

Blog からの街の話題抽出手法の提案

倉島 健[†] 手塚 太郎[†] 田中 克己[†]

[†] 京都大学大学院情報学研究科社会情報学専攻 〒606-8501 京都市左京区吉田本町

E-mail: †{ktakeshi,tezuka}@dl.kuis.kyoto-u.ac.jp, ††ktanaka@i.kyoto-u.ac.jp

あらまし Blog は地域を実際に体験した個人の情報発信の場となっている。Blog 上の個人の実体験に基づく情報は、街における対象の客観的な属性とは別の形で有意義である。本稿では、Blog から多くの人々が旅の目的としている「対象」と、個人の「体験」を街の話題語として抽出し、得られた語を地図上で提示するインタフェースの提案を行う。これにより、ユーザは今まで以上に効率よく Blog に反映されている個人の行動を把握することが可能となる。また、「体験」を示す語から Blog を参照することで、実体験に基づく情報と高い頻度で出会うことができる。街を歩く個人の姿が反映されたこのシステムは多くの人に街に関する発見をもたらすものであり、街に繰り出すきっかけづくりの場となると考えられる。本手法では多くの人々が旅の目的としている「対象」を地名・ランドマークと助詞との共起に着目して抽出する。また個人の「体験」は、行為を意味する動詞とサ変名詞に着目して抽出する。

キーワード GIS, Weblog, 時空間 DB

Geographic Topic Extraction from Blogs

Takeshi KURASHIMA[†], Taro TEZUKA[†], and Katsumi TANAKA[†]

[†] Department of Social Informatics, Graduate School of Informatics, Kyoto University

Yosidahonmati, Sakyou-ku, Kyoto, 606-8501 Japan

E-mail: †{ktakeshi,tezuka}@dl.kuis.kyoto-u.ac.jp, ††ktanaka@i.kyoto-u.ac.jp

Abstract Compared to usual Web page, Weblog has a lot of entry about an subjective impression and experience of an individual. These information is significant in a different meaning from an objective attribute of the object. In this paper, it is described the way to extract the "Object" of a town where a lot of people are going to visit and the "Experience" that people have in a town as a topic of the town in Weblog. As a result, user can grasp all the activities of individual reflected by Blog more easily, and refer to Weblog by using "Experience" word. It can be said that, for many people, it leads to the discovery of an attractive object and play of the town. We can make motive for going to the town by using this system. In this method, the "Object" of a town where a lot of people are going to visit is extracted aiming at the co-occurrence of a name of a place, a name of landmark, and the particle. And the "Experience" of people is extracted in the form of the sentence aiming at the verb meaning of an action and the specific noun.

Key words GIS, Weblog, Spatio-temporal database

1. はじめに

Blog は、地域を実際に体験した個人の情報発信の場となっている。地域ポータルサイトや地域情報誌上の情報は、単に街の客観的な事実を述べただけか、あくまで情報提供者側からの一方的な発信によるものである場合が多い。Blog における個人の实体験に基づく情報は、対象の客観的な属性とは別の形で有意義であるといえる。

しかし、現在、これらの情報は検索エンジンに「京都」などの地名をクエリとして得られた文書集合ひとつひとつに目を通

すという方法でしか得ることができない。また、得られた文書が必ずしも実体験に基づいて書かれたものとは限らないため、情報収集が困難であるのが現状である。

本稿では、実空間上の対象の探索履歴ともいべき Blog から、多くの人々が旅の目的としている「対象」と、個人の「体験」を示す語を抽出しユーザに提示する手法を提案する。これにより、ユーザは今まで以上に効率よく Blog に反映されている個人の行動を把握することが可能となる。また、「体験」を示す語から Blog を参照することで、実体験に基づいて書かれた Blog と高い頻度で出会うことが可能となる。街を歩く個人の姿

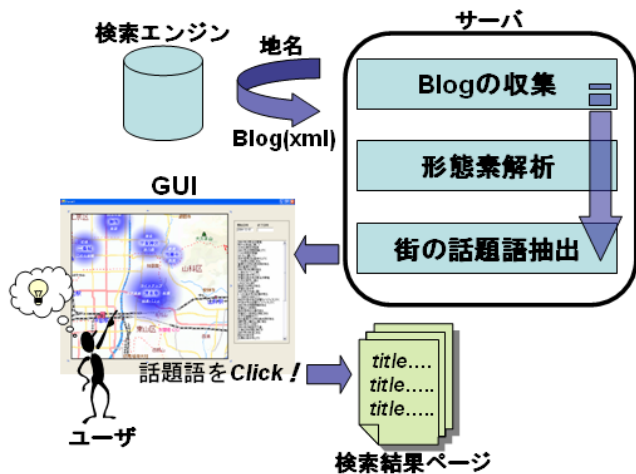


図1 システム構成図

が反映されたこのシステムは多くの人に街に関する発見をもたらすものであり、街に繰り出すきっかけづくりの場となると考えられる。本手法では、多くの人々が旅の目的としている「対象」の抽出を地名・ランドマークと助詞との共起に着目して抽出する。また人々の「体験」は、行為を意味する動詞とサ変名詞に着目して抽出する。

本稿の最後に抽出した街の話題語を提示するシステムの概要を示す。このシステムの特徴は街の話題語を地図インタフェース上で可視化し、眺めることができるという部分にある。ユーザにとっては「街に出かける」という部分で、地理的な要素が大きく関わってくることになり、時には足を運ぶきっかけにもなりうる。地図上で話題語を収集することができるため、「空間」という尺度での街の話題の参照が可能となる。また、街の話題は時間限定的なものほど魅力的な情報である可能性が高いのだが、その変遷は早い。眺めるという操作により、手軽に幅広く話題語を収集でき、その全体像を把握することが容易となる。

本稿の第2章では、本手法に関連する基本的事項や関連研究について述べ、3章では本手法の詳細について述べる。4章では、抽出した街の話題語をユーザに提示するシステムの概要と、その主要技術となる地図インタフェースについて述べ、5章では今後の課題、6章ではまとめを述べる。

2. 関連研究

2.1 話題語に関する研究

石井[1]らは掲示板から名詞句や固有名詞を話題として抽出し、語の発生密度に着目したスコアリング手法を提案している。この手法においては、その時点に向かってより密にその事象が発生している語ほど勢いが強いとみなされ、高いスコアを得る。奥村[2]らはBlogを収集・監視し、収集したBlogをマイニングすることで得られた情報を閲覧することが出来る「blogWatcher」を提案している。blogWatcherにおいては、藤木[3]らの提案するburst発見手法を利用して語の注目されている期間を抽出している。藤木らの手法はKleinbergの提案するburst検出手法に基づいており、Kleinberg[4]はburst状態を、ある期間でのBlogの出現間隔が通常よりもどの程度短く

なっているかを元に発見している。

これらの手法は主に、テキストに出現するすべての名詞句を話題語候補として抽出し、語の発生間隔に着目して話題語を抽出している。本稿で提案する、多くの人々が旅の目的としている「対象」抽出手法においては、テキスト中に出現する名詞句の周辺テキストまでもを考慮して語の抽出を行っており、その点で既存の研究と異なるといえる。具体的には、名詞句と助詞との共起に着目してである。また本稿で提案する「体験」抽出手法においては、名詞句よりも具体性のある文の形で「体験」を示す表現の抽出を行っており、この点においても既存の研究とは異なるといえる。

2.2 街 Blog に関する研究

上松[5]らは、Blogと、カメラやGPSといった機能が付加された携帯電話端末を用いて、各人が発信した情報を位置情報に基づいて整理し、新たな情報閲覧を可能にするシステム「場log」を提案している。情報閲覧方法としては、指定した地点から近いデータを時系列順、距離の近い順で提示する「場所Blog」と、データベースに登録された画像を、位置情報をもとに地図上に貼り付けて提示する「blog地図」がある。「blog地図」は、地図を用いて情報提示を行う点で本研究と共通しているが、可視化を行う対象が画像のみであり、テキスト情報に関しては扱っていない。画像に特化しているという点において、Blogのテキスト情報に着目している本研究のアプローチとは異なる。また、場logにおいては、Weblogツールが生成するRSSに位置情報を付加するという方法で位置情報を取得しているが、本研究においては、既存のWeblogツールが作成するBlogを対象とし、Blogに現れる地名・ランドマークに着目している点が異なる。

2.3 格助詞の分析に関する研究

手塚[6][7]らは、Webページ上における日本語の位置説明文を解析し、その解析結果が都市のイメージ、すなわち「イメージマップ」の描画に役立つことを示している。具体的には、格助詞、相対名詞、形容詞と地名との共起関係の分析を行い、地名のカテゴリと位置情報に用いられる品詞との間に相関関係があることを示している。本研究に関連のある結果として、格助詞に関する分析がある。まず、地名が、位置を示す格として用いられる「で/に」のどちらと多く共起するかの分析を行い、格助詞「で」と多く共起する地名は駅名・大学名などの人間の活動が頻繁に行われる場所、格助詞「に」と多く共起する地名は寺社や自然地形であると結論付けている。また、格助詞「へ」の分析も行っており、駅名などの移動の方向として重要な地名と多く共起することを示している。本研究とは、格助詞に着目しているという点で共通している。しかし、手塚らは地名を特性付けるために格助詞を利用している一方、本研究ではその場所に人が訪れているかどうかのスコアリングに格助詞を利用しているという点で異なるといえる。また、本研究は助詞の「の」や動詞、サ変名詞の分析も行っているという点で異なる。

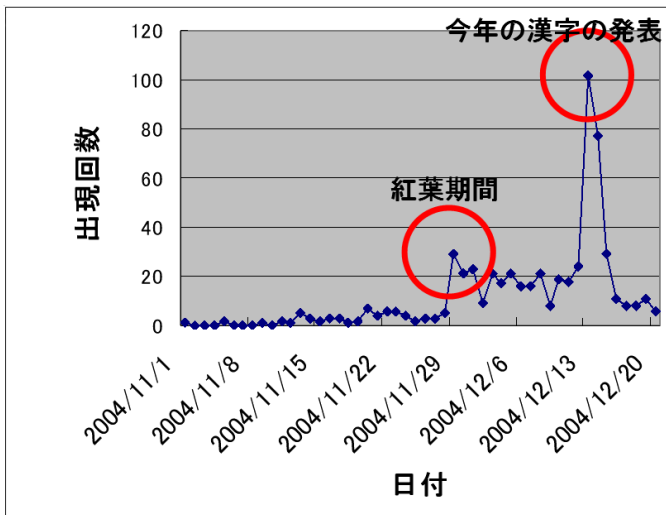


図2 「清水寺」の出現回数の推移
 期間:2004年11月1日~2004年12月20日

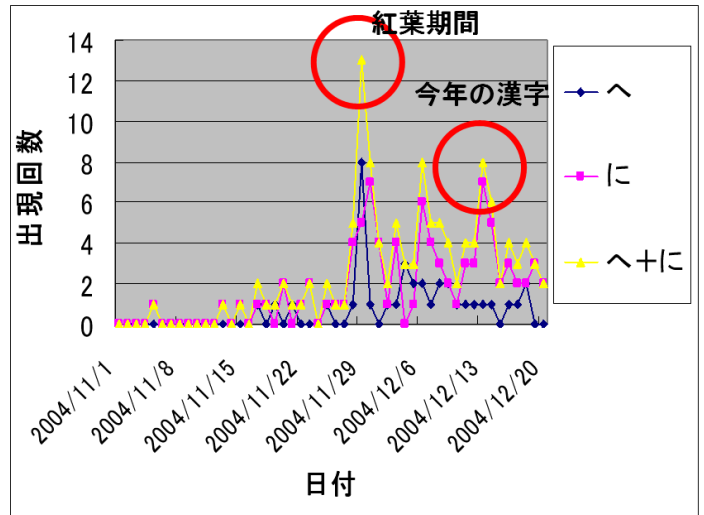


図3 「清水寺「へ/に」」の出現回数の推移
 期間:2004年11月1日~2004年12月20日

3. 話題語抽出手法の提案

3.1 多くの人々が旅の目的としている「対象」の抽出

人々の行動の履歴は、街における魅力的な対象を発見する際の確かな指標となる。本手法においては多くの人々が旅の目的としている「対象」を以下のようなプロセスで抽出する。

- (1) 多くの人々が訪れている場所を、地名・ランドマークと格助詞「へ/に」との共起に着目して抽出
- (2) 「対象」を、地名・ランドマークと連体化の助詞「の」との共起に着目して抽出

3.1.1 多くの人々が訪れている場所の抽出

本手法では、地名・ランドマークに伴って現れる格助詞「へ/に」に着目し、人々が実際に訪れている場所を抽出する。

格助詞「へ/に」は、共に移動に関わる格助詞である。格助詞「へ」は(1)のように移動の方向を示す。また、格助詞「に」は(2)(3)のように、「状態」述語と用いられる場合には「位置」を、「過程」あるいは「行為」の述語と用いられる場合には「着点」を意味する。地域に関して書かれた Blog は、日記のような形式で書かれているために、(3)の「過程」あるいは「行為」の「着点」として「に」が用いられる頻度が高い。

- (1) 加茂川 へ 行く
- (2) 嵐山 に 温泉がある
- (3) 今日は清水寺 に 行く

このことから、ユーザがその場所に訪れた体験をもとに Blog を作成した場合、(1) もしくは (3) のような表現を用いる場合が多く、地名と格助詞「へ/に」とが共起し出現する可能性が高い。そこで、格助詞「へ/に」との共起回数をその地名・ランドマークのスコアとする。

ブログホスティングサービスのひとつである goo ブログ [8] の検索エンジンを用いて RSS を収集し、実験を行った。以降、実験はすべて goo ブログから収集した RSS に基づいて行ったものである。

図2は「清水寺」の出現回数の推移、図3は「清水寺」が格

助詞「へ/に」と共起した回数の推移である。この期間に起こった主なイベントは、「紅葉拝観」と「今年の漢字の発表」である。図2においては、12月13日に行われた「今年の漢字の発表」で出現回数が増加しているが、図3においては、11月下旬の「紅葉拝観」の時期に出現回数が増加している。

「今年の漢字の発表」は、年末にその年を表す世相漢字を全国公募し、その結果を発表するというイベントである。清水寺の貫主が「今年の漢字」を色紙に揮毫する。多くの人がこのイベントについてのコメントを投稿したが、実際に清水寺を訪れた人が増加したというわけではない。一方で、「紅葉拝観」は実際に多くの人々が訪れ、その体験をもとに Blog を書いた人が増加した。「へ/に」格と地名との共起数をスコアとすることで、人が訪れている地名・ランドマークほど大きなスコアを得ることになる。

3.1.2 「対象」の抽出

連体化の助詞「の」は、名詞に接続して体言にかかるものである。「AのB」という名詞句は、「Aの」が「B」を修飾している。この修飾の意味としては「限定」、「属性」が考えられる。

イベントや地物などの街の「対象」は、この助詞「の」を用いて地理的属性を付加して現れる場合がある。例えば、「京都駅のクリスマスツリー」は、「京都駅」が「クリスマスツリー」を修飾し、地理的属性を付加している。本手法では、このような「地名・ランドマーク+の+名詞」という名詞句を手がかりにしてイベント名、店名、地物などの人々が旅の目的としている街の「対象」を抽出する。具体的には、形態素解析プログラム茶筌 [7] で得られる品詞に基づき、以下に示す品詞パターンにマッチする単語列を抽出し、さらに下線部分の名詞句を抽出する。?は直前の表現の0または1回の繰り返しを、+は直前の表現の1回以上の繰り返しを、|は選択を示す。地名と後続する名詞句をつなぐ助詞「の」は省略される場合があるので、その場合も抽出する。下線部分において、名詞ではなく名詞句を抽出するのは、名詞句は複数の名詞から構成されるため、語とし

ての具体性が増すと考えられるからである。

抽出する正規表現パターン

$Qa?((p?ns?) + | (p?ns?) + (a(p?ns?) +) +)$

Q: 地名, ランドマーク

a: 連体化の助詞「の」

p: 接頭詞

s: 名詞-接尾詞

n: 名詞(代名詞, 数, 非自立, 特殊, 接尾, 接続詞的, 動詞非自立は除く)

「AのB」という形の名詞句の中には, Bの名詞が位置を示し, 場所を示す名詞句を構成する場合がある。また, 時間の表現についても同様のことが起きる。例えば「清水寺の近く」や「清水寺の後」のような名詞句である。上に示す手法では「近く」「後」などが話題語として抽出されることになる。そこで, 「中, 外, 前, 後ろ, 横, そば, 近く, となり, あいだ, 前, 後, 時, 直前, 直後」などをストップワードとし, 話題語として抽出しないこととした。

表1に, 2004年12月1日から12月20日までの期間のBlog中から, 清水寺に関して「対象」抽出を行った結果を示す。具体的には, goo ブログの検索エンジンで「清水寺」をクエリーとして得られたBlog集合から, 本手法で提案する形式に該当する表現を抽出したものである。

表1 「対象」を示す表現の抽出結果

地名:清水寺, 期間:2004年12月1日~12月20日

話題語	出現数
貴主	17
ライトアップ	12
紅葉	8
夜間拝観	8
森清範貴主	3
お坊さん	2
今年の漢字	2
参堂	2
光の演出	2
舞台	2

3.2 人々の「体験」を示す表現の抽出

多くの個人がBlogに街での体験を記述している。これは, 既存のWebページや, 情報誌などの他の媒体にはない特徴である。また, Blog上に反映される個人の体験やその表現は, 多様なユーザの嗜好に応えるだけの多様性があり, さらに, 未だに多くの人が気づいていない街の楽しみ方が秘められている可能性がある。

既存の掲示板などからの話題語抽出手法[1]においては, 抽出する表現を比較的に単純な語の名詞句にすることで共起をとりやすくしている。しかし, このような手法では, 多様性に富んだ個人の「体験」は抽出できない。そこで本手法では, より具体性のある文の形で抽出する。

文は, どのような動詞で終わっているかで次の3つに分類す

ることができる。

- する文(行為)

(例)紅葉を拝観する

- なる文(過程)

(例)イベントが延期になる

- である文(状態)

(例)桜が五分咲きである

人々の「体験」を示す表現は, この中の「する文(行為)」に該当すると考えることができる。そこで, 本手法においては, サ変名詞および行為を意味する動詞(見る, 聞く, 食べるなど)に着目して文を抽出する。サ変名詞は「サ変名詞+する」のように用いられ, 行為を示す動詞的な役割を果たす品詞である。具体的には, 形態素解析プログラム茶筌で得られる品詞に基づき, 以下に示す品詞パターンにマッチする単語列を, 地名・ランドマークを含む一文から抽出する。

抽出する正規表現パターン

$((p?ns?) + | (p?ns?) + (a(p?ns?) +) +)j(S | m)$

a: 連体化の助詞「の」

p: 接頭詞

s: 名詞-接尾詞

j: 格助詞

S: サ変名詞

m: 行為を意味する動詞

n: 名詞(代名詞, 数, 非自立, 特殊, 接尾, 接続詞的, 動詞非自立は除く)

表2に, 2005年1月1日から1月10日までの期間のBlog中から, 清水寺に関して「体験」抽出を行った結果を示す。実験は「対象」抽出と同様, goo ブログの検索エンジンの検索結果に基づいて行ったものである。

表2 「体験」を示す表現の抽出結果

地名:清水寺, 期間:1月1日~1月10日

話題語	出現数
京都駅前の旅館を出る	1
京都の清水寺に行く	1
京都の清水寺に初詣	1
京都の街を散歩	1
清水寺のお店を冷やかす	1
清水寺の今年の一年を表す	1
清水寺の写真を載せる	1
清水寺の除夜の鐘を聞く	1
清水寺のところに行く	1
清水寺の様子を見る	1
交通安全のお守りを買う	1
故司馬遼太郎氏のお墓にお参り	1
写真に撮る	1
坂本竜馬の墓に立ち寄る	1
朱色の門が見える	1
念願の清水寺に行く	1

4. 街の話題語提示システム

3章で抽出した街の話題語を利用するシステムの概要を示す。このシステムの特徴は街の話題語を地図インタフェース上で可視化し、眺めることができるという部分にある。

現在、Blogの参照は検索エンジンを用いた手法が主流となっている。あるキーワードに関する情報を欲している能動的なユーザにとっては、検索エンジンは効率よく情報収集を行う手段となる。一方で、そうではない、情報に対して受動的な姿勢のユーザも存在する。街の活性化のためには、明確にこの情報がほしいという能動的なユーザだけでなく、情報に対して受動的なユーザに対しても街の魅力をアピールしその情報の存在を知ってもらう必要がある。

能動的に情報接触態度をとるような情報を持っていないユーザにとっては、ユーザの興味を引き出すような幅広い情報の提示を行い、その中から自分の興味に合致する情報を探してもらう場が有効であると考えられる。また、Blog上の情報は、時間限定的なものほど魅力的な情報である可能性が高いのだが、街の話題の変遷は早く、さらには多様性に富んでいる。そこで、本稿では地図上で街の話題語を提示し、ユーザの最小の操作である眺めるという方法で街の話題の全体像を把握できるような情報提示を行う。これにより、ユーザが街の魅力に気づく機会を増大させることができる。また、街の話題語を随時更新することで、街の話題の変遷を視聴することが可能となる。

システム処理の流れを以下に示す。本システムにおいては、以下のプロセスを定期的に繰り返して行き、常に新鮮な話題語をユーザに提供する。

(1) Blog 収集

サーバは対象領域内に存在する地名、ランドマークをクエリーとして Blog 検索エンジンに投げ、検索結果 (RSS) を取得する。Blog の収集は定期的に行う。

(2) Blog 文書の形態素解析

形態素解析プログラム茶笥を用いて RSS 中の Blog (タイトルと本文) 形態素解析する。

(3) 街の話題語抽出

本稿で提案した話題語抽出手法で、茶笥の実行結果と Blog の本文に基づき、街の「対象」と「体験」を抽出する。

(4) 話題語提示

地図インタフェース上で話題語を提示する。この地図インタフェースについては次項で詳しく述べる。

(5) 話題語の選択

ユーザが地図上で興味のある話題語を発見した場合、ユーザはその話題語をクリックすることでシステムにその語を含む Blog 集合を要求することができる。

(6) Blog 検索エンジンの検索結果の提示

ユーザがクリックした話題語と、その話題語が属する地名・ランドマークとを条件に、Blog 検索エンジンで AND 検索を行う。その結果をユーザに提示する。

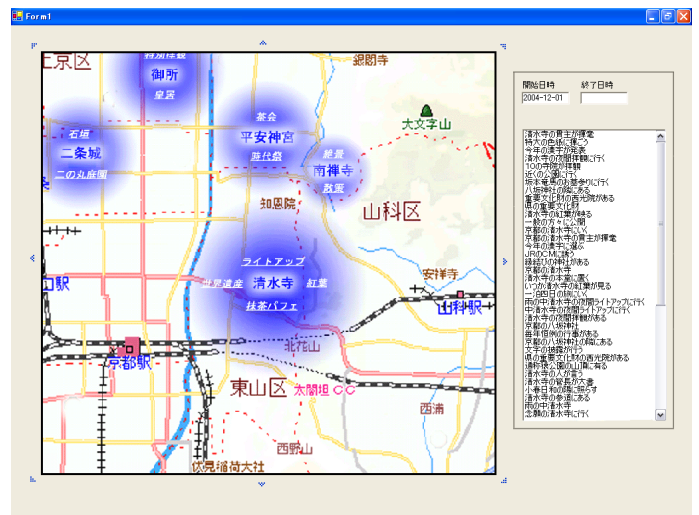


図 4 地図インタフェースによる話題語の提示

4.1 地図インタフェース

4.1.1 地名・ランドマークの表示

3.1.1 の手法においてウィンドウ内に見えている地図空間内に存在している地名・ランドマークのスコアリングを行い、スコアが上位の地名・ランドマークから数件を地図上で表示する。地名・ランドマークの周りの円形の領域の大きさが、地名・ランドマークのスコアに比例した大きさとなる。また、この領域内で「対象」に関する街の話題語提示を行う。ある地域に話題性の高い地名・ランドマークが集中すると、話題語提示領域の重複が起きてしまう。その場合は、話題語提示領域を動的に収縮させ、重複した領域を交互に情報提示するユーザが地名・ランドマークをクリックした場合、その地名・ランドマークに関連するすべての話題語を、話題語提示範囲を拡大することによって提示する。

4.1.2 話題語提示

抽出した「対象」に関する話題語の提示は、地名・ランドマークの周りの円形の領域で行う。話題語提示領域の大きさに応じて、出現数の上位の話題語から提示を行う。地図の右のリストボックスは、「体験」を示す話題語を提示する部分である。地図上で現在中心に位置している場所から地理的に近い地名・ランドマークの話題語から順に提示する。

ユーザがこれらの話題語をクリックした場合、「地名・ランドマーク」と「話題語」で Blog 検索エンジンを AND 検索した結果を、別ウィンドウで表示する。

4.1.3 期間の指定

このシステムにおいては、常に現在から一定の時間さかのぼった期間における Blog を対象にして街の話題語抽出を行っているが、ユーザは、話題語抽出期間を指定することによってその期間の話題語を閲覧することができる。街に関する話題語は四季に関連するものが多く、一年前、同じ場所で話題になった出来事は再び話題になる可能性が高い。去年の話題の推移を追うことにより、一定の割合で現在の話題の推移を予測することができる。

5. 今後の課題

5.1 手法の検証実験

本稿においては実験を行いデータを記しているが、その量は充分とは言えない。例えば、多くの人が訪れている場所の抽出手法の検証実験は「清水寺」についてのみ行っており、日数も50日間と短い。今後様々な場所におけるデータをとっていく必要があるだけでなく、街の話題は四季の変化に大きく左右されるため、四季をまたいだある程度長い期間におけるデータに基づいて実験を行う必要があると考えられる。また、駅などの場所は、話題の有無に関わらず普段から多くの人を訪れる場所であり、常にスコアが高くなると予想できる。今後、スコアの時間的変化を考慮した話題のスコアリング手法も検討していく予定である。

5.2 街の話題語抽出

表1において、「貫主」という「今年の漢字の発表」に関連する単語が上位にランクされてしまっている。「対象」の抽出は、(1) 場所の評価 (2) 場所における対象の抽出 という二段階から構成されているが、これは (1) の評価結果が (2) に反映されていないために起こった結果であると考えられる。今後、この2段階からなる抽出プロセスを再考する必要がある。また、街における「対象」は、Web、GIS、地域ポータルサイト、地域情報誌などからも抽出可能である。これらから得られる客観的な街の「対象」と、Blog から得られる主観的な「対象」とを比較してその差異をユーザにわかるように提示していく手法も模索していく予定である。

今回は、goo ブログで扱っている Blog のみを実験に用いたが、今後、他のブログサイトも含めて実験を行う予定である。しかし、このようになると抽出される「体験」が膨大な量になり、地図上で中心に位置している場所から地理的に近い順に提示するという現在の方法では対処できなくなると予想できる。この対処法のひとつとして、「体験」を抽出する条件を厳しくするということがあげられる。表2の「体験」を示す表現の抽出実験を行ったところ、格助詞「を」「に」を持つ文が多く抽出されている。現在は、「名詞句 + 格助詞 + (行為を意味する動詞もしくはサ変名詞)」という形で体験を抽出しているが、格助詞についての分析を進め、より精度の高い抽出手法の検討を進めていく。また、対象と体験は関連しているものもある。対象として抽出された名詞句が、体験の一部となっているものである。これらをどう扱っていくかも含め、検討をすすめていきたい。

5.3 地図インタフェースの検討

本研究においては、Blog 上の情報を街歩きに生かす第一段階として、街の話題抽出手法の検討を行ってきた。今後、取り組むべき大きな課題として、抽出した話題語をどのようにして地図上で可視化するかという問題がある。本稿では可視化のイメージを図に示したが、街の「対象」が可視化されている一方で「体験」は可視化できていない。この点について今後検討をすすめていく予定である。また、話題語抽出期間の指定の部分であるが、カレンダーインタフェースを用いて、時空間的な相互参照が可能なインタフェースにしていきたいと考えている。

5.4 Blog 独自の特徴を利用する可能性について

Blog は、更新が早いというだけでなく、リンクやトラックバックを持つという特徴がある。これらエントリー内のリンク、トラックバックなどを使って構造的に解析していくという展開も考えられる。また、今回の街の話題語抽出は「where」「what」「do」「when」に着目しているが、「who」、つまりプロガー検索のように書き手の属性から街の話題語検索をする仕組みも考えていきたい。

6. まとめ

本稿では、街の話題語として多くの人が旅の目的としている街の「対象」と、個人の「体験」を抽出する手法を提案した。また、得られた話題語を地図上で提示する地図インタフェースについても言及した。今後、現在わかっている問題点を分析し、それと並行して定量的な評価実験をすすめている。今回提案した抽出手法を修正していく予定である。また、地図インタフェースについても検討をすすめていく予定である。

謝 辞

本研究の一部は、《知的資産》文部科学省科学技術振興費プロジェクト「異メディア・アーカイブの横断的検索・統合ソフトウェア開発」(代表: 田中克己) および、平成16年度科研費特定領域研究(2)「Webの意味構造発見に基づく新しいWeb検索サービス方式に関する研究」(課題番号: 16016247, 代表: 田中克己) および、21世紀COEプログラム「知識社会基盤構築のための情報学拠点形成」および、平成16年度科研費基盤研究(A)(2)「モバイル環境におけるコンテンツのマルチモーダル検索・呈示と放送コンテンツ生成」(課題番号: 14208036, 代表: 田中克己) 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] 石井 恵, 中渡瀬 秀一, 富田準二: 名詞句と単語の勢いをを用いた話題抽出手法, 情報処理学会研究報告, NL160-12, 2004.
- [2] 奥村 学, 南野 朋之, 藤木 稔明, 鈴木泰裕: blog ページの自動収集と監視に基づくテキストマイニング, 第6回人工知能学会セマンティック Web とオントロジー研究会, 2004.
- [3] 藤木 稔明, 南野 朋之, 鈴木 泰裕, 奥村 学: document stream における burst の発見, 情報処理学会研究報告, 2003-NL-160, pp. 85-92, 2004.
- [4] J.Kleinberg: Bursty and hierarchical structure in streams, In Proc.of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1-25, 2002.
- [5] 上松 大輝, 沼 晃介, 徳永 徹郎, 大向一輝, 武田 英明: 場 log: Weblog 環境における位置情報利用の提案, 第6回人工知能学会セマンティック Web とオントロジー研究会, 2004.
- [6] 手塚 太郎, 李 龍, 高倉 弘喜, 上林 弥彦: ウェブ上の自然言語解析で描く都市のイメージ, DBSJ Letters Vol.1, No.1, pp.27-30, Oct. 2002.
- [7] 手塚 太郎, 李 龍, 高倉 弘喜, 上林 弥彦: Web の内容解析に基づく地理的領域の特性付けと地域情報検索への応用, 情報処理学会 DBS 研究会 & 電子情報通信学会 DE 研究会 合同ワークショップ (DBWS2002), 2002.
- [8] goo ブログ, <http://blog.goo.ne.jp>
- [9] 奈良先端科学技術大学松本研究室, 茶筌ホームページ <http://chasen.aist-nara.ac.jp/index.html>