

携帯情報端末における XML 代理サーバによる 属性情報管理機構の提案

新井イスマイル、和泉順子、中村豊、藤川和利、砂原秀樹

奈良先端科学技術大学院大学 情報科学研究科

A proposal of the attribute information management system using XML proxy server for mobile computers

Ismail Arai, Michiko Izumi, Yutaka Nakamura, Kazutoshi Fujikawa, Hideki Sunahara

Graduate School of Information Science, Nara Institute of Science and Technology

1 はじめに

携帯情報端末に無線 LAN、携帯電話、PHS (Personal Handyphone System) などの無線技術を用いる事でインターネット接続が容易になった。これにより、場所に依存することなくメールの送受信や WWW (World Wide Web) の利用が可能となりつつある。

現在多くの人々は多岐にわたる用途で WWW を利用した情報検索を行う。主に全文検索である。しかし、携帯情報端末を用いる場合、情報を表示する領域が狭く、多くの WWW コンテンツが一般のコンピュータに比べ満足に表示できない。またキーボードを持たないものが多いので速やかに文字を入力できないといったユーザインターフェイスの制限がある。

キーワードの入力を自動化する手法として文書の意味を示す属性情報を利用した概念検索がある。META タグを利用した HTML 文書、および XML (eXtensible Markup Language) は属性情報の取り扱いを考慮した設計がされている。ユーザおよび検索対象に時間情報、位置情報、嗜好情報といった属性情報を含ませると、お互いの属性情報が合致するコンテンツを抽出でき、ユーザに適した情報を提供する事が考えられる。本研究では 2 節で解説する将来性を考慮して属性情報を利用する文書の一つの XML 文書を利用する。ユーザの属性情報があらかじめ決められていて検索のキーとして利用すると検索時の入力文字数の減少が期待できる。しかし携帯情報端末の通信手段である無線通信は有線通信に比べて通信帯域が狭いため、1 回のリクエスト毎に大量の属性情報を送受信することは携帯情報端末で XML 検索することにおいて利便性を損なう。

本研究では携帯情報端末のユーザインターフェイスまたは通信速度の問題点を解決するために、XML を携帯情報端末向けにデータを縮退した HTML へ変換する XML 代理サーバを構築する手法を提案する。本手法を用いることで、XML 代理サーバが XML 検索によって得た情報を携帯情報端末の表示能力に適した HTML データに変換することが可能となる。また、XML 検索によって得た属性情報を多く含む XML コンテンツを、ユーザが得たい文字情報だけに縮退する事によって携帯情報端末と XML 代理サーバ間の通信容量を削減できる。

2 節では本研究で必要となる関連技術について解説し、3 節

では本論文における提案事項の設計について述べる。4 節では実際に提案した内容のプロトタイプ実装について述べ、システムの評価項目を設定する。

2 関連技術

2.1 XML における属性情報

WWW には現在数えきれないほどの情報が蓄積されている。そのたくさんの情報の中から有益な情報を得るには全文型検索エンジンなどでキーワードにより検索対象を絞り込む手法が現在主流となっている。しかし画像イメージやマルチメディアコンテンツなど、キーワードの組み合わせでは探すことのできない情報も WWW 上には多数存在する。キーワードの組み合わせによって検索できない情報を検索可能にするには、文書等の意味を示す属性情報を文書に付加し、WWW 文書に含まれる文字情報の全文検索と、直交した次元において属性情報をキーワードとして検索する方法が考えられる。

HTML 文書に付加させる属性情報に META タグがある [1]。META タグの情報は、ブラウザで見る実際の WWW ページには反映されず、検索エンジンがインターネット上の情報を収集する際に利用している検索ロボットにその WWW ページの情報として伝えたい情報を特定化することができる。一方、XML では属性情報のデータ形式が定義されている [2]。XML 文書の DTD (Document Type Definition) で記述した論理構造 (エlement) や Element の属性をキーにして、XML 文書を全文検索する XML 構造指定検索を用いると、XML のタグ構造を利用した全文検索や概念検索が可能となる。META タグを利用した HTML 文書と XML 文書の双方は概念検索に一般的に利用されている。META タグは属性情報の名前と値が常に一対となる仕様で、それに対し XML 文書における属性情報は構造化により一つの属性名に複数の属性情報を持たせる事ができ、属性情報を再利用できる面 [3] で将来性がある点を考慮して、本論文では属性情報を XML 文書で記述する事とする。XML において属性情報は要素の開始タグ内で定義する。図 1 にユーザの性別、年齢、所在地を示した例、図 2 に店舗の主な対象性別・

```
<?xml version="1.0"?>
<person sex="male"/>
<person age="23"/>
<person position-ns="n042018015" position-ew="135023045"/>
<shop address="Takayama-Cho XX-00, Ikoma-City, Nara"/>
```

図 1: ユーザ属性情報の XML 記述例

```
<?xml version="1.0"?>
<shop sex="male female"/>
<shop age-min="20" age-max="35"/>
<shop position-ns="n042020044" position-ew="135155034"/>
<shop address="Mayumi-cho oo-xx, Ikoma-City, Nara"/>
<book>
  Welcome to Restaurant OOX !!
</book>
```

図 2: 店舗属性情報の XML 記述例

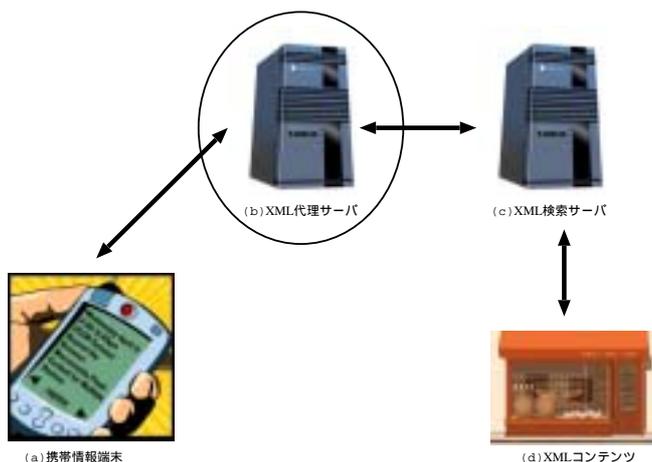


図 3: サービス全体図

対象年齢・所在地を記述した例を示す。

ユーザおよび検索対象に対応づけられた属性情報を含めると、検索時にお互いの属性情報が合致する XML コンテンツを抽出でき、ユーザに適した情報を提供する事が考えられる。しかし携帯情報端末の無線通信環境は通信帯域が限られており、XML 構造指定検索の度にユーザの大量の属性情報を送信することはユーザにとって望ましくない。

```
GET /cgi-bin/echoserver HTTP/1.0
Referer: http://scout.aist-nara.ac.jp/~index.html
Connection: Keep-Alive
User-Agent: Mozilla/4.7[en](isumai-a@is.aist-nara.ac.jp)
Host: scout.aist-nara.ac.jp:80
Accept: image/gif, image/x-bitmap, image/jpeg, image/png. */*
Accept-Encoding: gzip
Accept-Language: ja, en, en-US, en-GB, zh, zh-TW, zn-CN, ko
Accept-Charset: iso-8859-1, *, utf-8
```

図 4: HTTP ヘッダ情報

2.2 携帯情報端末向けプロキシサーバ

既存の研究にもユーザが受信する画像容量を小さくするモバイルプロキシサーバがある [4]。WWW には GIF,PNG,JPG 形式などの画像が利用される。モバイル通信においては画像の質よりもデータの容量を削減する事が優先されるため、GIF,PNG のような可逆変換可能な画像データは不可逆な JPG 形式のフォーマットに変換することによりデータの容量を削減できる。また XML 文書に含まれる属性情報には概念検索の為に必要なものとブラウザの表示に必要なものがあるが、これらが無作為に META タグに変換してしまうと、ブラウザに表示したい属性情報が表示されなくなってしまう。また概念検索に必要な属性情報はユーザが受信する際には冗長な情報となる。画像情報のデータ容量の縮退に加え、概念検索キーにのみ利用される属性情報も縮退する機能を付加したプロキシサーバを設計する必要がある。

2.3 既存技術の問題点

WWW にある数えきれない情報の中から有益な情報を得るには XML 構造指定検索による概念検索が有効である。携帯情報端末に XML を実装することによって、それは可能となる。しかし携帯情報端末には通信帯域およびユーザインターフェイスの問題点がある。XML を携帯情報端末上に実装すると全てのユーザおよび検索対象の属性情報を送受信し、画像容量も縮退されないため、通信帯域およびユーザインターフェイスの両方の問題が発生する。これらの 2 つの問題点を同時に解決するには、モバイルプロキシサーバにの機能に加えて、携帯情報端末

が不必要な属性情報を送受信する必要がなくなるような XML 代理サーバを設計する必要がある。

3 XML 代理サーバの設計

3.1 XML 代理サーバの概要

携帯情報端末上で XML 通信を行うには、通信帯域およびユーザインターフェイスに問題があり、通信帯域の問題点を解決するには、以下の機能が必要となる。

- ユーザが XML 通信の際に送信する必要のあるユーザの属性情報を XML 代理サーバにて管理し、XML 代理サーバが XML 検索サーバと属性情報をやり取りする。
- 検索に該当した情報に画像データが含まれていた時、そのデータ形式が可逆の圧縮形態の場合、不可逆な圧縮方式に変換する。
- HTML 変換を行う際にユーザにとって不要となる属性情報は削除する。

ユーザインターフェイスの問題点を解決するには、以下の対応が考えられる。

- XML 代理サーバに検索を依頼する際に、なるべくクリック数を減らす。
- 携帯情報端末の画面の大きさをはみ出ると思われる大きさの画像は縮小する。

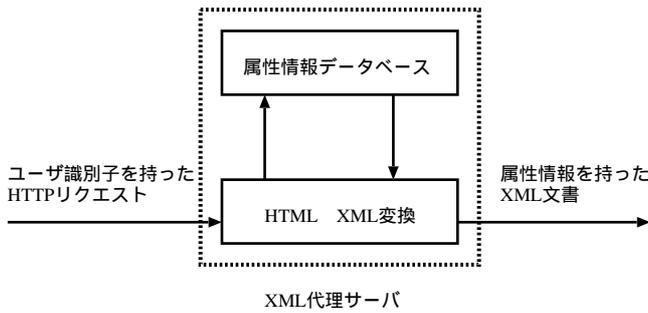


図 5: XML 代理サーバ

```
<?xml version="1.0"?>
<result id="01">
  <shop sex="male female"/>
  <shop age-min="20" age-max="35"/>
  <shop position-ns="n042020044" position-ew="135155034"/>
  <shop address="Mayumi-cho oo-xx, Ikoma-City, Nara"/>
  <Welcome to Restaurant OOX !!>
</book>
</result>

<result id="02">
  <shop sex="male female"/>
  <shop age-min="any" age-max="any"/>
  <shop position-ns="n042055032" position-ew="135100001"/>
  <shop address="Tomigaoka o-xx-oo, Ikoma-City, Nara"/>
  <book>
    Family Restaurant Cooking-Papa
  </book>
  <menu>
    <order>
      <name>Launch Set</name>
      <price>1,000</price>
    </order>
  </menu>
</result>
```

図 6: XML 検索結果

3.2 設計内容の詳細

XML 代理サーバを利用した XML 情報検索システムを図 3 に示す。システム構成要素としては、(a) 既存と仮定する携帯情報端末、(b) 提案する XML 代理サーバ、(c) 既存と仮定する XML の検索サーバ、および (d) 既存と仮定する XML コンテンツを蓄積するサーバが挙げられる。(a) の携帯情報端末は HTML のみを取り扱うことができるとする。以下動作の流れに沿って解説する。

1. ユーザはあらかじめユーザ登録をしている XML 代理サーバに、WWW コンテンツの提供を求める HTTP リクエストを送信する。
2. ユーザは HTTP 環境変数の User-Agent というフィールドの、'(' と ') ' で囲まれたコメント部分に、ユーザのメールアドレスなどのユーザを特定できる情報を埋め込む [5]。HTTP 環境変数にユーザの識別子としてメールアドレスを挿入した例を図 4 に示す。
3. XML 代理サーバは受け取った HTTP リクエストの中からユーザを特定できる情報を抽出する。
4. そのユーザを特定できる情報を元に XML 代理サーバはそれに該当するユーザの属性情報を参照する (図 5)。ユーザ

```
<HTML>
<BODY>
<UL>
  <LI>Address: Mayumi-Cho oo-xx, Ikoma-City, Nara
  <LI>Welcome to Restaurant OOX !!
</UL>
<UL>
  <LI>Address: Tomigaoka o-xx-oo, Ikoma-City, Nara
  <LI>Family Restaurant Cooking-Papa
  <LI>Menu
    <TABLE><TR>
      <TD>Launch Set</TD>
      <TD>1,000yen</TD>
    </TR></TABLE>
  </LI>
</UL>
</BODY>
</HTML>
```

図 7: XML 代理サーバが整形した HTML 文書

の属性情報は、あらかじめユーザによって XML 代理サーバに登録されていて、属性情報の変更についてもユーザのリクエストによって可能となっている。

5. XML 代理サーバは参照した属性情報とユーザの HTTP リクエストに含まれていた検索対象を組み合わせ、XML 文書を作成する。図 1 で示したような属性情報をユーザ isumai-a@is.aist-nara.ac.jp が持っていた場合、そのユーザが発する HTTP リクエストは図 4 のようになり、整形される XML 文書は図 1 のようになる。
6. そしてユーザの属性情報と一致する情報提供側のコンテンツを XML 検索サーバに問い合わせ、図 6 のような XML 形式の検索結果を受け取る。
7. この XML 形式の検索結果の XML 特有のタグを HTML のタグに変換し、さらに属性情報のうち、ユーザにとって不要となる属性情報を含む HTML タグを削除し HTML データを作成する。不要な属性情報は個々のユーザによって要求が異なる。例として、XML 代理サーバが図 6 のような検索結果を得た時に、ユーザが対象性別、対象年齢、経度、緯度を必要としない時、ユーザが得る HTML 文書は図 7 のようになる。ユーザが HTML 文書に必要としない属性情報はユーザのリクエストによりあらかじめ登録されている。
8. 整形された HTML 文書をユーザに送信する。

4 プロトタイプ実装・評価

4.1 XML 検索システムのプロトタイプ実装

本提案のプロトタイプとなる実装の環境として、クライアントマシンに SHARP 製の携帯情報端末 SL-C700、XML 代理サーバマシンに FreeBSD Release4.7-p10 を OS とする Pentium 1GHz の PC/AT 互換機、仮想的な XML 検索サーバおよび XML コンテンツ蓄積サーバにも XML 代理サーバと同様の計算機を用意する。通信路は想定される環境を実現するために dummy net を利用して帯域を制限する。ユーザと XML 代理サーバの間は



図 8: 実験環境

32kbps に、XML 代理サーバと XML 検索サーバの間は 10Mbps に帯域を制限する。実験環境を図 8 に示す。

4.2 評価項目

提案の有効性を確認するために、以下の評価項目を設定する。

1. 通信帯域の問題点に注目した評価

- ユーザが直接 XML 検索サーバに属性情報を送信した場合と、XML 代理サーバが送信した場合とのユーザのデータ送信量の比較
- ユーザが XML 文書を直接受信した場合と XML 代理サーバが生成した HTML 文書を受信する場合のユーザのデータ受信量の比較
- XML 代理サーバが取得した画像データと、その画像データを不可逆変換した場合のデータ量の比較

2. ユーザインターフェイスの問題点に注目した評価

- ユーザが XML 代理サーバを利用する場合としない場合のキー入力数の比較
- ユーザが得た情報を閲覧する際に画面スクロールをした回数

通信帯域の問題点に着目した評価は実際に送受信されているデータ量をモニタリングして保存し、総容量を測定する。ユーザインターフェイスに着目した評価は大量の実験参加者にシステムを利用してもらいアンケート形式により集計する。

5 今後の課題

今回のプロトタイプ実装ではユーザの属性情報の管理について詳細に設計していない。ユーザの属性情報の修正、追加、削

除といった変更方法を具体的に提案する必要がある。XML 代理サーバのユーザの判別においてはユーザが埋め込み、送信する、ユーザ判別情報を利用しているため、認証機構の設計も課題となる。さらに XML 代理サーバの過負荷によるサービス停止の対策として負荷分散を考えた場合に、ユーザの属性情報管理は今回の設計では実現し得ないため、設計の見直しが必要となる。

また本論文の扱う属性情報のスケーラビリティを考慮した場合に、属性情報は広く共通のフォーマットを利用することが望まれるので、Dublin Core Metadata [6] のメタデータの標準化についても調査し、適用できるか否かを調査する必要がある。

6 まとめ

本論文では携帯情報端末でインターネットを利用した情報検索における問題点を通信速度およびユーザインターフェイスの 2 点に整理した。それらの問題点を解決するためにユーザが XML 代理サーバを利用して XML 構造指定検索を行う事を提案した。提案を実現するための手法を整理し、詳細な設計を行った。そして、設計の有効性を示す評価指標を、整理した 2 つの問題点に着目して提案した。今後プロトタイプの評価により抽出された問題点を整理し、システムの再考と評価を行う。また、これに伴いユーザの属性情報の変更方法、ユーザの認証機構、XML 代理サーバの負荷分散およびスケーラビリティについて検討する必要がある。

参考文献

- [1] 杉本重雄. Dublin core metadata set に関して. 情報処理学会デジタル・ドキュメント研究報告, Vol. 18, No. 2, pp. 9-16, May 27 1999.
- [2] Erik T. Ray. 入門 XML, 第 3 章. オライリー・ジャパン, September 28 2001.
- [3] OHYA Kazushi. Necessities on a descriptive level for reusing metadata descriptions. *Proceedings of the International Conference on Dublin Core and Metadata Applications 2001*, pp. 97-100, 2001.
- [4] 辻順一郎, 小津浩直, 三浦敦史, 滝沢直喜, 水野忠則. モバイルプロキシサーバシステムの試作. 情報処理学会モバイルコンピューティング研究報告, Vol. 2, No. 12, pp. 67-72, July 25 2001.
- [5] 松村洋一, 山田宏之. CC/PP を利用した多種端末向け web アプリケーションの検討. 情報処理学会ヒューマンインタフェース研究報告, Vol. 86, No. 1, pp. 1-6, November 25 1999.
- [6] Dublin core metadata initiative. <http://www.dublincore.org>.