

## 住所情報を用いた Web リンク情報の実世界マッピング手法

加藤 義大† 櫻田 健† 河口 信夫‡

†名古屋大学大学院工学研究科

‡名古屋大学未来社会創造機構

## 1 はじめに

実世界に存在する施設や飲食店の情報は、住所や緯度・経度などの位置情報と紐付けされ、Web 上に無数に記述されている。例えば、施設や飲食店などは住所と共にサービスの詳細をホームページで公開している。また、SNS や個人のブログでは特定の店舗に対する評価やレビューが投稿されている。このような Web 上の実世界情報は、これから向かう飲食店や旅行先などを決定する際の行動支援に役立っており、適切なデータ整理により Web ページを分類する技術が必要とされる。

位置情報と結びついた Web ページを検索するシステムは複数存在し、NTT の Mobile Info Search では、GPS などの位置情報から電話帳を利用した検索を行い、周辺施設の Web ページ推薦を行った [1]。また、McCurley はページの持つ位置情報を利用し、あるページと地理空間上で近くに存在する別のページを探し出せるブラウザを開発した [2]。飲食店の検索を例に挙げると、食べログ、ぐるなび、ホットペッパーといった店舗や施設の検索サイト（以下、商用検索サイト）では地域やカテゴリで店舗を絞り込み、評価・評判と共に飲食店の情報を推薦する。

しかし、これらの位置情報検索システムでは、ページと位置情報の繋がりを利用するだけで、Web 上でのページ同士の繋がりまでは考慮していない。そこで、本稿ではページ同士の繋がりをマップ上に可視化することで、Web 上の情報をより詳細に実世界へマッピングする。具体的には、Web クロールにより収集した位置情報を持つ Web ページをマッピングし、それらのページをリンク情報に基づき結線することで、Web リンク情報がどのようにマップ上に可視化されるかを示し、その有効性を検討する。

## 2 Web リンク情報のマッピングシステム

Web リンク情報の実世界マッピングの流れを図 1 に示す。各処理の詳細を以下に述べる。

まず、Web クロールによるページの収集を行う。商用検索サイトで「都市名（本稿では名古屋市に限定し

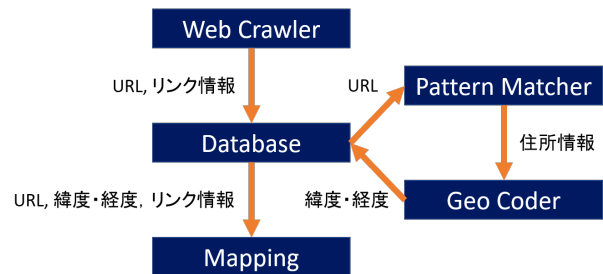


図 1: Web リンク情報の実世界マッピングの流れ図

た)」と検索して得られた検索結果一覧ページをクロールの開始位置とし、HTML パーサを用いて、そのページから辿れる URL を再帰的に取得し、データベースに保存する。Web クロールは幅優先探索として、開始位置から 5 ステップ以下で辿れる URL を収集した。また、Web クロールの際、ページ同士のリンク情報も同時にデータベースに記録していく。

次に、住所情報の抽出を行う。Web クロールにより得られたページに対して、正規表現によるパターンマッチングを行い、住所情報を取得し、Yahoo! Open Local Platform(YOLP)[3] のジオコードを使用して、住所を緯度・経度に変換する。

最後に、緯度・経度と結びついた Web ページをマッピングする。マッピングには OpenLayers3 を利用し、OpenStreetMap(OSM)[4] に描画する。Web ページは色のついた丸い点（以下、ポイント）で表し、ポイントを指定すると、そのページと Web 上でリンクした別のポイントへ結線されるシステムとなっている。

## 3 結果と考察

## 3.1 マップ全体図

図 2 に、名古屋市内に URL をプロットしたマップを示す。現時点で Web クロールにより収集した 1,343,656 件の URL のうち、名古屋市内の住所が記載されている 135,377 件のページをプロットし、そのページから辿ることのできるリンク数に応じてポイントを色分けした。また、同一住所および施設に関する情報を記載したページが複数存在する場合は重ねてプロットしているため、マップ上では 10,259 箇所にポイントが記載されている。i タウンページ [5] で名古屋市内の店舗・施

A Method of Hyperlink Network Mapping to Real World  
 †Yoshihiro KATO †Ken Sakurada ‡Nobuo Kawaguchi  
 †Graduate School of Engineering, Nagoya University  
 ‡Institute of Innovation for Future Society, Nagoya University

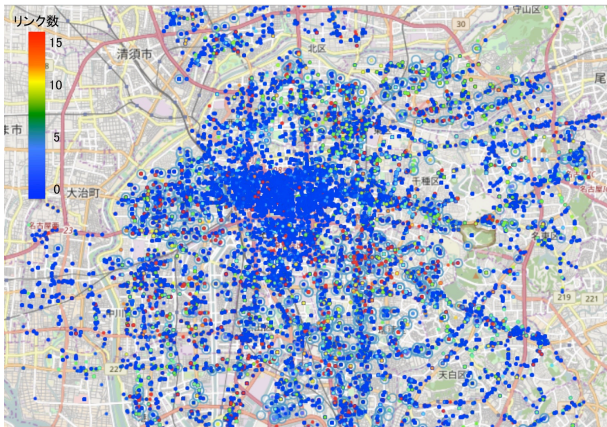


図 2: 名古屋市全体の URL マップ

設を探した結果である 119,470 件と比較すると、10,259 件という数は少なく思えるが、Web クローラの探索範囲を広げることでカバレッジの向上が見込まれる。

Web ページをマッピングすることで、ポイントの疎密から Web 上に多くの情報を記載されている地域が視覚的に捉えられるようになった。また、次節で注目する結果を見やすくするため、以後は商用検索サイトのポイントを非表示にしたマップを使用する。

### 3.2 支店と姉妹店のリンク構造

飲食店のホームページには支店同士が相互に結びついたリンク構造を持つものが多く、そのリンク構造をマップ上に可視化すると、図3が得られる。図3は、ある料理店1のA支店から辿れるリンクを結線し、そのリンク先のポイントからも順にリンクの可視化を行ったものである。また、図3の中で青く線が結ばれているポイントは、料理店1の支店ではなく、別の料理店2のホームページを示している。料理店1のリンクを広げることによって料理店2への繋がりが得られた原因を調べたところ、料理店2は料理店1と同一グループ会社の姉妹店の関係にあり、図4のようなリンクにより Web 上で結びついていることがわかった。

このような姉妹店の関係は Web 上で相互にリンクを持ち、尚且つ地理空間ではある閉じたエリア内に分散して配置された位置関係を築くことが多い。図3は、その関係性の特徴をよく表しており、これを類似度として定義し、他のリンク構造および地理空間の位置関係を持つ Web サイトと比較すれば、クラスタリング手法により、姉妹店の識別ができるのではないかと考える。

## 4 まとめ

本稿では、Web リンク情報の実世界マッピング手法を提案した。リンク情報をマップ上に可視化することによ



図 3: 料理店の支店とその姉妹店間のリンクの可視化



図 4: 同一グループ姉妹店へのリンク例

り、支店や姉妹店の繋がりと、その特徴をマップ上で視覚的に捉えられるようになった。今後は、このような地理空間と Web 上での特徴をクラスタリングに応用し、姉妹店の検出を目指す。

## 参考文献

- [1] Katsumi Takahashi, Seiji Yokoji, and Nobuyuki Miura. Location Oriented Integration of Internet Information-Mobile Info Search. In *Digital Cities*, pp. 364–377. Springer, 2000.
- [2] Kevin S. McCurley. Geospatial Mapping and Navigation of the Web. In *Proceedings of the 10th International Conference on World Wide Web*, pp. 221–229, 2001.
- [3] Yahoo! Open Local Platform (YOLP). <http://olp.yahoo.co.jp/>.
- [4] OpenStreetMap. <https://www.openstreetmap.org>.
- [5] iタウンページ. <https://itp.ne.jp/>.