは、検索結

果からの知識獲得のオーバーヘッドを削減でき、社会における文書データから獲得された知識の活用を活性化できるものと期待できる。

(貢献-2) モバイル情報化社会における知識獲得のためのコア・エンジン

現在、モバイル・コンピューティング技術が発展し、人間の空間的状况に応じた情報提供サービスが実現されている。本システムは、利用者への空間上の位置と文書データの緯度経度間の関連性評価機能を有しているため、今後、モバイル情報技術の発展によって迎えるであろうモバイル情報化社会における情報提供のためのコア・エンジンとしての価値がある。

6. 普及(または活用)の見通し

本プロジェクトにおいて実現したソフトウェアでは、その実行においてランドマークに関するメタデータを使用するが、本ソフトウェアの性能がこのメタデータの整備次第であることを考慮すると、ランドマークのメタデータが現実的に準備できるような情報サービスが開発できるかによって、本ソフトウェアの実用化が決まるといってもよい。幸い、GPS、PDA、無線ネットワーク技術などモバイル情報サービス実現のためのハードウェア技術が実用化段階にあるので、これらの技術と組み合わせて、特定地区限定のレストラン情報配信などメタデータを振らなければならないランドマーク数が少なくすむようなモバイル情報サービスの実現が本ソフトウェアの普及の鍵となると考えられる。

7. 開発者名(所属, e-mail アドレス)

細川 宜秀 (名古屋工業大学大学院工学研究科 助手 (おもひ領域),
hosokawa@nitech.ac.jp)

(参考)開発者 URL

開発者紹介: <http://tk-www.elcom.nitech.ac.jp/~hosokawa>

開発成果の紹介とサンプル: <http://tigerhill.elcom.nitech.ac.jp>