

目的を考慮した情報フィルタリング手法に関する基礎調査

松辻 智之[†] 王 軼群[†] 土方 嘉徳[†] 西田 正吾[†]

[†] 大阪大学大学院基礎工学研究科 〒560-8531 大阪府豊中市待兼山町1番3号

E-mail: [†]matsu@nishilab.sys.es.osaka-u.ac.jp, ^{†††}{hijikata,nishida}@sys.es.osaka-u.ac.jp

あらまし 近年、通信技術の向上や、移動 GIS・GPS の発展に伴い、ユビキタス環境が現実のものとなりつつある。ユビキタス環境では、ユーザはいつでも、どこでも Web 上の情報にアクセスすることが可能になる。そのような環境では、実空間を移動しているユーザに対する情報獲得支援が重要な研究課題となる。そこで我々は、実空間を散策しているユーザを対象として、地理施設に応じた利用目的を常識データベースから抽出し、ユーザの居る地理施設に関する利用目的を獲得し、携帯端末上で適切な評判情報の検索を可能にするシステムの提供を試みる。

キーワード 常識データベース, 目的指向, 情報フィルタリング, モバイルコンピューティング, GPS, GML

A Basic Survey on Information Filtering Method Dealing with User Goal

Tomoyuki MATSUTSUJI[†], Yiqun WANG[†], Yoshinori HIJIKATA[†], and Shogo NISHIDA[†]

[†] Graduate School of Engineering Science, Osaka University Machikaneyama-machi 1-3, Toyonaka-shi,
560-8531 Japan

E-mail: [†]matsu@nishilab.sys.es.osaka-u.ac.jp, ^{†††}{hijikata,nishida}@sys.es.osaka-u.ac.jp

Abstract The technologies for telecommunication, GPS, and mobile GIS have rapidly progressed in the last few years. They will allow us to access to the Internet at any time and any place. A user needs to acquire information easily and correctly through a web browser on a cell-phone or a PDA. Our purpose is to build a system providing blog pages written about a facility for a user who is standing in front of the facility. The user can see these blog pages through the web browser on his cell-phone. The system has two technical features. One is extracting users' goals of using facilities from a common sense database. The other is filtering blog pages based on the goal selected by the user.

Key words common sense database, goal-oriented, information filtering, mobile computing, GPS, GML

1. はじめに

近年、通信技術・移動 GIS・GPS などの基礎技術の発展に伴い、ユビキタスコンピューティング [1] やモバイルコンピューティング (または mCommerce) [2] が現実のものになりつつある。「いつでも、どこでも情報にアクセス可能」というこれらの環境においては、実空間を移動しているユーザに対する有効な情報獲得支援が重要な研究課題になってくる [3]。

実空間におけるユーザの移動は、決まった目的地への移動と気ままな散策の大きく二種類に分けられる。前者の場合、ユーザは目的地への移動にはっきりとした目的を持っている (例えば、「ニューヨークの本社からパークレイの研究所に研究の打ち合わせをしに行く」など)。そのため、ユーザは事前にデスクトップ環境で、その目的に沿った情報を収集することができる。よって、移動中のユーザにとって価値のある情報とは、目的地までの道中や、目的地で当日に発生したイベント情報 (例えば、渋滞情報や天気の情報など) であると考えることができる。

後者の場合は、ユーザははっきりとした目的地を持っていないことが多い。そのため散策中に何か興味のある施設の側を通ったり、何か面白そうなイベントをやっていたりすれば、自由にその施設に入ってみたり、そのイベントに参加してみたりする。その際、実際に入ってみるか (参加してみるか)、そのまま通り過ぎるかの意思決定は、その場で得られる情報から行われる。たとえば、観光地を訪れているときに、興味を惹くレストランが見つけたとすれば、メニューには何があるのかやそれらの値段などの情報から意思決定すると思われる。

本研究では上記のうち、散策中のユーザに対する情報提供を考える。現状では散策中のユーザが意思決定に利用できる情報は、その場 (実世界) で得られる情報に限られるが、携帯電話や PDA などの携帯端末を利用すれば、インターネット上に存在する情報も意思決定に利用することができる。特に Blog [4] と呼ばれる個人でも手軽に日記形式で記述できるページには、その場で得ることができないような評判情報 (レストランであれば、料理のおいしさや店員の対応など) が多く含まれるため、

意思決定には非常に有益である。

しかし、散策中のユーザが実際に Blog から地理施設に関する評判情報を検索するには、いくつか問題がある。まず1つ目は、Blog 上には様々な観点からの評判情報が混在している問題である。同じ地理施設であっても、その評判情報はユーザの利用目的により異なる。そのため、地理施設の名前だけで検索すると、様々な観点からの評判情報が検索結果に入ってしまう。例えば、デートで利用するには良いが、宴会で利用するには不向きなレストランがあったとする。この場合、デートで利用しようとしているユーザは、デートで利用した時の評判情報を獲得したいのであるが、宴会で利用した時の否定的な評判情報は、判断を迷わせる原因になりえる。

2つ目の問題は、利用目的の入力にかかる手間の問題である。1つ目の問題を解決するには、ユーザの利用目的を検索キーワードとして獲得する必要がある。しかし、携帯端末の入力装置はテンキーや十字キーに限られるため、利用目的を携帯端末から文字で入力させるのは困難となる。また、地理施設の利用目的は、「デートをする」、「友人と話す」、「お酒を飲む」など、動作すなわち動詞で表現されるが、それを正確に検索キーワードに反映しようとすると、例えば、「お酒を飲む」に対しては、「お酒」と「飲む」だけでなく、「飲み」や「飲んだ」などの動詞の活用形を含んだ単語を、AND や OR などの演算子で組み合わせる必要が生じる。しかし、このような複雑な検索キーワードを携帯端末で入力することは更なる困難となる。

3つ目は、ユーザの持つ潜在的な目的に気づかせることが困難な問題である。散策中のユーザはそもそもはっきりとした目的もなく散策していることが多い。しかし、そのようなユーザであっても、ある地理施設に興味をもったということは、何かしらの潜在的な利用目的を持っているはずである。例えば、恋人同士で散策中に、あるレストランを発見したとする。その場合、そのレストランの利用目的は、おいしい料理を楽しむためであったり、おしゃれな雰囲気を楽しむためであったりするが、決してビジネスの話をするためではない。しかし、そのような潜在的な利用目的は、提示されれば判断できるものの、自分から明示することは困難と言える。

これらの問題に対し本研究の目的は、実空間を散策している携帯端末を持ったユーザに対して、地理施設に応じた利用目的を選択肢として提示し、選択された利用目的を用いて評判情報を検索することのできる、目的指向な情報フィルタリングシステムを開発することにある。本研究の学術的な貢献は、地理施設の種類に応じた利用目的のリストを自動生成することと、選択された利用目的に含まれる単語の活用形を補完した上で検索クエリに反映させることの2点である。

以降の章では、2章で携帯端末向けの Blog 検索システムを開発する際に解決しなければならない他の問題について説明し、さらに本提案システムの技術的な特徴について述べる。3章でシステムの概要を説明する。4章では、目的指向な情報フィルタリングを実現する方法について説明し、5章で今後の課題を述べ、6章で本論文のまとめを述べる。

2. 研究アプローチ

携帯端末で地理施設に関する Blog 記事を検索しようとすれば、いくつかの問題が生じる。本章では、まずそれらの問題点とその解決方法について簡単に述べる。次に、目的指向の情報フィルタリングの方式を実現するにあたってもっとも重要となるユーザの利用目的の獲得の方法について述べる。さらに、獲得した利用目的を用いて評判情報を検索しようとしたときに起こる問題とその解決方法について述べる。最後に本フィルタリング方式の技術的な特徴をまとめる。

2.1 Blog 記事を携帯端末から検索する際の問題

散策中のユーザが携帯端末から地理施設に関する Blog 記事を検索しようとすれば、その利用目的だけでなく、地理施設の名前を入力する必要がある、その入力にかかる手間も問題となる。また、PC 向けに作成された Blog ページを携帯端末の小さな画面に表示することも問題となる。

これらの問題に対しては、ある程度の機能のものであれば、既存の技術により比較的簡単に実現することができる。1つ目の問題の解決方法としては、ユーザの絶対空間座標 (GPS データ) から GML 形式 [5] などの地図データに対する空間問い合わせを行うことで、ユーザの周辺にある地理施設の名前などの情報を獲得することが挙げられる。地理施設に関する絶対位置情報と意味的な情報を保持した GML のようなデータは、今後普及が進むと見込まれる。2つ目の問題に対しては、Blog ページから本文を抽出し、抽出した本文のみを携帯端末に表示することで解決が可能である。PC 向けのレイアウトで作られた Blog ページ全体は、不用意にレイアウトを変更すると、非常に見づらいものとなる。しかし Blog ページの本文だけを抽出できれば、携帯端末上で自由に改行ができるテキストデータとなり、レイアウトの変更を必要としない。

本文の抽出は、Blog ページから本文を抽出するテンプレートを作成することで可能となる。さらに、いくつかのテンプレートを作成することで多くの Blog ページから本文を抽出することが可能である。なぜなら、限られた Blog サービスプロバイダが Blog のシステム (フォーマット) を提供しているため、その HTML の構造は比較的正しく構造化されており、その多様性もそれほど高くないためである。

2.2 ユーザからの利用目的の獲得

一般にユーザから目的を獲得するには、明示的にアンケートを取る方法 (明示的手法) [6] と、ユーザの何らかの操作ログなどから自動推定する方法 (暗黙的手法) [7]~[11] が挙げられる。

明示的手法は、直接ユーザに自由記述で答えてもらう方式と、回答項目を用意しておき、ユーザにはその回答項目の中から選択してもらう方式の2種類が考えられる。前者は、直接に目的を入力しないといけないので、携帯端末には適用が難しい。また、一からユーザが目的を考えて入力しなければいけないため、ユーザの意識していないような目的は明示されない可能性がある。後者は、項目を選択する操作は必要ではあるが、直接入力するほど煩雑な操作は必要ない。また、あらかじめ用意された目的を閲覧することで、ユーザが意識していない目的に気づい

たり、もともと考えていた目的とは異なる目的に気づいたりする可能性がある。

暗黙的手法は、ユーザに入力の手間をかけないという利点がある。しかし、推定には多かれ少なかれエラーが混入してしまう問題がある。さらに、我々のように目的を持たずに散策をしているユーザを対象とした時には、気まぐれなユーザの行動を予測することができない問題もある。

本研究は、(i) ユーザに目的を文字で入力する必要がないようにすること、(ii) はっきりと目的を持っていないユーザでも、潜在的な目的に気づかせるようにできること、を満足させたい。そのため、ユーザの利用目的の獲得には明示的手法で回答項目を提示する方法を適用する。

2.3 利用目的の回答項目の作成方法

回答項目の作成方法は、人手で地理施設に関する利用目的を整理する方法が考えられる。しかし、この方法では、個々の地理施設に対して利用目的を考えるもの全てを調査するという膨大な作業を強いられてしまう。ゆえに、アンケート項目の作成は、既存に明示されている知識を元に、何らかの方法で自動的に行われることが望ましい。既存の知識源として、以下の3種類の情報源が挙げられる。

(1) Web

Web全体を対象として、地理施設の利用目的に関する情報を獲得することが考えられる。しかし、ある地理施設名が存在するWebページを獲得することは容易であるが、それが出現する文脈は様々であり、利用目的に関する情報だけを抽出することは困難である。さらに、Web全体には様々なレイアウトのページが存在しており、文章のスタイルも口語調のものがあつたり一文が長いものがあつたりと様々である。そのため、タグや表現の特定のパターンを抽出に利用することも困難である。

(2) 百科事典

電子化された百科事典にも、地理施設の利用目的に関する情報は含まれる。しかも、Web上に存在する情報と比べると記述される文脈が多様ではなく、さらに記述される型式も統一されている。しかし、そこに記述される内容は百科事典の著者の定義に基づいたものに限られるため、多くの利用目的を網羅することはできない。

(3) 常識データベース

社会における様々な人間の常識をデータベース化するという試みは、多く行われている。例えば“Cyc project”[12]や“Open Mind”[13]などがある。これらのプロジェクトでは、多くのユーザが協力して知識を入力することを前提としている。特に、Open Mindは、誰もがWebを通して入力でき、しかも特殊な知識表現形式ではなく自然言語で入力できるという点から、2007年2月時点で約70万ユニットもの常識が獲得されている。また自然言語ではあるが、比較的決まった言い回しが多いため、Web全体を対象とするよりは構造化された知識を抽出することは容易である。

これらを比較したときに、(3) 常識データベース、特にOpen Mindは、知識の抽出の容易さとその知識の多様性の観点から、本研究における地理施設の利用目的を獲得するのに適した情

報源と思われる。そこで本研究では、Open Mindを利用してユーザに提示する利用目的の回答項目を作成することとする。具体的には、Open Mindに記述された常識の記述形式を分析し、その記述形式をテンプレート化することで、地理施設の種類ごとに利用目的に関する記述を獲得する。また、全ての地理施設の固有の名称を保持しているとは考えにくいので、利用目的を地理施設の種類ごとに抽出し、回答項目を作成する。

2.4 回答項目の階層化と検索クエリの生成

Open Mindから獲得した利用目的を、そのまま列挙すると、その数は携帯端末の一画面上には収まりきらず、何度もスクロールをしなければならなくなる可能性がある。利用目的の主要な意味は動作、すなわち動詞で表されると想定されるため、本研究では抽出した利用目的を「動詞→名詞句」の階層構造で表示することとする。さらに、これらを段階的に展開表示できるようにし、限られた表示領域で、効率よく回答項目を閲覧することを可能にする。そこで、Open Mindから利用目的を抽出する際は、上記の構造を単位とすることとする。

また、Blogの情報を検索するためには、Blog用の検索エンジンを自分で実装するか、すでに商用で提供されているBlog用の検索エンジンを用いるかが選択肢として挙げられる。本研究では、実装コストの低さと商用の検索エンジンの性能の高さを考慮して、商用のBlog用の検索エンジンを用いることとする。しかし、一般の検索エンジンを用いる場合、検索クエリの生成に関する問題が発生する。利用目的を検索クエリに含める場合、抽出した動詞と名詞句から検索クエリを生成する必要がある。一般に動詞には、いくつかの活用形（現在形や過去形など）が存在するため、再現率を重視するのであれば、検索クエリにすべての活用形を含めなければならない。本研究では、オンライン辞書により全ての活用形をOR検索で検索することとする。

2.5 提案方式の技術的な特徴

本提案方式の技術的な特徴は以下のとおりである。

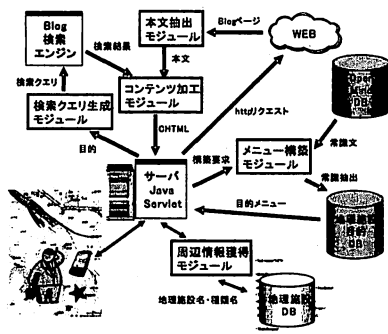
- 自然言語で記述された常識データベースOpen Mindに格納された文を分析し、地理施設の利用目的を「動詞→名詞句」の形式で抽出可能なテンプレートを作成する。
- 利用目的を「動詞→名詞句」の形式の階層構造で格納し、これらを段階的に展開表示できるようにする。
- 利用目的の動詞の活用形を補完した上で、検索クエリを生成する。

3. システム概要

本章では、目的指向の情報フィルタリングシステムの概要について述べていく。まず、実装するシステムがどのようなモジュールにより構成されているか述べる。次に、ユーザがシステムを利用する際のシステムの処理の流れについて説明する。

3.1 システム構成

本研究で提案するシステムの構成を図1に示す。システムは“周辺状況獲得モジュール”、“メニュー構築モジュール”、“検索クエリ生成モジュール”、“本文抽出モジュール”、“コンテンツ加工モジュール”から構成される。クライアントの端末とのや



り取りは Java サブレットで行う。

“周辺状況獲得モジュール”は、地理施設の名前と種類を保持したデータベースから、ユーザの絶対座標をキーとして、ユーザ周辺の地理施設に関するデータを獲得するためのモジュールである。空間問合せの高速化のために、事前に GML 形式の地図データの座標情報から多次元データの木構造を作成し、利用することも考えられる [14]。

“メニュー構築モジュール”は、検索対象の地理施設の種類に応じたユーザの利用目的の回答項目を構築するモジュールである。“検索クエリ生成モジュール”は、ユーザの選択した利用目的から Blog 検索エンジン用の検索クエリを生成するモジュールである。これらの詳細は 4. 章で詳しく説明する。

「本文抽出モジュール」は、Blog ページから本文のみを抽出するモジュールである。我々が 50 個の主要な Blog サイトを調査したところ、96% の Blog サイトでは、本文を `<div>` タグで記述し、さらに、`class` 属性に `"post"` や `"blog"` といった特徴的な単語を付していた。そこで、我々はこれらの情報を用いたテンプレートを作成し、本モジュールに実装した。本モジュールを用いて 200 の Blog ページから記事を抽出したところ、Blog 記事 175 ページから本文を抽出することができた。

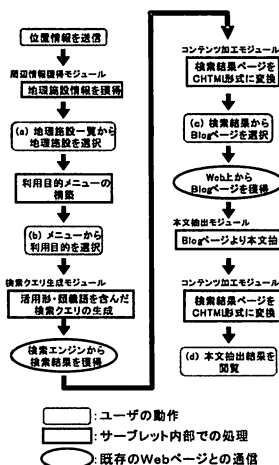
本システムでは、Web にアクセスする環境として NTT DO-COMO 社の i-mode を利用する。しかし、i-mode で利用できる Compact HTML(以下、CHTML) [20] は、HTML のサブセットであり、HTML で扱うことのできる機能をすべてサポートしているわけではない。そこで“コンテンツ加工モジュール”を用いて、抽出した HTML コンテンツを携帯端末で表示できるように変換を行う。

Blog ページのレイアウトは統一されたものではなく、さらに Blog の記事ごとの内容も、複数の話題に触れるものや他の Web ページへのリンクを集めたものなど、様々な記述形式が存在する。そこで“本文抽出モジュール”や“コンテンツ加工モジュール”には、ページの意味的な塊を推定して分割する手法 [15]~[18] や、ページのアノテーションとルールに基づいて Web ページを再構成する手法 [19] などの様々な表示手法を適用することも考えられる。しかし、本研究の技術的な特徴は、自然言語で記述された常識データベースから地理施設に関する利用目的を抽出して、それを検索クエリに反映させて、Blog を

検索することであるので、Blog ページをより携帯端末に適した形式で表示することに関しては、今後の課題とすることとした。

3.2 処理フロー

メニュー構築モジュールは、事前に Open Mind から地理施設の種類ごとに利用目的を抽出し、利用目的データベースを構築しておく。ユーザが興味のある地理施設の前で携帯端末を用いて本システムにアクセスすると、GPS などで取得した座標データを Java サブレットに送る。周辺状況獲得モジュールが、座標データから対応する地理施設の名前と種類を獲得する。地理施設の名前の一覧を提示し、ユーザは興味のある地理施設を選択する（図 2-(a) 参照）。ユーザが選択した地理施設の種類をキーとして、その地理施設における利用目的を階層構造のメニューとして構築、提示する。ユーザはそのメニューから利用目的を選択する（図 2-(b) 参照）。ユーザが選択した利用目的をもとに、検索クエリ生成モジュールが類義語・活用形を補完した検索クエリを生成し、検索を行う。その検索結果をユーザに提示し、ユーザは興味のある Blog ページを選択する（図 2-(c) 参照）。本文抽出モジュールが、選択した URL から対象の Blog ページにアクセスし、その Blog ページから本文が記述されている HTML のコードを抽出する。コンテンツ加工モジュールが、HTML のコードを CHTML のコードに変換する。CHTML はユーザの携帯端末に送信され、ユーザは地理施設に関する評判情報が書かれた Blog 記事を開覧することができる（図 2-(d) 参照）。



4. 目的指向な情報フィルタリング方式

本章では、目的指向な情報フィルタリング方式を実現するための、メニュー構築モジュール、及び検索クエリ生成モジュールの実装方法について述べていく。

4.1 メニュー構築モジュールの実装

地理施設に応じた利用目的を Open Mind から抽出し、それを用いてユーザに提示するメニューを構築する方法について述べる。そこでまず、Open Mind が利用目的を抽出する情報源

として本当に適しているかを調査する。次に、Open Mind に記述されている常識から利用目的の記述を抽出するためのテンプレートの作成方法について述べ、その精度と再現率を調査する。そして、抽出した利用目的を用いてユーザに提示する利用目的メニューの構築方法について説明する。

4.1.1 Open Mind の調査

我々は Open Mind について、地理施設に関する常識の中に、利用目的に関する常識がどの程度含まれているのかを調査した。3 人の調査員が、"casino", "church", "gallery", "hotel", "library", "restaurant", "shopping mall", "stadium", "university", "zoo" の 10 件の地理施設名に対して、Open Mind よりランダムに 100 文ずつを選び出し、その地理施設で何をするかという目的について述べた文（以下、「目的系の文」）の占めた割合を調べた。表 1 にその結果と Open Mind 内に格納されていた各施設種類ごとの常識の総件数を示す。その結果、100 文の地理施設の常識の中に 5~28%の目的系の文が含まれており、その平均は 14.8%であった。このことから、Open Mind 中に施設に関する常識がある程度格納されていれば、Open Mind にはその地理施設に関する目的系の文が存在することが分かる。

表 1 Open Mind における
地理施設の目的系の文の割合

地理施設名詞	目的系 [%]	総常識件数
casino	6	224
church	6	1052
gallery	22	531
hotel	6	832
library	23	1085
restaurant	15	3974
shopping mall	19	192
stadium	28	279
university	13	777
zoo	5	1964

4.1.2 利用目的抽出テンプレートの作成

Open Mind から地理施設の目的系の文を抽出するためには、利用目的の記述を判断できるテンプレートが必要となる。一般に、自然言語で書かれた文書からある記述を抽出するためのテンプレートの作成には、(i) 人間がそのテンプレートを定義する方法と、(ii) 教師信号によりテンプレートを学習する方法がある。Open Mind における常識は、一文が短く、しかも比較的決まった言い回しが多いので、人手でも実用的なテンプレートを作成できると考えられる。ゆえに、本研究では人手でテンプレートを定義し、地理施設の利用目的を抽出する。

そこで、我々は 4.1.1 節の調査で用いた 1000 件の地理施設に関する常識の中の目的系の文から、記述パターンを一般化することで利用目的を抽出するためのテンプレートを作成した。作成したテンプレートの一覧を表 2 の第 2 列目に示す。テンプレートの書式について説明する。〈term〉が地理施設の種類名、VP は目的系の文の動詞、VPL は目的系の文の動名詞、そし

表 2 テンプレートパターン

ID	表現文	精度 (%)	再現率 (%)
1	if you want to VP NP then you should go to 〈term〉.	96	24
2	you can use 〈term〉 to VP NP.	89	7
3	(going to) 〈term〉 is for people to VP NP. (going to) 〈term〉 is for VPL NP. 〈term〉 is (used) for VPL NP.	97	34
4	you would go to 〈term〉 because you VP NP.	100	1
5	people (can) VP NP [in, at, on] 〈term〉.	63	5
6	〈term〉 is a place to VP NP. 〈term〉 is (a place) where VP NP.	75	3
	全テンプレート適用結果	92	73

て NP はその動詞に続いた名詞句を表している。

4.1.1 節の調査で用いた 10 個の地理施設の種類のとは別の 5 個の地理施設の種類のに関する常識から各 100 件ずつ獲得し、それに 3 人の調査員が正解データを付与し、各テンプレートにより精度と再現率を求めた。その結果を表 2 の第 3 列目、第 4 列目に示す。全てのテンプレートを用いたときの抽出の精度と再現率はそれぞれ 92%、73%となった。この結果から、多くの目的系の文から利用目的を抽出可能であり、また誤抽出は 10%にも満たないことがわかる。今回作成したテンプレートは、先にも述べたように、Open Mind に含まれる常識は 1 文が短く、決まった言い回しが多いという傾向があるので、文の意味的な構造は考えず文字列として適合するかのみを考えるテンプレートを作成した。誤抽出となった文の多くは、文字列としての特徴が一致するものであったが、関係代名詞を用いるような文構造が複雑なものを抽出していた。また、抽出できなかった目的系の文を調査すると、「主語 動詞 前置詞 地理施設」のような単純な文構造のものが多くあった。このことから、精度と再現率をよりよくするための方法として、構文解析を利用し、文の意味的な構造を考慮したテンプレートを作成することが考えられる。

4.1.3 メニュー構築方法

本節では、利用目的メニューを携帯端末上での表示に適した形式で構築する方法について述べる。携帯端末の画面の問題上、メニューが横方向に長くなることは避けたい。そこで、本方式でのメニューの構築方法は、動詞 → 名詞句の形式で抽出された利用目的を、そのまま二階層で構築することとする。つまり、第 1 階層には利用目的の動詞となる単語を置き、第 2 階層にはその動詞の目的語となる名詞句を置く。表示に際しては、そのまま最初は第 1 階層である動詞のみを表示し、ユーザの展開操作により第 2 階層である名詞句を表示する方法をとる。その際、抽出した動詞は様々な活用形があること想定されるが、これらはすべて Porter Stemming Algorithm [21] を使用し、原型に戻し表示する。

4.2 検索クエリ生成方法

選択された利用目的から Blog 検索エンジンに入力する検索クエリを生成する。検索クエリには選択された利用目的の VP

と NP (4.1.3 節で集約したもの) から単語群を切り離し, stop word の削除を行う ([22] を利用した). さらに, オンラインの類義語辞書を用い, それぞれの単語に対する類義語を獲得し, 動詞に対してはオンラインの言語辞典を用い, 動詞の活用形を獲得する. これらの単語から以下のような検索クエリを生成する.

Term AND (VP OR S_{VP}) AND (NP_[i] OR S_{NP_[i]}) ...

上の式の中では, *Term* が地理施設名, *S_{VP}* が *VP* の活用形・類義語群, *NP_[i]* が *NP* 中の *i* 番目の名詞, *S_{NP_[i]}* が *NP_[i]* の類義語群を表している. 単語群 *S_{VP}* と *S_{NP_[i]}* は, *S_{VP}*[1] OR *S_{VP}*[2] OR... のように各類義語を OR の形式で並べた形になっている. このようにして生成した検索クエリを Blog 用検索エンジンに入力し, 地理施設に関する適切な評判情報を検索し, "本文抽出モジュール"と"コンテンツ加工モジュール"を用いて検索結果を閲覧する.

5. 今後の課題

今後の課題としては, 4. 章で作成した各抽出テンプレートに対するより詳細な調査を行い, 構文解析を用いた抽出テンプレートやテンプレートの組み合わせ方を考慮し, より高い精度と再現率を得られるものを旨とする. また, それらを用いて提案システムを実装した後に, 被験者実験を行い提案システムや各提案手法の有効性を検証する必要がある. 被験者実験を行う際は, 実際に被験者に散策を行ってもらい, 提案システムを使用してもらうことが望ましい. しかし, 提案システムが英語専用であること, 精度のよい GPS データの獲得が困難なこと, GML データが現段階では普及していないといった様々な要因があり, 上記のような被験者実験を行うことは困難である. そのため, 携帯端末を用いて散策している状況を想定した被験者実験を行わなければならない, 慎重な実験状況の設定が必要である.

6. まとめ

本研究では, 携帯端末を持った散策中のユーザに対し, 現在いる施設に応じた評判情報を獲得するためのシステムを提案し, 実装のための調査を行った.

さらに, 自然言語で記述された常識データベース Open Mind から地理施設の種類に対応した利用目的を動詞 → 名詞句の形式で抽出するためのテンプレートを作成した. また, 抽出した利用目的を動詞 → 名詞句の階層構造で段階的に提示するためのモジュール, 選択した利用目的の類義語や動詞の活用形を補完した検索クエリを生成するモジュールを作成した.

文 献

- [1] Kalle Lyytinen, Youngjin Yoo: Issues and Challenges in Ubiquitous Computing, Communications of ACM, Vol.45, No.12, pp.62-65, 2002.
- [2] Andrew Urbaczewski, Joseph S. Valacich, Leonard M. Jesup: Mobile Commerce Opportunities and Challenges Introduction, Communications of ACM, Vol.46, No.12, pp.30-32, 2002.
- [3] Gordon B. Davis: Anytime/anyplace computing and the future of knowledge work, Communications of ACM, Vol.45, No.12, pp. 67-73, 2002.
- [4] Andrew Rosenbloom: The Blogosphere, Communications of ACM, Vol.47, No.12, pp.30-33, 2004.
- [5] GML - the Geography Markup Language: <http://www.opengis.net/gml/>
- [6] A. N. Dragunov, et al.: Tasktracer: A Desktop Environment to Support Multi-Tasking Knowledge Workers, Proc. of IUI'05, pp. 41-50, ACM Press, 2005.
- [7] E. Horvitz, et al.: The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users, Proc. of UAI'98, pp. 256-265, 1998.
- [8] T. Fawcett and F. Provost: Activity Monitoring: Notice Interesting Changes in Behavior, Proc. of KDD'99, pp. 53-62, 1999.
- [9] M. Philipose, et al.: The Probabilistic Activity Toolkit: Towards Enabling Activity-aware Computer Interfaces. Technical Report IRS-TR-03-013, Intel Research Lab, Seattle, WA, 2003.
- [10] Mark Dredze, Tessa Lau, Nicholas Kushmerick: Automatically Classifying Emails into Activities, Proc. of ACM IUI'06, ACM Press, pp.70-77,2006.
- [11] Jianqiang Shen, Lida Li, Thomas G Dietterich, Jonathan L Herlocker: A Hybrid Learning System for Recognizing User Tasks from Desktop Activities and Email Messages, Proc. of ACM IUI'06, pp.86-92, ACM Press, 2006.
- [12] OpenCyc: <http://www.cyc.com/>
- [13] Open Mind: <http://commonsense.media.mit.edu/opensource-commonsense.html>
- [14] 玉田隆史, 門馬啓, 瀬尾和男, 土方嘉徳, 西田正吾: XML 空間インデックスを用いた携帯端末における G-XML データの効率的な管理方式, 電気学会論文誌, Vol. 123-C, No. 1, pp.19-25, 2003.
- [15] Buyukkorkten, O., Garcia-Molina, H. and Paepcke, A.: Focused Web Searching with PDAs, Proc. of 9th Int. World-Wide Web Conf., WWW9, pp.213-230, 2000.
- [16] Buyukkorkten, O., Garcia-Molina, H. and Paepcke, A.: Accordion Summarization for End-Game Browsing on PDAs and Cellular Phones, Proc. of the Conf. on Human Factors in Computing Systems, ACM CHI'01, pp.213-220, 2001.
- [17] Trevor, J., Hilbert, D.M., Schilit, B.N., Koh, T.K.: From desktop to phonetop: a UI for web interaction on very small devices, Proc. of the 14th annual ACM symposium on User interface software and technology, ACM UIST'01, pp. 121-130, 2001.
- [18] Xinyi Yin and Wee Sun Lee: Using Link Analysis to Improve Layout on Mobile Devices, Proc. of WWW'04, pp. 338-344, 2004.
- [19] Hori, M., et al.: Annotation-based Web Content Transcoding, Proc. of WWW-9, <http://www9.org>, 2000.
- [20] <http://www.w3.org/TR/1998/NOTE-compactHTML-1998020>
- [21] M.F. Porter: An algorithm for suffix stripping, Program, 14(3) pp 130-137, 1980.
- [22] <http://dvl.dtic.mil/libstopword>
- [23] NTT DOCOMO 社: <http://www.nttdocomo.co.jp/>