F-027

所在地に着目した観光地情報検索支援システム
Support System for Searching Sightseeing Area Information Using Location Data

小宮山 博之 † 延澤 志保 ‡ 太原 育夫 §
Hiroyuki Komiyama Shiho Nobesawa Ikuo Tahara

1. はじめに

近年，旅行などで観光を行う場合に旅行者自らが情報を収集し，計画を立てることが多くなった．情報収集の手段として Web 検索を用いることが多くなってきている．Web 上での観光地情報検索は気軽に観光地を調べられる反面，観光地間を移動する距離感がつかみにくい．例えば，何かの用事で普段は行かないような場所を訪れた際に，少し周辺の観光をしたいと思うときがある．近場の観光地を検討するためには，観光地が距離の近い順に並んでいる方が便利である．距離感を分かりやすく提供するために，山田 [1] は観光地とその周辺情報を地図上に表示するシステムを提案している．本稿では，主にユーザが徒歩で移動することを想定して，現在地を入力すると，周辺観光地の情報をその所在地に着目することで一括して取得・表示する手法について述べ，評価実験によりその有用性を示す．

2. 観光地情報抽出手法

本稿で提案するシステムの目的は，Web 上の観光地情報を含むページからユーザの入力した現在地の周辺観光地の情報を抽出し，その距離順にまとめて表示することで，ユーザが訪れたい観光地を検討する際の支援をすることである．したがって，各周辺観光地の詳細情報を抽出することは冗長である．なぜなら，詳細情報については，ユーザが各周辺観光地の名称などから興味を持ったときはもっと詳しい情報を調べればよいし，そうでないときは詳細情報を知る必要はないからである．そこで，周辺観光地の名称，所在地を抽出し，参考になるページの URL とその距離・徒歩でかかるおよその時間を加えて表示することにする．

こうした簡易的な周辺観光地情報の抽出・算出を以下のように行う．

1. 所在地情報の抽出
2. 観光地名の抽出
3. 観光地名照合
4. 距離・時間算出

2.1 所在地情報の抽出

所在地抽出アルゴリズムを図 1 に示す．まず，観光地情報ページを茶筌 [2] で形態素解析する．形態素解析で得られた品詞をもとに「市」「区」「町」「村」の形態素を探索する（茶筌では「名詞-接尾-地域」の品詞に分類されているので，それを用いている）．そこから前後方向に「名詞」が連続している部分を走査する．最後に「何丁目何番地」にあたる「未知語」を付加する．これで得られた範囲を所在地として抽出する．

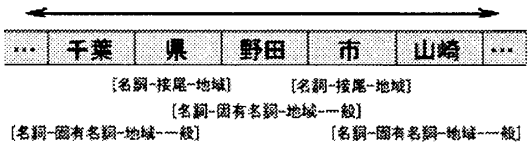

図 1: 所在地抽出アルゴリズム

2.2 観光地名の抽出

一般に，観光地について解説されたページでは観光地名があって，その後に所在地が書かれることが多く，所在地の後に観光地名が書かれることは少ない．そこで，所在地に対応した観光地名は所在地より前の部分から探索を行う．形式はテーブル形式やリスト形式など様々であるが，観光地名が文ではなく単独で置かれ，かつタグや記号で前後を囲まれていることがほとんどである．また，観光地情報ページの中には，1 箇所の観光地についてのみ述べられたものだけではなく，複数の観光地について述べられたページもあるので，所在地と観光地名の対応を考慮する必要がある．

このようなことから，抽出された所在地から先頭方向に走査し，

1. 所在地より前の部分にある
2. 「固有名詞」を含む複合名詞である
3. 前後を「未知語」の形態素で挟まれている

の 3 つの条件を満たすものの中で，所在地との距離が最短のものを対応する観光地名として抽出する．ここで，抽出した観光地名と所在地はそれぞれ組にしておく．

2.3 観光地名照合

あらかじめシステムのデータベースに登録された観光地名に，抽出した観光地名が含まれているかを照会し，含まれていれば引き続き距離と時間の算出を行い，含まれていなければその組は除外する．

2.4 距離・時間算出

距離の算出には Web で公開されている Geocoding を用いた API によって得られた距離を利用する．Geocoding は，ランドマーク（その土地の目印や象徴となる建造物）名か住所を入力すると，その地点の緯度と経度を返す API である．この緯度と経度を用いた「A 地点と B 地点の住所から距離などを割り出す API」（http://blog.yusukeooki.com/mylab.html）では 2 地点間の距離や徒歩での時間などを XML 形式で出力する．

†東京理科大学大学院理工学研究科情報科学専攻
‡武蔵工業大学知識工学部情報科学科
§東京理科大学理工学部情報科学科

ユーザが入力した場所の所在地と抽出した各周辺観光地の所在地をシステムが入力し，出力されたXMLから距離と徒歩での時間（分速80mとしたときの時間）を取得する．

3. 実験

実験は，全国5箇所（札幌テレビ塔，東北大学，日本科学未来館，みなとみらい駅，通天閣）について手動でGoogle検索を行い，得られた検索結果上位から観光地情報（掲載地域は問わない）を含むページを10件ずつ取得して，本システムを用いて観光地検索を行った（観光地情報を含むページとは，観光地の公式サイトやそれに準ずるページ，観光地のデータを載せているページやブログなどを指す）．システムの性質から，入力する地名はある程度周辺に観光地があるところを選んでいる．観光地名照合には，Yahoo!トラベル（http://travel.yahoo.co.jp/）に掲載されている「見る」の分類の観光地名を用いた．

3.1 評価方法

入力した地名から半径800m以内の観光地を対象として，適合率P，再現率R，F値Fを用いて評価する．抽出された正解観光地数をn_t，抽出された観光地数をn，正解観光地数をNとすると，各算出式は以下の通りである．

$$P = \frac{n_t}{n} \quad (1)$$

$$R = \frac{n_t}{N} \quad (2)$$

$$F = \frac{2PR}{P+R} \quad (3)$$

本稿では適合率，再現率ともに同程度重要であると考え，重要度の比は等しいとしている．

3.2 実験結果

各入力地名に対する評価値は表1のようになった（評価値は小数第4位で四捨五入した値）．

表1: 5件の入力地名に対する各評価値

地名	評価値		
入力地名（都道府県名）	適合率	再現率	F値
札幌テレビ塔（北海道）	1.000	0.500	0.667
東北大学（宮城）	1.000	1.000	1.000
日本科学未来館（東京）	1.000	0.429	0.600
みなとみらい駅（神奈川）	1.000	0.500	0.667
通天閣（大阪）	1.000	0.750	0.857
平均	1.000	0.636	0.758

4. 考察

4.1 適合率について

適合率についてはすべての入力地名に対して1.000と良い結果が出た．照合は観光地名でしか行っていないので，誤った抽出をすると予想していたが，照合は観光地名だけでも十分であることが分かった．

4.2 再現率について

再現率については入力地名によってばらつきが出た．抽出できなかった観光地について詳しく見てみると，大きく分けて2つの特徴があった．

- 観光地名に「固有名詞」が含まれていなかった場合

  観光地名を抽出するアルゴリズムでの抽出条件に「固有名詞を含む複合名詞である」というものがあるので，そもそも観光地名に固有名詞が入っていないと抽出されない．

- あまり有名な観光地ではなかった場合

  システムに渡すURL群はGoogleで検索を行い，上位から観光地情報を含むページを人手ではあるが機械的に取得している．したがって，照合に用いたYahoo!トラベルには掲載されているが，与えたURL群のページの中にその観光地の情報が載っていなければ抽出はされない．

これら2つの特徴の観点からすると，前者はアルゴリズムにまだ改良の余地があることを示唆している．例えば，「科学館」などのキーワードを含む場合を考慮するなどが考えられるだろう．後者は本研究の目的から，さほど有名ではない観光地は他に周辺観光地が検索されていれば必要ないと言ってもよい．解決策を挙げるとすれば，フィルタリングの条件を改良するか，抽出対象とするページ数を増やすことが考えられる．

また，抽出はできたものの，観光地名照合を行った際に削除されてしまった観光地も少ないながらあった．本システムでは観光地名照合を完全一致にしているため，観光地名に別称があってそれが括弧書きされている場合などは一致せず削除されてしまっていた．この点もまだ改善の余地はあるだろう．

5. まとめ

本稿では，観光地情報検索におけるWeb検索の手間と時間の削減を目的とした情報抽出法の提案を行い，システムの有効性を実験により示した．その結果，適合率に対する観光地名照合の十分性やF値から，観光地検索において所在地を用いた情報抽出法は有効であるという結果が得られた．しかしながら，観光地名の抽出手法にはまだ改良の余地があり，性能の向上はまだ可能であることも分かった．

参考文献

[1] 山田忍，中原知也，清水明宏，“観光情報収集に適したWeb型配信システム，”電子情報通信学会技術研究報告，vol.104，No.714，OIS2004-104，pp.55–58，2005.

[2] 松本裕治，北内啓，山下達雄，平野善隆，松田寛，高岡一馬，浅原正幸，形態素解析システム『茶筌』version2.3.3使用説明書，2003.