DEIM Forum 2009 i1-32

地域依存検索のための地域特徴語に基づくクエリ生成支援システム

藤坂 達也† 北山 大輔‡ 李 龍† 角谷 和俊†

†兵庫県立大学環境人間学部 〒670-0092 兵庫県姫路市新在家本町 1-1-12
‡兵庫県立大学大学院 環境人間学研究科 〒670-0092 兵庫県姫路市新在家本町 1-1-12
E-mail: †{nc06h211, ne07p001}@stshse.u-hyogo.ac.jp, ‡{leeryong, sumiya}@shse.u-hyogo.ac.jp

あらまし 近年，スマートフォンなどのモバイル情報端末の普及によってアウトドア環境での情報検索の要求が増えてきている．アウトドア環境の情報検索では，場所に依存したクエリが相対的に多くその検索の際に位置情報などのコンテキストを使って情報検索の質問生成支援や検索結果の精度を高めることが可能である．しかし，位置情報のようなダイレクトで低レベルのセマンティクスのみでは，一般ユーザに対して情報検索の時にまだ多くの情報フィルタリングや再検索，クエリの追加入力の手間がかかるといった問題がある．そこで，本研究では一般ユーザの地域依存検索におけるクエリ生成のコストを最小限に減らすことを目標とし，そのために地域ごとの特徴語を利用することに着目する．具体的には，地域ごとの特徴を表す地域特徴語を定義し，モバイルユーザの情報検索におけるクエリ入力の際に，ユーザのコンテキストに応じて地域特徴語をユーザに推薦する．それにより，モバイル端末上でのクエリ生成を容易にさせるシステムを開発する．本論文では，地域特徴語を抽出手法とそれに基づくクエリ推薦方式について述べる．

キーワード 地域依存情報検索，地域特徴語，クエリ生成支援

A Query Generation Support System with Local Characteristic Words for Location-aware Web Searches

Tatsuya FUJISAKA†, Daisuke KITAYAMA‡, Ryong LEE†, and Kazutoshi SUMIYA†

† School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

‡ Graduate School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

E-mail: † †{nc06h211,ne07p001}@stshse.u-hyogo.ac.jp, ‡{leeryong,sumiya}@shse.u-hyogo.ac.jp

Abstract Recently, demands for easy web searches in outdoor environments are rapidly increasing due to the spread of mobile devices such as web-accessible smart phones. In such environments, queries users need to make are thought to be highly dependent on places in which they are located. Therefore, if search devices can use local contexts such as geographic location, it would be a great help to reduce users' efforts especially in their generating queries and to improve search results. However, the low-level context of location is still too rough to help users directly generate web search queries generally requiring high-level semantics. Even though users become to know the exact position on the map, they still have to need additional time and cost to filter out a lot of unrelated search results and to re-generate appropriate queries to repeat next searches until finding satisfying results. In this research, we aim to reduce the user's efforts and cost for generating textual queries on mobile devices for the outdoor web searches. For that, we propose to use local characteristic words about specific local areas as a knowledge base for the query generation support. With a local characteristic word dictionary constructed from local web pages beforehand, the user's activity to input query words on mobile devices will be supported by recommending strongly relevant local characteristic words or automatically completing the full-strings inputs with the most relevant candidate. We are also developing a prototype system to utilize the dictionary to facilitate users' web searches outdoors as well as indoors.

Keyword Location-Aware Search, Local Characteristic Word, Query Generation Support

1. はじめに

近年，インターネットの普及によってあらゆる分野に対する膨大な情報がウェブに蓄積され，ユーザは知りたい情報に対して容易に検索が行えるようになってきている．また，それらの情報はデスクトップ環境に留まらず，iPhone や G-Phone などのモバイル端末でも必要な情報をウェブから検索できるようになっている．しかし，モバイル端末上で行なわれる情報検索は，PC 上で行なわれるものと比べ機能性の面や検索傾向において異なりがあると考えられる．

まずモバイル端末の場合，ブラウザの大きさや数に制限があり閲覧がしにくいことや入力しやすいキーボードがないため入力操作などに時間がかかることなど，操作性の面で明らかに PC 上より劣る．こういった要因があるため，ユーザはモバイル端末でのウェブ検索が面倒になり，あまりモバイル端末上での情報検索が行われないと考えられる．

また，モバイル端末でのウェブ検索は場所や時間といったユーザのコンテキストに依存したクエリが相対的に多くなることが予想され，それに応じた検索支援が必要となると考えられる．例えば，ユーザが観光地へ旅行に行き，そこで初めて見るものや聞く言葉について知識を深めることで，より旅行を有意義にすることが考えられる．そして，現地で情報を得る手段としては，モバイル端末を用いた検索が有効であり，ユーザはそれを活用すると想定される．

しかしながら，ウェブ上で行われている過去の検索履歴や個人の入力履歴などによってクエリを推薦する従来の検索支援方式では，ある場所，ある時間に依存するような場合の検索支援には不向きであるといえ，十分でないと考えられる．

そこで，ユーザがモバイル端末上で検索する際，位置情報などのコンテキストを用いた質問生成の支援方式が必要となる．しかし，単なる位置情報のような低レベルの情報のみではユーザに十分な検索支援ができないと考えられる．例えば，GPS で位置情報を取得し，地図を用いてその周辺地域の情報を提示するだけでは，地域依存レベルの低い情報しか提示できない．

そこで，地域に依存するような単語による情報のフィルタリングやキーワード同士の関連を考慮したキーワード推薦を行い，ユーザの検索時になるべく手間がかからないような検索支援をすることが重要であると考え，我々はそれに基づいた一般ユーザのモバイル端末上におけるユーザコンテキストに依存したクエリ生成支援システムを提案している．具体的には，地域ごとに依存したその地域の特徴を表すような『地域特徴語』を提案し，それに基づきモバイル端末上のユーザの情報検索おけるクエリ入力の際に，ユーザのコンテキストに関連ある地域特徴語を推薦し，モバイル端末上でのクエリ生成を容易にさせるシステムを開発する．

本稿では以下の構成で述べる．2 節で地域特徴語の概念について述べる．3 節で本研究の概要と関連研究について，4 節で地域特徴語の抽出について述べる．5 節で地域依存検索支援システムを説明し，6 節でまとめを述べる．

2. 地域特徴語の概念

我々が生活している地域には，その空間に限定したものが多数存在する．例えば，名産や駅弁といった食べもの，方言などの言葉，祭りなどの伝統行事，生活習慣や考え方などの文化的なことなどその範囲は多岐に渡る．しかし，こういった「地域に依存する」という考え方の基準は，①ある単語からそれに依存する地域を特定する場合（単語→地域），②ある地域を定めそれにより依存する単語を特定する場合（地域→単語），の 2 つの見方のどちらにおいても人によって異なる．

例えば，単語を「ラーメン」，地域を「北海道」として①の場合を考えると，“ラーメンの有名な地域はいくつかある中で北海道が特に有名であるから地域に依存するものである”と考える人や“ラーメンの有名な地域は北海道以外にもたくさん存在するため北海道に依存するものとはいえない”と考える人もいる．②の場合を考えると，“北海道の中でラーメンが特に有名だから地域に依存するものである”と考える人や“ただ単に北海道にしか存在しないものが地域に依存したものである”と考える人もいる．

このように①，②両者のどちらの場合においても，地域の特徴を表す単語になるという明確な定義がないと考えられる．そこで我々は「単語を組み合せる」ということに着目し，『1 語では違うが単語を組み合わせることで特定の地域を表すような対象（＝単語の組み合わせ）に豹変するもの』を地域特徴語として定義している（図 1）．

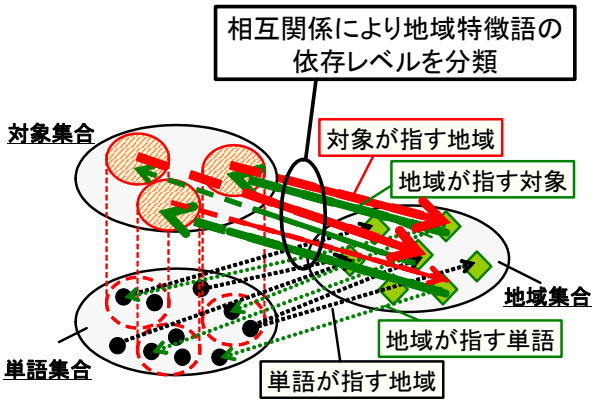

図 1　地域特徴語の概念

例えば，3 つの単語“ラーメン”，“みそ”，“ブラック”があった場合，“ラーメン”や“みそ”自体はある多くの地域群を指すような単語で，“ブラック”は全然地域を指さないような単語であり，どれも地域特徴語にはなりえない．しかし“ラーメン∧みそ”と単語を組み合せた場合，これはラーメンでも“みそラーメン”で有名な地域，つまり北海道を指すような地域特徴語になると考えられる．また，“ラーメン∧ブラック”と単語を組み合せた場合，これはラーメンでも富山県で有名なスープが真っ黒の“ブラックラーメン”を指すような地域特徴語になると考えられる．

また地域特徴語を「対象→地域」と「地域→対象」による 2 つの観点から抽出することができると考えられ，この 2 つの観点における相互の地域依存の関係により地域特徴語の依存レベルの分類も行なうことできると考えられる（表 1）．

表 1　2 観点による地域特徴語の地域依存レベル

		地域側からの観点（地域→対象）	
		高←地域依存度→低	
対象側からの観点（対象→地域）	高↑地域依存度↓低	双方とも高い（強い結束）	対象側からのみ高い
		地域側からのみ高い	双方とも低い

例えば両観点においてどちらも地域依存度が高い値の場合，お互いがお互いを支援しあうような強い結束を持った関係が成立すると考えることができ，この時は地域依存レベルが特に高い“地域の名物”などがそれに当たると考えられる．

このようにして，2 つの観点を用いて評価することにより，地域特徴語の地域依存レベルを指定して抽出を行なうことが可能であると考えられるが，今回はまず「対象→地域」と「地域→対象」による 2 つの観点から地域特徴語をどのようにしてウェブから抽出するかを考え，さらに抽出した地域特徴語を用いてモバイル端末上でのクエリ生成を容易にさせるシステムを考えている．

3. 本研究の概要と関連研究

本節では，地域特徴語に基づくクエリ生成支援システムの概要ついて述べる．我々が提案するシステム(図 2)では,「2 つの観点による地域特徴語の抽出」を行い，それに基づいた「地域依存検索におけるクエリ生成支援」を行う．以下，それぞれについて説明する．また，ここでは前提条件としてアウトドア環境で，ある地域の空間にユーザがいるとき，その地域空間で何か地域に依存するようなものを検索することを仮定としている．

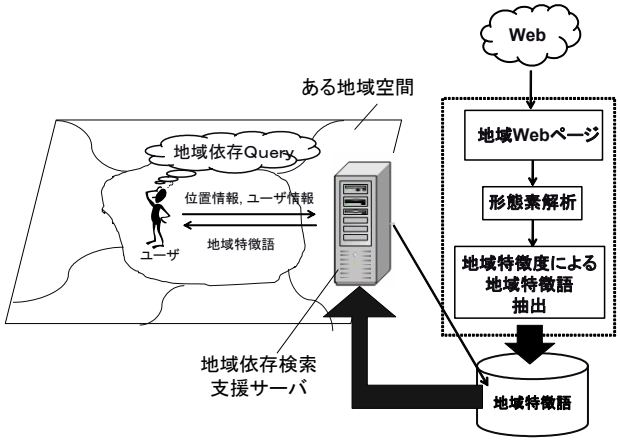

図 2 システム概要

3.1.「地域→対象」による地域特徴語の抽出

まず地域名をクエリとして得られたウェブページの解析を行い，名詞のみを抽出する．そしてクエリと各名詞の出現位置の関係，さらに地域同士における名詞の出現頻度の関係の 2 つを用いた地域の依存度を表すような「地域特徴度」から名詞のフィルタリングを行い，名詞のリストを選定する．

次に，選定した名詞のリストにおいて単語の組み合せを作り，単語の組み合せとそれ構成する単語 1 つ 1 つの地域内と地域間での出現頻度の比較より地域特徴語を判定し抽出を行なう．

3.2.「対象→地域」による地域特徴語の抽出

まず入力キーワードからそれに関連するような単語を yahoo!関連語検索を用いて抽出する．次に“入力キーワード”と“入力キーワード∧関連語”をクエリとしてそれぞれウェブページを取得し,解析を行い固有名詞のみを抽出する．そして,“入力キーワード”における地域名の出現数と“入力キーワード∧関連語”における地域名の出現数を比較することで地域特徴語かを判定し抽出を行なう．

3.3. クエリ生成支援

本章では，地域特徴語に基づくクエリ生成支援システムの概略について述べる（図 3）．ここでは，地域に依存したものを検索のクエリとして用いることを前提とする．まず，ユーザの位置情報やコンテキストを取得し，それに合致する地域特徴語をユーザが入力する前に既にそれを地域特徴度順に提示しておく．

次に，ユーザの入力した文字に対して，最初に提示した地域特徴語にフィルターをかける．

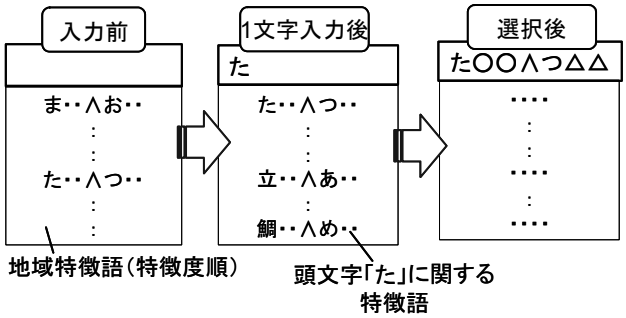

図 3 クエリ生成支援手法

例えば，図のようにユーザが“た”と 1 文字目を入力したら，最初提示した地域特徴語の中でも頭文字が“た”に当たるもののみをユーザに提示する．これより,最初よりも絞った地域特徴語をユーザに提示でき，この操作を繰り返し行なうことでユーザのクエリ生成を常に支援し，最終的にユーザがあまり文字を打ち込まなくてもユーザが検索したい地域に依存した検索クエリを生成することができる．

3.4. 関連研究

地域の特徴に着目した研究として，馬ら[1]のローカル度算出や近藤ら[2]の地域的支持度の算出手法が挙げられる．馬らは，ウェブページのコンテンツに出現する地理オブジェクトやその関連語句とその領域の出現密度からそのウェブページがどの程度地域に特化したトピックについて述べているかを算出する手法を提案している．また近藤らは，あるウェブページにリンクしているウェブページのコンテンツ発信地点をもとに，ウェブページの地域的な支持度を算出する手法を提案している．本手法では，ウェブページそのもののローカル度や地域的支持度でなくそのコンテンツ内の解析により，その地域に依存して出現するような特徴語を算出するという単語の評価を行う点で異なる．

また，地域とユーザコンテキストに着目した研究として，郡ら[3]の BlogCarRadio システムや手塚ら[4]の車窓システムが挙げられる．郡らは自動車内のユーザに対し,位置情報に基づいた地域性の高い Blog コンテンツを抽出し，それを可聴化して提示する手法を提案している．手塚らは，ユーザが地図上に描画した経路に沿ってウェブ空間を仮想的に移動しながら地域のウェブページを閲覧できるシステムを開発している．本手法では，ある地域領域にいるユーザがモバイル端末で地域依存検索を行う際の支援を行う点で異なる．

情報検索時の質問生成支援の研究として，若木ら[5]のトピック指向単語抽出による検索語提示手法や加藤ら[6]の語の共起を用いた研究が挙げられる．若木らは，ユーザが検索で得たいトピックに特化した検索結果が得られるように，それに特化した単語を抽出し提示することで，質問にあった曖昧性を解消するための手法を提案している．加藤らは，ウェブから 2 つの語 A,B の関係語を発見し，それとある語 C を用いて検索を行うことにより，A と B の関係と類似する関係を持つような語 C に対する語 D を発見する手法を提案している．本手法では，地域に依存するような検索の際にユーザコンテキストによって，今のユーザに合った地域のクエリを，地域特徴語を用いて提示することで情報検索の支援を行うことで異なる．

4. 2 つの観点による地域特徴語の抽出

本章では,「地域→対象」と「対象→地域」の 2 つの観点においてウェブからどのように地域特徴語の抽出を行なうかを具体的に述べる．

4.1.「地域→対象」の観点における地域特徴語の抽出

我々は「地域→対象」の観点より地域特徴語を抽出するために「地域特徴度を用いた名詞のフィルタリングによる名詞の選定」と「選定した名詞の単語の組み合せによる地域特徴語判定」を行っている．

(1) 地域特徴度を用いた名詞のフィルタリングによる特徴語候補の選定

まず地域レベルを定め地域名をクエリとしてウェブページを収集する．次に収集したウェブページを形態素解析し，名詞のみを抽出する．そして抽出した名詞のリストを用いてすべての名詞が取得したウェブページの文書内でどの位置に出現しているかを算出する．

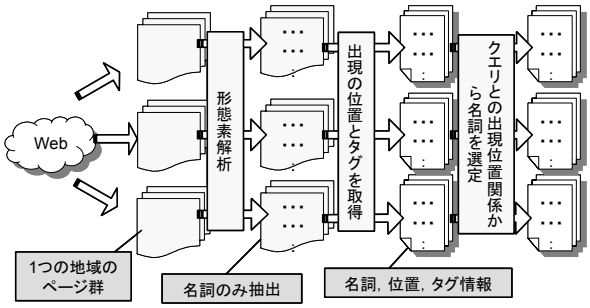

図 4 クエリとの位置関係による地域特徴度算出に基づく名詞のフィルタリング

それより地域単位ごとにクエリと名詞の出現距離を測り，ある閾値を定めそれ以降に出現する名詞は地域には関係ない名詞と考え除外し名詞の選定を行う．ここではクエリや同じ名詞が何回も出現する場合があることが大いに考えられるため，下記の式でクエリと名詞の距離における地域特徴度 DLC（Degree of Local Characteristic）を算出する．

$$DLC_{dis(q,n)} = \frac{\sum_{i=1}^{freq(n)} \min(loc(q_i), loc(n_i))}{freq(n)} \qquad (1)$$

ここでは，出現しているすべての名詞 n から最小位置に出現しているクエリ q を探し距離を算出し，それを n の出現頻度 freq(n)で除算することで q と n の出現距離における $DLC_{dis(q,n)}$を決定している．

次に，図 5 のようにクエリ単位で収集した地域のウェブページを 1 つのドキュメント文書に置き換え，選定した名詞リストのドキュメント内とドキュメント間の出現頻度における地域特徴度を(2)式より算出する．

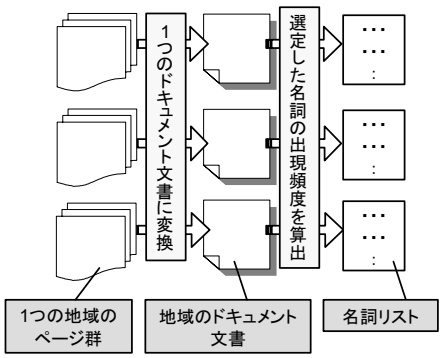

図 5 名詞の出現頻度による地域特徴度算出に基づく名詞のフィルタリング

$$DLC_{tf-idf} = tf \times \frac{1}{df} \qquad (2)$$

ここでは地域内での名詞の出現数を tf，地域間での名詞の出現数を df として tf-idf 値を算出し，ある閾値以上の名詞のみを選定している．

(2) 選定した名詞の単語の組み合せによる地域特徴語判定

ここまでである程度ノイズとなる名詞を除去できたと考えられるが，残った名詞の集合においてどの組み合せが地域特徴語になるか不明であるとともに，ここまで選定した名詞の組み合わせで地域特徴語になるかは分からない．そこで選定した名詞の組み合せにおける地域特徴語判定を行う(図 6)．

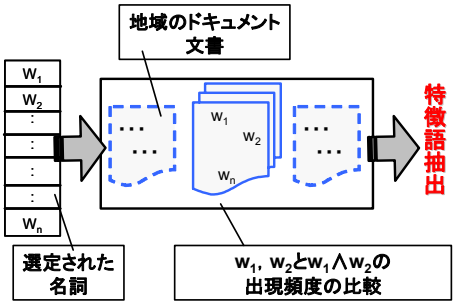

図 6 地域特徴語判定

ここでは，単語の組み合せとそれ構成する単語 1 つ 1 つの地域内のウェブページと地域間で出現頻度の比較より地域特徴語かを判定している．例えば，単語 w_1, w_2 の場合で考えると，まず w_1 や w_2 単体での地域内ウェブページと地域間の出現頻度を算出する．次に w_1 と w_2 が共起する地域内ウェブページと地域間の出現頻度を算出する．それより，w_1 と w_2 が地域内ウェブページで共起する数が多く，他の地域では共起する数が少なくなるような単語の組み合わせの場合地域特徴語として抽出を行なっている．

4.2.「対象→地域」の観点における地域特徴語の抽出

我々は「対象→地域」の観点より地域特徴語を抽出するために「入力キーワードにおける関連語取得」と「入力キーワードと入力キーワード∧関連語をクエリとしたウェブ解析の比較による地域特徴語判定」を行っている（図 7）．

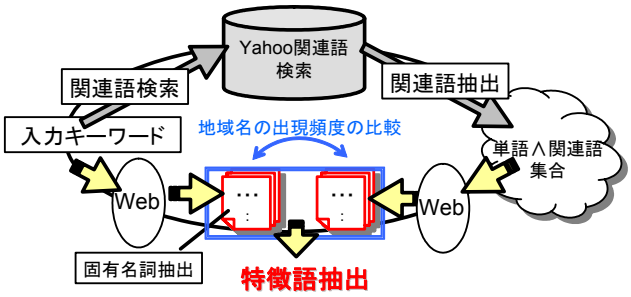

図 7「対象→地域」の観点における特徴語の抽出

まず入力キーワードを用いてそれに関連する語を yahoo 関連語検索により取得する．次に，入力キーワードと入力キーワード∧関連語をクエリとしてウェブページそれぞれ 10 件取得し，形態素解析を用いて固有名詞（地域名）のみを抽出する．そして，入力キーワードの固有名詞の出現数と入力キーワード∧関連語の固有名詞の出現数の比較を行い，出現数の上昇により地域特徴語判定を行なう．例えば，入力キーワードではある地域名の出現数が 5 だったのに対し，入力キーワード∧関連語の場合，出現数が 30 に上昇していたら入力キーワード単体よりも明らかに地域に依存するものと考えられる．このような出現数の比較により地域特徴語の抽出を行なう．

5. 地域依存検索支援システム

我々は，提案手法を評価するために下記に示したシステムを構築した（図 8）．操作としては，入力キーワードと位置を入力し，地域特徴語ボタンをクリックすると入力キーワードに関連する単語の組み合わせからなる地域特徴語とその地域を提示する．この際，位置による順位付けを行なうことで，指定した地域に近い

順に地域特徴語を並べて表示している．

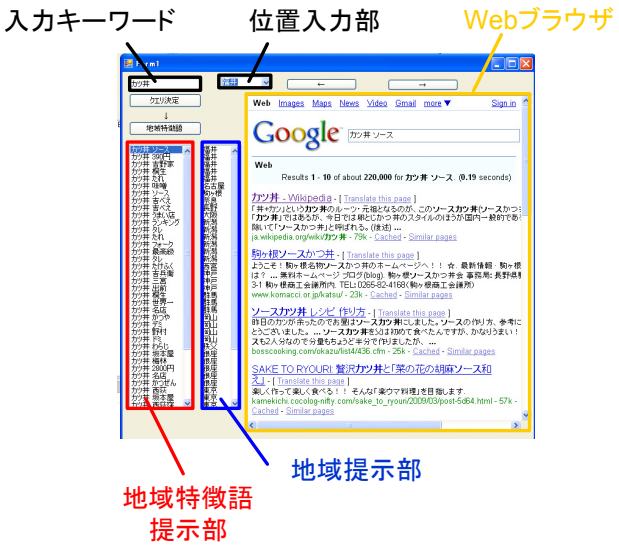

図 8 地域依存検索支援システム

これにより，ユーザはあまり文字を打ち込まなくても意図した地域に依存したクエリを生成しモバイル端末上でも検索が容易になると考えられる．

6. まとめ

我々はユーザのモバイル端末上におけるユーザコンテキストに依存したクエリ生成支援を行なう地域依存検索支援システムを提案した．提案手法では，モバイル端末上で地域依存検索を行うユーザに対して，地域特徴語という地域に依存して出現するような単語を抽出し，それをユーザクエリ入力における質問生成の支援を行う．今後の計画は，システムの評価実験と地域特徴語辞書の構築を行う予定である．

謝　辞

本研究の一部は，平成 20 年度科研費基盤研究(B)(2)「ユーザの潜在的意図を用いたレス・コンシャス情報検索基盤の構築」(課題番号：20300039) によるものです．ここに記して謝意を表すものとします．

文　献

[1] 馬強,松本知弥子,田中克己: ページ内容と位置情報に基づく Web コンテンツのローカル度検出とその応用,情報処理学会研究報告,No.67,pp. 515-522(2002)

[2] 近藤浩之,手塚太郎,田中克己: 支持の地域性と局所性を用いた Web ページのりランキング,No.65,pp. 435-440(2007)

[3] 郡宏志，手塚太郎，田中克己:地域 Blog 情報の可聴化インタフェースの提案と音声化に適したテキストコンテンツの抽出手法, DEWS2006

[4] 手塚 太郎,田中 克己:ウェブからのランドマーク抽出に基づくクエリフリーな地域情報閲覧, 日本データベース学会 Letters, Vol.4, No.1, pp.141-144(2005)

[5] 若木裕美，正田備也，高須淳宏，安達淳：検索語の曖昧性解消のためのトピック指向単語抽出および単語クラスタリング，情報処理学会論文誌データベース，Vol.47 ，No.SIG19，pp. 72-85(2006)

[6] 加藤 誠，大島 裕明，小山 聡，田中 克己: 語の共起を用いた Web の類似関係検索，Web DB Forum 2008