

Geo マイクロフォーマットを用いた 住所自動検出・地図表示システムの開発

福田拓真[†] 力宗幸男[‡]

兵庫県立大学 大学院 応用情報科学研究科^{†‡}

はじめに

近年、ネットワークの転送量の増大やパソコンの処理能力の向上等から、Web 上で大容量のデータを取り扱えるようになり、サーバ上の地図や動画等にアクセスして情報を得る Web サービスが増加している。さらにこれらの Web サービスの多くは Web API を公開しており、外部の Web ページ内にそれらのコンテンツを組み込む事が出来る。

これによってユーザは文章とそれを補足する地図や動画、画像等を組み合わせたリッチなコンテンツを構築出来るようになった。このようなコンテンツは、閲覧者にとっても文章だけで構成されたコンテンツよりも内容を理解しやすく、ある地点からコンテンツ中に記述されている場所までの交通手段等の、関連情報を得る等の拡張的な利用も見込む事が出来る。

しかしながら、現状では文章と画像だけで構成されたコンテンツが一般的であり、そのようなリッチなコンテンツは少ない。

そこで、それらの通常のコンテンツをリッチなコンテンツへと変換する事が考えられる。

本研究では、ブラウザ上に表示される様々な情報の中でも住所情報に着目し、兵庫県立大学元木らが開発した住所自動検出・地図表示システム^[1]のアルゴリズムを応用し、ブラウザ上に表示された情報の中から住所情報を検出して緯度経度情報に変換、マイクロフォーマットとして当該 XHTML 内に埋め込み、Operator や、Minimap Sideba 等の Firefox アドオンで読み込んで地図等を表示するシステムを開発した。

先行研究

本研究では、兵庫県立大学元木らが開発したシステムのアルゴリズムを応用している。元木らのシステムは、図 1 のように Web ブラウザ上に表示されている

る住所を自動検出し、Google Map 上でのその住所の地図へのリンクを生成して、その地図のサムネイル画像を表示させるというものである。

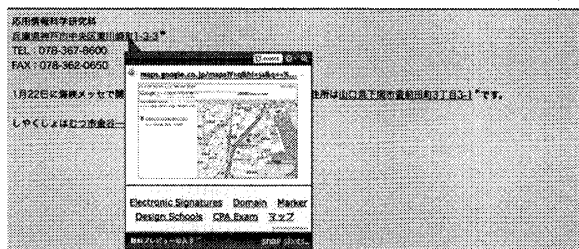


図 1. 元木らのシステム画面

このシステムには、Google Map の地図しか表示出来ない、住所情報の誤検出・検出漏れが起きる等の問題点がある。

住所自動検出

本システムでは、まず Web ブラウザ上に表示されている住所情報を自動抽出し、それらを緯度経度情報に変換する。

住所情報の自動抽出には、元木らのシステムのアロリズムを応用したアロリズムを作成した。元木らのシステムのアロリズムでは、都道府県名の表記の無い住所を検索する場合に、郡または、市または村または町から始まり、30 文字以内に“アラビア数字”、“漢数字”、“アルファベット”、“番地”、“の”、“号”、“丁目”、“一 (ハイフン)”、“棟”、“無番地”、“先”、“通り”、“条”、“地割”、“地”、“甲”、“乙”、“丙”の文字で終わる文章や、郡または、市または村または町から始まり、“アラビア数字”、“漢数字”、“アルファベット”、“番地”、“の”、“号”、“丁目”、“一 (ハイフン)”、“棟”、“無番地”、“先”、“通り”、“条”、“地割”、“地”、“甲”、“乙”、“丙”がなく、郡、市、町、村以降に漢字のみが用いられる文章を住所として検出していた。

[1] An Automatic Address Detection and Map Display System using Geo Microformat.]

[†] [Fukuda Takuma · Graduate school of applied informatics University of Hyogo]

[‡] [Yukio Rikiso · Graduate school of applied informatics University of Hyogo]

しかし例えば、「古本市場 2 号店」というように、「市」や「郡」などの文字を含み数字で終わるような文章も住所として検出されてしまう為、誤検出が起きる。また、逆に、「店舗所在地加古川市加古川町北在家 2000」というような、地名の前後に漢字や数字が付随した文字列が住所として認識されない為に検出漏れが起きていた。

この問題点を解決するため、上記のアルゴリズムを廃止し、全国の全ての市、郡、および一部の町村の名称を登録して照合を行うアルゴリズムを追加した。

ジオコーディング

元木らのシステムでは、検出した住所情報をそのまま Google Map へのリンクに組み込んで使用していた。しかし検出した住所情報は日本語によって表記されたものであるため、日本語に対応したサービスでしか利用する事が出来ない。

そこで、検出した住所情報をジオコーディングする。ジオコーディングとは、住所情報を緯度経度情報へ変換する事である。検出した住所情報を、数値でユニークに地球上の位置を表現する事が出来る緯度経度情報へと変換を行って、あらゆる言語のサービスに対応出来るようにする。

マイクロフォーマット

ジオコーディングして生成した緯度経度情報は、Geo マイクロフォーマットとして当該 XHTML ファイルに埋め込む。

マイクロフォーマットとは、Web ページ上に様々な情報を記述する際に用いるマークアップの法則である。文章中の単語や語句が、どのような意味体系を持ったものであるかをコンピュータが理解出来るようにする為に用いられる。マイクロフォーマットの中で、地理情報を表すのに用いられるものの一つが、Geo マイクロフォーマットである。マイクロフォーマットは通常の Web ブラウザでは利用する事が出来ないが、Operator や Minimap Sidebar 等のアドオンを組み込んだ Firefox を用いることで、Geo マイクロフォーマットの情報を利用して地図情報を表示させる事が出来る。

システム概要

本システムは、Firefox のアドオンとして開発した。そのため、Firefox が起動する環境ならば、OS やハードウェア等の使用環境に依存せず利用する事が出来る。

本システムを起動すると、前述のアルゴリズムを用いて Web ブラウザ上に表示されている住

所情報を自動抽出する。抽出した住所情報を Google Maps API を用いてジオコーディングし、Geo マイクロフォーマットとして当該 XHTML ファイルに埋め込む。そして、図 2 のように Operator もしくは Minimap Sidebar を使用し、埋め込まれた緯度経度情報から地図情報を検索・表示する。

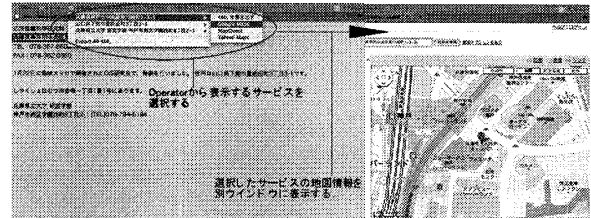


図 2. 本システムの実行画面

まとめ

本研究によって、住所自動検出・地図表示システムシステムの、住所情報の自動検出アルゴリズムが大きく改善された。「市」や「郡」を含む文字列を住所として誤検出するケースや住所表記の前に漢字が連続して書かれている文字列を住所として検出出来ないケースについては、誤検出が大幅に低下した。

また、Google Maps API を用いて緯度経度情報に変換して、Geo マイクロフォーマットとして XHTML ファイル内に埋め込む機能を実装し、Operator や Minimap Sidebar 等の Firefox アドオンを併用する事で、Google Map だけでなく様々な Web サービスの提供する地図情報を利用出来るようになった。これにより、テキストや画像等のみで構成された単純な Web コンテンツであっても、コンテンツ内に住所情報が含まれていれば、その住所情報から様々な Web サービスが提供する地図情報を検索・表示するリッチな Web コンテンツへと変換する事が可能となった。

しかしながら、住所表記の直後に数字等が付随している場合、誤検出や検出漏れを起こしてしまうケースについては対処する事が出来なかった。今後は、広くシステムを公開し、ユーザの評価を受けながら、システムの機能と検出アルゴリズムを更に改善し、より正確性の高いシステムを構築したい。

[1] 元木諒介, 力宗幸男, 「住所自動検出・地図表示システムの開発-Firefox アドオンとして-」, 電子情報通信学会, オフィスインフォメーションシステム研究会, 信学技報, Vol. 108 No. 397, pp. 1-6, OIS2008-65, 2009