ヒューマンインタフェース 86－7
モーバイルコンピューティング 11－7
（1999．11．26）

位置指向の情報構造化と情報フィルタリング ～モーバイルインフォサーチ3実験～

三浦 信幸† 横路 誠司‡ 井上 香織‡ 高橋 克巳‡
高橋 健司‡ 島 健一†

† NTT移動通信網(株) マルチメディア研究所
‡日本電信電話(株) 情報流通プラットフォーム研究所

本稿では，インターネット上に存在する雑多な形式の情報を，位置に応じて適切に提供するための情報構造化や情報フィルタリングを行う手法を検討する．このような情報を適切に提供するためには，雑多な形式の情報に対して構造化を行い，構造化された結果を利用して，位置を含めた様々な観点から情報を分類・フィルタリングする必要がある．検討した手法では，情報構造化に際してパターンマッチや特定分野の辞書を用いた形態素解析などを行う．また，情報フィルタリングに際しては，構造化された情報と構造化されなかったHTMLファイル中の名詞や固有名詞の中から*tfidf*値を参考に頻出する情報を抽出する．さらに，検討した手法のプロトタイプである，モーバイルインフォサーチ3実験(**MIS3**)について紹介する．

Location-oriented Structuring and Filtering of Information – Mobile Info Search 3 Experiment –

Nobuyuki Miura † Seiji Yokoji ‡ Kaori Inoue ‡ Katsumi Takahashi ‡
Kenji Takahashi ‡ Ken'ichi Shima †

†NTT DoCoMo Multimedia Labs.

‡NTT Information Sharing Platform Labs.

E-mail : †{miura, kshima}@mml.yrp.nttdocmo.co.jp
‡{yokoji, inoue, takahasi, kt}@slab.ntt.co.jp

We have developed techniques for location-oriented structuring and filtering of heterogeneous information on the Web. To provide information which is relevant to specific locations, we need to structure heterogeneous information and categorize or filter structured information from various viewpoints. In our techniques, for structuring, morphological analysis and pattern-matching are used. For filtering, *tfidf* values are used to determine information that best describe a location.
Based on these techniques we also have prototyped *Mobile Info Search 3*, a system to provide useful information for users interested in a particular location.

1 はじめに

昨今，GPSや位置情報対応PHS等により，モバイル環境下で位置情報を利用できる機会が増えてきており，様々な利用の仕方が検討されている．そのひとつとして，自分の現在いる場所の周辺の情報を得るための情報検索の検索条件として利用することが行われている．検索対象としては，人間の手によってまとめられた，地図・時刻表・レストランガイドのようなものが用いられる場合が多い．このような情報は，店舗名：○○，営業時間：××～△△といった具合いに，個々の検索対象について属性と属性値の組が付与された構造化情報である．一方で，WWW上に存在する個々のHTMLファイルのような非構造化情報は，日々増大し，サーチエンジン等の力を借りて日常的に利用されており，非常に魅力的な情報源である．このような情報源には，特定の場所について言及している情報が実際に数多くあり，例えば、WWW上で言及されている場所を白紙の上にプロットすると日本地図ができあがるほどである[1]．しかしながら，モバイル環境下で現在地周辺の情報を得るための検索対象としてWWW上の非構造化情報が用いられることは少ない．それは，従来の一般的なサーチエンジン等では，位置を検索条件とすることがうまく行えないからである．また，モバイル環境下で現在地周辺の情報を得た場合には，実際にどのレストランに行くかなどの意思決定を行うために，得られた情報どうしを様々に統合して判断するが，WWW上の非構造化情報に対してそのような作業を支援してくれるものも少ない．我々は位置に関連した情報を探しだし，それを位置という観点から統合することを位置指向の情報統合と呼んでいる．WWW上の非構造化情報をモバイル環境下で生かすには，非構造化情報に対する位置指向の情報統合の仕組みが必要である．

我々は，位置指向の情報統合を行うためにモーバイルインフォサーチ(**MIS**)という研究開発を行っている．**MIS**は，図1のように，モバイルユーザに現在地周辺に関する情報を提供することによって，実世界とネットワークの両方から情報を得ることができるようにし，さらに実世界とネットワークの間で相互に情報が行き交うことができるようにすることを目的とするプロジェクトでもある．これまでに，あらゆるWWW情報に対して，位置を検索条件とするための手法を検討し，そのプロトタイプを作成し，モーバイルインフォサーチ2実験(**MIS2**)[1]として公開し評価を行ってきた[2]．

MIS2では，WWW上の非構造化情報に対して，その中から位置情報を抽出しておくことで，任意の位置を検索条件として検索することができた．しかし，

[1] http://www.kokono.net/

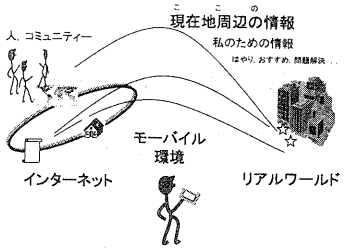

図 1: MIS構想

位置以外の情報を検索条件として指定したり，位置以外の観点から情報を分類したり，取捨選択したりすることはできなかった．位置指向の情報統合を行うためには，位置以外の観点を取り扱えるようにすることが必要である．また，WWW上の情報は膨大であり，また，利用者の状況によって欲しい情報は様々に異なるため，位置という情報だけからではWWW上の非構造化情報の利用はやや不十分である．位置以外の情報を取り出して検索条件として指定できるようにすること，すなわち，情報構造化が必要である．また，その構造化された情報を使って，WWW上の情報を分類し，取捨選択できるようにすることも必要である．

本稿では，位置に応じた情報を提供するための情報構造化や情報フィルタリングを位置指向の情報構造化・位置指向の情報フィルタリングと呼び，そのための手法を検討する．また，その手法のプロトタイプである，モーバイルインフォサーチ3実験(**MIS3**)について紹介する．

2 位置指向の情報構造化

位置指向の情報構造化では，位置指向のWebロボット[3]が収集したHTMLファイルに対して，後段の位置指向の情報検索や情報フィルタリングのために，位置をはじめとする様々な情報を抽出する(図2)．

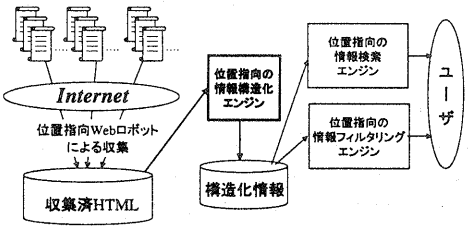

図 2: 位置指向の情報構造化の概略

現在地周辺の情報を提供するための情報構造化においては，店舗・施設・名所旧跡・特定場所でのイ

ベントや事件・地域特性といったものが構造化対象になる．それらに対して以下のような要素が抽出される必要がある．

- 対象の名称
- 対象の種類
- 対象の場所
- 対象にたどり着くための情報
- 対象の時間的な分布
- 対象のその他付帯的な情報

一方，非構造化情報，特に文書情報からの情報抽出については以下のような方法がある[2][4]．

a. 記号や文字のパターンマッチ

b. 特定の文字列や記号列を識別子とするパターンマッチ

c. 特定分野の辞書を併用した形態素解析

d. 表形式など，情報の記述形式が一定の部分に対する，出現位置によるパターンマッチ

抽出方法**c.**について，様々な分野について分野毎の辞書を揃えていけば，より多くの情報が抽出できると考えられるが，今回はまず，最も需要の多そうな飲食店に関する情報の詳しい構造化を狙って，店名と業種名の対応辞書と飲料・食品に関する辞書のみを用いることとした．

抽出すべき要素・実際に行える抽出方法の両面から，ここでは以下のような情報を抽出することとした．

- 対象の名称
  主に店舗名 (抽出方法 **c.** および **b.**)
- 対象の種類
  主に業種名 (抽出方法 **c.** および **b.**)
- 対象の場所
  住所，駅名，ランドマーク名
  (抽出方法 **c.** および **b.**)
  郵便番号 (抽出方法 **a.**)
- 対象にたどり着くための情報
  特に抽出しない
- 対象の時間的な分布
  営業時間・利用可能時間 (抽出方法 **a.**)
- 対象のその他付帯的な情報
  - 電話番号 (抽出方法 **a.**)
  - 座席数 (抽出方法 **a.**)
  - 価格 (抽出方法 **a.**)
  - 人名 (抽出方法 **c.**)
  - 提供メニュー (主に飲食品) (抽出方法 **c.**)
  - E-mailアドレス (抽出方法 **a.**)
  - 文書中の名詞・固有名詞
    (一般辞書のみの形態素解析による)

[2] 日々増大する大量の情報に対して構造化を行うには，ある程度高速な処理が必要であるため、今回は意味解析までは行わず，形態素解析程度までの範囲で構造化処理を行うこととした．

対象にたどり着くための情報としては，地図上の道路や鉄道路線などの経路というのもありえるが，それらは非構造化情報から抽出しなくても，対象の場所が抽出できれば、地図データや経路探索用のデータを用いて付与できるため，ここには含めていない．

後段の様々な情報検索・情報フィルタリングで柔軟に利用できるよう，抽出された情報をXMLの形式で構造化情報として出力する．以上の位置指向の構造化の流れを図3に示す．また，その具体例を図4に示す．

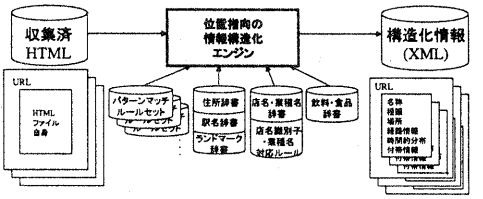

図 3: 位置指向の情報構造化の詳細

このようにして位置指向の構造化を施した情報に対して，構造化した要素をすべて自由に検索条件として指定できるようにした検索を我々は構造化検索と呼んでいるが、モバイル環境下においてはことこまかに検索条件を指定することは，端末の表示能力・インタフェース等の制約や検索のために費やせる時間的制約などから考えて，すべてのユーザが簡単に行えることではない．ことこまかに検索条件を指定することが困難な状況においては，我々は，検索結果を自動的に分類して提示することを考えている．現在，開発中の検索プロトタイプでは，検索結果は「業種」と「営業時間」で自動分類され，かつ各々の分類の中で現在地との距離が近いものから順に出力される．

3 位置指向の情報フィルタリング

膨大なWWW上の情報の中から現在地周辺の情報を提供しようとすると，そのままでは非常に膨大な情報量になる．例えば、東京都中央区銀座4丁目付近を検索すると，たった半径200メートル以内に200件以上の情報が存在する．2章の最後で述べたように，位置指向に構造化した情報に対して，様々な検索条件をユーザに付加してもらって検索を行ったり，検索結果を自動的に分類して提示するというのも情報を取捨選択する方法であるが、ここでは，サーバ側で何らかの基準で情報をフィルタリングする方法を考える．この方法はユーザの負担を軽減するだけでなく，モバイルユーザの限りあるネットワークリソースを節約できるという意味でも重要な仕組みである．

現在地周辺の情報を提供するための位置指向の情

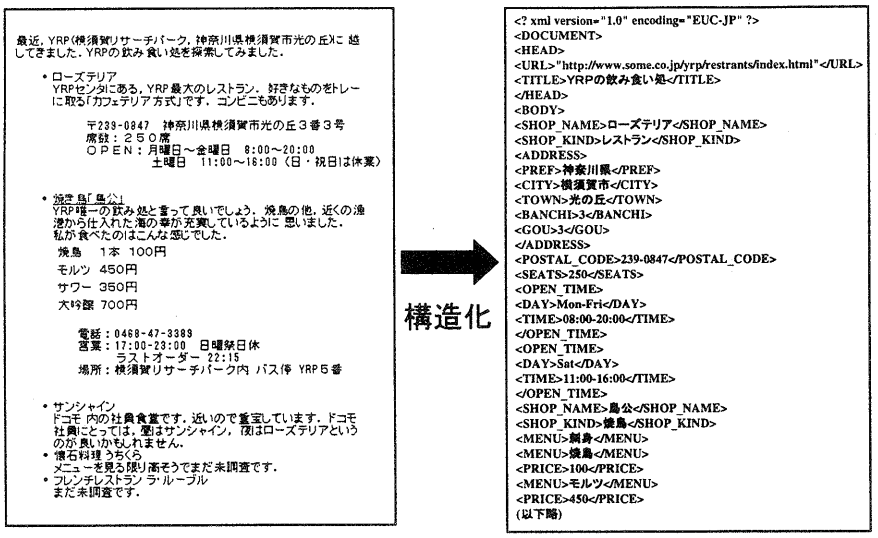

図 4: 位置指向の情報構造化の例

報フィルタリングを考えた場合、現在地周辺の特徴的な情報をフィルタリング結果として提供するのがごく自然な方法と考えられる．例えば、秋葉原といえば電気街，三浦半島の三崎港といえばマグロや寿司屋が真っ先に思いつく．実際，秋葉原には電気に関連する店舗や施設が多いし，三崎港には寿司屋が多い．様々な観点からのフィルタリングが存在するが，まずは現在地周辺で最も多い情報をフィルタリング結果として提供することを考える．

一方で，フィルタリングの方法のひとつとしてSocial Filtering[5] がある．これは，多くの人が勧める情報や多くの人が参照した情報を他の人にも有益な情報と考えて，それらを選び，他を捨てるというフィルタリングである．WWW上で情報を発信するということは，それなりに手間のかかることであり，WWW上に発信されている情報は，単に参照されたということよりも一般に，より強い動機，より強い興味が示された情報である．WWW上に多く発信されている情報を取り出すことは，Social Filteringの手法の一つと考えられる．

次に，WWW上に多く発信されている情報という場合の情報の単位を考える．2章で行った構造化された情報要素はこの単位になりうる．また，2章の方法では構造化しきれなかった，HTMLファイル中の単語(名詞・固有名詞)もそのような情報の単位として考えることにする．また，現在地周辺という場所の単位を各住所や各駅，各ランドマークといった単位で考え，この単位をここではエリアと呼ぶことにする．このような情報の単位のそれぞれについて各エリア毎に，WWW上で発信されている回数を例として図示したのが図5である．

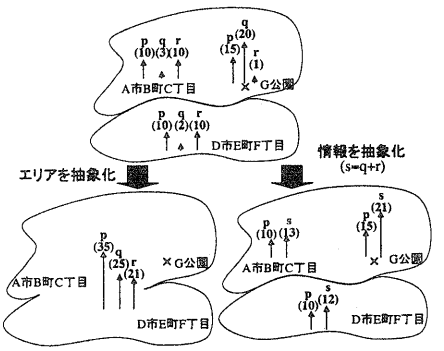

図 5: 位置指向の情報フィルタリングのイメージ

この例では，G公園において情報qが顕著であり，G公園だけに限って言えば、情報qを選択し，他の情報を捨てるというフィルタリングが考えられる．その他のエリアについては特に顕著な情報はないように思える．しかし，エリアを抽象化し，A市B町C丁目，D市E町F丁目，G公園をひとつのエリアとして考えた場合には，情報pが顕著である．また，情報qと情報rは類似の情報で情報sとして抽象化が可能であるとすれば、情報を抽象化した場合には，A市B町C丁目やD市E町F丁目においても情報sが優位であることになる．このように，位置指向の情報フィルタリングにあたっては，情報の単位，エリアの単位それぞれについて，最小単位での各情報の

優位性で判断するのみならず，情報の単位，エリアの単位それぞれについて抽象化を行った場合についても考える必要がある．

エリアの単位については隣接しているものどうしで抽象化していけば良い．ここでは，現在地を中心とするある半径の円に入っているエリアに対して各情報の優位性を判断することとし，半径を徐々に大きくしていって，優位な情報が現れたところで半径の拡大を止めて、その優位な情報を選択することにする．

一方，情報の単位の抽象化にあたっては，シソーラスを利用することとする．様々なシソーラスが考えられるが，ここでは，2章において特に特定分野の辞書を用いて構造化を行っている，業種名と飲食物についてシソーラスを用いた抽象化を行うこととする．

さらに，ある単位の情報が優位であるかどうかの判定基準として，$tfidf$ 値[6]を用いることとする．$tfidf$ 値は，文書に含まれる特徴的な単語を統計的に取り出すための値である．文書全体の集合を $\{d_k|1 \leq k \leq n\}$，ある文書 d_j における単語 w_i の出現回数を $wc(w_i, d_j)$，ある文書 d_j 中に単語 w_i が含まれるとき1，含まれないとき0を返す関数を $we(w_i, d_j)$ とすると，ある文書 d_j における単語 w_i の $tfidf_{w_i,d_j}$ は次のように定義される．

$$
\begin{aligned}
tf_{w_i,d_j} &= \frac{wc(w_i,d_j)}{\sum_{k=1}^{n} wc(w_k,d_j)} \\
df_{w_i} &= \frac{\sum_{k=1}^{n} we(w_i,d_k)}{n} \\
tfidf_{w_i,d_j} &= \frac{tf_{w_i,d_j}}{df_{w_i}}
\end{aligned}
$$

ある文書におけるある単語の $tfidf$ 値は，その単語が，文書集合全体としてはあまり出現しないが，その文書において多く出現する場合に，高くなる．

位置指向の情報フィルタリングの場合には，情報を単語，エリアを文書と考えて $tfidf$ 値を求め、あるエリアにおいて $tfidf$ 値が高い情報をそのエリアにおいて優位な情報であると判断することにする．エリア全体の集合を $\{a_k|1 \leq k \leq n\}$，あるエリア a_j における情報 w_i の出現回数を $wc(w_i, a_j)$，あるエリア a_j 中に単語 w_i が含まれるとき1，含まれないとき0を返す関数を $we(w_i, a_j)$ とすると，あるエリア a_j における単語 w_i の $tfidf_{w_i,a_j}$ は次のように定義される．

$$
\begin{aligned}
tf_{w_i,a_j} &= \frac{wc(w_i,a_j)}{\sum_{k=1}^{n} wc(w_k,a_j)} \\
df_{w_i} &= \frac{\sum_{k=1}^{n} we(w_i,a_k)}{n} \\
tfidf_{w_i,a_j} &= \frac{tf_{w_i,a_j}}{df_{w_i}}
\end{aligned}
$$

このような $tfidf$ 値が一定の値を越えた場合に，優位な情報が見つかったと判断することにする．以上をまとめたのが図6である．

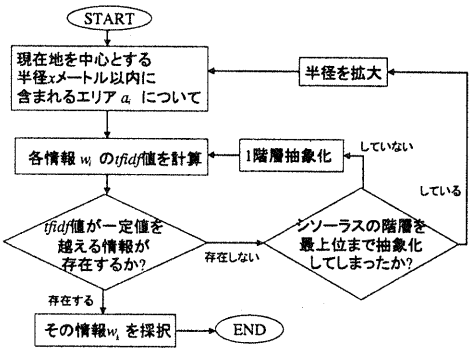

図 5: 位置指向の情報フィルタリングのフローチャート

4 モーバイルインフォサーチ3実験

モーバイルインフォサーチ実験は，検討した位置指向の情報統合手法のプロトタイプの評価の場で，http://www.kokono.net/ にて1997年9月より公開実験を行っている．本稿で検討した内容を盛り込んだプロトタイプを1999年内にモーバイルインフォサーチ3実験(**MIS3**)として提供予定である．本章では，**MIS3** について概略を紹介する．

2章で検討した位置指向の情報構造化を用いた，構造化検索と位置指向の情報分類は，「ここのサーチ」という機能として提供する．これは指定された現在地周辺の情報をユーザからの付加検索条件で検索を行い，検索結果を業種や営業時間といった観点から自動分類し，検索結果を提示する．

また，3章で検討した位置指向の情報フィルタリングのプロトタイプを「ここのおすすめ」という機能として提供する．これは現在地周辺について言及している情報の中から頻出する単語や業種等を抽出し，その単語等について言及しているURLへのアンカーとともに情報を提示する．ユーザは，現在地周辺では提示された単語や業種等がこのエリアのおすすめ情報であることを知り，さらに詳しい情報を知りたいときにアンカーをたどってWWW上で公開されて

いるHTMLファイルを参照する．

MIS2まででは，現在地を住所リスト等から手動で指定する以外に，位置情報対応PHSとMS-Windowsマシンあるいは MS-WindowsCEマシン，パソコン用のGPSとMS-Windowsマシンの組合わせで，現在地を自動的に**MIS**サーバに伝達し，現在地周辺の情報を検索することができた．**MIS3**からは他の組合わせについても順次対応していく予定である．そのひとつとして，図7にGPSアンテナ付PDAでの利用イメージを示す[3]．PDAに内蔵しているGPSアンテナと携帯電話の先のネットワーク上にあるGPS測位サーバとが連動して測位した現在地の緯度経度を**MIS3**のサーバにURLとして伝達し，図7のような画面イメージを表示する．

画面イメージは上から，位置指向の情報フィルタリングである「ここのおすすめ」の結果，位置指向の検索と情報分類である「ここのサーチ」へのボタン，さらに各コンテンツプロバイダが提供している駅の時刻表やホテル情報，天気予報，地図，鉄道の経路探索，お店等のタウン情報へ接続するためのボタンである．ユーザは，ここのおすすめを参考に情報を参照したり，ここのサーチなどのボタンを押して求める情報を参照する．

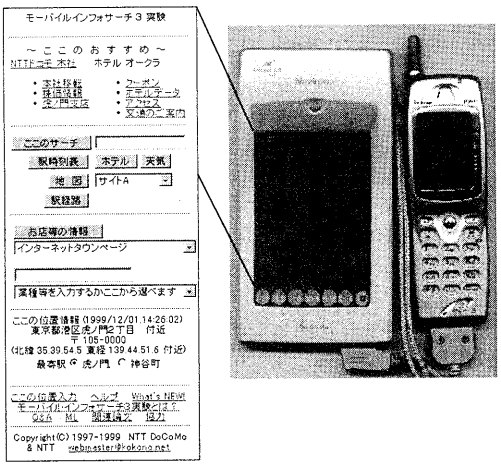

図 7: **MIS3**の利用イメージ

5　おわりに

本稿では，WWW上に存在する個々のHTMLファイルのような非構造化情報から現在地周辺の情報を検索するための位置指向の情報統合のうち，特に，位置指向の情報構造化と位置指向の情報フィルタリングについてその手法を検討した．また，それらの手法のプロトタイプである，モーバイルインフォサーチ3実験(**MIS3**)について紹介した．

MISはモバイルユーザへの現在地周辺の情報を提供する基盤としてだけでなく、地域情報を流通する基盤としてもとらえることができる．WWWがいわば，全国版(世界版)新聞であったのに対し，昨今は，地域版新聞に相当するような地域の情報を発信する場として広がりつつある．様々な地域の情報が個々に発信される中で，そのような情報をどのように統合し，流通させるかが鍵となってきている．2章で検討したようなXML形式の構造化情報はそのひとつの解になりうると考えている．

また，iモードのようなマイクロブラウザが普及してきている．3章で検討したような情報フィルタリングをより厳選した情報を選びだすような方向に発展させていく必要があり，検討中である．

今後は，**MIS**をあらゆるモバイルユーザ，さらに，地域情報を求めるあらゆるユーザへの情報提供基盤とすべく，研究開発を進めていく．

[3] NTTドコモから発売予定のGPSアンテナ内蔵PDA「Naviewn」での動作イメージ．
協力：ドコモ MCビジネス部 技術開発第三担当

謝辞

日頃，モーバイルインフォサーチ実験にアクセス頂いている皆様に感謝致します．

また，モーバイルインフォサーチ実験にコンテンツプロバイダとしてご協力下さっている各社の皆様に感謝致します．

東芝情報システムの増田康明氏，ニチメンデータシステムの筒井俊晴氏，NTTソフトウェアの金山博明氏に対し，日頃の研究開発業務支援に感謝致します．

参考文献

[1] Katsumi Takahashi, S. Yokoji, and N. Miura. "Location Oriented Integration of Internet – Mobile Info Search –". In *Designing the Digital City*. Springer-Verlag, 2000.

[2] 三浦信幸, 横路誠司, 高橋克巳, 島健一. "位置指向の情報統合 ～モーバイルインフォサーチ2実験～". 情報処理学会 第57回 全国大会, Vol. 3, pp. 637–638, Oct. 1998.

[3] 横路誠司, 三浦信幸, 高橋克巳, 島健一. "特定分野のリソース収集を行うWWWロボットの性能評価". 情報処理学会 第57回 全国大会, Vol. 3, pp. 163–164, Oct. 1998.

[4] 井上香織, 横路誠司, 高橋克巳. "広告の自動構造化". 第132回 自然言語処理研究会. 情報処理学会, Jul. 1999.

[5] W. Hill, L. Stead, M. Rosenstein, and G. Furnas. "Recommending and Evaluating choices in a virtual community of use". In *CHI '95*, pp. 194–201. ACM, 1995.

[6] G. Salton and M. McGill. *"Introduction to Modern Information Retrieval"*. McGraw-Hill, 1983.