

Webから目的地の最寄り駅の探索

内藤 啓介[†] 馬 青[‡] 村田 真樹[‡]

[†] 龍谷大学大学院理工学研究科

[‡] 情報通信研究機構

1 はじめに

現在 Web 上には、指定した駅から駅までの行き方を検索できるサービスが多くある。しかし、そのようなサービスは利用者が駅名を入力することで初めて利用することができる。そのために、前提条件として利用者は目的地の最寄り駅が分かっている必要がある。最寄り駅を探索するときに目的地に関するページにその記述があれば容易に得ることができるが、それに関する記述がないもしくは目的地のページが存在しない場合は取得困難である。

最寄り駅の探索を行うことが出来るものとして Yahoo のサービスに地図情報の検索 [1] がある。これは、目的地を入力すると目的地の周辺の地図と最寄り駅を表示するものである。利用してみた結果、一文字違うだけで、またはごく普通の言い換えでうまく検索できなくなるケースが多くあった。例えば、「シンガポール総領事館」と入力すると正しく答えが返されるが、「シンガポール領事館」と入力すると「キーワードに一致する情報は見つかりませんでした」と表示され検索に失敗してしまう。また、「龍谷大学瀬田学舎」と入力すると正しい答えが返されるが、「龍谷大学瀬田キャンパス」と入力すると「キーワードに一致する情報は見つかりませんでした」と出力される。これは、目的地と最寄り駅の対応があらかじめ人手で登録されているデータベースがあり、そのデータベースと照合することで結果を出力していると考えられる。したがって、そのような場合は以下の問題点があると考えられる。

1. データベースの更新が遅い、コストが高い

2. 有名なところや大きなところの情報しかない
3. データベースの登録情報と完全に合致するように入力する必要がある

Yahoo 地図情報のほかに最寄り駅を検索できるサービスとしては、goo 地図 [2] などもある。しかし、これらも人手でデータベースを作成していると考えられるので、上で述べたような問題点があると考えられる。

われわれは、上記サービスや検索エンジンを利用することで目的地の最寄り駅を容易に知ることができないときに、目的地の住所やビル名から関連ページをたどることによって最寄り駅を探索するシステムの開発を試みている。提案手法は、あらかじめデータベースを作成しておくのではなく直接 Web データを利用するので上で述べた問題点を解消できると考える。また、本研究は最寄り駅の情報だけでなく最寄り駅から目的地までの、よりローカルなアクセス情報も探索できるようなシステムの開発を目指している。例えば、「龍谷大学瀬田キャンパス」という入力要求に対し、システムは「瀬田駅」という最寄り駅の情報だけでなく「瀬田駅から龍谷大学までバスで8分」のような情報も提供する。

本研究の関連研究として [3] [4] がある。[3] では与えられた固有名からその固有名に対する住所情報を Web データから探し出す手法を提案している。一方、[4] では固有名が与えられていない状況から固有名と住所情報に関するデータベースを Web データから自動的に作成する手法の開発を試みている。これらの研究は本提案手法の中の一処理ステップである「手がかりの抽出」に相当するものと考えられることができる。

2 人間の場合

本研究は、目的地の固有名からその目的地の最寄り駅を探索することを目指しているため、ここで、人間が実際に検索エンジンを利用し目的地の最寄り駅を探す場合を考えてみる。

人間の場合、まず、目的地を検索エンジンにかけ最寄り駅が記述されていそうなページを見つけてそれらをチェックする。チェックした結果は大きく二つの場合に分かれる。ひとつは該当ページからアクセスマップなどの交通情報が容易に見つけられる場合である。この場合はそこから、最寄り駅を見つけることができるため簡単に目的が達成する。一方、該当ページからアクセス情報が容易に見つけられない場合もよくある。その場合は、該当ページから何らかの手がかりを見つけ、その手がかりをもとに再探索を行う。その手がかりは例えば目的地がテナントとして利用しているビル名または住所などである。

本研究では、このような人手での探索手法にならない目的地の最寄り駅の自動探索手法の開発を行っている。

3 提案手法

3.1 基本的な考え方

提案手法の基本的な考え方は以下の通りである。まず、検索エンジンから得られた Web ページの情報をもとに、駅候補を探す。探した結果、駅候補が見つかったならば最寄り駅としての相応しさに応じてそれらを順位付けする。その結果、1位と2位の候補駅に対し、目的地との対を用いてそれぞれ再度 Web 上のヒット数を調査し、その差が決められた閾値を超えたならばその駅を最寄り駅とし、そうでなければ「回答不能」とする。「回答不能」とした場合、または最初から候補駅が1つも見つからなかった場合は Web ページから住所やビル名などの何らかの手がかりを抽出し、再探索を行う。

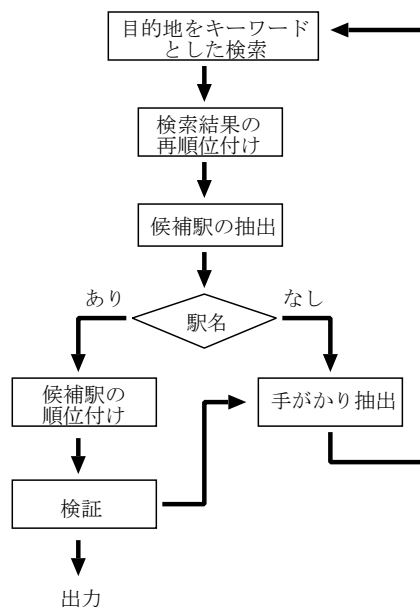


図 1: 処理の手順

3.2 処理手順

実際の処理手順は図 1 に示す。以下、その詳細について述べる。

3.2.1 目的地をキーワードとした検索

この処理は、Google を検索エンジンとし、目的地をキーワードに情報検索を行い、目的地に関する情報を含めそうな Web ページの取得を目的としている。

3.2.2 再順位付け

Google の順位付けは PageRank と呼ばれる指標を主に、ページ全体の情報の関連度で行われているため、われわれがほしい目的地のみに関する重要度で行われているわけではない。したがって、目的地情報のみを考えた新しい順位付け（再順位付け）を行う必要がある。再順位付けは Google 検索で得られた上位ページのタイトル（見出し）に絞りその中に出現した目的地情報の頻度を用いて行う。

3.2.3 候補駅の抽出

本処理では再順位付けされた上位ページから部分パターン照合を用いてすべての駅名を候補駅として取り出すとともにそれぞれの出現頻度も取得する。

3.2.4 候補駅の順位付け

複数の候補駅が見つかった場合、まず、それらの上位ページでの出現頻度で順位付けを行う。さらに、再順位付けで得られた1位のページに対し、そのページ内に候補駅が含まれているか否かを調べる。候補駅が見つかった場合、以下に述べるルールを適用し、その中からもっとも最寄り駅の可能性の高い候補駅を1位とし、その他の候補駅の順位を1つずつ繰り下げる。

ルール

「〇〇駅から□□方面、△△駅下車」というパターンがあれば

「〇〇駅」より「下車」に近いほうの「△△駅」を最寄り駅とする。

3.2.5 手がかりの抽出

候補駅が見つからなかった場合、再検索のために新たな手がかりを抽出する必要がある。新たな手がかりの見つけ方としては、TABLEタグによって対応する住所や建物名が見つかることができた場合は、それを優先して利用する。TABLEタグのように一対一の対応が取れているタグを用いることでより関連性の高い手がかりを見つけることができると考えられる。TABLEタグが見つからない場合は、再順位付けで得られたページに対し、その上位から順に「アクセス」、「住所」、「所在」、「マップ」などの単語を見つけ出し、その近くに現れているものを新たな手がかりとする。

3.2.6 検証

順位が付けられた候補駅に対し、その1位に順位付けられた候補駅を回答としてよいかについてWeb情報を再度利用することにより検証を行う。具体的には、1位に順位付けられている駅と2位に順位付けられている駅のそれぞれと、目的地を対に同時に検索エンジンにかけ、そのヒット件数

表 1: 使用した目的地

スイス総領事館	安曇川電子工業	小泉薬品
しか屋	あづま食品	キステム
山正	タキロン	寿福
日本銀行	オーストラリア大使館	比叡山鉄道
クオリカ	原信	森観光
トラスト	平和堂	うつぼ屋
比叡山鉄道	母恵夢	藤村薬品
弓場建設	江若交通	内閣法制局
武海建設	シグマソリューションズ	ソユール
平和堂	最高裁判所	新来島
広島	島ドック	オウミ住宅

を得る。1位のヒット件数と2位のヒット件数の差が決められた閾値より大きければ、1位に順位付けられた候補駅を回答とする。一方、その差が閾値よりも小さいときは「回答不能」とする（新たな手がかりとなるものを取り出し、最初から再度検索を行う）。

3.2.7 出力

上記処理手順を経て回答が見つかったならば回答を出力し、決められた時間内に回答が見つからなければ回答が見つからないというメッセージを出力し、探索処理を終了する。

4 実験

実験には目的地として30個の固有名を用いた(表1)。目的地をキーワードとした検索処理からは上位10件の検索結果を利用した。実験の結果、現段階では最終結果のみで判断した場合その正解率がまだ非常に低かった。しかし、下に示す例の通り、途中まで処理が順調に進んだ例もあった。また、処理のプロセスを調べると、新たな手がかりをみつけ、再び探索を行うことにより、正解の最寄り駅が得られた、というような期待通りの結果もあった。以下、結果の一部及び問題点について述べる。

1. 「スイス総領事館」を目的地とした探索

1回目の検索では候補駅を取得することはできな

かった。そのため、次に手がかりの抽出処理が行われた。その結果、手がかりとして「堂北ダイビル」を取得することができた。この新たな手がかりを用いて再度検索を行った結果、「北新地駅」、「淀屋橋駅」を取得することができた。これらは、どちらも目的地の最寄り駅となり得る駅だがヒット数を用いての検証において不正解となってしまった。

2. 「小泉薬品」を目的地とした探索

この場合も1回目の検索では候補駅を取得することはできなかった。そのため、次に手がかりの抽出処理が行われた。その結果、「宮城県仙台市若林区卸町」という住所を新たな手がかりとして取得することができた。この住所は正解であり人手で行ったとしてもまず初めに使うであろうと考えられる。しかし、新たに抽出できた手がかりを用いた再探索も正しい答えを得ることはできなかった。

3. 問題点

上に述べたように、手がかりをうまく抽出できたとしても残念ながら正しい最寄り駅を取り出せないケースがあった。また、本実験では手がかりは抽出できたもののそれ自身が間違っている例もあった。それは上位のサイトから順番に手がかりの抽出を行っているため、本来は目的地とは関係の無い建物などが新たな手がかりとして選ばれてしまったためと考えられる。更に、1位に順位付けされた候補駅が、ほんとうに目的地の最寄り駅であるかは1位と2位の候補駅と目的地をそれぞれ対にしたWeb上のヒット件数の差を用いて検証しているが、そのため、本来正解であったはずのものも正解として判断されない例もあった。また、目的地と候補駅の対で検索を行うと当然ながら、目的地での単独検索に比べ、ヒット件数が極端に少なくなり、検索のヒット件数が1桁のケースもあった。1位の候補駅に関するヒット件数が極端に低い場合は、2位のものに関するヒット件数もさらに低い場合が多かった。そのため、両者の差は大きいままで1位の候補駅が正解として相応しくなくても正解として出力されてしまう例が

あった。

また、今回の実験結果においては再順位付けの効果を確認することができなかった。その理由として利用している検索結果の上位件数（本実験では10件）が少なすぎたことを第一に挙げることができる。したがって、利用件数を増やした再実験を行う必要があると考える。

5 おわりに

本稿ではYahoo 地図情報サービスや検索エンジンを利用することで目的地の最寄り駅を容易に知ることができないときに目的地の住所やビル名から関連ページをたどることによって最寄り駅を探索する手法の提案を行った。提案手法は、あらかじめデータベースを作成しておくのではなく直接Webデータを利用するので従来システムに存在するデータベースの更新が遅いこと、有名どころや大きなところしか取り扱えないこと、さらにデータベースの登録情報と完全に合致するように入力する必要があることなどの問題点を解消できる。しかし、研究はまだ初期段階にあるので、実用レベルにはほど遠い。今後は関連研究の手法も参考に手がかりの抽出法や関連ページの順位付けに更に工夫を重ね提案手法の性能の大幅な向上を図るとともに、最寄り駅の情報探索に加え、最寄り駅から目的地までの、よりローカルなアクセス情報も探索できるシステムの開発を目指す。

参考文献

- [1] yahoo 地図情報 : <http://map.yahoo.co.jp/>
- [2] goo 地図 : <http://map.goo.ne.jp/>
- [3] 佐藤理史 : ワールドワイドウェブを利用した住所探索, 情報処理学会論文誌, Vol. 42, No. 1, pp. 59-67, 2001
- [4] 村山紀文, 南野朋之, 奥村学 : メタデータ付与のための住所録自動生成, 2004-NL-162, pp. 41-47, 2004