

地図インタフェースを利用した多種情報源の検索

村本達也 永末壽宏 北角智洋 池田哲夫
NTTサイバースペース研究所

1. はじめに

我々は従来、住所の選択や地図の操作で選択された店舗などの施設情報を施設情報DBから検索して検索結果を地図上にアイコンの形で配置表示し、更にアイコンを選択することにより詳細情報を参照できるシステムを提案した[1]。一方、Web上のサービスとして緯度経度等の空間情報等を検索キーとして該当する空間あるいは近隣の空間に位置する施設の情報を提供するサイトが存在する。

地図情報提供サービスの今後を考えたとき、アプリケーション掛け付けの施設情報のDBだけではなく、WWW上で提供される多種の情報をも横断的に検索したいというニーズが生じるものと考ええる。

本稿ではこのようなニーズに応える為に、既存の施設情報DBと位置関連情報を提供するWWW情報等の情報源とを連携させることで、利用者が多種の情報源を統一的に検索することを可能にする拡張方式を提案する。

2. 多種情報源検索の課題

従来提案したシステムでは単独の情報源をシステム内部で管理し、この情報源に適応した検索インタフェースを提案していた。

しかしWWW情報を含む他者が管理する位置関連情報情報源をも検索対象とするには、多種情報源のDBの異種性を考慮する必要がある。具体的には(1)検索キーのフォーマットが各情報源で異なる場合がある、(2)各情報源でデータを説明するデータ項目(カラム名)や単位等が異なる場合がある、(3)情報源によって欠落するデータ項目がある、という3点の課題が存在する。

3. 多種情報源の連携方式

多種情報源を検索する為に基本的にはクライアントからの検索要求をサーバ内のDBと他のDBとに渡し

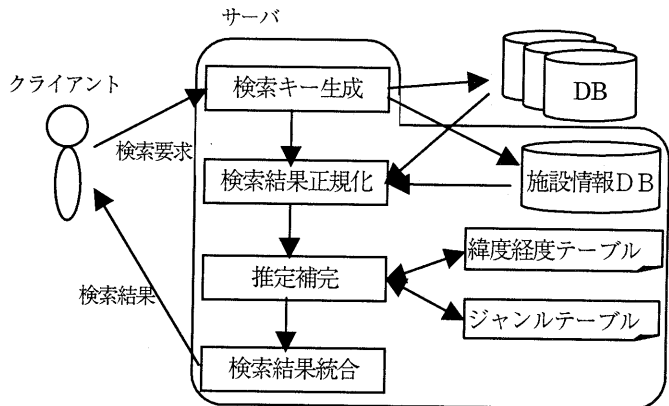


図1：機能構成の概観

て検索結果を受け取り、これらを統合してクライアントに返す構成となる。ここで上記3点のDBの異種性から来る課題を図1に示すモジュール構成で対処し、多種情報源の連携を可能にする。

検索キー生成のモジュールでは、クライアントからの検索要求を予め用意してある各情報源の検索キーのフォーマットに合わせて検索キーを生成し、第1の異種性を解消する。例えば、クライアントからの検索要求が「北際=y1, 南際=y2, 東際=x1, 西際=x2」という書式であり、情報源Aに対しての検索キーフォーマットが中心座標を表す「経度=X, 緯度=Y」であるならば、

$$(X,Y) = ((x1+x2)/2, (y1+y2)/2)$$

という変換を行い検索キーを生成する。

検索結果正規化のモジュールでは検索結果のデータ項目とデータの単位をクライアントが要求する形式に統一し、第2の異種性を解消する。この部分はDBSENA[2]の機能を用いる。

推定補完のモジュールではある情報源の検索結果に、地図上へ配置表示する為に必要な緯度経度の情報や検索の絞り込みに必要な施設情報のジャンル情報が無い場合にその情報を推定補完し、第3の異種性を解消する。

緯度経度を推定するには緯度経度テーブルを用いるが、これはサーバ内に保持する施設情報DBから(住所, 緯度, 経度)のデータ項目のみを抽出して作成したものである。具体的に推定方法を説明すると、検索結果に含まれる住所を用いて、(1)緯度経度テーブルに完全マッチする住所があればその緯度経度を返す、

A method of retrieving multiple information sources in GIS

Tatsuya MURAMOTO, Toshihiro NAGASUE,
Tomohiro KITAKAKU, Tetsuo IKEDA
NTT Cyber Space Laboratories

1-1 Hikarinooka Yokosuka-Shi Kanagawa 239-0847

(2) 完全マッチする住所が無ければ前方最長一致する住所群の緯度経度を差異部分の逆数を係数とした加重平均を取って返す、という方法をとる。例えば検索結果が「A市B町1-4」で、緯度経度テーブルには①「A市B町1-1」、②「A市B町1-2」、③「A市B町1-8」があった場合、

緯度経度 $= (1/3① + 1/2② + 1/4③) \div (1/3 + 1/2 + 1/4)$
 という加重平均で推定する。この方法で住所に対応する正確な緯度経度が得られる訳では無いが、地図にアイコンを表示配置する上で実用上問題ない精度で推定補完が出来る。

ジャンルを推定するにはジャンルテーブルを用いるが、これはサーバ内に保持する施設情報DBが参照するコンテンツを形態素解析して頻度の高い単語（キーワード）を抽出し、ジャンルとジャンルに対応するキーワード群とをテーブル形式で整理したものである。具体的に推定方法を説明すると、①検索結果のタイトル、施設名称、紹介文、ソーステキスト等に含まれるジャンルテーブルのキーワードをカウントし、②ジャンル毎にカウントの総和を取り、③総カウント数の最も多いジャンルを推定ジャンルとする。この方法により人手を介さず実時間で検索結果にジャンルを決定し付与することが出来、検索結果の絞り込みに利用することが出来る。

検索結果統合モジュールでは上記モジュール群で複数情報源の異種性を解消した検索結果を統合してクライアントに提示する。

この連携方式により多種情報源の異種性を解消することができる。即ち、緯度経度テーブル・ジャンルテーブルと各情報源毎の検索キーフォーマットとを用意するだけで、クライアントの検索インタフェースを変更すること無く、またDBのスキーマ等を変更すること無く多種情報源の情報を提供することが出来る。

4. 地図アプリケーションへの実装

提案方式の有効性を確認するために、検索キー生成・検索結果正規化・緯度経度の推定補完・検索結果統合を実現した図2に示すプロトタイプを作成した。本プロトタイプは、従来提案したシステムに対して多種情報源を検索可能なように拡張したものである。本プロトタイプでは他の情報源としてモバイルインフォサーチ[3]、GeoLink[4]を適用した。モバイルインフォサーチは巡回ロボットが収集したWebページ群を緯度経度や郵便番号などの空間情報を検索キーとして検索することが出来る位置指向検索エンジンであ

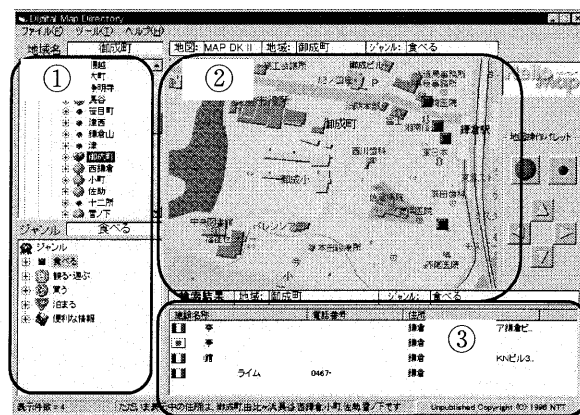


図2：プロトタイプの画面例

り、GeoLink はジャンル毎に整理されたWebページを地図操作によって検索できるシステムである。本プロトタイプの動作を簡単に説明すると①ディレクトリ部分の操作で検索する地域・ジャンルを絞り込んだり②地図表示部分の拡大・縮小・スクロールの操作で表示領域を指定することによって利用者の検索要求が生成される。これを各情報源毎の検索キーに変換して各情報源の施設情報を検索する。推定補完され、統合された検索結果は③検索結果部にリストアップされ、推定補完した緯度経度情報を用いて②地図表示部分にアイコンが配置表示される。

これにより既存の施設情報DB、モバイルインフォサーチ、GeoLink という3種の異なる情報源を統一したインタフェースで検索・閲覧することが可能になる。

5. おわりに

本稿では既存の施設情報DBと他の情報源とのDBの異種性を解消し連携することで、多種の情報源を统一的に検索することを可能にする拡張方式を提案した。今後は本方式の評価を行う予定である。

参考文献

- [1] 星野他：ディレクトリを用いた情報提供に関する一手法～地図情報への適用～，情報処理学会研究会資料，2000-DBS-120 (2000)。
- [2] 星野他：“DBSENA：マルチデータベース環境における情報資源管理と管理方式”，情報処理学会研究会資料，98-DBS-114 (1998)。
- [3] <http://www.kokono.net/>。
- [4] <http://www.digitalcity.gr.jp/openlab/kyoto/>。