

地理情報部品同定モジュールの実現による コンテンツ検索結果の Google Maps 表示システムの開発

齋藤功 遠峰隆史 嶋津恵子

慶應義塾大学デジタルメディア・コンテンツ統合研究機構

1. はじめに

現在地図を用いた情報表示手法は、様々なアプリケーションで用いられている。例えばインターネット上のあるサービス¹は、施設や観光名所などのコンテンツを地図上に配置する。またカーナビゲーションシステムは、現在地から入力した目的地に到達するための経路情報を地図上に出力する。この手法の実現には、(a)コンテンツの地理情報(緯度経度値)をキーにして地図上に配置する仕組み、(b)コンテンツの地理情報を正確に漏れなく特定する仕組みが必要である。前者は Google² が 2005 年 6 月に公開した Google Maps API³ を利用することで、低コストで実現できるようになった。これにより 2005 年 6 月から 2006 年 11 月までの日経 BP 社発行のビジネス・IT 関連誌 41 種類には Google Maps API の利用事例の記事 63 件が掲載された。後者(b)は具体的には、b-1) コンテンツ中の地理情報を示す文言の特定と、b-2) それらの緯度経度値への対応付けの機能により実現される。Google 自身が提供する地域情報検索サービス⁴では、b-1) の機能に関し、地理情報を示す文言のうち住所として特定できるものだけ取り扱い、緯度経度を算出する。この方法はコンテンツが住所以外の地理情報(“横浜”, “宇都宮”等)を持つ場合は処理の対象とならない[1]。

一方、例えば福澤諭吉の海外の滞在都市(名)を利用したい場合は、“イギリス”や“サンフランシスコ”程度の粒度の表現で充分である。また、市町村名や字等の住所の一部から、ある程度正しい緯度経度をひも付けられれば、[2]の主張する地図を用いた感覚的な UI による情報検索がより有効になる。そこで我々は、住所として成立しない粒度の情報および住所の部品である情報(具体的には国名、地方自治体名、字、建物名、

駅名、道路名、路線名、特殊地形名)を“地理情報部品”とし、これを元にコンテンツ検索結果を地図上に表示するシステムを開発した。

2. 地理情報部品を使ったコンテンツの地図上への表示

2.1. 地理情報部品の特定

特定の語をテキストから抽出するには(ア)辞書を備えた形態素解析エンジン[1]か、もしくは(イ)前後の文脈から名詞の種類を推定する高度な構文解析器のどちらかの利用が考えられる。一方、実用システムの構築には運用コスト(辞書のメンテナンス工数)の最小化と語の特定に関する程度の精度の保証とを両立させたい。そこで我々は、両方の機能を備えた半構造データ処理技術である UIMA⁵を地理情報部品の抽出に利用した。この技術は製薬における臨床試験データベースからの情報取得⁶や、企業の顧客対応部門におけるアンケート分析⁷において、前者は医学専門用語の、後者は顧客が望む製品/サービス用語の特定に利用されている。

UIMA は場所を示す語を発見すると、最長単位を“lex”としてタグを付ける。次にこの語を最少単位に分解しそれぞれに別のタグをふる⁸。我々は、この最小単位の語を地理情報部品として利用し、コンテンツ検索テーブルを更新した(図の(i))。

2.2. 緯度経度のひも付けと地図上への表示

地理情報部品の緯度経度へのひも付けは、地理上の何らかの名称、対応する経度緯度、アルファベット表記の値から成る参照テーブルを利用した⁹。(ここで、我々のいう地理情報部品は“地理上の何らかの名称”の部分集合。)同定された地理情報部品の緯度経度を使って検索結果のコンテンツを、Google Maps API 経由で Google の

Isao Saito, Takashi Tomine, Keiko Shimazu
Research Institute for Digital Media and Content, Keio University

¹ <http://www.mapfan.com/> や <http://www.mapion.co.jp/> 等

² <http://www.google.co.jp/>

³ <http://www.google.com/apis/maps/>

⁴ <http://maps.google.co.jp/>

⁵ <http://www.research.ibm.com/UIMA/>

⁶ http://messier51.blogspot.com/2006_04_09_messier51_archive.html#114487609986862839.

⁷ <http://jisi.dreamblog.jp/blog/235.html>

⁸ 実際のシステム詳細は、UIMA をライブラリとする固有名詞抽出ソフトウェアがこれらの処理を行う。

⁹ <http://gpscycling.net/fland/fla/flapna/flapna.htm>

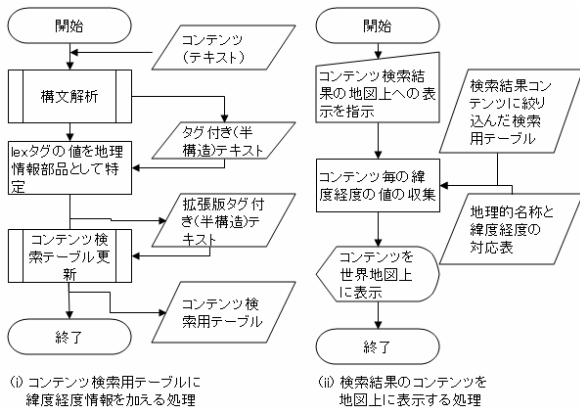


図. 検索結果コンテンツの地図上への表示

サーバに送信し 世界地図上に表示する (図の (ii)). またこのシステムでは, 同 API の提供機能である地図の縮尺の変更を利用することで, 表示コンテンツ数の絞り込み支援を行う.

3. 実験と考察

コンテンツ検索結果を地図上に表示するシステムの有効性を確認する実験をおこなった. これは, 従来のキーワードの追加によるコンテンツの絞り込み検索との比較である. 具体的には 50 件の検索事例を用意し, (1)結果の絞り込みの度合い (最初の検索実行時に出力されるコンテンツ数を 100 としたときの, その後続く再検索で出力数が削減する割合)と, (2)絞り込み結果に含まれる適性コンテンツ (検索の目的に合致したコンテンツ)の割合を計測した. 特にコンテンツ検索結果を地図上に表示するシステムの場合は, コンテンツ内の住所表記だけを対象とするものと, 我々の提案である地理情報部品すべてを用いた場合の両方を実験した.

表 1 は(1)と(2)の結果である. キーワードの入力による検索は急激に絞り込まれ, 地図の縮尺の変更によるものは比較的なだらかに絞り込まれている. また, 適性コンテンツの出現割合は, キーワードの追加による絞り込みは高精度を維持し続けるが, 地図の縮尺の変更によるものは特に住所だけを利用したときは精度が低い.

表 1 の適性コンテンツの出現割合は全 50 の実験の結果の平均を表現しているが, これに対し, 実験ケースごとの精度のばらつきを表したのが表 2 である. (a)はキーワードの追加による絞り込みの全結果, (b)は地図の縮尺の変更による絞り込み (地理情報部品を利用)の全結果である. 前者は実験事例によってばらつきが大きく, 後者は小さい.

これらのことから, 検索の緩やかな絞り込みと, 安定した精度を期待できる点で我々の提案手法

表 1. 実験ケースごとの絞り込みの度合いと精度

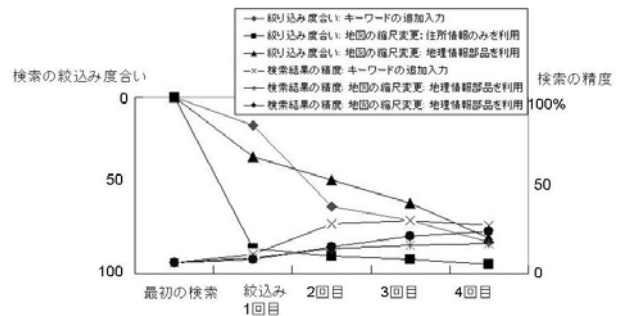
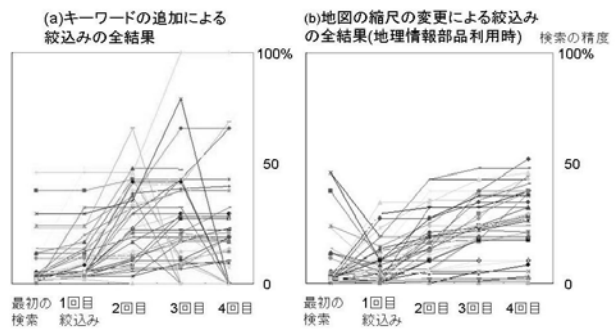


表 2. 実験ケースごとの精度のばらつき



が有効であることがわかった.

[2]は, 滑らかな検索用に地図上への検索結果表記を主張し, [1]はそれを有効に活用するために住所を付加したコンテンツの量産を前提としている. 我々の提案する手法は, この両者に対し貢献するものである.

4. まとめ

我々は, 住所として成立しない粒度の情報および住所の部品である情報 (地理情報部品) を元にコンテンツ検索結果を地図上に表示するシステムを開発し, その有効性を実験した.

謝辞

本稿は, 文部科学省科学技術振興調整費の支援による研究の一部である.

参考文献

- [1] 横路誠司, 高橋克巳, 三浦信幸, 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 Vol.41, No.7 2000年7月
- [2] 水口充, 増井俊之, ボーデンジョージ, 柏木宏一. なめらかなユーザインタフェースによる地図情報検索システム. インタラクティブシステムとソフトウェア III: 日本ソフトウェア科学会 WISS'95, pp. 231-240. 近代科学社, 1995