

ローカルフード検索システムの提案

岡村 雅仁[†] 鈴木 彰真[†] 佐藤 永欣[†] 村田 嘉利[†]
 岩手県立大学ソフトウェア情報学部[†]

1 はじめに

旅行先や出張先には、地域特有のローカルフードがあり、それらはテレビ番組で取り扱われるなど注目を浴びている。このうち、ガイドブックにないマイナーなローカルフードを検索するためには、大量のメニューデータの効率的な収集が必要であり、通常の検索システムでマイナーなローカルフードを見つけることは難しい。そこで、これまでにマイナーなローカルフードも検索できるシステムについて研究してきた [1]。本稿では、最適なウェブスクレイピング対象の選定、データベースの構造化による検索効率の向上についてそれぞれ検討し、以前より高速かつ網羅性の高いシステムを構築した。また、旬や流行に応じたローカルフードが存在するため、その抽出可能性についても検討した。

2 関連研究

唐沢らの研究では、料理教材テキストから素材、調味料、調理方法、代替可能な素材を抽出する手法を提案している [2]。料理名が与えられた際、形態素解析を用いることで料理の分類が可能であることを述べている [3]。浜田らの研究では、料理映像とテキスト教材の対応付けをするシステムを開発している。固有の辞書を構築することで、調理手順の高精度な構造解析が可能であることを述べている。一方、本研究ではローカルフードの検索を対象に行ってきた。

3 ローカルフード検索システム

ローカルフード検索システムは、ウェブスクレイピングによってメニューを取得し、地方で扱われている料理から東京で扱われているものを取り除くことで郷土料理を抽出している。システムの概要を図1に示す。図1のように、提案するシステムは、ユーザが利用する検索用ページ、料理情報を格納したデータベース、データの受け渡しを行うサー

バ、スクレイピングを行うクローラで構成される。ユーザが検索用ページで検索したい地域を選択し検索ボタンを押すと、サーバを通してデータベースから料理の情報を受け取り、メニューが表示される。サーバは、データベースから検索した地域で扱っている料理と料理名、東京で扱っている料理を取り出す。また、表示された料理を押すことで、その料理を扱っている店舗が表示される。図2にユーザ用の検索ページを示す。料理の地域性を表すために、一般性と抽出度のパラメータが用意されており、ユーザは二つのスライダを使って料理を絞り込む。一般性は、東京で扱っている料理数を表し、値が小さいほど東京で扱われていない料理のみ残る。抽出度は、検索した地域で扱っている料理数を表し、値が大きいほどその地域では有名な料理が残る。クローラは、Webサイトをスクレイピングし、料理名、店名、URLを取得する。その後、取得したデータにフィルターによって不適切な情報を取り除き、データベースへ格納する。

4 メニュー抽出元の検証

ローカルフードを検索するため、構造化されたメニューを持つ既存のWebサイトからメニューを抽出する。このようなWebサイトとして食べログの

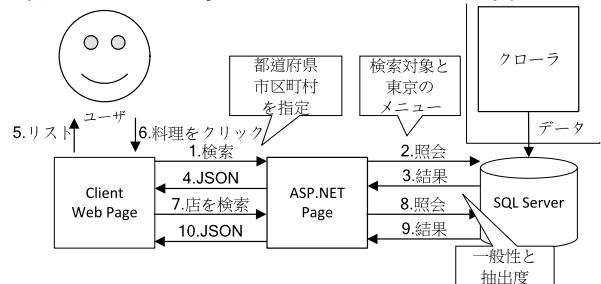


図1 システム概要図



図2 検索ページ

Proposal for searching local food [†] Okamura Masahito, Suzuki Akimasa, Sato Nobuyoshi, Murata Yoshitoshi, Iwate Prefectural University

コメントと食べログ, Retty.me, ヒトサラの各メニューがある。ローカルフードをより多く網羅するため, これらの抽出元から, テレビ番組で紹介された郷土料理の抽出率を比較した。表1に都市の規模の違う北海道, 山形, 岩手, 神奈川, 群馬, 三重, 大阪, 岡山, 鳥取, 高知, 福岡, 沖縄の12の道府県で郷土料理が抽出できた平均を示す。この結果, 食べログのコメントによる郷土料理の抽出率が最も高く, ローカルフードの抽出に向いていることが分かった。

表1 郷土料理の抽出率の平均 (%)

	食べログ メニュー	食べログ コメント	Retty .me	ヒトサラ
平均	31.75	54.49	32.37	13.77

5 データベース構造の再検討

検索の高速化のため, 先行研究における料理情報の構造をテキストベースのデータから関係データベースに見直した。データベースの構造を図3に示す。市区町村ごとに店舗を格納し, すべての料理の中からその店舗で扱っているものを示すメニューを持つ。検索した地域で扱っている料理はメニューのみで調べることができ, 料理から扱っている店舗を検索することができる。表2に都市の規模の違う4都市を検索したときの検索時間の差を示す。データベースの構造化によって, いずれも検索時間が構造化前と比べて短縮された。都市の規模が大きいほど東京で出される料理を扱うため, ローカルフードの件数が少ない傾向にある。また, 検索時間は検索対象の料理数に比例する。そのため, 都市の規模が大きいほど検索時間は短縮された。

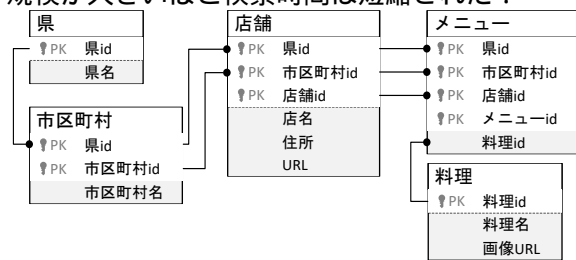


図3 データベース構造

表2 DBの構造化による検索時間の比較

	構造化後 [s]	構造化前 [s]
盛岡市	33.52	89.71
札幌市中央区	35.19	90.1
名古屋市千種区	6.87	109.21
京都市北区	3.64	89.89

6 季節限定メニューの提案

ウェブスクレイピングでは静的なウェブページから料理を抽出するため流行や季節による料理の抽出は難しい。そこで, Twitterを利用してメニューの流行の抽出を試みた。その中でも一例として, 季

節による移り変わりがあるスイーツを対象に絞った。ツイートの中からスイーツ辞典 [4][5]にある語句が使われた件数を1日ごとに数えたところ, 図4に示すようにフランボワーズとレーズンの語句を含むツイートが, 新商品の発売に合わせて増加していることが確認できた。このことから全国的なスイーツの流行について, 語句のみでの抽出できることが分かった。

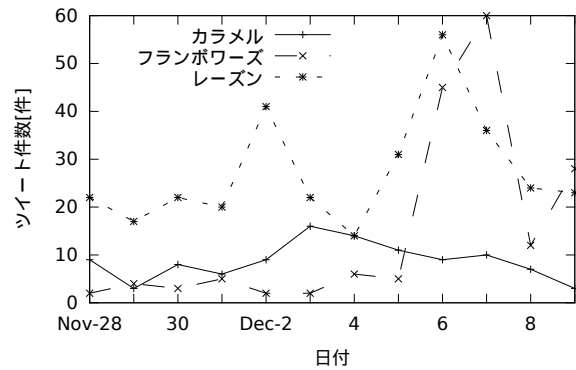


図4 1日当たりのツイート件数

7 おわりに

ローカルフードの抽出元の検討, 検索の効率化と季節限定メニュー抽出方法の提案をした。郷土料理抽出の比較によって, ローカルフードの抽出に適しているのが食べログのコメントであることが分かった。また, データベースを構造化することによって検索速度を向上させることができた。さらにTwitterを用いることで, 日本全国ではスイーツの流行についての抽出が可能であることが分かった。今後は地域での流行のメニューの抽出方法の検討を行う。

参考文献

- [1] 宇部雅彦, 村田嘉利, 鈴木彰真, 「郷土料理検索エンジンの提案」(2015) マルチメディア, 分散, 協調とモバイル (DICOMO2015) シンポジウム p.1309-1315
- [2] 唐澤 隆, 浜田 玲子, 井手 一郎, 坂井 修一, 田中 英彦, 「料理教材テキストからの素材と調理法に関する知識の抽出」
- [3] 浜田 玲子, 井手 一郎, 坂井 修一, 田中 英彦, 「料理 テキスト教材における調理手順の構造化」(2002) 電子情報通信学会論文誌, Vol.J85-D-II, No.1, p.79-89
- [4] お菓子の辞典, 入手先: <http://www.webl.io.jp/category/food/ishro>
- [5] スイーツの用語辞典, 入手先: <http://gakuen.monteur.co.jp/library/word/>