

観光案内のためのコンテンツサーバ構成法の検討

嶋 俊光 永瀬 宏

金沢工業大学 工学部 情報工学科
〒921 石川県金沢南局区内野々市町扇が丘 7-1

観光情報をインターネット上で提供するために情報の発信源からコンテンツを収集して、必要な情報の提示並びに検索支援を行う方法を検討した。コンテンツの収集ではHTML文書に新規タグを導入して、必要な情報をコンテンツサーバが抽出する方法を提案すると共に意味の分野依存性等の問題を議論する。次に収集したコンテンツのユーザの視点に基づく検索が有効であることを述べ、その場合の検索画面の構成、ならびに検索画面の総数を低減する方法を論ずる。

Design of Contents Server for Tourist Guide

Toshimitsu SHIMA Hiroshi NAGASE

Department of Information and Computer Engineering
Kanazawa Institute of Technology
7-1, Ougigaoka, Nonoichi-Chou, Ishikawa-south area, Ishikawa, 921, Japan

In order to provide tourist information through the Internet, a method is proposed to collect contents from information resources, and to support users for searching appropriate contents. For contents collection, several tags are newly introduced to HTML texts. Then, some semantic problems are discussed arising from domain oriented terms and so on. Next, it is shown that information search based on user's view is effective to accommodate user's motivation. With this respect, it is discussed how to design windows to search arbitrary contents and to decrease the number of those windows.

1. はじめに

観光産業はしばしば地域の基幹産業として重要な位置付けにあり、観光情報をいかに適切に市民に供給するかが検討されつつある。その観光情報を提供するシステムがセンタ中心で運営される場合、システム管理者は情報を提供するひとの要望にこたえきれず情報を更新できないことが問題点として指摘されている。⁽¹⁾ その結果、情報が陳腐化し利用されないことになる。一方、情報の提供が分散して行われているインターネットについて考えると情報の管理者が情報の提供者に近く、その情報量もあまり多くないため常に新しい情報が提供できることが期待できる。反面、情報発信はそれぞれの個人、組織などがバラバラに行なっているので実際に情報を得る為に情報の中をいろいろとみて回らなければならない。

そこでナビゲーションの支援が必要となる。ナビゲーション支援の1つの方向はインターネット上で提供されている広範囲の内容を分類して、適切なWebページにユーザを導くものであり、Yahoo⁽²⁾等が知られている。もう1つの方向として、地方公共団体或いは業種毎にホームページを作るといふ動きがある。^{(3),(4)} これは分類だけでなく、コンテンツも提示する機能をもつ。現段階では独立したWebページを作成しているケースが多いが、将来的には個々の情報提供者のWebページをもとにコンテンツを抽出して、新たにWebページを開くというスタイルが望ましいと思われる。

このような背景から本研究では、個々に発信された情報をまとめ、要約したコンテンツを用いたシステムについて検討を行なった。この場合、コンテンツとページの内容を同時更新するようなリアルタイム性の情報収集機構の実現、ユーザへのコンテンツ提示方法、データベースへの格納方法などの問題がある。本稿ではコンテンツサーバを用いたシステムについてこれらの問題を検討したので報告する。

2. 本研究の狙い

2.1 インターネットによる情報提供の現状

電子化文書を記述する方法にはSGML等が知られているが、検索の際インターネットではWWWで使われているHTMLが普及している。そこでHTMLを基礎とした情報提供について本研究では考える。WWWの特徴は次に挙げる。

- ① 情報を保有している人が情報を発信する。
- ② 情報の形態が多様である。(文字、画像)

- ③ リンクを辿って別の情報にアクセスできる。
(時間的余裕がある場合し楽しみながら検索できる)

この問題点として次のものが挙げられる。

- ① 情報と情報のリンクが情報提供者側の意図に依存する。
- ② どこに目的の情報があるか解らない。(どこにリンクを張っていいか解らない)
- ③ 目的意識を持って情報を探してもその情報がすぐには得られない。(情報が整理されているとは限らない)

ところで現在、様々なナビゲーションツールが存在している。Archieはファイル検索に用いるツールである。これによってFTPサイトにあるファイルのアドレスを探することができる。メニュー形式であるGopherはVeronicaを使うことにより、検索のための環境を作り上げている。Archie, Gopher, Veronicaの他にWAIS等の検索ツールを目的とする情報を得るために組み合わせることで有効に検索できる。これらのナビゲーションツールによってユーザは目的とするファイルを得られる。従ってこれらをWWWサイトについて拡張すれば問題点の①、②は簡単ではないがある程度解決できる。しかし実際目的とするファイルの場所が解ってもそれらを取り寄せるためにFTPやHTTPの操作は一般の人には必ずしも容易であるとは限らない。また別の問題としては実際欲しい情報があるとは限らないことである。これは発信する情報に関するガイドラインが存在しないので情報が欠落しているかもしれないからである。

2.2 要求条件

これまでに述べたインターネットに関する種々の問題を解決するために必要と思われる課題を整理すると以下ようになる。

[情報の提供者側]

- (1) すでに様々な形で存在するデータベースの流用。
- (2) HTMLによるリンクの張り方に対する支援。
- (3) 情報の質(内容の充足性、最新性)の保証。

[情報を探す側]

- (1) データベース/HTMLを意識せずに情報への統一的なアクセス手段で提供して欲しい。
- (2) 情報が存在するアドレスを意識せず具体的な情報を直ちに入手する場合と、アドレスを意識したより詳細な情報を入手する場合の2通りのケースに対応して欲しい。

- (3) 利用者は利用者自身の視点に合った検索をしたい。

[システムを構築する側]

- (1) 現在普及しているHTTP/FTP等の標準的のプロトコルを流用してシステムが普及しやすい形態で構築する。
- (2) 知識の集積として現在普及している関係データベースやオブジェクト指向データベースを使用する。

2. 3 提案するシステム構成

ここではこれらの課題を解決すべく図1のようなシステム構成を提案する。各ブロックの機能は次の通りである。

(1) コンテンツサーバ

- ① 利用者が観光案内システムに最初に入るホームページである。
- ② 利用者の意図を対話により理解する。
- ③ 最適な情報の入っているサイトの選択、或はコンテンツサーバ自身による情報の提供。
- ④ サーバは目的に応じて複数設置する場合もある。

(2) サイトtype A (情報提供者)

- ① これは現存するWWWであり利用者はネットスケープなどのビューワによってみることができる。
- ② テンプレートの内容をコンテンツサーバに通知する。

(3) サイトtype B (情報提供者)

- ① データベースの内容をHTML文書に変換してWWWのデータのように見えるようにする。すなわちHTTPプロトコルのメッセージをサイト内のデータベースクエリに変換する。インターフェースについては通常のHTTPプロトコルに従う。
- ② テンプレートの内容をコンテンツサーバに通知する。

コンテンツサーバは情報の内容から検索を可能とするためにテンプレートの内容をもとにして構成されたデータベースシステムである。基本的な動作手順は次の通りである。

ステップ1

利用者の意図を理解する。

ステップ2

既にコンテンツサーバにデータが蓄積されていればそれを表示する。

ステップ3

コンテンツサーバに必要なデータがなく、データがあるサイトのアドレスが既知のとき

は、そのデータベースにメッセージを送る。(又はHTML文書にリンクする)

ステップ4

サーバに必要なデータがなく、データがあるサイトも確定できないときは可能性のある複数のサイトに放送形式で問い合わせる。

この中で最も重要かつ困難な問題となるのはステップ1の利用者の意図を理解することであり、本報告では主にこの点を中心に議論する。

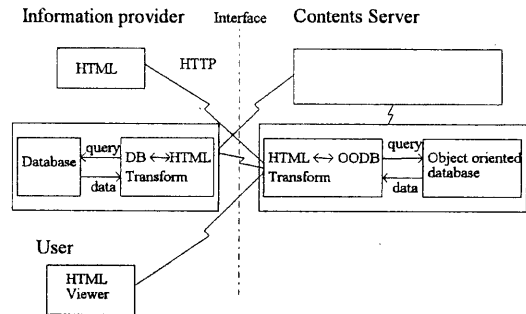


図1. 観光情報システムの構成

3. 情報の自動収集

3. 1 情報の分類

一般に存在する図書情報検索の方法は、まず本のタイトル、本の著者による検索がある。また他には記事によって分類された多数の図書から探している図書を見つける場合、本で使用されている単語を特徴付けることによって行なっている。これらは言わば大規模な知識ベースから所望の情報をいかに効率的に探索するかに関心が寄せられている。これに対して、観光情報システムでは、Webページを公開する対象がホテルや公園等と限定されており、始めから情報が分類されラベル付けされていると考えられる。これによって個々の詳細情報を抽出する時、何が欲しいかはっきりさせている。

3. 2 情報の収集方法

コンテンツサーバに情報を収集する方法には利用者に直接キーワードとなる単語をコンテンツ側が用意した、情報を収集するためのフィールドに入力してもらいそれをコンテンツサーバ側に送る方法がある。また情報提供者が個々に持っている実際の情報の上に中心的な用語にタグを使って囲ったものをコンテンツ側に渡し、そのタグで囲んだ用語をコンテンツ側が解析して情報を収集するものが考えられる。まず

フィールドを用いた情報収集の場合の長所は、質問の標準化が可能であり、問い合わせしやすい。しかし短所として情報提供者に実際のページによる情報とフィールドにキーワードとなる情報を提供させるので負担がかかる。また情報を提供する分野が多く存在するので、フィールド項目として何があればよいのかは明らかでない。一方、タグを用いた情報収集の長所は情報提供者が、用意された任意のタグを使ってキーワードとなる情報を選ぶことができる。短所として意味的なタグの組織立てをHTMLにおいて強要される必要がある。本稿ではタグを用いた情報収集方法について考える。

3.3 タグ

現在のHTML文書は文書構造、書式制御などのためにタグを用いている。タグは予約語<>の中に<BODY>...</BODY>のように記述し、タグ中に属性値を含む場合は<FORM METHOD=***>のように記述する。これに対して、情報収集ではWebページに含まれる意味内容を標準化して規定するものであり、表1にホテルページに関するタグの例を示す。

表1. タグの例

タグ	意味	例
<<NAME= >>	ホテル名	<<NAME=金沢ホテル>>
<<ADR >>	住所	<<ADR (〒)=921 (県)=石川県 (市)=金沢市 (町)=金沢町 (番地)=1-1 >>
<<TEL= >>	電話番号	<<TEL=0762-12-3456>>
<<FAX= >>	FAX番号	<<FAX=0762-12-3457>>
<<TRFC >>	交通情報	<<TRFC <<(バス)=金沢駅から 金沢町10分 (徒歩)=金沢町から5分>> <<(タクシー)=金沢駅から10分>> >>
<<PARK >>	駐車場	<<PARK (有無)=有り (台数)=200台 (期間)=1日 (料金)=1000円>>
<<ROOM >>	部屋	<<ROOM (シングル料金)=7000円 (ツイン料金)=10000円 (シングル数)=50室 (ダブル数)=40室 >>
<<CHK >>	チェックイン時間 チェックアウト時間	<<CHK (IN)=15:30 (OUT)=10:30>>
<<PAY >>	支払い	<<PAY (時間)=チェックアウト時 (クレジット)=使用可 (カード種類)=VISA >>
<<RSV= >>	予約方法	<<RSV=電話>>

以下では、タグに対する構文並びにセマンティクス上の問題点を論ずる。

(1) 構文

タグの構文には次のようになっている。

<<NAME=Kanazawa>>

<<ADR ***...>>

タグの構文の問題として挙げられるのは単語で

はなく文章等の長い部分をタグで囲った時どうやってその中の必要な部分だけを抽出できるかである。表1の例では、駅前から現地までの交通手段を提示するのに、乗継ぎや別の交通手段があることをタグの構文から判定できる必要がある。具体的には

<<TRFC <<(バス)=*** (徒歩)=***>>

<<(タクシー)=***>>

のように記述すると、バスと徒歩は継続する交通手段、タクシーは別の交通手段という解釈になる。

一般にタグ内の意味の識別可能な最少の単語に分割する場合、タグの構文規則が複雑になる。上記の例ではバスの停留所名、徒歩の場合に規定される様々な状況(～を右に曲がる等)を全てルール化する必要がある。従って実際にはコンテンツサーバに入れる情報を適度に取捨選択することが大切である。

(2) セマンティクス

タグには意味表現の多様性がある。そのため次のことについて考える。

① 用語の統一 (ページ内)

時間を表わす同じ情報には8時30分や8:30また8時半といった様々な表現方法がある。この場合、予めどの表現を使用するか決めておく必要がある。

また例えば動物を表わすものとして犬、ネコ、ライオン等の個々の動物の種類や、哺乳類、魚類等で分けるものといったレベルの違った分け方がある。ネコと哺乳類の両方の情報を動物を表わすタグを使用して画一的に使用するとおかしくなる。この問題を解決するために表現形式の多様性をどう克服するか問題となる。この問題はタグをどこまで詳細化するかにも依存している。

② タグの分野依存性

これは分野毎に専門用語が多数存在するため起こる問題である。それぞれ別の分野で用いられている異なる用語が同一の意味であることがある。例えば遊園地などでは中へ入る時に必要となるお金の呼び方として入場料である。しかし美術館の場合、入館料となっているかもしれない。そこでタグの名前の抽象化をどこまで進めるかが問題となる。

③ タグの文脈依存性

これはタグがページの文章中の情報に異なる意味に重複して使われるかもしれない問題である。例えば博物館に関したページで、時代を表わすタグを用いるとき、それが博物館に置かれているものに歴史がある場合と建造物である博物館に歴史がある場合で意味が異なる。

4. 視点に基づく探索

4.1 ユーザの意図

現在インターネット上にある情報を探すシステムとしては分野毎に情報を整理してそれを木構造のような情報のつながりによって表現しているものが多数ある。この方法は分類を徐々に絞って目的の情報にたどりつこうというものである。しかしこの方法では情報を整理する人の感覚によって作られている部分が多いので人によって自分の欲しい情報が得られないかもしれない。このことは探索が1つのノードを根とする木構造ではなく、複数の根を持つ森構造で行なわれることが望ましいということを意味している。言い換えれば、同じ項目に利用者独自の多様な視点から到達できるということが大切である。

視点(ビュー)とはある関連性を持った概念の存在する領域(ドメイン)である。図2に視点の例として動機(Intention)、分野(Category)、行動(Action)、位置(Location)、時間(Time)を提示している。例えば行動という視点では見る、食べる、行く、買う、習う等の人間の行為を表現する概念が含まれる。情報検索とはこれらの視点の中のもの(=検索項目)を指定することで対象を限定する行為である。

ところでユーザの情報検索の一般的特性として

- (1) 部分的で不完全な概念により指定する場合がある。
- (2) 直接的に概念を指定せず、関連性のみ限定する場合がある。

ということが想定される。例えば図2で、ユーザがCategoryビューからArtを選び、ActionビューからLookを選んだとする。このままでは検索できないので、ユーザの意図を限定するためにIntentionビューを表示する必要がある。この結果、ユーザがInquire facilityという項目を指定すれば、ユーザの目的を達成するための方法がInquire museumと表示される、という具合に検索が進む。これが上記(1)の不完全な問合せに対するユーザのガイダンス方法である。

さて以上の例でさらに具体的な美術館名を限定するまでのプロセスでは、どういう美術館に行きたいのかというユーザの希望を明確化する必要がある。このためには絵画、工芸等の分野を直接的に限定する場合と、～駅に近い美術館というように対象との関連性を指定する場合がある。後者は特に地理条件、時間条件等が関連する場合が多く、コンテンツサーバに格納されている全ての対象について関連性を計るための

距離が定義されている必要がある。

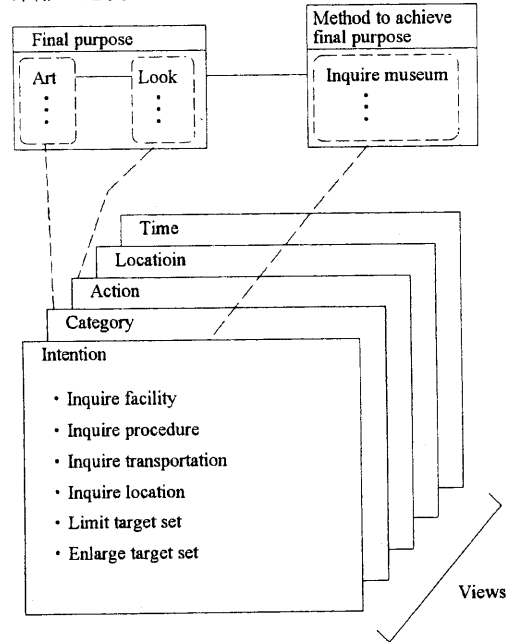


図2. 視点からの探索

4.2 ビューの設計

普通のホームページの場合、一般に目次の役割を果たすページを紹介しているところは木構造のように情報がつながり利用者は自分の現在探している位置の確認がわかりやすくなっている。またYahoo等のナビゲーションツールも同様に探索木の構造を有している。一方、視点を用いた探索ではユーザは様々な視点を指定することになる。このため、検索項目の追加指定や削除を行う過程で、現在、自分が何を指定して情報を探しているかが分かりにくくなる恐れがある。そこで現在、我々が設計中のコンテンツサーバでは図3(次項)に示す検索画面のようにユーザが指定した視点の有効な範囲を表示するようにしている。

ナビゲーションツールの構成では検索画面の数、ならびにその深さ(何回の選択で目的とするWebページやコンテンツに到達するか)が重要である。いま多分木構造の探索木でn分岐でm段の深さがあると仮定する。またn分岐毎に1枚の検索画面を構成するものとする。このとき、全体で n^k 個の情報に辿りつき、検索画面数は

$$\sum_{k=0}^{m-2} n^k = \frac{n^{m-1} - 1}{n - 1}$$

となる。この場合、検索画面数は深さ m に関して指数的に増大するが、提示可能な情報（ページやコンテンツ）の総数とほぼ等しい。

一方、複数の視点から検索を進める方式では視点の伝意性（興味ある視点のみ選択）と各視点内での検索項目の伝意性がある。指定された任意の組合せに対して新たな検索メニューを提示する場合、検索画面数は膨大になる。いま視点が n 個 ($V_1 \sim V_n$)、各視点での項目数を T_i 個 ($i = 1 \sim n$) とすると、全ての視点について項目を指定する組合せは $\prod_{i=1}^n T_i$ である。この中

には分野 = “美術”、行動 = “食べる” のような無効な組合せがある。これらを除くと、目的とする情報やコンテンツと同程度、すなわち探索木並みの検索画面数になる可能性がある。これに関しては視点の一部しか指定されなかった場合の扱いが問題になる。検索画面数の増大させない方策は

- (1) ユーザに視点の指定を促す。
- (2) 既知の検索画面の内容を動的に合成する。が考えられる。ここで (1) は未指定の視点をユーザに提示するという自明な方法があるが、ユーザの選択範囲を始めから限定してしまうという強制的な側面がある。一方、(2) の方法は曖昧なユーザの要求を少しずつ明確化するという、より穏やかな方法である。ここで視点 $V_1 \times V_2$ (V_1 と V_2 の組合せ) から視点 V_1 の検索画面を得るには、関係データベースにおける射影のような操作を行なう。これが視点の合成の基本的な考え方である。このとき、多数視点を単一の視点に射影すると、検索画面中の項目数が増大するという問題がある。今回、コンテンツやメニューを格納するデータベースの構成を論ずることはできなかったが、上述したような問題に対処するには問合せに対して該当する項目の総称を返答するような知的な機能が必要になると考えている。

5. むすび

本稿では観光情報をインターネット上で提供するためにコンテンツを収集、提示する方法を議論した。各ホームページはの内容を自動的に収集するためにHTML文章中の各項目に内容を識別するためのタグを付与すべきことを提案した。

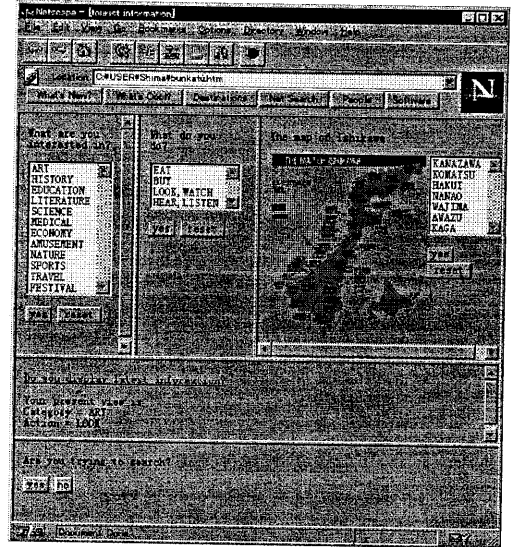


図3. 検索画面の一例

また情報の検索ではユーザの複数の視点から対象を絞り込むブラウザが重要と考え、部分的に試作した例を示した。これらの検討を通じて多様な課題が明確になった。例えばタグによる内容抽出ではタグが規定する内容の範囲や、タグの分野依存性、文脈依存性が問題である。また複数の視点からの検索では途中段階での検索画面の膨大化、或いは目的とする情報への到達性の検証などが課題である。本報告で検討した内容は必ずしも観光情報に対象が限定されている訳ではないが、以上のような課題に答えていくには今後も対象をある程度、限定することが大切と考えている。

謝辞

本研究の発端は北陸テレコム懇談会が主催した調査研究に筆者の一人（永瀬）が加わったことにあります。貴重な議論をいただいた上野貴弘 前北陸電気通信監理局電気通信部長を始めとする調査研究会の関係各位、並びに飯島泰裕 金沢大学経済学部助教授を始めとする作業部会の関係各位の皆様には感謝いたします。

参考文献

- [1] マルチメディア時代の観光情報等に関する調査研究報告書、北陸テレコム懇談会、1996
- [2] <http://www.yahoo.com/>
- [3] <http://www.nsknet.or.jp/ishikawa/>
- [4] <http://www.nsknet.or.jp/kcity-stay/>