

喜田 弘司 石黒 義英 山田 洋志 福島 俊一 松田 勝志

NEC

1. はじめに

近年、iモードなどの普及により、携帯情報端末を使って、WWW のページを閲覧することが簡単にできるようになった。これにより、ユーザが携帯端末を持ち歩き、現在の居場所に関する周辺情報を検索するといった使い方(モバイル位置指向サーチ)が増えると予測できる。この目的に特化したサーチエンジンとしてモバイルインフォサーチ(MIS)などがすでに提案されている[1]。

一方、我々は、より汎用的な視点から、いろいろな目的にカスタマイズすることで高精度な検索・分類機能を提供する目的特化型 WWW サーチエンジン「」を研究・開発している[2]。本稿ではこのフレームワークの一例として試作したモバイル位置指向サーチエンジンに関して述べる。

2. モバイル位置指向サーチの要件

一般的なサーチエンジンや MIS を使ってモバイル位置指向サーチを行った場合、以下の点が不十分である。

- (1) モバイル用以外のコンテンツも検索されてしまう: 通常、携帯端末では、すべての WWW コンテンツを見ることができるわけではない。モバイル用に作られたコンテンツだけを検索する仕組みが必要である。
- (2) 位置の検索が柔軟にできない: 例えば、奈良県に関するコンテンツを検索するために、検索キーワードに「奈良」を追加した場合、奈良の電話番号が書かれていても「奈良」と明示的に書かれていないコンテンツは検索することはできない。あるいは、奈良時代のことについて書かれている別の場所のコンテンツが「奈良」というキーワードにマッチしてしまい検索されてしまう。各コンテンツを位置で分類しておく必要がある。

これら要件に対し、我々は、以下の2つの手法を開発した。

- (1) インターネット上に存在するWEB 文章の中から、iモードのページを自動抽出する手法
- (2) 抽出されたiモードのページを位置情報で分類する手法

3. システム構成

処理の流れを以下の図に示す。

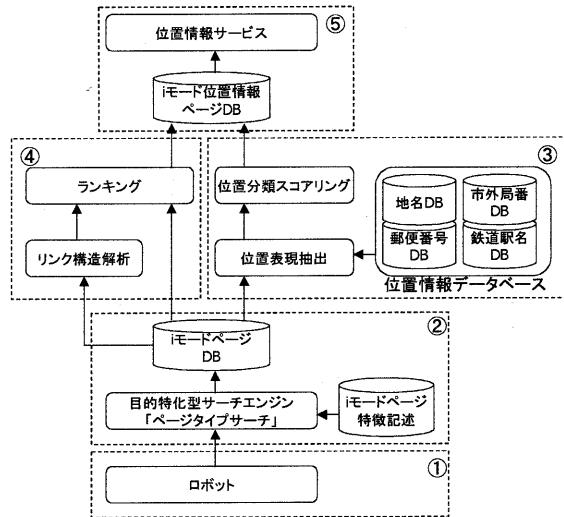


図1 システム構成図

- ① WWW コンテンツをロボットで収集
- ② 収集した WWW ページの中から、iモードのページを抽出
- ③ 抽出されたiモードページに対して、位置情報で分類
- ④ 抽出されたiモードページをランキング[3]
- ⑤ ユーザは、位置で分類されたiモードコンテンツを検索

以下では、モバイル位置指向サーチ用に今回新たに開発した、②、③に関して説明する。

3.1. iモードページの自動判定方法

目的特化型 WWW サーチエンジン「ページタイプサーチ」[4]をi-modeのページだけを抽出するようにカスタマイズして実装した。このカスタマイズは、検索したいWWWページの特徴をキーワードの出現、特定タグの使用などで記述することで行う。iモードページの特徴を以下に示す。

加点の条件

- 絵文字の使用、ディレクトリに/i-mode/など、
- href先が電話番号、...

減点の条件

- <FRAME>など特定タグの使用、
- テキスト部分のサイズが2500Byte以上、...

3.2. iモードページの位置情報分類方法

iモードのページを入力に、そのページがどの位置に関連しているかをインデックス付けする方法を説明する。この処理は以下の2つのフェーズからなる。

位置表現抽出

iモードページのテキストを走査し、地名、市外局番、郵便番号、鉄道駅名のデータベースに登録されたキーワードにマッチする文字列を抽出する。データベースの各キーワードは、都道府県→区→市群→町村の階層で表現された特定の位置に対応づけられており、マッチしたiモードページの文字列を、そのキーワードの位置に関連づける。例えば、ある市外局番にマッチした文字列は、その局番の市の階層までに関連づける。また、複数の位置に同一の地名が存在するため、文字列を複数の場所に関連づけることもある。以下に位置表現を抽出した例を示す。

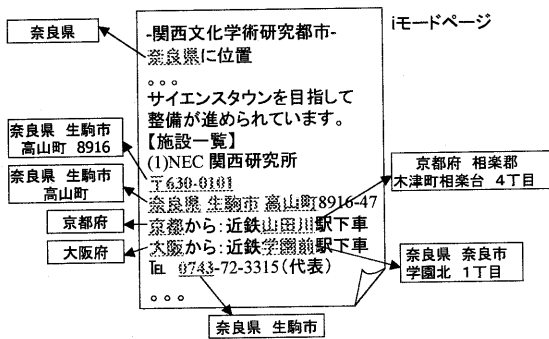


図2 位置表現抽出例

位置分類スコアリング

あるページが位置Xであるスコアは、そのページ内で位置Xと関連のある位置表現の出現頻度と、位置X以外の位置と関連のある位置表現の出現頻度の差で計算する。図2の例の場合、このページが奈良県であるスコアは $5-3=2$ である。また、この計算は、位置表現の各階層毎に行う。図2の例の場合、市の階層に注目すると、生駒市であるスコアは1であり、奈良市や相楽郡であるスコアは-3である。このスコアは、検索する際の優先度として利用する。

4. モバイル位置指向検索ポータル

ユーザが位置を入力するとその位置に関する情報が検索されるモバイル位置指向検索ポータルを試作した(図3)。なお、GPSなどを使って位置情報を取得すれば、ユーザは位置を入力する必要はなく、居場所に応じて情報をプッシュすることも可能である。

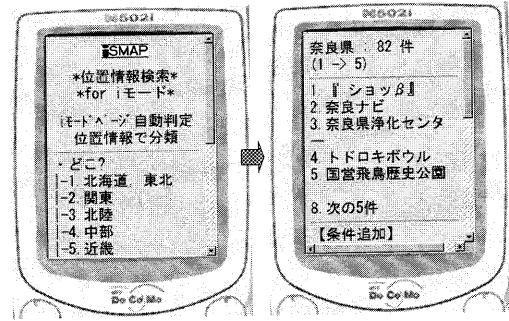


図3 モバイル位置指向検索ポータル画面例

5. 評価実験

5.1. iモードタイプサーチの評価

NETPLAZA[5]で収集したWWWページに対して、iモードタイプサーチの評価実験を行った。

適合率 iモードページと判断した977件中、実際にiモードのページであったのは966件(98.9%)

再現率 iモードページ1833件に対してiモードと判定されたのは1339件(73.0%)

適合率、再現率ともに十分な精度であるといえる。

5.2. 位置情報分類の評価

モバイル位置指向検索ポータルを使って、東京都、奈良県、沖縄県をサンプルに(1957件)、分類精度を調べた。なお、適合率は、スコア毎に無作為にデータを選択して、正しい分類ができているか調査した。

スコア	適合率	再現率
1以上	14/15 (93%)	1480/1957 (76%)
0以上	17/22 (77%)	1704/1957 (87%)
全スコア	18/28 (64%)	1957/1957 (100%)

この結果からスコアが1以上の分類を正しい分類とすれば適合率93%、再現率76%という実用的な精度が得られる。

6. おわりに

本稿では、WWWのページの中からiモードのページを自動抽出し、抽出されたページを位置情報で分類するモバイル位置指向検索エンジンを提案した。今後、抽出精度をさらに改善し、実用化に取り組んでいく予定である。

参考文献

- [1] 三浦 他" モバイルインフォサーチ: 移動環境下でのユーザ指向WWW検索", 情処MC研, MC-3-7(2000)
- [2] 福島 他" 目的特化型WWW検索エンジンの開発", 第61回情処全大講演論文集, 1U-01(2000)
- [3] 高野 他" 情報構造解析を用いたWWW検索ランキング方式", 第59回情処全大講演論文集, 3P-07(2000)
- [4] 松田 他" 文書タイプ分類による問題解決向きWWW検索システムの開発と評価", 情処FI研, FI-53-2(1999)
- [5] <http://netplaza.biglobe.ne.jp/>