

研究論文

ローカルフード検索システムの実用化に関する研究

岡村 雅仁^{1,†1,a)} 村田 嘉利^{1,†1} 鈴木 彰真^{1,†1} 佐藤 永欣^{1,†1}

概要: ローカルフードにおける地域性の違いは興味深く、ローカルフードを満喫することは旅行者や出張者において大きな楽しみである。また、観光や地域振興の面でもローカルフードは重要である。そこで、これまで通常の Web 検索では困難な、一般に認知されていないローカルフードも検索可能なローカルフード検索手法を提案し、手法としての有用性を評価してきた。筆者らは、ローカルフード検索システムの実用化に向けて、検索速度と検索精度を向上させた検索方法とその手法を実装したシステムを提案し、これらを実験的に評価した。また、実際に構築した web システムを公開し、検索における再現性や使用感を主観評価によって評価した。

キーワード: ローカルフード, Web スクレイピング, 形態素解析

Practical study for Localfood searching system

MASAHITO OKAMURA^{1,†1,a)} YOSHITOSHI MURATA^{1,†1} AKIMASA SUZUKI^{1,†1} NOBUYOSHI SATO^{1,†1}

1. はじめに

旅行先や出張先には、その地域特有の食文化があり、旅行者や出張者がそれらを食べることは楽しみの一つである。また、対象地域以外ではなじみの薄い食文化について紹介するテレビ番組が放送されるなど、地域の食文化の違いは一般的な関心が高く、現地の人々がまだ気づかないメニューを地域特有の食文化として認知することで観光資源として活用できることが考えられる。観光庁の統計 [1] によると、2016 年の年間国内旅行者数は 64,108 万人となっており、飲食費としての消費額は 2,379,320 百万円となっている。また、日本政府観光局の統計 [2] によると、2016 年の訪日外国人は 2,403 万人となっている。このことから、旅行先や出張先における地域性のあるメニューの飲食はその地域での経済効果が見込まれ、地域の活性化につながり地方創生に良い影響を与えられると考えられる。

しかし、地域特有の料理を一般的な検索エンジンを利用する場合、地域の料理として紹介されないマイナーなメ

ニューは、地域との関連を示すサイトが存在しないため、検索が困難である。そこで、マイナーな地域特有の郷土料理を検索するために、メニューの地域性と都市部との比較によって導出できる地域特有の料理をローカルフードと定義し、ローカルフードの検索が可能なアルゴリズムを提案し、研究を進めてきた [3]。また、Web スクレイピングによってローカルフードを抽出できることを示してきた。

筆者らは、ローカルフード検索システムの実装にあたり、情報の抽出元として最適な Web ページの選定、料理を含む雑多な情報から明らかに料理でないものを減らすことに適したフィルタリングの検討、検索速度を考慮したデータベース構造の提案を行い、実用性を考慮したシステムを構築し、提案システムの有用性について評価した。

2. 関連研究

地域の料理については、B 級グルメのに対する取り組みと成功事例について空間や組織の観点から考察し、B 級グルメがもたらした経済効果等の成果と課題について述べているものがある [4]。この論文では、B 級グルメによって、大きな経済効果があったこと、観光客への知名度の向上があったこと、また、街づくりとしてその地域の住民が B 級

¹ 情報処理学会

^{†1} 現在、岩手県立大学

Presently with Iwate Prefectural University

a) g231p005@s.iwate-pu.ac.jp

グルメを意識するようになったことが述べられており、地域特有の食文化について周知することで、その地域の活性化に大きく貢献できると考えられる。

Morlandら [5] は、ローカルフードの抽出を目的としたものではないものの、食料品店や料理店を住所からジオコーディングし、土地の価値や黒人の比率との関連を評価している。また、Powellら [6] は食料品店、Blockら [7] はファストフード店をそれぞれ郵便番号と結び付け、人種や所得との関連を評価している。これらの論文では、地域の所得や人種によって食料品店の種類や店舗数に特徴が出ることを述べている。

一方、Aileenら [8] は、Trip Adviserなどの観光レビューサイトを Web スクレイピングして、店舗につけられるレビューを抽出することで、評価の低い店舗に対して評価の高い店舗のレビューを比較することによって、改善案を提案することができると述べている。また、杉本ら [9] は口コミデータを活用するための感情の調査やデータベースの開発を行っている。ここでは、口コミを形態素解析にかけることで感情情報の抽出を行えることを述べている。これらの論文は、Web スクレイピングによって地域における店舗の特徴が抽出できることを示している。

それらに対して本研究は、Web スクレイピングによって得られた料理店の住所を利用し、料理店の扱っている料理を地域ごとに分類することによって、地域性のある料理の抽出を行う、ローカルフード検索システムの実用化に向けた取り組みを行っている。

3. ローカルフード検索エンジン

本研究におけるローカルフードは、ある一定範囲の地域において多くの料理店でメニューとして提供されており、大都市においてメニューとして多く提供されていない料理と定義する。また、地方の料理に注目し、東京の料理は含まないものとする。一方、地域性の持つ料理のうち、一般的な料理と同じ名称で呼ばれているものはシステムの検索対象に含まない。また、検索を行う地域の定義としては、出張者や旅行者の利用を考慮し、市区町村の単位で抽出を行う。先行研究として、宇部らは郷土料理の抽出についての提案を行ってきた [3]。インターネット上の地域と料理が紐づけされたサイトから、Web スクレイピングによって料理の情報を抽出を行い、料理以外の情報をフィルタリングによって削除し、大都市と比較を行うことによって郷土料理の抽出を行うことができることを示してきた。本研究では、情報の抽出にあたっての適切な Web サービスの検討、Web サービスとして構築する際の効率的な検索可能なデータベース構造をについての検討を行った。

3.1 ローカルフード検索アルゴリズム

表 1, 表 2 にローカルフードと一般的な料理を扱っている

© 2017 Information Processing Society of Japan

る料理店の店舗数を示す。表 1 は、複数のローカルフードに関して、地域と東京で扱っている料理店の数を示している。ローカルフードは、全国で扱う店舗数に限らず、地方の店舗数が東京の店舗数よりも多くなっている。また、表 2 は、一般的な料理に関して地域と東京で扱っている料理店の店舗数を示している。一般的な料理の場合、地方の店舗数は東京の店舗数と比較して少なくなっている。このことより、図 1 にローカルフードと一般的な料理の分布の考え方を示す。左の集合は大都市での料理店の集合を示し、右の集合は地方での料理店の集合を示す。その中で、あるローカルフードを提供する料理店を●で表し、ある一般的な料理を提供する料理店を×で表している。ローカルフードは、アンテナショップを除きそれぞれの地域の店舗で提供されている。一方、一般的な料理は、全国的に広く提供されており、さらに大規模な都市では、提供している店舗が多く密集していると考えられる。そこで、東京都で多く提供されている料理を一般的な料理とし、それぞれの地域で提供されている料理の中から、一般的な料理を除くことでローカルフードの検索ができる考えた。そこで、本研究で用いるアルゴリズムでは、最初に与えられた検索対象の地域で提供されている料理のリストアップする。次にリストアップされた料理の中から、東京都で一定以上提供されている料理を取り除くことでローカルフードを含むリストを生成する。検索地域で提供する店舗の数が少ないメニューほどマイナーな料理が含まれる一方、料理名ではないものが抽出されやすくなる。また、比較対象である東京において対象のメニューを提供する店舗の数が少ないほど郷土料理である可能性が高いが、同時に郷土料理以外の料理も抽出されやすくなる。この基準は、検索地域の規模によって存在する店舗数と料理数が違うため、あるメニューを提供する店舗数や東京で扱っている店舗数の閾値を一定の基準することは難しい。そこで、東京の店舗数を一般性、検索対象地域の店舗数を抽出度として、二つのパラメータを用意することで、都市の規模の差によって全く検索されなかったり、必要ではない情報が表示されたりしてもパラメータ調整で対応できるようにする。

ローカルフード検索システムでは、地域の料理と東京の料理を比較するため、あらゆる料理を地域ごとに網羅していることが望まれる。料理についての情報は、料理店のレビューサイトにメニューとして多く投稿されると考えられる。そこで、地域情報と料理情報を既存の Web サービスからスクレイピングを用いることによって抽出する。

3.2 先行研究におけるローカルフード検索アルゴリズム

先行研究 [3] で提案するローカルフード検索システムのアルゴリズムを図 2 に示す。システムは、他の Web サービスから料理店と扱うメニューの抽出を行うクローラ、メニューや住所などの情報を格納するデータベース、利用者

表 1 ローカルフードを扱う
地域の店舗数と東京の店舗数 (件)

料理名 (地域)	地方	東京
タレかつ丼 (新潟)	47	28
鮭のルイベ (北海道)	12	7
沖縄ぜんざい (沖縄)	4	1
じゃじゃ麺 (盛岡)	27	25

表 2 一般的な料理を扱う
地域の店舗数と東京の店舗数 (件)

料理名 (比較地域)	地方	東京
豚しゃぶ (沖縄)	10	261
ハンバーグ (新潟)	11	1770
ラザニア (富山)	5	373
アイスクリーム (秋田)	6	443

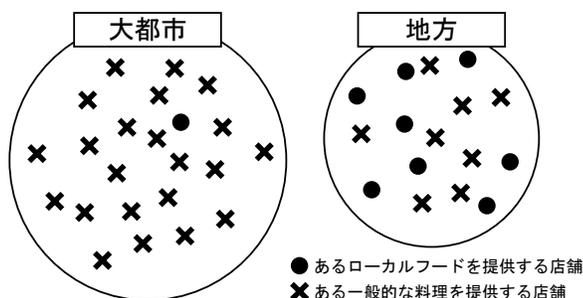


図 1 ローカルフードと一般的な料理の分布

にローカルフードの情報を提供する Web ページによって構成される。まず、図 2(a) に示すように、大量の料理店やメニューの情報を持つ既存の Web サービスから料理店の持つメニューと住所を Web スクレイピングによって抽出する。取得した Web ページには、住所や扱っている料理に関する情報がタグによって記述されている。そこで、料理店や料理の情報を HTML タグから正規表現を利用して抽出する。次に、図 2(b) で抽出したデータに対し、必要な情報を消さない程度のフィルタリングをかけ、不要な情報を削除し、地域ごとのメニューとしてデータベースへ格納する。利用者は、図 2(c) の検索ページでは地域を指定して検索することで、データベースから検索地域のローカルフードと料理を取り扱っている店舗の情報を表示する。

4. 実用システムに向けた検討

先行研究 [3] では、実用化にあたって検索のためのデータベース構造について検討が行われておらず、実用的な速度での検索が難しい。また、料理の情報について、適切な情報の抽出元の検討が行われておらず、抽出した情報に料理以外の情報が含まれている場合、検索精度に影響を及ぼす。そのため、実用的な速度で検索が可能であり、料理以外の情報を含まないローカルフードの抽出を可能にする必要がある。そこで、料理情報の適切な抽出元、料理以外の情報を取り除くフィルタリング、データベース構造の検討

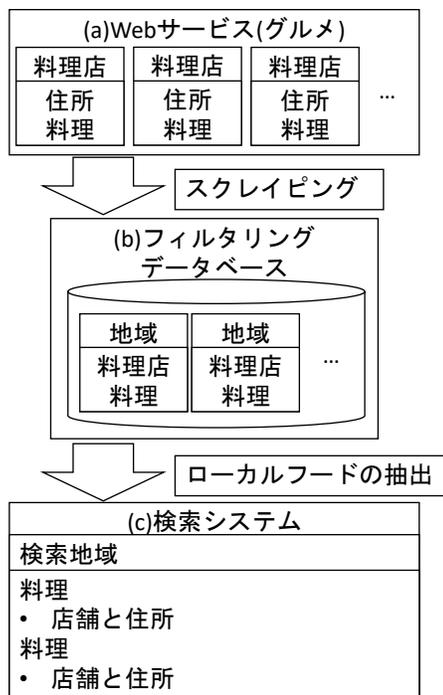


図 2 システムの流れ

を行う。

4.1 適切な料理抽出元の選定

本提案システムにおいて、情報の抽出はスクレイピングによって行う。スクレイピングによって情報を抽出するにあたって、対象となる Web ページは料理や地域の記述が定型に沿っていることが好ましい。そこで、定型的な料理ページを持つ Web サービスとして、食べログ [10]、Retty.me [11]、ヒトサラ [12] のメニュー、食べログのコメントを候補とする。食べログのメニューは、`<p class="name">(.)</p>` の正規表現によって抽出することができる。食べログのコメントは、`` の正規表現によって抽出できる。Retty.me のメニューは、`<div class="txt" itemprop="menu">\n\t*<p class="name">(.)</p>` の正規表現によって抽出できる。ヒトサラのメニューは、`<div class="color-orange mb5 bold">(.)</div>` の正規表現によって抽出できる。これらのサービス同士では、料理店やメニューについて競合している。しかし、店舗や地域の扱い方が違うため、情報を統合するのは難しい。そこで、最も料理を網羅している一つに絞って情報を抽出する。

食べログにはメニューのほかに写真をコメント付きで投稿できる機能があり、レビューの補助として投稿されることがある。必ずしもメニューとして投稿されるものではないが、メニューを含む投稿が多いため、利用できる可能性がある。そこでサンプルをとり、メニューを含むコメントがどの程度含むかを調べ、抽出率の比較のためのデータを作成した。サンプル数は下記の数式 1 に基づいて決定した。

$$n = \frac{1}{\left(\frac{e}{Z}\right)^2 \frac{1}{P(1-P)}} \quad (1)$$

P は、取得した情報の中にメニューを含む確率を表している。予測が困難な場合、50%が最も安全なサンプルサイズになる。 e はサンプルと P との差の範囲を表している。 Z はサンプルが e の範囲に収まる確率を表している。 Z は信頼度より標準得点を計算した結果になる。どの程度メニューを含むか判断するにあたり、 $P = 0.5$, $e = 0.3$, 信頼度 95%より $Z = 1.96$ として、抽出する食べログにおけるコメントのサンプル数を求めた。これは、抽出したコメントの中には、50%の確率でメニューを含むと仮定し、サンプルの抽出を行った場合、95%の確率で抽出したコメントの 50%± 3%がコメントの中にメニューを含むサンプル数を表す。式 1 より、求めたサンプル数は 1067 件となった。1067 件のそれぞれ異なる店舗と異なる投稿者のコメントを抽出し、主観的評価によってメニューを含む割合を求めた。判断の基準として、料理名が書かれているものはメニューを含むものは料理名以外の情報が含まれていてもメニューを含む情報とし、メニューを含まないものは料理名を表すものがないものを数えた。主観的に評価を行った結果、76.67%± 3%が 95%の確率でメニューを含むコメントであることが分かった。

表 3 に食べログ、Retty.me、ヒトサラのメニュー、食べログのコメントにおいてローカルフードの抽出率を示す。テレビ番組 [13] でローカルフードとして紹介された料理と比較し、テレビ番組で紹介されたローカルフードの何割が、その地域の全メニューの中に含まれているか調査した。規模の違う 12 の道府県において食べログのコメントの抽出率が最も高かった。コメントは、レビューの補助として使われる写真に料理名を付加することが多い。一方、メニューは、レビューの補助として使用できず、投稿者が少なくなる。このため、コメントの投稿数がより多く投稿され、ローカルフードを含むコメントが投稿される可能性が高くなったためと考えられる。抽出率の結果から、食べログのコメントを料理の抽出元として利用することにした。

4.2 抽出された料理以外の情報のフィルタリング

抽出元として選定した食べログのコメントは、メニューの投稿を目的としたものではない。そのため、メニューを含む投稿には、料理名のほかに様々な情報が含まれていることがある。また、メニューを含まないコメントも同時に抽出される。そのため、様々な情報を含むコメントの中からメニューではないものを取り除く必要がある。

表 4 に抽出した情報の中で料理以外の情報を含むものを示す。表 4(a) は、料理とは関係ない文章である。同様に、表 4(b) も料理の感想として美味しいと述べているが、料理自体の情報はない。表 4(c) や表 4(d) は、料理に撮影日

表 3 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
北海道	27.14	50.00	35.71	14.29
山形	18.42	31.58	18.42	05.26
岩手	40.00	53.33	40.00	20.00
神奈川	45.45	72.73	54.55	18.18
群馬	35.29	58.82	23.53	11.76
三重	36.36	63.64	54.55	18.18
大阪	43.48	60.87	52.17	26.09
岡山	14.29	42.86	14.29	07.14
鳥取	18.18	45.45	09.09	09.09
高知	25.00	75.00	12.50	00.00
福岡	35.00	45.00	30.00	05.00
沖縄	42.42	54.55	42.42	30.30
平均	31.75	54.49	32.37	13.77

表 4 抽出したメニューの例

- (a) まるでリアル感の無い商品サンプル
- (b) 新鮮 美味しい
- (c) 焼魚弁当 [サバ] (2013 / 04 / 23 撮影)
- (d) いらそば得もり 1000 円
- (e) ポテトが黒い (´▽` ;)
- (f) セットのサラダとスープ

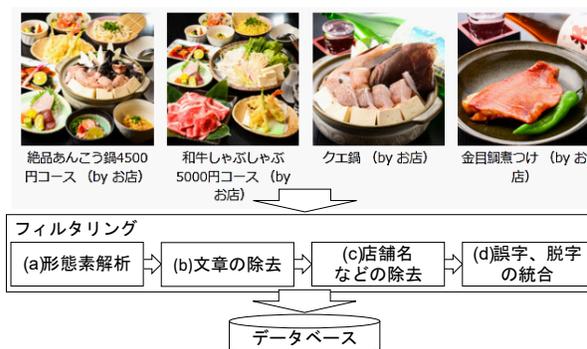


図 3 フィルタリングの流れ

時や値段等の情報が含まれている。表 4(e) は、料理の感想のほかに顔文字が含まれている。表 4(f) では、複数の料理が一つの投稿に含まれている。また、投稿内容に同じ料理を示していても誤字や表記の仕方によって異なるものが存在する。これらの情報によって同じ料理でも違う料理として扱われるため、検索精度が低下する要因になる。そのため、表 4 に示される料理以外の情報を取り除き、同じ料理の表現の違いを統一する。また、ローカルフード検索システムとしての利便性向上のため、普通の料理を落とさない程度のフィルタリングにする。形態素解析器は、MeCab を用いた。図 3 にフィルタリングの流れを示す。

図 3(a) では、抽出した料理情報を形態素解析器にかけ、品詞ごとに単語を分解する。図 3(b) で単語の品詞から、品詞の種類と数によって文章を取り除く。表 5 に文章として取り除く条件を示す。表 5 の条件を一つ以上満たす場合、

表 5 長文として扱う条件

品詞	条件
助詞	三つ以上
助詞または助動詞	三つ以上
感動詞	二つ以上
空白以外の記号	一つ以上
助動詞「です」	一つ以上
助動詞「ます」	二つ以上

文章として除外を行う。助詞を三つ以上含む条件は、「マグロとしめ鯖とサーモン」のような、料理名を複数列挙しているものを想定している。「カボチャのパイ」のような、料理を特徴付け、食材を目立たせる目的で食材と料理を助詞で結ぶ場合があるため、複数の助詞がある場合に限り取り除くようにした。助詞または助動詞を三つ以上含む条件は、「5つの部位から選べる『ステーキ』」のような、料理を説明する文章を想定している。助詞が三つ以上含む場合、同時にこの条件に該当する。感動詞を二つ以上含む条件は、「まあ満足です」のような、料理の感想を書かれたものを想定している。助動詞「です」や助動詞「ます」は、「背脂煮干は中太麺です」のような、ですます調で書かれた文章を想定している。助動詞「ます」では、文章以外にも鱒をひらがなで書いたものが存在するため、一つの「ます」では取り除かないようにした。

図 3(c) では、店舗名や値段といった料理を説明するために付加された情報を取り除く。空白によって料理名を分割し、それぞれを一つの料理名として扱う。分割された料理名を、さらに接続詞によって語句を分割する。この際、「と」と「の」の接続詞がこの順で出現している場合、これらが後の料理名に係ることが想定されるため、分割しないようにした。また、単体の接続詞に関して、接続詞「の」のみ直後の料理名に係ることが考えられるため、分割しないようにした。分割した結果から店舗名、値段、日付や形容詞単体のものの順番に取り除く。店舗名の除外は、語句がデータベースに格納されている店舗の名前と一致しているもの、語句が「店」で終わるものについて除去を行う。次に値段についての削除を行い、日付の除去を行う。値段や日付は数字の次に「円」や「年月日」が並んでいる語句についての除去を行った。また、日付を除去する際に「1201」や「12/01」のような表記を想定し、4文字数以上の数字は取り除いた。これらの除去を行う際、接続詞「の」が存在する場合、同時に除去を行う。

図 3(d) では、誤字や脱字や表記ゆれについての統合を行う。統合について先頭の文字が同じかつ、Jaro-Winkler 距離 [14] が 0.9 以上のものに対して行った。Jaro-Winkler 距離の計算では、「カキ」と「かき」など料理名をひらがなで書かれている場合とカタカナで書かれている場合が存在するため、ひらがなとカタカナを同じ文字として計算を行う。また、料理の統合を行う際、統合先がすでに他の料理

表 6 入力した料理名とフィルタリング後の料理名の例

入力した料理	フィルタリング結果
岬かき揚げ丼 600 円	岬かき揚げ丼
2014 年 11 月 島っ子丼 950 円	島っ子丼
美味しい	(削除)
真っ二つに	真っ二つに
品目を極力ひかえたサラダだよ	(削除)

表 7 統合されたメニューと Jaro-Winkler 距離の例

料理 A	料理 B	距離
かきの天麩羅	カキの天麩羅	1.0000
ガトーショコラ	ガトショコラ	0.9619
ゴーヤ天	ゴーヤ天麩羅	0.9333
サンマー麺	サンマー麺大盛り	0.9250

と統合されている場合、統合先がすでに統合している料理との統合を行う。フィルタリングによる情報の除去が行われた例を表 6 に、Jaro-Winkler 距離による料理の統合が行われた例を表 7 に示す。フィルタリングによって取り除かれたものは 4,425 件、コメントの一部除去が行われたものは 178,545 件になった。また、Jaro-Winkler 距離によって統合されたメニューは 257,130 件になった。情報の除去を行った結果、値段や日付の除去が行えていることが分かった。一方、料理以外の情報の中で、削除されていない情報も存在するため、除去する条件について改善の余地がある。また、Jaro-Winkler 距離による料理の統合では、前方の文字列に重みを置くため、「大盛」などの盛量についての記述は取り除けることが分かった。

4.3 データベース構造の検討

本システムでは、ローカルフードの抽出に検索地域で提供されているメニューを列挙し、取り扱う店舗数を数える。次に、東京都内で提供されるメニューを列挙し、取扱店舗数を数える。その後、地域のメニューと東京都内のメニューを組み合わせローカルフードの抽出を行い、料理の列挙をする。このため、高速な検索を行うには地域と店舗、店舗と料理の関連付けができる構造を持つデータベースが必要とされる。図 4 にデータベース構造を示す。ローカルフード検索システムでは、料理店は必ず市区町村ごとに検索されるため、県と市区町村から直接店舗を検索できるような構造にした。また、料理店の持つメニューも同様の検索方法で抽出する形になった。

表 8 に検索時間を示す。データベースは、Microsoft SQL Server を Azure 上の S2 Standard プランにて実装した。規模の違う 4 都市で検索をかけたところ、件数が多い地域で 40 秒以内の検索時間になった。検索時間は、その地域で提供されている料理数に依存している。また、統合する際の料理の件数は検索地域の料理数に依存するため、東京で提供されているすべての料理を毎回検索していても、検索時間に影響しないことが分かった。本研究のデータベース

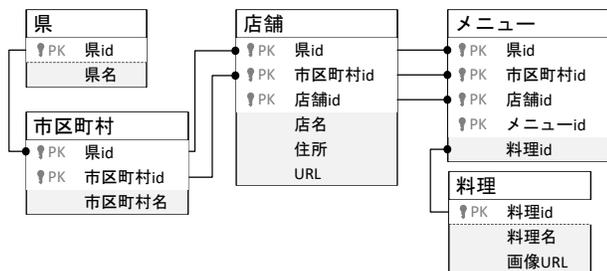


図 4 構築したデータベースの E-R 図

表 8 地域ごとの検索時間とメニューの件数

	検索時間 [s]	件数 (件)
盛岡市	33.52	1579
札幌市中央区	35.19	3031
名古屋市千種区	6.87	1118
京都市北区	3.64	217

構造では、ローカルフードの抽出をデータベース上でやっているため、件数に応じた検索時間になった。今後、あらかじめローカルフードの抽出を行った結果をデータベースに格納することによって、検索速度のさらなる高速化をめざす。

5. 提案システムの構成

図 5 にシステムの構成を示す。システムは、実線で示されているクローラ、データベース、Web システムより構成されている。クローラは、他の Web サービスから店舗の住所と扱うメニューを抽出しフィルタリングを行う。フィルタリングの結果残った料理を、店舗の地域と関連付けてデータベースへ格納する。データベースは、関連付けられた料理と店舗の情報からローカルフードの検索を行う。Web システムでは、都道府県と市区町村を指定して検索をかけると、データベースからローカルフードの情報を取得することができる。

図 6 に Web システムの機能を示す。図 6 は、Web ページごとにページ名、保持しているデータ、動作を示している。Web システムは、利用者が検索するために使用する検索ページ、ローカルフードをデータベースに問い合わせをする料理抽出ページ、ローカルフードを取り扱っている店舗をデータベースに問い合わせをする店舗抽出ページを持つ。検索ページは、検索地域、料理や店舗のリスト、先行研究 [3] で紹介した一般性や抽出度のデータを持ち、料理や店舗の検索、一般性や抽出度の調整を行う機能を持つ。料理抽出ページ、店舗抽出ページは、それぞれ検索地域のデータを持ち、それぞれ、料理、店舗の問い合わせを行う機能を持つ。また、図 6 に示す通り Web システムは、ASP.NET によってコーディングし、Microsoft Azure 上に Web App プランで実装した。

図 7 に利用者が検索するための Web ページを示す。検索ページは、検索地域を指定するためのプルダウンメニュー
© 2017 Information Processing Society of Japan

と検索開始ボタン、料理の一般性や抽出度の調整バー、検索結果表示画面を持つ。利用者は、道府県と市区町村を指定することでローカルフードの検索を行う。図 8 に Web システムのシーケンス図を示す。まず、図 8(a) に示す、検索ページで料理の検索を行ったとき、料理抽出ページに検索地域の都道府県と市区町村を送る。料理抽出ページは、図 8(b) のデータベースに検索地域と東京の料理を扱う店舗数を統合したデータを問い合わせる。その後、図 8(c) に示すようにデータベースから料理抽出ページ問い合わせた結果をリスト形式で受け取り、図 8(d) で示すように、検索ページに JSON 形式でローカルフードのリストデータを返す。

ローカルフードのリストには料理名、一般性、抽出度がデータとして含まれる。次に、表示されたローカルフードをクリックすることで、図 8(e) の検索地域でその料理を扱う料理店のリストの検索を行う。店舗抽出ページは、検索地域の情報と料理を受け取り、図 8(f) でデータベースに検索地域で料理を扱っている料理店を問い合わせる。料理抽出ページは、図 8(g) で料理店のリストを受け取り、図 8(h) で検索ページに JSON 形式で料理店のリストを返す。料理店のリストには店舗名、住所、紹介ページへのリンクがデータとして含まれる。ローカルフード検索システムでは、旅行中や出張中の人を対象とするため、店舗を案内する際に紹介ページと住所も同時に表示する。利用者がローカルフードを扱っている店舗の情報が必要な場合、紹介ページへアクセスすることで、扱っている店舗についての情報を得ることができる。

図 7 に示す検索ページでは、一般性、抽出度の値は、スライダーによって調整することができる。一般性は、その料理を扱っている都市部の店舗数を表す。一般性が高いほど都市部でも多く出されるような一般的な料理であるといえるため、値を下げることで東京で提供されている料理が絞られる。抽出度は、その料理を扱っている検索地域での店舗数を表す。抽出度が高いほど、その地域で多く出されるような一般的な料理であるといえるため、抽出度を上げることで検索地域であまり提供されていない料理が絞られる。初期値では、最も多く料理の表示できるように一般性を最大値に、抽出度を最小値に設定した。検索結果が多い場合は、これらのバーを調整することで、必要な情報を絞ることができる。バーを移動させると動的に検索結果が変化する。

6. システムの有用性の評価

6.1 再現率による有用性の評価

定量的なローカルフードの抽出率を判断するため、テレビ番組 [13] で紹介されたローカルフードとの再現率を求めた。表 9 に再現率を示す。再現率を出すための正解集合は、表 3 で用いたもの同様のものを用いている。本研究の

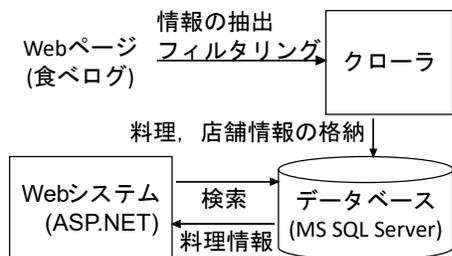


図 5 システムの構成図

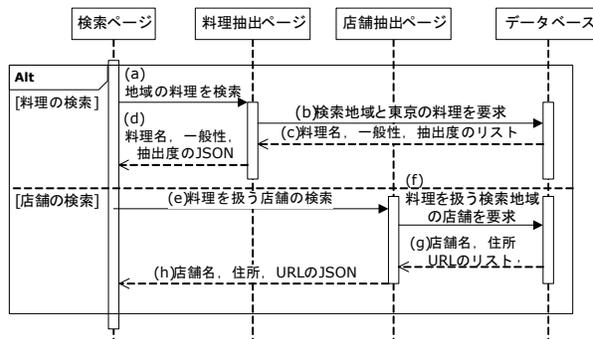


図 8 Web システムのシーケンス図

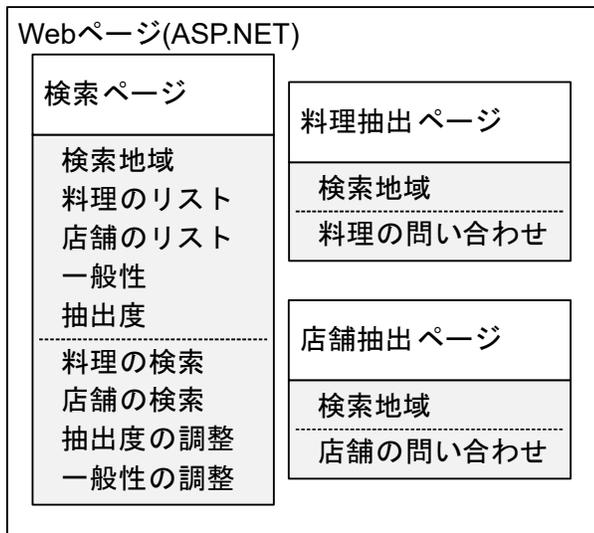


図 6 Web システムの機能

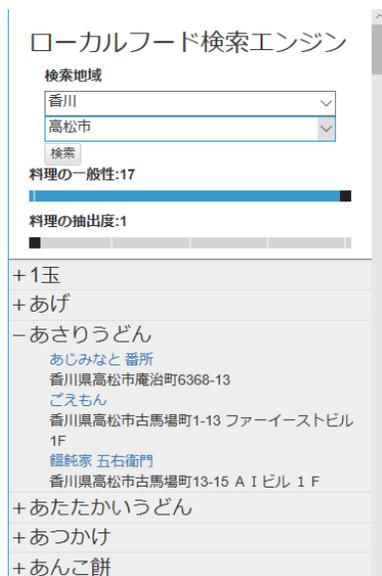


図 7 検索ページ

表 9 先行研究との再現率の比較 (%)

	岩手	新潟	静岡	広島
本研究	58.33	50.00	63.64	80.00
先行研究	44.44	33.33	46.15	80.00

スクレイピングが後に行われたため、抽出した情報が多くなり再現度が向上したと考えられる。

6.2 アンケートによる有用性の評価

Web システムの主観的な有用性を、すでに認識しているローカルフードが抽出されているか、また、マイナーなローカルフードが抽出されているかとして評価する。大学生 52 人にシステムを使って居住地のローカルフードを検索してもらいアンケートを取った。表 10、表 11 にアンケートの結果を示す。表 10 は、「自分の知らないローカルフードが出てきたか」と質問し、「全く出てこなかった」「出てきたものがあった」「よく出てきた」の選択肢の中から得られた回答を示している。表 11 は、「知っているローカルフードで出なかったものはあるか」と質問し、「0 個」「1-4 個」「5 個以上」の選択肢の中から得られた回答を示している。また、同時に自由記述の形式でアンケートを取った。

アンケートによって、既知のローカルフードの抽出に関して抽出されない例があることが分かった。料理店で扱われる料理以外のローカルフードについては、家庭料理や地域の食材そのものが挙げられる。旅行者や出張者を対象とする本システムでは、料理店で扱われない料理は検索の範囲外としている。また、料理店で出されるローカルフードでも、ある地域で一定数以上の店舗が扱っているメニューでないと抽出されない。本システムでは、4.1 節において最も検索件数が多いスクレイピング元を検討しているが、既存の Web サービスに寄せられる投稿をもとに料理を抽出しているため、その地域の店舗数が少ない、また、スクレイピング元に投稿する人が少ない場合、比較するための料理の情報がないため、ローカルフードの抽出を行うことができない。そのため、今後料理店の少ない地域では、検索範囲を市区町村ではない、より広い範囲で検索可能なシステムに拡張する。一方で、「げんげの天ぷら」や「ホヤ刺し」など、知名度の低いローカルフードに関して出てきたという回答が多く得られた。また、アンケートの自由記述の欄に書かれたシステムの感想では、居住していたがローカルフードと認識していなかった料理が出てきたという回答も得られた。

7. まとめ

ローカルフード検索システムの実用化に向けて、料理情

表 10 知名度の低いローカルフードの抽出に関する主観的調査

	人数 [人]
全く出てこなかった	4
出てきたものがあった	25
よく出てきた	4

表 11 既知のローカルフードの抽出に関する主観的調査

	人数 [人]
0 個	7
1-4 個	16
5 個以上	4

報の抽出元についての検討, 実用的な速度での検索性能を考慮したデータベース構造の検討, 料理以外の情報を取り除くためのフィルタリングについて検討を行い, システムとして実装してきた。また, その有用性をテレビ番組で紹介された料理の抽出率とアンケートによる主観的評価によって評価を行った。12 道府県でテレビ番組で紹介された料理の抽出率を比較した結果, 四つの Web サービスから食べログのコメントが最も情報の抽出に向いていることが分かった。また, 提案したデータベース構造により, 地域を指定することで地域のローカルフードの抽出を行えることが分かった。また, 検索時間は検索対象地域における飲食店やメニューの数に依存し, 40 秒以内に検索できることがわかった。さらに, 新たなスクレイピング手法の検討により, 秘密のケンミンSHOWで紹介された料理を基準に, 先行研究より高い再現率を發揮した。また, 提案システムの実用性についてアンケートをとった結果, その地域の住人が知らないローカルフードの提案が行えることが分かった。

今後の展望として, 県単位や地域単位といった検索地域の選択が可能システムへの拡張, SNS を利用した旬や流行のあるローカルフードの抽出についても検討を行う。また, 検索速度のさらなる高速化をめざし, あらかじめローカルフードの抽出を行った結果をデータベースに格納することによって, さらに, 検索の結果, ローカルフードではない一般的な料理が多く出てきたという評価が得られたので, 東京の料理以外の基準を設けて取り除くことも検討する。

参考文献

- [1] 観光庁. 旅行・観光消費動向調査. <http://www.mlit.go.jp/kankocho/siryou/toukei/shouhidoukou.html>, jun 2017.
- [2] 日本政府観光局. 統計データ (訪日外国人・出国日本人). http://www.jnto.go.jp/jpn/statistics/visitor_trends, 2017.
- [3] 宇部 雅彦, 村田 嘉利, and 鈴木 彰真. 郷土料理検索エンジンの提案. マルチメディア, 分散, 協調とモバイル (DICOMO2015) シンポジウム, pages 1309–1315, 2015.
- [4] 浩祥 佐野. 富士宮市における b 級グルメによる中心市街地

の活性化にむけた課題. 立教大学観光学部紀要, 13:59–69, mar 2011.

- [5] Kimberly Morland, Steve Wing, Ana Diez Roux, and Charles Poole. Neighborhood characteristics associated with the location of food stores and food service places. *American journal of preventive medicine*, 22(1):23–29, 2002.
- [6] Lisa M Powell, Sandy Slater, Donka Mirtcheva, Yanjun Bao, and Frank J Chaloupka. Food store availability and neighborhood characteristics in the united states. *Preventive medicine*, 44(3):189–195, 2007.
- [7] Jason P Block, Richard A Scribner, and Karen B DeSalvo. Fast food, race/ethnicity, and income: a geographic analysis. *American journal of preventive medicine*, 27(3):211–217, 2004.
- [8] L.C. Aileen. From ratingrestaurantvenues: The impact of social media on the restaurant industry. Master's thesis, New York University, 2013.
- [9] 杉本祐介, 土井千章, 中川智尋, 太田賢, 稲村浩, 水野忠則, 菱田隆彰, et al. 口コミデータを活用するデータベースシステムの実現. 研究報告ユビキタスコンピューティングシステム (UBI), 2014(44):1–6, 2014.
- [10] 株式会社カカクコム. 食べログ. <https://tabelog.com>, 2016.
- [11] Retty 株式会社. Retty グルメ. <https://retty.me>, 2016.
- [12] 株式会社 USEN. ヒトサラ. <https://hitosara.com>, 2016.
- [13] 株式会社カカクコム. 「秘密のケンミン show」で紹介された情報. <http://kakaku.com/tv/channel=4/programID=10547/>, 2016.
- [14] William Cohen, Pradeep Ravikumar, and Stephen Fienberg. A comparison of string metrics for matching names and records. In *Kdd workshop on data cleaning and object consolidation*, volume 3, pages 73–78, 2003.