

GPS による位置情報を利用した地域情報の自動検索システム

Information Retrieval from GPS data for a Car Navigation System

大森俊秀^{*1} 平石広典^{*2} 溝口文雄^{*1*2}

OHMORI Toshihide, HIRAIISHI Hironori, MIZOGUCHI Fumio

^{*1} 東京理科大学 理工学部

^{*2} 東京理科大学 情報メディアセンター

Faculty of Sci. and Tech., Tokyo University of Science Information Media Center, Tokyo University of Science

This paper describes Information Retrieval from GPS data for a Car Navigation System. Our system extracts homepages concerned with current area from the web. This is realized by translating from GPS position to area information. Furthermore, in order to make the result list suitable for car navigations, our system rearranges the list by using browsing history.

1. はじめに

現在, www は急速に発達しており, それに伴い膨大な量の情報が web 空間上に存在している. また, カーナビゲーションシステムではネットワークに対応している機器が生産されている. しかし, この www とネットワークの 2 つを結ぶ検索の部分は従来と同様に手動で行っているのが現状である. 現在, www と位置情報を結びつける研究としては www 上にある位置情報を含んだページを地図上に表示させる研究などがおこなわれている[1]. これは, 住所を含んだページを緯度経度に変換し, 地図上に表示するものである.

我々は, 現在急速に発展している www に存在している, 地域に関するページと現在地の位置情報である緯度経度を組み合わせることにより, 車で移動中に現在地周辺の地域に関するページを自動的に検索するシステムを構築した[2]. しかし, このシステムでは直接 www にアクセスするため, www に存在しているその地域に関係のないページや多くの人が興味を示さないページに対しても検索結果として表示してしまうという問題点があった.

そこで本論では, このシステムを利用するユーザすべてから検索結果に対する個々の履歴を収集し, その履歴を利用することにより, 検索結果のページの順位付けを行い, その地域のユーザの関心の高いページを検索結果に反映するシステムの設計を行う.

2. 設計

本研究では, 直接 www にアクセスせずに Proxy サーバを間に入れて検索をおこなう. これにより, システムを利用するユーザすべての履歴情報を Proxy サーバで取得することが可能になり, それを利用することにより各地域のユーザが好んでいるページを検索結果の上位に表示し, 多くのユーザが好まないものは下位に表示することができる. ユーザにとっては Proxy サーバを導入しても依然と変わらない操作でサービスを受けることができる. もともと, セキュリティーのために設置される Proxy サーバは, 現在さまざまなシステムで利用されている[3]. 以下にシステム原理, システム概要, 履歴収集について述べる.

2.1 システム原理

ある地域 X においてその地域情報を www から検索したユーザ A, B, C, D... がいたとする. このユーザ A の地域 X に関する地域情報のページの履歴を利用することにより, その履歴の回数が多いページは上位に, 少ないページは下位に順位付けることが可能である. これは, 地域 X の地域情報において A の興味があると思うページは上位に, 興味がないページは下位に検索結果を順位付けていることと同じである.

そこで, ある地域 X においてその地域情報を検索したユーザ A, B, C, D... すべての履歴を利用することにより, 地域 X の地域情報に関するページの中で, ユーザの興味が多いものを上位に, 興味が少ないものを下位に順位付けることが可能である. 本システムはこの考えをもとに設計をおこなう.

2.2 システム概要

本システムでは Proxy サーバをとおして www とのやり取りをおこなう. この Proxy サーバを用いることにより, www の検索結果に対して順位付けをし, 多くのユーザが興味を持っているものを上位に, 興味のないものを下位に表示するという付加価値をつけることができる. 図 1 は個々のユーザと Proxy サーバのやり取り, そして Proxy サーバと www のやり取りの様子を示している.

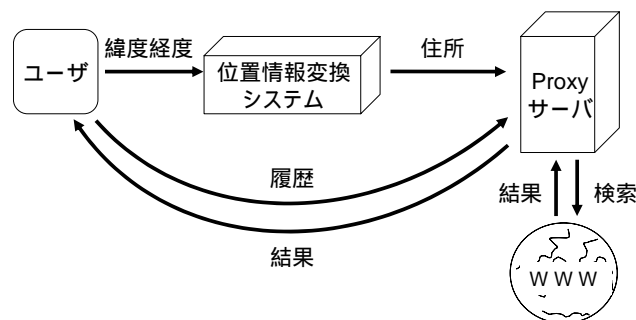


図 1 システムの概要

システムの動作としては, 左上のユーザから始まり次のような経過をたどっている.

- 緯度経度を住所に変換する.
- 住所をもとに Proxy サーバを通して検索を行う.
- www からの結果を順位付ける.

- 順位付けられた検索結果をユーザに返す.
- ユーザのブラウジングの履歴を Proxy サーバに返す.
- 履歴をもとに検索結果の順位付けの基準を変更する.

この個々のユーザと Proxy サーバのやり取りを全体的にみると図 2 のようになる. 個々のユーザがそれぞれ Proxy サーバとやり取りをし, その個々の検索結果に対する履歴を Proxy サーバに集めることにより, 特定の地域のユーザが好むページの履歴が加算され, また好まれないページは履歴が少ない状態になる. そして, この履歴の数の違いを利用することにより, 特定の地域のユーザに好まれるページは上位へ, 好まれないページは下位へ順位付けされるという仕組みになっている.

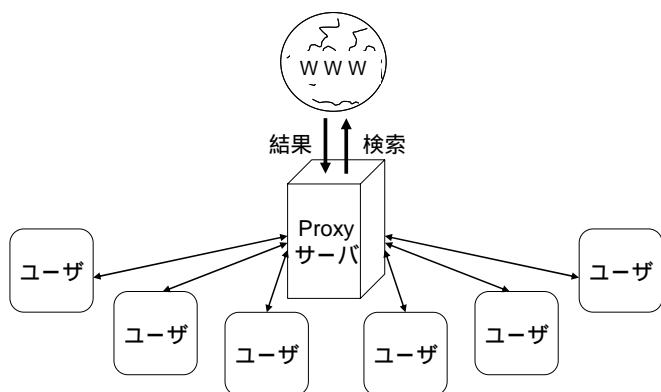


図 2 システムの全体像

2.3 履歴収集

履歴収集には Proxy 機構を用いた. 通常, Proxy サーバはネットワーク資源の共有化, 情報の変換, フィルタリング, 情報の付加などのサービスにおいて利用されている. www 上のページを一度 Proxy サーバに通して閲覧することにより, 普通のブラウジングに付加価値をつけることが可能になる. 本研究では, URL, ユーザ名, ページタイトルを履歴としてデータベースに記録するために用いる. この記録した情報からページのアクセス回数の履歴などを算出し, それをもとに以下に示すようなページの評価をおこなうことにより, その評価に伴った順位で検索結果のページの表示をおこなう.

3. ページの評価方法

ページの評価はもとの順位と全体の履歴をひとつに集めた回数による順位の 2 つをもとにおこなう. これにより, もとの検索エンジンの順位付けを利用しつつ, 個々のユーザの履歴による動的な順位付けが可能になる. ここでは, それぞれのページを偏差値で評価したうえでスコアリングをおこなう.

3.1 順位付け

u を順位, n をページの全体の数, X を点数, D を偏差値, S をスコアとする. 順位 u のスコアは以下の式(1)のように求める. まず, $1 \sim n$ 位の点の合計を計算し, それに対して順位 u の全体での偏差値の割合をかけることにより全体の合計点を割り振っている.

$$S_{(u)} = \sum_{i=1}^n X_{(i)} \times \frac{D_{(u)}}{\sum_{i=1}^n D_{(i)}} \quad (1)$$

また, 以下の手順でスコアをもとめることにより, 順位付けをおこなっている.

- もとの順位 $1 \sim n$ には基本点として $n - 1$ の点を与える.
- 基本点, 履歴それぞれの偏差値を求める.
- 偏差値ごとにスコアを求める.
- それぞれのスコアの合計をもとに順位付けをおこなう.

表 1 は具体的な例である. ページの全体の数が 10, タイトルが T1, T2, T3, T4... , 履歴の回数が 5, 0, 3, 1... と与えられている.

まず, 始め 1 位で履歴も 1 位の T1 はもちろん順位付けが 1 位であり, 始め 10 位で履歴も 10 位だった T10 はもちろん順位付けが 10 位になっているのが分かる.

そして, 始め 2 位で履歴が 0 回だった T2 はこの順位付けにより 5 位になり, 始め 8 位で履歴が 5 回だった T8 は 3 位に順位付けられた. この結果から, 履歴を検索結果のページの順位付けに反映させることが可能であることが分かる.

表 1 順位付けの例

順位	タイトル	基本点	偏差値	順位スコア	履歴	偏差値	履歴スコア	計	順位付け
1	T1	10	64.86	7.13	5	65.91	7.25	14.39	1
2	T2	9	61.56	6.77	0	39.39	4.33	11.10	5
3	T3	8	58.26	6.41	3	55.30	6.08	12.49	2
4	T4	7	54.95	6.04	1	44.70	4.92	10.96	6
5	T5	6	51.65	5.68	3	55.30	6.08	11.77	4
6	T6	5	48.35	5.32	1	44.70	4.92	10.23	7
7	T7	4	45.05	4.96	0	39.39	4.33	9.29	8
8	T8	3	41.74	4.59	5	65.91	7.25	11.84	3
9	T9	2	38.44	4.23	1	44.70	4.92	9.15	9
10	T10	1	35.14	3.87	1	44.70	4.92	8.78	10

4. おわりに

本研究では, ある地域の地域情報を検索したすべてのユーザの履歴を用いることにより, その地域情報の検索結果に対して順位付けをおこなった. これにより, 以前の問題点である地域に関係のないページやユーザの多くが興味を示さないページを検索結果の上位に示してしまうことを防ぐことが可能となり, また, 多くのユーザが興味を持っているページを上位に示すことが可能となる.

参考文献

- [1] 相良 毅, 有川 正俊, 坂内 正夫, 2000. ジオリアレンス情報を用いた空間情報抽出システム. 情報処理学会論文誌, 41/SIG6(TOD7), 69-80
- [2] 大森俊秀, 平石広典, 溝口文雄, GPS を利用した地域情報の自動検索システム, 第 66 回情報処理学会全国大会, 2004 年 3 月 4S-1
- [3] 岩下文彦, 西山裕之, 溝口文雄, 個人履歴情報を利用したリンク構造アルゴリズムの研究, 第 66 回情報処理学会全国大会, 2004 年 3 月 3U-5